SYNOPTIC REPORT TOWARDS THE COMPLETION OF PHD BY PUBLICATION

BRUNEL UNIVERSITY **MURRAY DICK**
NOVEMBER 2015

TITLE

News Values in Online and Visual Data Journalism

INDEX OF CONTENTS

1. INTRODUCTION

This submission explores tensions in the (Gansian) 'news values' evident in the working practice and outputs of online and data visual journalists; caught between discourses of 'tabloidization' and 'the fourth estate' ideal.

I submit the following six pieces (included in appendix) – a monograph, and five peer-reviewed journal articles, for the award of PhD by publication:

1. 2010 (I). 'Vanishing Point: On the consequences of lost online news in the UK.' *Ethical Space*, 8(1-2): 14-19.
2. 2011. 'Search Engine Optimisation in UK News Production.' *Journalism Practice*, 5(4): 462-477.
3. 2012. 'The Rebirth of the Beat: A Hyperlocal News-gathering model.' *Journalism Practice*, 6(5-6): 754-765.
4. 2013. *Search: Theory and Practice in Journalism Online*. London: Palgrave MacMillan.
5. 2014. 'Interactive Infographics and News Values.' *Digital Journalism*, 2(4), 490-506.
6. 2015. 'Just Fancy That: An analysis of infographic propaganda in The Daily Express, 1956-1959.' *Journalism Studies*, 16(2), 152-174.

Collectively this work fulfils the criteria for PhD by publication, in terms of volume and academic substance. In particular it:

- Forms a distinct contribution to knowledge in journalism studies;
- Evidences a broad knowledge and understanding of journalism studies;
- Is the product of the successful application of appropriate methodologies pertinent to high-quality research in this field;
- Is satisfactory in its literary presentation, and;
- That it is (self-evidently) suitable for publication.

The body of work presented here is a hybrid PhD submission; partly arising from research-as-praxis developed over my ten years of working in journalism, and partly arising from empirical cultural studies and social scientific research epistemologies. Each output within this submission represents a point along the phenomenological arc of my contribution to new knowledge and understanding in the practice (and study) of online journalism. Collectively it represents a rigorous and significant challenge to the conventional way in which journalism is often understood within the academy; compromised between vocational and liberal arts approaches (Skinner et al 2001).

My research is concerned with how news values are operationalized in two (relatively) new and related fields; online journalism and visual data journalism. Each work is, in turn, concerned with the interconnection between technology and process in the working cultures of these two fields. I found that the news values of the professionals who work here, defined with reference to Gans (1979), are broadly the same 'fourth estate' news values embraced within more traditional journalistic cultures. However, some non-journalistic values find voice too, amidst the constraints of organisational norms. While the online environment represents new ethical challenges for journalists (Dick 2010 I; Dick 2013), in national newsrooms across the UK, editorial judgement continues to drive journalistic work online; whether in the fields of trend-spotting (Dick 2011), news-gathering, beat-reporting (Dick 2012), interactive design or investigative journalism driven by 'big data' (Dick 2014). But just as these journalistic norms are seen to guide approach and judgement today, this was not always the case in visual data journalism. On the contrary, I find that the origins of infographics in UK news lie in cold war propaganda, rather than in the 'fourth estate' ideal (Dick 2015). The rise of visual culture, and of tabloidization; and the growing sense of self-confidence and autonomy amongst those specialists working in networked newsrooms, all represent essential context for our better understanding of contemporary practice in online and visual data journalism in the UK.

1.1 RESEARCH AIMS AND QUESTIONS

Research questions must be answerable, that is capable of being answered within the context of the time-frame of a research project (Andrews 2003: 3). Throughout the research, and in each of the discrete research projects presented here, I sought to identify a series of questions towards answering one general, over-arching research aim, in the hope that this approach may yield nuanced understanding (and new knowledge) of news values in online and visual-data journalism. In essence this process involved a refining process [or 'tightening of aperture' (Andrews 2003: 10)] that in turn defined "the shape, direction and progress of the research" (Andrews 2003: 4). The primary influences over this refining process may be understood in terms of 'orienting decisions' (Cohen et al 2000:74); that is, each research project was informed by a series of structural factors which influenced its shape and design. In distinguishing my general aim from various research questions, I sought to ensure that each discrete research project was both manageable and workable (Andrews 2003: 24).

Two approaches to identifying research questions were employed across the body of works presented here; some represent an attempt to identify, isolate and interrogate a particular manifestation of the general aim; others arose during the various literature reviews undertaken. In moving from a broad aim oriented around 'an investigation' to a series of 'what are' questions, my intention has been to tighten the 'aperture' to focus on applied news values in different contexts. Though 'what are' questions may be answered in several ways (using different methodological approaches), in terms of qualitative research, I have found them to be particularly well suited to yielding 'thick description' (Geertz 1973).

In the remainder of this section, I will set out and justify my research aim and each of my research questions, in each case factoring in (where appropriate) the following key elements:

- Inspiration (and where appropriate, influences in extant literatures)
- Key definitions (that both contextualise and set the bounds of this research)
- 'Answerability'
- 'Workability'
- 'Aperture'
- 'Orienting decisions'

1.1.1 RESEARCH AIM

To investigate news values in online and visual data journalism; and the extent to which these values tend towards 'tabloidization', or the 'fourth estate' ideal.

Inspired in the first instance by my own experiences in the field, and latterly by a long-established body of organisational studies of the newsroom (Tunstall 1971; Schlesinger 1978; Tuchman 1978; Gans, 1979; Fishman 1980), this research aim seeks to tease out the tensions between theory and practice in two relatively new fields of news-work.

By delimiting all but one of the studies presented here to a particular theorisation of news values (Gansian 1979), my intention was to tighten the theoretical aperture in what is a contested (and not to say disputed) field of study (Wahl-Jorgensen and Hanitzsch 2009: 10). I position my work amongst other studies of the phenomenon as a socially constructed phenomenon in the newsroom (in terms of values and decision-making). There is one exception to this approach within the works I present here; my study of infographic propaganda may more appropriately sit within a foundationalist approach to news values, what may otherwise been described as a 'uses and gratifications' approach.

For the sake of this research, 'online journalism' may be defined as:

> *News and other forms of journalistic output that is distributed via websites, social media, RSS feeds, email newsletters, and other forms of online communication.* (Harcup 2014: 220)

Though much of my working life has involved training journalists working across various media (from radio, to television, print and online), the process of research carried out online remains consistent; and indeed all but one of the works presented here are exclusively concerned with discrete manifestations of journalism that can only be enacted online (database research, newsgathering and beat-building; un-publishing; search engine optimisation and interactive infographics).

In the absence of a formal definition of 'visual data journalism' in the literature, for the purposes of this study this concept is explored in the context of the working practices of journalists responsible for the creation of infographics in news. Infographics may be defined as:

> *A way of presenting information visually online, often involving elements of interactivity whereby website users can click to change which information is displayed, open up videos or other ways of telling the stories contained within the data, or search for particular terms or places* (Harcup 2014: 138)

The open-endedness of my over-arching aim reflects the ongoing (ostensibly intractable) nature of defining news values; it reflects a broad span of journalism practice, and my own shifting pedagogy during the course of the research presented here. News values are prone to change, influenced by changing patterns of (and influences upon) both production, and consumption. In order to ensure that my approach to studying this phenomenon was both answerable and workable, it was therefore necessary to tighten the 'aperture' (Andrews 2003: 10) of this study by establishing parameters for the concept in the form of a definitional spectrum. This made it possible to establish the meaning of news values in the context of theories of 'tabloidization' and the 'fourth estate' ideal, presented as poles in the spectrum of 'news values'.

It is acknowledged that there are serious conceptual (and not to say ideological) problems inherent to commonly accepted definitions of these two concepts (which I will discuss in depth during the review of literatures presented here). Nevertheless, both in terms of establishing the answerability and workability of my research aim, and while addressing issues arising in the literature, such as fears attendant to the 'rise of the click-stream' (Currah 2009), I found it beneficial to work within this conceptual framework.

Some of the following research questions arose from professional experience of little-explored fields; in some cases there was little literature to rely upon. Exploratory empirical work was necessary in all cases; and I found a range of paradigmatic approaches were required across the project. The research questions I outline here may be considered 'contributory questions' towards (collectively) resolving the broad research aim outlined above. My approach takes after the non-conventional ‘inverted pyramid’ (Andrews 2003: 45) approach to research question-setting, in so far as in most of the research I present here, I started with a (methodologically speaking) small initial sample (while also relying on my own experience and understanding), in order to develop a wider methodology accordingly.

1.1.2 RESEARCH QUESTIONS

RQ1 *What methods and sources can be employed to optimise online newsgathering?*

RQ2 *What methods can be employed to ensure the veracity of information sourced in online newsgathering?*

The work I present here and in particular my book *Search: Theory and Practice in Journalism Online* (2013) emerged from a sustained attempt to challenge two early crises in online journalism practice: information overload and the problem of verification in online news sources.

Information overload is not a new concept, but it became a pressing challenge during the 1990s as a result of the increasing importance of e-mail and the internet in professional life (Bawden et al 1999). In many professions, not least journalism, this led to the perception of a growing need for training in online research skills (Garrison 2000). My interest in this concept arose during my studies in Information Science at City University, London (2002-2003), informed by the works (and teaching) of leading figures in the field, including Professor David Bawden. Here I learned that traditional positivist approaches to theorising information retrieval (i.e. mathematical set theory) are severely limited without reference to information seeking behaviour (from education and organisational studies perspectives) (Bates 2002). This realisation had a lasting impact on my work, both as practitioner and academic; and as a result, my praxis is situated firmly in the field of participatory research (Cornwall and Jewkes 1995). The various theories I learned during my MSc, as applied to the field of online journalism, form the theoretical spine of the monograph presented in this submission.

I first started to apply search theory to the practice of news journalism in taking up a position as researcher/trainer at the BBC, in 2004. This professional experience would later introduce me to a second major theme in the emergence of online news: the crisis of verification online. In December 2004 BBC World broadcast an interview with a situationist hoaxer pretending to be the Chief Executive of Dow Chemicals, who claimed responsibility (on Dow's behalf) for the Bhopal disaster of 1984 (Wells and Ramesh 2004). The production journalist responsible for organising the interview (I discovered later, through my own contacts within the corporation) had fallen prey to a hoax website, purporting to be that of Dow Chemicals (www.dowethics.com). The events leading up to this broadcast represent the continuation of a discourse emerging from the origins of online news reporting in the 1990s; that in its rise to primacy as the leading platform for breaking news, a crisis of verification was perceived as undermining the reliability of online news (Allan 2005: 72-74; Allan 2006: 21). My working life and experience over the past decade represent a sustained attempt to challenge this; and this experience in turn, represents a means of tightening the aperture on my wider research aim.

The first two research questions I present here contribute to making my research aim answerable and workable in so far as they both inform the socially constructed nature of news values. Both represent manifestations of news in this context; both are concerned with the routinised human processes that are inherent to deciding what is news.

The term 'newsgathering' as employed in both of the questions above, may be defined as:

The process of obtaining information and other material for the purpose of transforming it into items of news (Harcup 2014: 202)

The 'orienting decisions' that informed the research projects I established to answer these questions, were primarily structural. For example, I was paid to undertake and develop my praxis, and so at its origins this corpus of knowledge owes a debt to (but was also constrained by) the influence of corporate needs at the BBC during the time of my employment. The lack of time I had to conduct this research, and the limited nature of my contacts within the corporation also contrived to limit the replicability and reliability of my means of answering these questions.

RQ3 *What are the consequences of un-publishing online-only news?*

During my time at the BBC, I first encountered the issue of unpublishing online news as a practitioner. The consequences of this phenomenon posed a serious challenge to my training, but it was equally clear to me that if online news which (by degrees) reflects our life and times may be prone to disappear without warning, that this may represent a serious issue of much wider public concern. Given the contacts I had made across the BBC by this time, it was a relatively easy process to pitch a features article on this topic to the technology features desk at bbc.co.uk, which I duly did in mid-2008. The article was published the following December (Dick 2008), and on passing from industry to academia in 2009, this topic became the subject-matter for my first piece of formal academic research. The form of this research question would later be influenced by the (mostly US) practitioner literature on un-publishing (Maier 2002, 2007; Nadarajan and Ang 1999; Shepard 1998).

I made this research question both answerable and workable by developing a case study of best practice that (at the time) offered a relatively transparent means of challenging the phenomenon. Because The Guardian was one of a small number of UK newspapers with a readers' editor, and a relatively public (and transparent) policy on the treatment of requests to unpublish, and because the readers' editor agreed to contribute, it became relatively easy to set the bounds of this study. However, these factors in turn constrained this study too – a much wider interrogation of unpublishing at other UK newspapers, and indeed newspapers around the world, would have yielded a much wider, and more reliable sense of the state of the phenomenon across industry more generally.

In keeping with my first two research questions, my third question is also informed by the socially constructed nature of news values (and by the operationalization of newspaper policies with regard to unpublishing). As such it also represents a manifestation of 'news values' as a process of decision-making, that in turn makes it possible to yield a more nuanced sense of what news values are and how they are practised. Given both the nature of the problem, and the limited source of contributions, I decided to devise a semi-formal interview case study in this project. This in turn served as a helpful means of grounding wider policy advice; albeit given the limitations explored above, it is acknowledged that a much wider survey of industry should inform future policy guidance in this field.

RQ4 *What is the optimal way to develop an online news beat?*

This question was inspired in part by my own praxis in online search, and in part by concerns in the literature raised about the decline of journalistic beat-reporting in UK local government (Morrison 2010) and courts (Davies 2008: 77/78). The structure of the project developed was the product of grounded, experimental, empirical research.

For the purposes of this project, the following definition of a 'news beat' is most fitting:

> *...the subject matter or geographical area that a reporter might be assigned to cover as in 'the crime beat' or 'the Washington beat'.* (Harcup 2014: 34)

The news beat as expressed online is conceived here as a socially-organised construct, informed by journalistic action; and so it speaks to news values (and decisions about the 'newsworthiness' of sources online) in this respect.

This research question became answerable as a consequence of the 'action research' approach, and the experimental methodology I developed for this little-studied field. Here I sought to create a normative framework that may be of practical use to online news workers who may be mapping an offline beat online. I developed (based upon my own experiences, a posteriori) an iterative prototype for an online news 'beat', and tested it over a series of weeks in two experimental real-world case studies; one rural (covering Ilminster) and one urban (covering Borehamwood). I engaged two journalists; one working in each area (both former journalism MA students at Brunel University), to test these systems, which led to the discovery that this automated system approach to online newsgathering can identify potential news stories that local journalists may otherwise miss. The 'orienting decisions' that defined this project included limitations attendant to local beat reporting in the UK; time and resource constraints, and obsolescence in contemporary freeware technologies; which in turn set limits on the replicability, reliability and representativeness of my approach (in both cultural and temporal terms).

RQ5 *What effect is Search Engine Optimisation having on online news?*

Having undertaken three discreet projects on the processes involved in newsgathering in journalism online (i.e. ‘inputs’), and having acquired an understanding of the consequences these elements have in shaping journalistic news values, it became apparent that in order to better understand the nature of news values in online journalism it would be necessary to engage with journalistic ‘outputs’ too. This decision was reinforced when I read of the so-called ‘culture of the click stream’ (Currah 2009) in online news; an argument that journalists and editors are beginning to allow user metrics to influence their editorial decision-making. In addressing this problem, it was necessary to situate 'the culture of the clickstream' within existing (and long-established) literatures on quality and standards in news; and so two concepts ('tabloidization' and the 'fourth estate' ideal) emerged as poles across a spectrum, that may allow for a more nuanced (and theoretically robust) interrogation of this ostensibly new phenomenon. This in turn represents a particular manifestation of news values in online journalism that contributes towards answering my wider research aim. In order to adequately interrogate this issue I felt it necessary to return to the newsroom, and to bear witness to practice in the field.

For the purposes of this study, Search Engine Optimisation (SEO) may be defined as:

> *Writing copy and selecting photographs for online journalism in such a way as to increase the likelihood of items being found by and ranked highly in search engine results pages.* (Harcup 2014: 274)

As is true of my earlier research questions discussed above, 'what are'-type questions may be interpreted (and answered) in a range of ways. However, ethnography is a particularly useful approach in understanding online journalism in this particular context, because some of the practices undertaken (especially as they relate to computer work) represent new ways of *doing*. In asking the above research question it is necessary to understand not only what are the concept's theoretical precepts, but more importantly to observe how what is understood is actually practised. Google tells us that up to 200 'signals' are used in the ranking of content in its algorithms (Google 2015); but it does not say how they are weighted (with good reason – otherwise the system would be open to corruption). In this vacuum of informed opinion the concept of SEO takes on a slippery quality; such that if not handled carefully, it may dissolve into meaninglessness, and so any attempt to understand how contemporary online journalism is conducted may miss (or crucially misunderstand) a key element in its’ practice.

While researching SEO I encountered two ethical dilemmas that challenged my assumptions and findings (and that may in turn be considered 'orienting decisions'). Firstly, I found myself contributing my own understanding (and expertise) in the field whilst engaging with some of the subjects of this research during participant-observation. Of course this approach (as with research-as-praxis) may be considered to challenge the detachment or objectivity of the study, but I make no claims in these directions. On the other hand, in taking this approach, it was possible for me not only to de-mystify some of the attributes of SEO-as-practiced, but also to shed light on what those in the industry understand of the concept and its application. This approach in turn enriches my findings; and underpins the originality of my contribution to knowledge in this field.

It was also necessary for me to iteratively change the content and nature of my questions over time, according to the nature of the themes I identified. On interviewing an SEO expert at The Guardian for example, I asked three questions that were declined on grounds of commercial sensitivity:

- *What proportion of your website's traffic comes from:*

    1. *search*
    2. *third party 'editorial' pages*
    3. *your home page (or hub pages)*

    … and how have these proportions fared over time?

- *What proportion of your traffic comes from people typing your domain into a browser, and how has this fared over time?*

- *What proportion of your traffic comes from people typing your title/domain into a search engine, and how has this fared over time?*

In this research I was concerned with how SEO informs decision-making (and news values) in the newsroom, and so this information I considered secondary to that understanding. But it is nevertheless worth reflecting upon that not only must researchers seeking to understand the potentially slippery concept of SEO grapple with the secrecy of search engine companies, they must also negotiate the secrecy of the news organisations seeking to mediate search engines.

RQ6 *What effect are interactive infographics having on our online news?*

Moving on from the relationship between SEO practice and news values, I began to question whether other web-native practices may similarly influence story selection. This led me to the field of interactive graphics (and visual data journalism). I continued the ethnographic approach employed in my work on SEO because of the technologically esoteric nature of working practice here (and because of the fluid nature of 'best practice' in this field). In addition, I found depth interview to be a particularly useful approach to better understanding practice here. News 'interactives' are created in a range of ways by a cast-list of individuals from very different professional, technical and skilled backgrounds; and there was little coherence or consistency of approach amongst those I interviewed. To this end, it was necessary to engage with these individuals on their own terms; to iteratively change the format and content of questions in order to improve clarity and understanding; in order that a common 'language' of interactive graphic composition could be established.

This question was by degrees influenced by a series of organisational studies of non-interactive graphics in the newsroom (Lowrey 1999, 2002), but also by an ongoing debate around the concept of 'chart junk' (Tufte 1983) in infographic news design. In order to fully interrogate 'chart junk', it was necessary to situate it within the same existing (and long-established) literatures on quality and standards (and spectrum) in news as discussed above. Again, this represents a particular manifestation of news values in online journalism that contributes towards answering my wider research aim. The answerability of this question was established in part due to limitations in access (in the same senses as were experienced in my study of SEO).

H7 *Infographics emerged in the UK press as a means of producing post-war propaganda at the Daily Express.*

The final research question I present here differs from all the others not only in that it is a hypothesis, but also because it does not speak to a socially-constructed definition of news values. I sought in this final research project to contextualise the knowledge I had derived up to this point in relation to how interactive design (and data visualisation methods) inform news values. I sought to establish what were the news values of those who originated the visual data journalism form in UK news. Because this topic is not defined by contemporary practice (indeed many of the key figures involved in the origins of news infographics are no longer alive); two issues became apparent:

- A constructivist approach would be well-suited to exploring this field; and;
- A fully representative data sample of infographics would be necessary.

For the purposes of this study, propaganda may be defined as:

> *...the deliberate, systematic attempt to shape perceptions, manipulate cognitions, and direct behaviour to achieve a response that furthers the desired intent of the propagandist." (Jowett and O'Donnell 2006: 7). These perceptions, cognitions and/or behaviour may be "perceptual, cognitive, behavioural or all three (Jowett and O'Donnell 2006: 8).*

I was heavily reliant on online databases as sources of secondary data during this project. In practice, this approach represented a key 'orienting decision'. Limitations in technology may be seen to inhibit a full interrogation of the field by database search alone. For example, the Optical Character Recognition software used to digitise the archive cannot be fully relied upon; corruptions in word-spellings are not uncommon, which can limit the reach (and robustness) of keyword-based search. However, I established various measures to counter this. Aside from several generic keyword searches, I undertook a grounded approach to testing content in the system; using those infographics found initially as sampling for other keywords (including bylines). Similarly, once significant hypotheses were arrived at in this study, in an iterative fashion (such as that no infographics on the theme of Britain remaining a global power seemed to have been published between August 1956 and April 1958), I manually browsed the archive between those key dates, in order to evaluate the validity of these contributory hypotheses.

By focussing on outputs and not the 'inputs' of news values, a valid critique of my approach may be that it doesn't represent 'news values' in any verifiable sense of the term (and indeed it may be argued that this project represents an inconsistency in the wider project that undermines the over-arching validity of my wider research aim). However, I would dispute this on the same grounds that I have justified my approach; in terms of paradigm interplay (which I will explore in some detail in a subsequence section).

2 LITERATURE REVIEW

2.1 SITUATION WITHIN ONLINE JOURNALISM LITERATURE

Academics have attempted to critically map the terrain of journalism studies, in order to, amongst other things, provide an epistemological justification for its presence within academia. Wahl-Jorgensen and Hanitzsch (2009) suggest that developments in this field are best conceived of in the form of four chronologically sequential epochs;

- The nineteenth and early twentieth (normative) theories;
- The mid-twentieth century empirical (and largely positivist) sociological studies,
- The 1970s twin sociological and cultural studies scholarships, and;
- The 1990s 'Global Comparative Turn'

The limitations in each of these approaches is evident in turn. In bending the human elements of journalism to ideology, the first movement of theorists failed to engage with the process of negotiation and settlement that is central to most human (and so professional) endeavour. The empiricists of the mid-twentieth century failed to accommodate the system that holds each individual together, and the unseen influences of the (individualist) culture within which they were embedded. The sociologists of the 1970s carried (by degrees) the limitations of the empiricists into a new era, while cultural studies theorists carried the ideological mores (and blind-spots) of the earlier normative theorists.

Much of the work I present here has at its origins the sociological narratives developed during the 1970s; albeit as seen through the prism of contemporary and post-1990s research (especially the work of Wilson Lowrey). The organisational studies of this earlier period have taught us much about how news values are (and the functioning of the newsroom is) negotiated, by foregrounding the individual in the newsroom, and by selecting and applying suitable methodologies (primarily ethnographical) in order to better understand the practice, process and function of news-work. My work in infographics also has the 'radical turn' at its heart; particularly the semiotic analysis of Stuart Hall (albeit viewed through the scientific positivism of Edward Tufte). Here, in the absence of a clearly defined epistemological framework, I undertook to use experimental methodologies towards better understanding the forms (both monosemic and polysemic) of propaganda sign-systems in mid-twentieth century newspaper infographics. In so doing I have sought to develop an alternative way of thinking about propaganda to those deterministic theories that minimise or dismiss the role of the journalist (more particularly the visual journalist) in the modern, networked newsroom.

The rise of the internet has established the primacy of the network in contemporary journalism studies. Scholars drawing upon theories from the sociology of networks (Castells 1996; Volkmer 2003) have advanced new understandings of online journalism by conceiving of it as comprising a network of nodes (information outlets) interdependently connected. I too have accommodated network theory in my own research because of its value in helping us understand the importance of working pressures in online journalism. That news organisations are increasingly looking to the internet in producing interactive journalism has had profound consequences for the relationships between journalists and their audiences.

Where hierarchy used to impose hegemony in organisations, today it is argued that power relations are transacted across networks. The nodes that comprise this network are not uniform; some may be 'more important' than others; but relationships are defined by inclusion within the network (Castells 2000). Herein journalists may act as 'supernodes' within the network, given their privileged access to others; with a competitive premium placed on scale and interconnectedness within and across these networks (Heinrich 2011). Of course in reality this may not be the case; the specialists I have identified as key to modern interactive journalism need not, in practice, have any more or less direct contact with their audience than their offline peers. Nonetheless, exclusion from the network may be seen to pose a challenge to public interest journalism: a journalist disconnected from the network may struggle to keep abreast of those who wield power, and so may be unable to fulfil their fourth estate function. In reality online networks are self-selecting, and so it is necessary that journalists make concerted efforts to map their offline networks online, otherwise they may fall into the trap (indeed some may say re-produce the failings) of elitism that are seen to damage journalism. An ethical space can therefore be seen to open up concerning the extent to which publics are included in those news and information networks that are central to modern life. My normative 'beat' was designed prescriptively in order (in part) to give journalists control of their research. But just as importantly, it is intended to map offline journalistic networks to an online network, in the hope that this process may make up for the diminution of 'beat reporting' commonly found in contemporary research in the field (Hansen et al 1998; Machin and Niblock 2006).

In terms of the shifting relationship between journalist and audience, the assumption underpinning contingent theories of online (or converged) journalism, that consumers of online news are 'active' users, as opposed to primarily passive consumers of offline news (Deuze, 2004) is problematized in several respects in my work. The notion of 'active' online news audiences is predicated upon journalists having the time, desire and incentive to develop a wide-ranging online public (social) network, and indeed to privilege this network over off-line, or private networks. This need not, in my research findings (and experience), be the case.

The internet has inspired both optimistic and pessimistic views of the future of news. Allan (2005) shows that positive discourses concerning the potential and realised value in online interaction and up-to-dateness, were countered by negative discourses concerning the provenance and reliability of online information. Raising his own concerns that online journalism is embracing the "attractive wrapping of television news" (2005: 80) he calls for more research, and deeper historical perspective. This critique echoes the work of Harrison (2006), who construes approaches to journalism studies more generally in two forms; one that perceives news as an ideological problem, and another manifested in normative approaches that are concerned with what news *should* be. Predicated on value judgements, these approaches tend towards either optimism or pessimism; the latter far outweighing the former in the literature (Harrison 2006: 186). In terms of online news, Harrison's approach may be re-cast as: "there is no doubt that 'high speed news' is in the ascendant; but is it any good?" (Pavlik 2000: 232). My own research findings in SEO suggest that the online environment can help journalists to identify news stories and coverage that may otherwise be missed, and none of the journalists I interviewed suggested that they were pressured into writing up stories in order to chase social media trends.

Pessimism and optimism are otherwise predicated across a range of binaries in online journalism studies; informed by, for example, themes of "speed and space; multiplicity and polycentrality; interaction and participation" (Fenton in Allan 2010: 555). Here the deterministic notion that the internet may bring about more and greater diversity within news has been challenged. Redden and Witschge (in Fenton 2010) tracked coverage of five stories across various online news outlets, finding that these stories (and much of the content contained within them) had been reproduced widely across the web. That this approach does not acknowledge the full range of news published by particular news outlets over the course of time and that it doesn't consider some key social (i.e. Twitter) and 'dark social' (i.e. email) means of online sharing, limits its reach and conclusions. My own experience and research findings indicate that ideological theorising struggles to accommodate the realities of practice. Taking a theory-building approach, my findings include both cause for optimism and pessimism within the organising themes established by Fenton.

The continuing popularity of traditional news brands online has been cast in both positive and negative perspectives. On one hand, the same ideological forces that dominate (and circumscribe) offline journalism are seen to continue their monopoly of attention online (Fenton et al, 2010). Alternatively the dominance of these news brands indicates that "audiences want the core journalistic disposition to truth-telling and its corollaries, accuracy and sincerity" (Harrison 2006: 205). Elsewhere Harrison argues that "it is not true to say that the internet is solely in the hands of a particular set of a few powerful news providers, or suffers from an ideological rectitude placed upon it by such a group" (Harrison 2006: 189). A logical consequence of this approach is that the internet is not (yet) leading to a redefinition of pluralism in news production in any thorough-going sense. But conversely, it cannot be doubted that some assumptions inherent to traditional media online (such as journalistic authority) are coming under pressure as never before (Robinson 2007). This issue has generated pessimistic and optimistic discourses too. Some question the future of professional journalism, for if journalists lose their status as gatekeepers; as breakers and makers of 'news', then surely they lose their sense of purpose. On the other hand, despite the rise in online tools, journalistic skills, insight and ethics remain crucial to the developing news ecology. Indeed, the importance of these skills today, and for the future, has been expressed compellingly by several authors, conceptualised as the rise of Networked Journalism (Beckett, 2008) or Grassroots Journalism (Gillmor, 2004) in context of a shift in roles from gatekeeper to gatewatcher (Bruns, 2004).

Throughout my academic research, my general aim has been, while acknowledging the value of past discourses in journalism studies, to move on from the ideological battles of yesteryear in order that journalism studies continues to be seen to be relevant in society. For all the value and insight created in past studies, much in the field still talks “past journalism rather than addressing it” (Calcutt and Hammond, 2011: 2). Blumler and Cushion (2013) outline six dominant forms of analysis in journalism studies, decrying the fact that the journalists whose working lives and actions form the basis of this large (and increasing) body of study, seem so little influenced by its findings. In order that future work may influence events outside of the academy, they propose a series of normative challenges that may be addressed collaboratively by academics and journalists in their 'civic capacities'. One area of future collaboration they identify is 'the challenge of excluded voices'. This is an area I have sought to influence in my own research (in the creation of a normative news beat). It was only by understanding the locus of local beat journalists at close hand, when married to my own practitioner knowledge of the forms and methods used in online media consumption, that I was able to build a beat system with any potential for impact in the profession.

Some perceive academic and journalist as competitors (Hartley, 2000). This approach may be dismissed as paranoid (here journalists are framed as *redactors*; by connotation, *censors*) or dystopian (journalism is constructed as a profession of *violence*); but more damagingly it is blinded to human endeavour by ideology; the human elements of journalism are reduced to a footnote - journalists are dismissed as mere search engines (Hartley 2000: 43). My work and my experience represent a refutation of this position. Over the past ten years I have trained (and worked with) journalists from various mediums, and across continents. This experience informs and authenticates the research I present here; and it does so from an epistemological basis that cannot be matched without the residual knowledge I have acquired along the way. That journalism studies lacks a common epistemological basis makes it all the more crucial that practitioner-academics find their voice, and find a means of influencing debate on the future of journalism. That some discourses within the field are ideologically (even pathologically) opposed to this position, only serves to reinforce its vitality to the future of the field.

2.2 SITUATION IN AND CONTRIBUTION TO NEWS VALUES LITERATURE

> “[news values] are the (somewhat mythical) set of criteria employed by journalists to measure and therefore to judge the 'newsworthiness' of events.” (Franklin et al 2005: 173)

> “A subjective set of criteria that journalists use to assess the newsworthiness of events or topics. Though rarely written down or codified, journalists are thought to invoke news values in deciding what merits attention in news.” (Zelizer and Allan 2010: 89)

While for journalists news values are sometimes considered to be inherent, sociologists working in this field have long observed that they are more to do with the adoption of arbitrary, institutionalised values than they are the manifestation of instinct or a 'nose for news'. The detailed study of news values may be traced back, on Wahl-Jorgensen and Hanitzsch’s terms, to the ‘the empirical turn’ of journalism studies; and here two schools of thought, two normative approaches have been used to make claims to knowledge in this field.

The first, what may be described as a *uses and gratifications* approach, conceives of news values on different terms to the definitions introduced above. This approach considers news values to be foundational, and intrinsic to the empirical truth of journalism, traces of which may be detected in published output. This approach conceives of news values as noun; rather than news values as verb. Galtung and Ruge's (1965) taxonomy of 'news values' is widely considered to be the first major defining work on news values, and certainly it is a classic exemplar in this first school of thought. The approach taken here was a content analysis forged around thematic analysis of published outputs across Norwegian media output.

Criticisms of this approach (and the resultant twelve news factors) are many. There is a tendency in this model to privilege large-scale international news (Tunstall 1971). The authors themselves accepted a central limitation of their study was that audience effect was not considered; a factor that subsequent researchers would seek to build into their research design (Schulz 1982). When it comes to the functioning of the newsroom, others have argued that Galtung and Ruge's approach may tell us more about coverage than it does about why stories are selected in the first place (Hartley 1982). Anyone who has witnessed at first hand the negotiated practices and the clash of norms inherent to newsrooms, may feel that any model that fails to engage the individuals involved in these decisions would necessarily be limited in its conclusions. As with all normative approaches, these values are vulnerable to obsolescence. More recent scholarship indicates that some of the criteria identified require re-appraisal to bring them into contemporary news language and culture (Harcup and O'Neill 2001). Similarly, in conceptual terms, the significance of frequency as a news value, in terms of the extent to which news events occur within news production schedules, is challenged by the realities of 24-hour news (Wahl-Jorgensen and Hanitzsch 2009).

There is some irony in that this seminal work doesn't actually engage with what we understand news values to be in any thorough-going sense; that it is not actually concerned directly with decision-making (Harcup and O'neill 2001). Indeed, where it is necessary to seek out academic precedent in terms of looking at the decision-making involved in news selection, then the real origins of the study of news values may rather be found in Manning-White's 'gate-keeper' theory of news production (1950), informed in turn by the positivist sociology of Kurt Lewin (1943); or in Breed's functionalist analysis of the newsroom (1955).

These early approaches to the practice of news are somewhat limited by various aspects of their composition; by the hubristic notion of being able to 'capture' culture; by the hegemonic relationship between field researcher and 'subject'; and by the rigidity of a totalizing scientific method that seeks to drive out, rather than engage with difference and nuance; problems inherent to all pre-critical sociological qualitative research (Denizen and Lincoln 2005). Consequently, a less structured, more consensual school of inquiry into news values emerged via organisational studies; seeking to claim knowledge of news values from within the 'thick description' (Geertz 1973) of news culture as practised, experienced, and lived.

Tunstall (1971), influenced by the irreverence in the work of classic organizational studies scholar Everett Hughes, produced the first critical empirical organisational study of news culture amongst specialist reporters in the UK, in which he challenged foundationalist theories of the organization that cannot accommodate organizations with several types of goal. In the US, Epstein (1973) showed how the reality we experience in TV news is a myth forged by the twin pressures of corporate policy and budgetary requirements that similarly inform the news values of broadcast journalists.

Drawing upon Chicago School sociologies of the workplace, namely that people categorize work in order to control it (Becker et al 1961); Gaye Tuchman (1973) sought out the practical routines or classifications that journalists use in defining events-as-news, before seeking to establish whether these were sufficiently consistent to allow for routinization (she found they reduce variability of the matter that comprises the news). In a subsequent study (1978) Tuchman sought deliberately to avoid setting out a normative list of news values; an approach that differs markedly from the study most clearly associate with 'news values' today; Herbert Gans' *Deciding What's News* (1979). For Gans, reporting is based on inherent assumptions, which in turn are informed by both enduring and topical values that may be observed (after Galtung and Ruge) in the published outputs of news. Gans sets out a basic normative schema for thinking about news story selection; and he additionally (and separately) sets out criteria for story suitability. While accepting that these normative lists have some analytical value (not least because they accommodate much wider factors than merely decision-making (Brighton and Foy 2007: 11), they are nonetheless limited in their outlook; not least as they are still overly prescriptive, and too far removed from reality – they are fundamentally informed by a hegemonic subject-relationship.

In the work I present here I have drawn not so much upon Gans' normative theories, but in his conceptualisation of news values. Throughout my work (and in this report) I use the term 'news values' as an attribution of judgement amongst those specialists who affect (or produce) online and data news. I am mindful that "the values in the news are not necessarily those of journalists" (Gans 1979: 39), but nevertheless, it is important to understand how these individuals (some of whom do not come from traditional journalistic professional backgrounds) understand what they do, and why; and indeed what this means for an industry in which such specialists have an increasing role to play in influencing news production. My findings show that some non-journalistic professional values form a challenge to decision-making in the networked newsroom – that non-journalists (including graphic designers and computer programmers) openly question the coverage of certain stories, and the ways in which certain stories are covered. This is an important issue given the agency these professionals now enjoy in the newsroom, unlike their non-interactive, offline predecessors (Lowrey 1999; 2002). This in turn represents a potentially fruitful field for future news values studies.

Both normative and organisational approaches to news values are vulnerable to political economy theorisations. For example Schlesinger (1979) found *picture value* to be a selection criterion in television news; but for later scholars, normative lists were sometimes considered a transitory shell for wider issues in the occupation and ideological assumptions of journalists (and perceived ideological assumptions of news audiences) (Golding and Elliott 1979).

In terms of ideological interrogations of news values, Hall (1973) (whose own definitions lack systematic empirical basis) maintained that conventional approaches to defining news values, whether by means of content analysis of output, or interviews with journalists, are prone to underestimate (or elide altogether) the overarching ideological realities which shape the media. McChesney (2000), taking this theme further, argues that the emphasis on discrete events, and on individuals (rather than communities) in conventional, commercial, western news, speaks more to the need to perpetuate capitalistic norms, than it does to notions of democracy, freedom or truth. Herman and Chomsky (1988) do not assume that the news industry is entirely monolithic, but nonetheless, their approach is criticised because of its reductionism (see later section on propaganda). Others have questioned the value of the concept of 'news values' at all, suggesting that the little training many journalists experience, in addition to the lack of understanding many generalist journalists bring to sometimes complex news issues, leads to contradictory, and inconsistent findings (Lewis 2006). My own experience was rather that BBC journalists receive a good deal of journalistic training in-house, albeit I fully acknowledge that this may be exceptional within the UK media industry.

A division between 'news values' and 'commercial values' has been proposed (Allern 2002), the latter factoring in the economic realities manifested in costs associated with the pursuit of a story, the preparedness of material passed on to the journalist, and the exclusivity of distribution afforded. More recently, informed by the ideological school of news values, scholars have raised concerns about the news values of journalists now working in converged newsrooms (Williams & Franklin 2007; Fenton et al 2010). My own research findings would suggest that distinctions between 'news values' and 'commercial values' are in reality rather arbitrary; at least in so far as these factors do not dominate decision-making to the detriment of editorial values. But on the other hand, Allern's suggestion certainly finds its voice in the contemporary news practice of SEO: where, for example at the Guardian, press agency stories are expressly not optimised for search by journalists in-house, for fear that these stories may compete with original (costly) copy generated within the organisation, on related news issues.

Taking a different approach, Harrison (in Allan 2010) re-casts the process of gate-keeping and news selection, embodied elsewhere as the struggle between 'professional' and 'organisational' norms (Lowrey 1999), as a product of competing working contexts: “the newsroom background from which the fore-grounded application of rules is applied in a mundane and daily fashion according to how news stories are judged to fit into that background” (Harrison in Allan 2010: 199). This approach most clearly connects to my own experiences, and to a key finding in my research, which is to say that the extent to which ‘trending’ stories are pursued in online newsrooms is contingent upon the editorial and professional mores evident in those newsrooms; that no monolithic exemplar fits all.

The 'comparative turn' in this field has some rather early forebears; Chaudhary (1974) found that English language news journalists share common approaches to news judgement despite significant geographical and socio-cultural distance; while Lange (1984) found that government sanctions were found to shape news values in some developing countries. More recently, the role of gender in the application of news values has been interrogated (Lavie & Lehman-Wilzig, 2003).

Contribution

> “Ethnographic studies of news production provide invaluable insights into the nature and determinants of news production and a necessary corrective, therefore, to grand speculative claims and theories about the news media” (Cottle 2007)

Informed by the critical turn, the work I present here contributes toward a second wave of organisational studies and newsroom ethnographies concerned with the converged newsroom (Boczkowski 2004; 2005; Bruns 2005; Deuze 2007; Usher 2014). My work is unique amongst these studies in so far as I draw upon ten years' personal experience to explore modern practice.

In terms of the literature in news values, the research I offer here contributes to the phenomenon's eternal re-conceptualisation (news values being subject to ongoing change, and contested claims) (Wahl-Jorgensen and Hanitzsch 2009: 10). I offer an up-to-date re-appraisal of news values in the converged newsroom, as distinct from various ideological studies of the field [Williams and Franklin 2007; Fenton (ed) 2010]. I also provide 'thick description' of the operationalization of news values in the contemporary converged newsroom.

In my research I have sought to address specific calls for further study. It has been suggested that the converged nature of the newsroom (Harcup and O'neill 2001) and the increasing availability of metrics in converged newsrooms (Wahl-Jorgensen and Hanitzsch 2009: 171) are both important new topics for field work studies such as those I present here. In the work I present here, I find that the traditional tensions between organisational and professional norms have not been unbalanced by the availability of metrics in the online newsroom. The ‘culture of the click’ (Currah 2009) represents (at the time of writing) more of a moral panic, than a representative model of conduct across UK newsrooms.

My research also offers a grounded framework upon which future theories may be tested; and indeed, it has already informed one study based around Field Theory (Tandoc 2013; 2014). This is important given that news values are today being subjected to alternative theoretical and methodological critique, such as Critical Discourse Analysis (Bednarek and Caple 2014).

2.3 SITUATION WITHIN LITERATURE ON THE 'FOURTH ESTATE' IDEAL

> “Classical liberal theory views the press as a defender of public interests and a 'watchdog' on the workings of government... It is derived from the notion of 'estates of the realm'. The traditional three

are the Lords Spiritual (clergy that sits in the House of Lords), the Lords Temporal (other peers) and the House of Commons... Described as arrogant and grandiose by some, and satirised in an 1855 novel The Warden by Anthony Trollope, the notion refuses to lie down and die." (Franklin et al 2005: 84)

"a term historically used in classical liberal theory to denote the press, now broadened to refer to the news media – in particular, their role in ensuring a functioning democracy." (Zelizer and Allan 2010: 47)

Society and media production have long moved on since the mid-Victorian era, and so it has been argued, the ideal of the Fourth Estate created during this period, and which still finds voice today, should in turn be subjected to a more contemporary scrutiny (Curran I in Curran and Seaton 2003: 347).

For Hampton (2010) the concept is seen to be expressed in two different (albeit overlapping) discursive approaches, both underpinned by utilitarian liberalism. The predominant model is defined as the 'educational' ideal; here the press facilitate engagement in a process of "politics by public discussion" (2010: 4), allowing the 'truth' to emerge from a diversity of opinion, in the market-place of ideas. This approach was largely abandoned during the 1880s in favour of the second model (Hampton 1994: 106) that may be termed the 'representation' ideal. This approach doesn't integrate the public in political debate, but rather, in publishing a broad range of views, newspapers are considered as a collective proxy for public discussion.

In both of these models, the press serve as link between politicians and polity; but in some reading the press was even more. In the words of one of the more celebrated editors of the mid-nineteenth century, the Fourth Estate performed the political role it had been attributed by Carlyle as "a constitutional check upon the conduct of ministers" (Hampton 2004: 110); while to W. T. Stead, the press represented the "Chamber of Initiative" (Hampton 2004: 112).

In modern times, and reflecting upon the rise of the mass media during the early 20th century, the 'political myth' of the Fourth Estate has been challenged by those who, for example, perceive structural problems around the concentration of ownership within the commercial press. During the second half of the 20th century, concentration in media ownership arose as a concern in government; in particular during the 1977 Royal Commission into the newspaper press that spoke of a need to "...'protect editors and journalists from owners'. The exercise of proprietorial power no longer appeared to be legitimate as a basis for securing press diversity, at a time when only three men controlled over half of total daily and Sunday newspaper sales in Britain" (Curran I in Curran and Seaton 2003: 349). The fear of centralised control of newspapers was shared too amongst journalists: "a problem exacerbated by the decline of editorial autonomy caused by the actions of aggressive proprietors." (O'Malley 1997: 143)

Boyce outlines the origins of the Fourth Estate as a political myth that emerged from the utilitarian theory of (and perceived significance of) public opinion in public life (Boyce 1978: 21). From Bentham via James Mill, the campaign against taxes on knowledge, it was argued, was waged in order to establish an 'independent press' that might give voice to public opinion (Boyce 1978: 22).

However, contemporary scholarship has observed an anomaly between the Fourth Estate ideal of a press performing an important, political public function, and the realities of what the news press actually publish (Curran I in Curran and Seaton 2003: 347). The concept is in the academy, conceived variously as the "metonymic character of the dominant discourse" (Dahlgren in Dahlgren and Sparks 1991: 7); as a "contribution to the discursive formation of journalism" (Conboy 2004: 109), and even as undeservedly 'canonical' (Sparks 1991: 59). It is seen, on the one hand, to serve as a convenient political myth to lend credibility to an industry grounded in entertainment rather than news, but at the same time it represents a limitation on our understanding of the civic potential this material might have.

Zelizer (2004) champions an alternative approach vested in the cultural inquiry into journalism, and its interests “in the more recent transmutations by which journalists act as journalists, including the Internet, cybersalons, newsgroups, and newzines” (Zelizer 2004: 103). A decade on, this terminology may look rather quaint, but nevertheless, Zelizer is right in arguing that critical inquiry into modern practice yields new ways of thinking about journalism that challenge and create new meanings around the 'metonymic' 'God-terms' of journalism; facts, truth, and reality.

The Fourth Estate ideal comprises two topical components that are of direct concern to the work I present here, and the contribution this work makes towards wider academic debate. These are:

- The relationship between advertising and press freedom
- The rise of professionalism in modern journalism

The Fourth Estate, Advertising, and SEO

The emergence of the Fourth Estate occurred in the context of diminishing alternatives to the dominant commercial newspaper model which was “reinforced by the triple pressures of technology, capital and distribution” (Conboy 2004: 112). In turn, a central component underpinning the concept of independence amongst the Fourth Estate was the rise in significance of advertising. It is in the context of this commercial impetus going back to the 1820s, when advertising revenue began to dominate newspaper incomes, that my study of Search Engine Optimisation in the modern news room is situated. The rise of the popular press after 1900, of the press barons during the 1920s, and increasing concentration of ownership during the 1930s, led to a high tide in newspaper competition; and the rise of systematic marketing as a means of sustaining and growing audiences (Curran and Seaton 1991). In this reading then, SEO is merely a contemporary manifestation of this trend, in early 21st century journalism.

The early 19th century saw “the beginning of the end of the egalitarian potential of the hand-press” (Conboy 2004: 111), leading eventually to a commercially concentrated journalism during the first half of the 20th century (Boyce 1978: 36). In this trajectory, power shifted decisively away from the celebrated editors of the late 19th century; Delaney, Levy, Greenwood and Stead; towards the press barons of the early 20th century; Northcliffe, Beaverbrook and Rothermere. During this transition, editors “declined in power and prestige as the business managers came into their own. The press lost the mystique of being regarded as an estate: it was now described in down-to-earth terms as an industry” (Boyce 1978: 36). The Fourth Estate is, in this reading, a political myth used as a means of addressing the paradox of a commercial press that may hold to account commercial (as well as political) interests (Boyce 1978: 27).

The notion of the Fourth Estate as a unifying concept is critiqued by some who argue that the move away from a 'dependent' press, towards an advertising-maintained 'independent' press, merely represents “its adaptation to the conditions of bourgeois democracy” (Sparks 1991: 60). On the other hand, if we accept that there is some interconnectedness between the notions of financial viability and independence from direct party-political influence, then, it has been argued, the increasing dependence on advertising and declining dependence on political subsidy and patronage during the early 20th century, led to an independence of sorts. It has been suggested that Baldwin's denunciation of the press barons in March 1931 may be read as “a politician’s natural anger at newspaper that were no longer amenable to his wishes” (Boyce 1978: 35).

Professional standing of the journalist

The rise in status amongst interactive graphics designers of the present age, by comparison with their print forbears, represents a very modern manifestation in the much earlier emergence of journalism as a respectable profession, which is contingent to the Fourth Estate myth (Conboy 2004). The rise of the professional journalist may be traced back to the mid-Victorian period; before the 1850s, journalists were held in low esteem, but the market evened out discrepancies in this perception, with the rise of “professionalized norms of balance and objectivity, tailored towards the desires of readerships identified by an awareness of the requirements of advertisers" (Conboy 2004: 113). The Times cemented its political significance and contribution to civic debate with its independent (and critical) reporting of the Crimea War (Conboy 2004: 118), while later the Telegraph developed a more participatory form of journalism that was seen to facilitate wide-ranging debate with its readership (Conboy 2004: 121). However, these two reportorial approaches were buttressed by a range of technological and social factors, in the creation of journalism as a respectable profession, including:

- The introduction of steam printing in 1830s.
- The emergence of web rotary presses (during the 1860s), then private wires and telephone.
- The combination of emergence of news agencies (during the 1850s) and the invention of the Telegraph (during the 1870s) that influenced news discourse, and the notion of news markets.
- The emergence of communication and distribution networks; the telegraph, railways, the penny post, and WH Smiths newsagents.
- The emergence of Analysis, discussion and opinion in the evening press
- The emergence of house style under empowered and emboldened editors
- The emergence of Press clubs from the 1870s
- The emergence of reading rooms in large public buildings
- The emergence of the modern lobby system in 1885

One element missing from the above analysis is the journalists' self-perception, which is key to maintaining public perceptions; and these are in turn formed within the newsroom, between competing sub-groups of professionals (Lowrey 1999). In Lowrey's time, infographics designers served an auxiliary role, however today, with the advent of the internet, and the range of media now used by modern news consumers, the interactive graphics designer is afforded more prestige and respect in the newsroom, as my research attests to (Dick 2013).

Conclusion

In the context of monopolistic challenge to editorial freedom, my study of the Expressograph offers new knowledge to an understanding of the Fourth Estate myth. In this research I provide systematic evidence that Beaverbrook (in keeping with the other press barons of the early 20th century) used his newspapers "not as levers of power within the political parties, but as instruments of power against the political parties" (Curran II in Curran and Seaton 2006: 44).

The Fourth Estate then, is a myth; but as with all myths, it contains a kernel of truth. As my research into SEO shows, a serious press continues to put editorial agendas ahead of marketing techniques when it comes to setting the news agenda. This finding throws into relief the moral panic arising in 2009 about a threat to serious news from the 'culture of the click' (Currah 2009). In truth, the Fourth Estate ideal represents a conceptual challenge to the realities of modern news production; SEO is employed as a linguistic method, rather than as an editorial tool in the modern UK newsroom. This in turn speaks to an interpretation of the media more nuanced than the Victorian Fourth Estate ideal, in short: "it is likely that to speak of a single unified category of 'the press' or 'newspapers' or 'journalism' conceals very much more than it reveals" (Sparks 1991: 62).

I offer no new contribution to the theory of The Fourth Estate in these works per se. However, I would argue that this political myth that refuses to die continues to represent an important normative counter-narrative to much of the critical engagements with journalism practice that have tended to dominate journalism studies. In particular, it represents a continuing challenge to (or at least a problematisation of) the perception of tabloidization, and to other ideological critiques of industry, and of news values.

3.4 SITUATION WITHIN AND CONTRIBUTION TO 'TABLOIDIZATION' LITERATURE

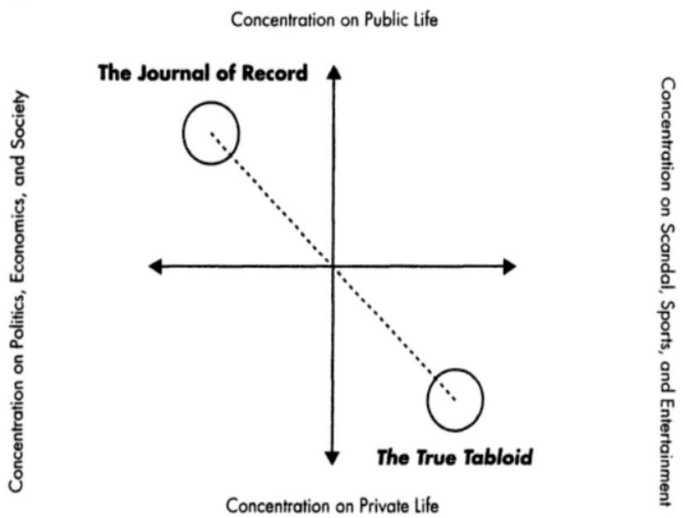
Before introducing a definition of tabloidization as verb, it is helpful first to consider the phenomenon as noun. To this end, Sparks (in Dahlgren and Sparks 1992) offers a useful definition (and wider theorization) of the tabloid. In this account, the popular press differ from the ‘serious press’ in a range of ways, they are: “different in physical size, in pictorial content, in news values, in language in readership and in price from the 'quality press” (Sparks in Dahlgren and Sparks 1992: 37). In terms of content, there are some long-term differences in structure and language, bound up with “a conception of politics which concentrates on the everyday at the expense of the historical… achieved by means of a direct appeal to personal experience. The popular conception of the personal becomes the explanatory framework within which the social order is presented as transparent” (Spark in Dahlgren and Sparks 1992: 39).

Yet while there may be some common ground amongst critics on what constitutes a tabloid newspaper, the nature of the process of 'tabloidization' (where this agreement that it is happening at all) is much more hotly contested. Based on a series of perceived problems, Sparks attempts to define tabloidization in three senses:

- “The first sense is specific to newspapers and the journalistic output of broadcasting. From this perspective, the tabloid is a form marked by two major features; it devotes relatively little attention to politics, economics and society and relatively much too much attention to the personal and private lives of people” (Sparks in Sparks and Tulloch 2000: 10)
- “The second sense involves a shift in the priorities within a given medium, away from news and information toward an emphasis on entertainment.” (Sparks in Sparks and Tulloch 2000: 10/11)
- “The final usage concerns the shifting boundaries of taste within different media forms. There, the obvious and dramatic example is Jerry Springer, whose name seems indissolubly linked to the phrase 'tabloid television'” (Sparks in Sparks and Tulloch 2000: 11)

A diagram (or more specifically a matrix) is then offered which delineates a spectrum of tabloidization; from *Journal of Record*, to *True Tabloid*:

Figure I.1 The Two Axes of Journalism



Sparks acknowledges some limitations in this model; not least the lack of space for the ‘common sense’ definition of the term (the actual physical size of the news product); and indeed all design matters; a serious limitation given the more recent rise of scholarship in the importance in the form of news (Barnhurst and Nerone 2002). But in any case, contrary to Sparks’ protestations, there is a way around this problem – the inclusion of a third dimension (or axis) to the above model would allow for a third variable (news form, or 'visual dimension' in Sparks' terms). This could span sober expression in text, image and design at the top, down to sensational expression in text, image and design at the bottom. Trying to make sense of Sparks' five kinds of newspapers typology using a three-dimensional model would be impossibly difficult using these convenient groupings; but in a sense, this in turn highlights the pre-converged nature of the concept (and the difficulty in holding tightly defining characterisations of news media as they migrate online). Barnet's (2011) contention that important public issues are often conveyed on personal terms, represents a further problem for Sparks' model – the notion that public life and private life polls at either end of a spectrum is open to challenge.

Alternatively, McLachlan and Golding (in Sparks and Tulloch 2000: 76-77) (with respect to newspapers) identified four 'indicators' in newspapers, used to define the tabloid form, namely:

- Range
- Form
- Mode
- Market structure

The research value in this approach is that it supports investigation of the phenomenon on a quantitative level, in so far as these all represent variables that may be measured. What McLachlan and Golding's underlying data shows, however, is that there is no corporeal base level of 'tabloid' on any of these measure; it is a relativistic notion.

Those sympathetic to the civic potential in tabloids have taken alternative approaches to defining the phenomenon. What, for example, is meant by Tabloid, in the context of tabloidization? Barnett (2011) offers a useful re-appraisal of Spark's (2000) definition which he summarises thus:

> "Tabloidization is, first of all, a process in which the amount and prominence of material concerned with public economic and political affairs is reduced within the media. It is, secondly, a process by which the conventions of reporting and debate make immediate individual experience the prime source of evidence and value" (Barnett 2011: 142)

In terms of this binary approach, two further questions arise which in turn form bounds to the debate around tabloidization, namely;

- Is triviality and sensation actually replacing serious news material?
- Is the treatment of our news leading to oversimplification or distortion of content just to maximise its intended audience?

On these terms Barnett offers a revised definition of the original, along the following lines:

> "Tabloidization is the progressive displacement of citizen-enhancing material with material which has no other purpose than to shock, provoke, entertain or retain viewers; and the progressive erosion of professional journalistic values in favour of televisual techniques involving sensationalism, distortion, misrepresentation and dramatization of the trivial" (Barnett 2011: 169)

This definition is a useful starting point for the exploration of the phenomenon of tabloidization, but again it overlooks the proposition that form and content in news may be considered mutually dependent, and hence non-separable entities (Barnhurst and Nerone 2002).

Fiske (in Dahlgren and Sparks 1992), an American, writing from a news culture which perceives tabloid culture as in some way alien, and lacking both in integrity and truth (Sparks 2000: 9), constructs a robust critique of the tabloid form, based on a binary division of alliances of interest; the power-bloc and the people. It is in this context, that the notion of popular news consumption must be considered:

> "The last thing that tabloid journalism produces is a believing subject. One of its most characteristic tones of voice is that of sceptical laughter which offers the pleasure of disbelief, the pleasures of not being taken in. This popular pleasure of 'seeing through' them (whoever constitutes the powerful them of the moment) is the historical result of centuries of subordination which the people have not allowed to develop into subjection." (Fiske in Dahlgren and Sparks 1992: 49)

To Fiske, sensationalism represents not merely the transgression of news norms, but the exceeding of these norms; creating "a sort of mega-normality which writes large and visibly that which is normally taken for granted and whose political effectivity depends precisely upon its status as uninspected common sense" (Fiske in Dahlgren and Sparks 1992: 53). In terms of critiquing this position, it may be said that taking the risks of tabloidization too lightly, as some may perceive to be the case with this distinctly post-modern reading, risks obscuring the fundamental importance (and truth) in our serious news.

In returning to the notion of the display elements of tabloidization, Becker offers an important contribution that sets this issue firmly at the centre of debate. The incorporation of photography into news culture, during the mid-19th century, it is argued, represented a "threat to reason, and the journalistic institution's Enlightenment heritage." (Becker in Dahlgren and Sparks 1992: 129). The rise of the sensational form of imagery we now associate with the tabloid press emerged during the 1920s in the US (Becker in Dahlgren and Sparks 1992: 132). This led to a change in public perception: "…the leading role photography was playing in the tabloids' abuse of press credibility made it increasingly difficult to see the photograph as a medium for serious news." (Becker in Dahlgren and Sparks 1992: 134)

As is clear then, 'tabloidization' is an inherently slippery concept (Turner 1999), that has both detractors and defenders; as well as those who recognise distinctions between conventional 'high' and 'low' definitions as ideologically loaded by the context from which they arise (Hartley 1996). Rejecting the term as sociologically unsound, indeed better suited to comment in newspapers, Turner nonetheless sketches out two competing discourses in the field; focussing on Franklin (1997) and Langer (1998). On one hand it is argued that tabloidization represents a 'dumbing down' of journalism's public function; a squandering of its fourth estate legacy. Paraphrasing Bourdieu (1998), Allan suggests that "If sensational news equals market success, then professional standards cannot help but be influenced by audience ratings in a detrimental way" (Allan 2010: 249). It may be countered that such hostility to the form is snobbery "grounded in a conventional and long-standing hostility to popular culture itself" (Turner 1999: 63). Tabloidization need not necessarily be a bad thing (Gripsrud in Dahlgren and Sparks 2000). But then there is a tendency in some of its defenders (or advocates) to view tabloid media as entirely benign. Failure to adequately contain the normative definitions and determinants of tabloidization has led to the misappropriation and abuse of the term, so that it has become "a miscellany of symptoms for a cultural malaise" (Turner 1999: 68).

Fears about tabloidization may be read as a theoretical anxiety concerning how some communication forms do violence to the fourth estate function of the press, or they may be bound up in concerns of the marketisation of news conceived as a product, and those steps taken in industry to ensure as wide a 'reach' for the product as possible (Rowe in Allan 2010: 359). These fears speak to similar anxieties that plague notions of 'news values' in the newsroom; for indeed, concerns about tabloidization stem from the perceived spread (and dominance) of tabloid news values in the wider news ecology. Harrison (in Allan 2010) posits that the functions of the newsroom are refracted through a duality of symbolic models (discourses) of news; namely 'informed public opinion' and 'homogeneity': that constructs journalists as either "vigilant watchdog" or "compliant lapdog" (Harrison in Allan 2010: 194).

As theories of the network find voice in the current 'global comparative turn' of journalism studies, so the tabloidization debate has arisen in critique of online news. It has been argued that the increasing choice of news across converging media represents no more than a cynical marketing ploy leading us further down the road of tabloidization; towards greater profit for the greedy corporations who make our news, at the expense of the public interest (Fenton in Allan 2010: 562). But online journalism poses a challenge to the definitions and manifestations of tabloidization long-established in the literature. Rowe (in Allan 2010) outlines some perceived manifestations of tabloidization in print media, including the 'space budget' of the principal news pages (cited in Franklin 1997). While decreasing amounts of text in print may imply 'dumbing down', online landing pages (home and index pages – those most important pages on any website) are predominantly optimised for navigation, not for reading – in keeping with long-established empirical research in online user design (Neilsen 1997). Similarly, the accommodation of user behaviour in online news provision driven by best practice in industry (Krug 2000) has wide-ranging consequences for print-era definitions of tabloidization, covering aspects of the presentation of text. If we are to problematize tabloidization in online news it is therefore necessary for us to identify elements of its manifestation online that reflect the fundamental character of the medium.

One contemporary manifestation of tabloidization in online news may be found in Currah's (2009) formulation of 'the rise of the clickstream'. Contemporaneous studies in this field (Outing 2005; MacGregor 2007) have shown that online editors exert caution over assumptions made through the use of online audiences. However, in his study, Currah found anecdotal evidence of a "growing tension between editorial values and knowledge of what will actually generate revenue" (Currah 2009: 48). This study embodies a duality that informs our understanding of SEO in the newsroom; while data may be used to improve coverage (by helping editors see more accurately what their readers' interests are), less popular (but nonetheless civically important) news stories may suffer exclusion in favour of "populist, click-friendly topics" (Currah 2009: 48). To this end, my work on SEO is presented as a theory-building launch-pad for future studies of the phenomenon. By bringing literature from the nascent (and definitively non-academic) field of SEO into journalism studies, and by generating empirical data on its manifestation, I have forged a path for future researchers to interrogate theories about the marketisation of news in a networked context, based upon the ideals and practice employed in modern newsrooms.

The tabloidization debate has some bearing on the manifestation of infographics (and by degrees interactive graphics) in news too. Summarizing the views of McChesney (2007) on tabloidization in newsprint, Rowe explains: "if there was a corresponding increase in the size of visual material (especially so-called 'impact' photographs, as well as diagrams, tables and cartoons) it might similarly be claimed that textual depth is being sacrificed in the interests of visual attraction and, perhaps, distraction" (Rowe in Allan 2010: 353). The notion that infographics (in the form of diagrams or tables) are intrinsically indicative of 'dumbing down' may be a relatively new claim amongst academics, but such opinion has long been noted as a prevailing ethos in UK newsrooms (Sullivan, 1987). Similarly, the assumption that clever people think in words and stupid people think in pictures, is predicated on the sort of hostility to popular culture noted by Turner above. This assumption is challenged by a key narrative elsewhere in journalism studies literature: Sir Harold Evans' tenure as editor of Sunday Times during the early 1970s coincides with what is widely considered to be a high-point in fourth estate British investigative journalism (Spark 1999; De Burgh 2008; Keeble 2008), and yet visual journalists working at this broadsheet newspaper (including Peter Sullivan) pioneered the infographic form in UK news.

Contribution

My contribution to the literature on tabloidization is two-fold. Firstly, I offer a critique of the suitability of what may be considered a manifestation of the form (infographics), as expressed, for example, in McLachlan and Golding's (in Dahlgren and Sparks 2000) model as *Form*: 'greater use of easily understood illustration' (Sparks 2000: 76/77). Firstly, I show that infographics emerged in regular form in UK newspapers as a means of editorial propaganda. It is important to note here a clear distinction between 'dumbing down', and the deliberate misuse of standards in infographic design. The originality in my approach is to offer a non-ideological critique of tabloidization as propaganda in infographic form.

Secondly, I show that interactive infographics are employed in modern news in order to offer a different way of understanding some of the most serious, and complex of news stories, as well as some of the most personal, and trivial. Here, I show what may traditionally be considered a manifestation of tabloidization, is capable of expressing both serious and trivial news, with little awareness of the classificatory distinctions between news 'types' often expressed in the literature. My findings around the use of 'chartjunk' suggest that, sometimes, journalists may opt for graphical forms of less numerical integrity than 'best practice' might otherwise permit. But similarly, it is clear that these decisions are made reflexively, and with the audience in mind; in keeping with Barnett's (2011) thesis.

I find that tabloidization is a pre-converged term that lacks specificity in the world of online news; it struggles to accommodate the realities of modern, networked news production, and the needs (and expectations) of networked, online audiences. My findings challenge the notion that important civic matters are ideally expressed verbally within the public sphere (Habermas 1991). On the contrary, interactive graphics today engage news audiences on serious matters by a range of means of engagement that text or verbal communications could not achieve.

2.5 SITUATION WITHIN AND CONTRIBUTION TO PROPAGANDA LITERATURE

> “The systematic propagation of political beliefs... The negative connotations associated with propaganda stem mainly from government campaigns during the First World War – particularly those of the British government...” (Franklin et al 2005: 204/205)

The work of Walter Lippmann (1922) has been cited as being central to our understanding of propaganda; and indeed, in his writing may be seen the origins of seminal communications theories including gate-keeping and agenda-setting. Informed alternatively by American pragmatism (William James) essentialism (George Santayana); British Fabianism (Graham Wallas) Marxism and Freudian psycho-analysis, Lippmann theorised propaganda as relying upon a barrier between the public and an event (Lippmann 1922: 28) that is in turn informed by a reductive, limiting mass media.

The socially damaging nature of this limiting function was further engaged with in the work of Harold Lasswell, who may be considered the first post-analytical contributor to the field. Undertaking empirical research (both qualitative and quantitative – and pioneering new methodological paradigms in communications along the way) Lasswell sought to problematize the phenomenon of propaganda, and measure its success by means of effect on the mass audience (1927). Lasswell concluded that the prevalence of propaganda sapped optimism amongst the public; confronted with a limited world view in their mass media, they give up hope of ever achieving a true understanding of political life; becoming cynical and withdrawn.

These early theorists wrote at a time when ‘the new propaganda’ had been employed to devastating effect, in America and Britain during the first world war (Miller In Bernays 2005: 11/12). Edward Bernays, drawing upon elitist notions expressed in Lippmann’s earlier work (and influenced similarly both by pre- and post-analytic theorists of the crowd) propagated the necessity of propaganda in a confusing modern world where a mass public literate but “quite innocent of original thought” (48) are ruled by ‘invisible governors” (Bernays 2005: 37); where “Propaganda is the executive arm of the invisible government” (Bernays 2005: 48). During this time, newspapers were a major source of war propaganda (Jowett and O’Donnell 2006: 105), which in turn led to a rising tide of concern, amongst progressives like Lippmann and Dewey, with respect to the dangers inherent to propaganda in the mass media, which in turn led to a wider program to challenge propaganda in “news, religion, entertainment, education and government” (Jowett and O’Donnell 2006: 101).

Drawing upon the psycho-analytical approach of Freud, Adorno perceived an ideological impetus in propaganda based upon the political aims of those employing it (Adorno 1970). Adorno drew in particular upon Freud’s Group Psychology and the Analysis of Ego (1922); and in so doing, he set out the ways in which fascist propaganda plays on the irrational mind, using the symbolism of childhood.

Jacques Ellul (1965), eschewing the psycho-analytic dominance that influenced studies hitherto, instead offered a common-sense logical approach; framing propaganda as a particular technique that is entirely necessary in modern life; needed by public and propagandist alike (serving, in a sense, as a proxy for Lippmann's educated elite). Distancing his approach from the psycho-analytical definitions and precepts that abound in earlier research that tend towards a form of reductivism, his approach is instead informed by technological determinism, as he avows: "…the study of propaganda must be conducted within the context of the technological society" (Ellul 1965: xvii). To Ellul, propaganda was a sociological phenomenon, innate, and distinct from human intent.

Following on from Ellul's notion of propaganda as something pervasive, later scholars, such as Terence Qualter (1962) have attempted to explain the success of the phenomenon according to its adaptation to the communicative process: "influencing attitudes, anticipating audience reaction, adapting to the situation and being seen, remembered, understood, and acted on are important elements of the communicative process" (Jowett and O'Donnell 2006: 6).

More recently, scholars have suggested (Combs and Nimmo 1993) that propaganda is simply the highest form of palaver; of sleight of hand (Jowett and O'Donnell 2006: 6). This approach is similar to Ellul's in so far as there is a perception that propaganda may be so pervasive, that it can be difficult to distinguish what is propaganda from what is not. In my own study, this is not necessarily the case; for it is the absence of universally agreed (or taught, or understood standards – across society at the time) that is the problem, not language per se. If basic standards are adhered to, then many of the anomalies used would immediately become apparent.

Propaganda in data form is little explored in the literature. Thomson (1999) references 'illustrative propaganda' in relation to photography, and outlines three types of propaganda; the purely rational, the quasi-rational and the purely emotional (1999: 46). It may be said that the first of these most clearly defines the nature of data propaganda: "using purely rational material the style is factual, informative or logical, and the act of persuasion lies mainly in the selection of facts favourable to the argument and the leaving out or discounting of those which are not" (Thomson 1999: 47). Such propaganda demands audience attention, which is not necessarily easy to achieve (though 'chart junk', it could be argued, may aid in this respect). However, the problem with Thomson's approach is that the data artefacts may contain elements of any combination of these classifications; from my own study, all statistical graphics require readers to interpret them; but they also use "allusions and associations" (Ibid: 47) in the form of loaded labelling; and indeed some also use "subjective ideas" (Ibid: 47).

A more sociologically rigorous approach to the phenomenon is offered by Jowett and O'Donnell (2006), that steers a path between the over-emphasis of psychological manipulation (and the inherent presumption of universality of interpretation), and the unwillingness of some theorists to define the phenomenon. Their definition is as follows:

> "Propaganda is the deliberate, systematic attempt to shape perceptions, manipulate cognitions, and direct behaviour to achieve a response that furthers the desired intent of the propagandist." (Jowett and O'Donnell 2006: 7).

These perceptions, cognitions and/or behaviour may be "perceptual, cognitive, behavioural or all three" (Jowett and O'Donnell 2006: 8).

In this analysis it is the propagandist who is the intended beneficiary of propaganda; not necessarily the audience; this, along with the presence of "elements of a deliberate intent and manipulation" (Jowett and O'Donnell 2006: 16) is what distinguishes propaganda from education. Such a detailed definition poses a challenge for my own study, which had no means of exploring the intent or the systematic plan, because no such things exist in the public domain, and because key figures involved in these events (in particular editorial and artistic figures at The Daily Express) are not even alive to account for the materials uncovered.

Jowett and O'Donnell establish the characteristics of propaganda as being; “context, sender, intent, message, channel, audience and response” (1999: 5). As with any attempt to define visual culture, defining propaganda in terms of a long list of characteristics in this way radically expands the field and as such it is important to state that the work I present here is concerned only with the first five of these characteristics; the others I hope to interrogate in future research. In their definition of propaganda, Jowett and O'Donnell shift focus away from people and their interpretations, and towards more nuanced communication processes:

Just as journalism, and indeed visual media cannot ignore the context of media convergence, so too is it helpful in setting the bounds of the study of propaganda to refuse to tie the definition to particular media. Moreover, the centrality of 'representations' here has a clear echo in the study of visual culture, namely to “the use of language and images to create meaning about the world around us” (Sturken and Cartwright 2001: 12).

Our understanding of propaganda in journalism studies is influenced in large part by Herman and Chomsky’s much debated five-filter model (1988). This approach has been critiqued as being “highly deterministic with a straightforwardly functionalist conception of ideology” (Schlesinger in Scannell et al 1992: 307), and has been unfavourably compared with other, less publicly visible studies of the phenomenon (i.e. Morrison and Tumber 1988) on account of its lack of engagement with the actual practice of journalism, and not least its lack of engagement of the body of organisational studies. Further critiques have been advanced concerning how generalizable this theory is beyond the American media landscape; around inconsistency in the author’s reasoning in terms of audience effect (Schlesinger in Scannell et al 1992: 311) and the questionable presumption on the inevitability of ‘elite consensus’ (Schlesinger in Scannell et al 1992: 312).

These criticisms have in turn been challenged by the model’s acolytes, who argue that various accusations of conspiracy theory, determinism, lack of engagement with audience effect and presumption about the monolithic nature of ruling class interests, do not stand up to scrutiny (Klaehn 2002). This discourse is framed in a telling way – the author seeks to engage “those who are academically and/or politically opposed to such an analysis” (360); those whose professional or lived experience represents a challenge to this analysis are not afforded a voice in Klaehn's critique (just as journalists are reduced to drones in the propaganda model).

In turn, it has been suggested that the “heated and highly polarized” (Corner 2003: 371) debate around this model, speaks to a bifurcation in theoretical approaches, along political economy and cultural studies lines, and not least by more recent postmodern scholarship, which “has been seen by some to undermine or displace that emphasis on the ‘brokering’ of public knowledge by the media and the relation of their output to ideas of power, truth and justice which is also the guiding imperative of Herman and Chomsky’s writings” (Corner 2003: 371). Corner goes on to outline two further contradictions concerning the model; an unresolved dissonance in the author’s assumptions about conscious and unconscious engagement in the production and consumption of news, and the nature of just how ‘brainwashed’ an audience may be considered, according to opinion polling data (Corner 2003).

Contribution

Lippmann theorised propaganda in terms of representing a barrier between the public and news events. My approach sits within this outlook; the barriers in my reading represented both by the journalist as gatekeeper in 1950s news culture, but also in terms of a much wider access (and understanding of) best practice in information design amongst the general public at the time. That said, I do not share the pre-critical notion of universal interpretation that limits Lippmann's approach.

Elsewhere, my findings accord with Ellul's thesis that propaganda is necessary in modern life, but only in so far as this concerns the selection of which data (and variables) to present in news infographic; not with respect to the (normative) means by which they are presented. Ellul's notion that it may be difficult to distinguish propaganda from non-propaganda is normatively challenged in my findings, and more generally by best practice in the field of infographic design; which though not universal, nonetheless embodies a range of replicable, valid relatively robust guidelines.

Previously, theorists found that British newspapers during World War I were a a major source of war propaganda (Jowett and O'Donnell 2006: 105), my findings bring this conclusion up-to-date with reference to a single news title (coincidentally owned and influenced by one of the key publishers active during the previous period; Lord Beaverbrook. Just as Jowett and O'Donnell (2006) offer a long, normative list of propaganda characteristics, so do I offer a normative approach based around best practice. My approach represents a nuanced means of identifying propaganda in a particular news medium, without the presumptive ideological baggage other approaches are circumscribed by (Herman and Chomsky 1988).

2.6 SITUATION WITHIN AND CONTRIBUTION TO INTERACTIVITY LITERATURE

> "A process involving the multi-directional flow of information between agents, which might include computers, the media and the audience, is seen as interactive. The non-linear qualities and technological capabilities of the world wide web have given interactivity a new dimension with almost unlimited potential." (Franklin et al 2005: 115)

Sociological inquiry into the phenomenon of interactivity, as is true of many concepts connoted with the rise of the computer, owes a debt (albeit one diminished by a certain lack of clarity) to Marshall McLuhan, and his classically positivist binary of 'hot' and 'cold' media:

> "There is a basic principle that distinguishes a hot medium like radio from a cool one like the telephone, or a hot medium like the movie from a cool one like TV. A hot medium is one that extends one single sense in "high definition." High definition is the state of being well filled with data. A photograph is, visually, "high definition." A cartoon is "low definition" simply because very little visual information is provided." (McLuhan 1964: 22)

For McLuhan, cool media encourage interaction in the form of oral dialogue; and this is deigned the apogee of interactivity in media; albeit what McLuhan himself considers worthy of conversation seems, at times idiosyncratic, rather than self-evident. McLuhan's typology has been problematized in recent times; not least by the emergence of 'home cinema'; but more generally by the contested notion of privileging a single sense over others. Yet despite these problems, the notion McLuhan sets out about 'high definition' is interesting in the context of the ongoing debate in infographic design about 'chart junk', alluded to in my work (2013). This approach offers a positivist communications theory groundwork for the minimalism of Tufte et al; the less information presented in chart form (i.e. the lower the *data/ink ratio*) the more engagement is required of the audience. More generally however, and crucially (and as I will demonstrate with all subsequent theorisations of the phenomenon) there is no accommodation in McLuhan's thesis, of the human processes behind the message (or, on McLuhan's terms, the medium).

Alternatively, Manovich (2001) instead theorises interactivity around the generation of ideas and mental images as a consequence of human computer interaction, rather than human interaction following engagement with media. This approach is a challenge to the certainty of positivism around the ways people engage with media; albeit it has been suggested that it continues the technologically determinist paradigm established by McLuhan (Gane and Beer: 2008: 93).

Manovich's scepticism about the novelty of interactivity, and his challenge to what is 'new' about new media, are key to improving our understanding of these phenomenon. The precision with which he uses the term, rightly observing that it is essentially tautologous in the context of human computer interaction (55), and his alternative ways of describing and thinking about the phenomenon also add much-needed precision to our understanding of the concept. In his view, the interactives explored during my study fall either into categories of *image-instruments*, or *visual-simulation*.

My findings accord with Manovich's misgivings around the nature of interactives that may be categorised as of a branching tree variety; the use of templates in interactives is seen to stifle creativity, or rather constrain it within the news schedule (2013). Similarly, my findings across two studies accord with his scepticism about the novelty of new media; namely that analogue calculators published as newspaper infographics during the 1950s (intended to help modern consumers calculate the correct rate of payment on higher purchase goods) (2014) may be considered to be early precursors to contemporary news interactives that allow audiences to calculate which Olympics athlete most clearly matches their own body-type (2013).

Shifting focus from technology to user, Kiousis (2002) offers a human-centric theorisation of interactivity; namely, that the phenomenon is, rather than a product of technology, must be understood in the thoughts, attitudes and reflexions of the users engaged in technologies.

This approach finds voice in other theorists, including Manuel Castells:

> "Culture was historically produced by symbolic interaction in a given space/time. With time being annihilated and space becoming a space of flows, where all symbols coexist without reference to experience, culture becomes the culture of real virtuality. It takes the form of an interactive network in the electronic hypertext, mixing everything, and voiding the meaning of any specific message out of this context, except that is for fundamental, non-communicable values external to the hypertext. So, culture is united in the hypertext but interpreted individually (in line with the 'inter-active audience' school of thought in media theory). Culture is constructed by the actor, self-produced and self-consumed. Thus, because there are few common codes, there is systemic misunderstanding." (Castells, 2000: 21)

In this reading, news interactives are integrated in a culture of *real virtuality*, in the space of flows; and interactivity is experienced in the form of an infinite range of individual experiences; though not necessarily positive (nor well-informed or understood) experiences. A sub-strand of this field literature is comprised of audience-centred interpretations of the phenomenon; concerned in turn with audience uses and gratifications (Morris and Ogan 1996) and audience ethnography (Lindlof and Schatzer 1998).

Subsequent critical and political economy readings of interactivity contend that the phenomenon must be understood both linguistically and structurally. In terms of the former, the phenomenon may be understood in the discourse of neoliberalism:

> "…the advent of digital television, according to its most enthusiastic supporters, will rapidly expand the sheer amount of information available and provide greater interactive involvement for the viewer. Digital butlers will order and sort this information as the new smart television sets 'learn' about the preferences of their viewers. This promises no more boring nights at home slumped in front of the box, but an interactive universe where we might spend the evening e-mailing a loved one in Australia, downloading material from the Washington Post or creating our own camera angles at a Derby County home game (Gilder 1992, Negroponte 1994)" (Stevenson 2000: 202/203)

In terms of the latter, interactives are merely techniques of power, used to mediate an unequal balance of power between a hegemonic state, and dis-empowered citizen; they are a Foucauldian *disciplining technology* (Barry 2001). In this reading, modern citizens must make what sense they can of the interactive tools offered to them, and in turn 'interactivity' becomes a euphemism for a limited range of 'options' presented as an alternative to genuine democratic engagement. It is hard to reconcile the pessimism of this approach with the production of news interactives, not least because they are often not presented as an alternative to, but as an addition to news reportage (Dick 2013). One further political economy reading of interactivity online (Thrift, 2005), bound up with notions that commercial media may monetise web users' engagement with online sources, is of no direct relevance to my own study, in so far as the only news organisation who recorded information about user engagement with their interactives, was a public service broadcaster, who use the information to improve future infographics, and expressly not for commercial gain (Dick 2013).

More recently, what may be describe as a cybertheory critique of interactivity has emerged; Lunenfield's (1993) differentiation between 'extractive' and 'immersive' manifestations of the phenomenon, for example, offer insight into online distinctions, but are ahistorical in their treatment of the term, and are in any case subject to the same criticisms Manovich has outlined with such clarity.

Interactivity is a major theme in online journalism studies literature (Boczkowski 2004; Bruns 2005; Deuze 2008), but in terms of interactive visual data journalism, research has been limited to two output studies (Schroeder 2004; Quandt 2009); an audit of pre-existing material on routines in the field (Giardina and Medina 2012) and an end-use-focussed analysis from a gaming perspective (Bogost et al. 2010).

Interactivity is a concept with several meanings, applications and epistemological precedents in communications studies literature. Some a priori definitions conceive interactivity in the sense of being discrete; whether bound up in a relationship between communications media and audiences, or in a gradation of engagement, an approach which lays greatest emphasis on the audience (McMillan, 2002). It has been suggested that Foucault's theories on the nexus of power and knowledge, which encourage the framing of interactivity as a "continuous process of becoming more and more interactive" (Fuery 2009: 29) may yield a more rounded understanding on the parties (and technologies) engaged therein. In turn this has led some to acknowledge the value of distancing interactivity from the technology which makes it possible (Cover, 2006). Here my work is concerned only with how perspectives of 'truth' and 'objectivity' are constructed by (specialist) practitioners in the field.

A particularly compelling (purposive) definition of interactive graphics is set out by Nichani and Rajamanickam (2003) for whom the process of interaction yields explanation. In my study I use the term 'interactives' to denote that these items exist within the regular contexts of news production, albeit I fully accept that this definition is challenged by the rise of non-journalists in this domain; such as may be found in 'Hacks and Hackers' events around the world, and as embodied in the Data Journalism Handbook (Gray et al 2013). This is another area for future research that I had no space to consider in the works presented here.

In terms of how interactivity has been approached in journalism studies, there are broadly two schools of thought. Functional studies usually take the form of content analyses whereby authors identify (or audit) interactive elements in online news. This approach has tended to group interactive graphics with those often textual online interactive media, such as reader blogs, forums or bulletin boards (Rosenberry 2005; Schultz 1999; Tankard and Ban 1998), and as such, interactive graphics tends to disappear from analysis. Indeed one construction of a normative taxonomy of interactivity in online news (Kenney et al 1999) makes no mention of interactive graphics whatsoever.

Alternatively, perceived studies are concerned with how interactivity is considered by the audience. Larsson's (2012) study, in keeping with many in this field, is concerned with manifestations of interactivity as defined by Chung's (2007) typology of interactive features, which conceives of four interactive types: human (user-to-user); medium (site content interaction); human-medium and medium-human (this last classification is how most interactive graphics may be classified). Larsson found there to be a correlation between news budgets and levels of medium-human interaction in infographics on news sites. Hujanen & Pietkikainen (2004), in seeking to address the discursive nature of the relationship between technology and culture, opt instead for an analysis of interactivity based upon three discrete discourses; the centrality of the market, the foregrounding of author and audience, and the emergence of a 'new' journalism. Both functional and perceived approaches to the study of interactivity (and the analysis of news 'interactives') are lacking in the same fundamental way: neither explores the 'why' and 'how' decisions central to the production of news infographics, and those who conceive, design and produce them, and who negotiate the presence of infographics in our news. In seeking to answer these questions, I present my work on interactives as a source for future theoretical analysis.

Conclusion

In making sense, then, of the dominant approaches to interactivity outlined in this section; what may be classified as:

7. *The structuralist* (after McLuhan, but more specifically Manovich);
8. *The symptomatic technology* (after Kiousis); and
9. *The critical* after Shultz (2000) and Barry (2001)

…Gane and Beer (2008), outline four normative theoretical approaches to the phenomenon, namely:

> "The first is a technically informed or structural vision of interactivity in which interactive potentials are built into the hardware and software of different media systems. The second defines interactivity in terms of human agency, and sees human involvement and freedom of design or use as the defining variables. Third, interactivity can be used as a concept to describe communication between users which is mediated by new media, and which gives rise to new possibilities for interpersonal communication. And finally, interactivity can be seen as a political concept that is tied to broader changes in governmentality and citizenship." (Gane and Beer 2008: 97)

While my own study (2013) draws upon a structuralist notion of interactivity (certainly in so far as audience use was not explored), and upon a new media conceptualisation (looking as it does at the potential for story telling in interactive infographics, if not concerned with novelty per se), it is nonetheless clear that, I offer a way of understanding interactivity that is not represented here. Schudson's (1997) critique of early positivist research into journalism practice, (1997); on its tendency to leave the actual substance of which news is comprised "sociologically untouched", is relevant here. In all four of these theorisations, emphasis is placed upon agents engaged in the communicative process (however conceived); whether technological, or human (be it designer, or user, in turn). But in framing the phenomenon in this way, the actual substance being relayed; the actual news stories being conveyed, are ignored, as is the process of negotiation, human engagement, and competing norms (i.e. organisational, professional) of those involved in producing this news. My unique contribution to this field is therefore to offer groundwork for an organisational studies theory of interactivity, which challenges all of the above to re-consider the message, not as medium, nor as user (nor even as a combination of the two) but as a discrete entity in its own right, informed by a complex of human negotiations, norms, and behaviours.

2.7 SITUATION WITHIN ONLINE LITERATURES ON VISUAL DATA JOURNALISM

The radical sociological and cultural studies movements that shaped journalism studies also influenced the emergence of visual studies (or visual culture or visual cultural studies) (Walker and Chaplin 1997: 35). The mediation of contemporary life through television; the ubiquity of video surveillance; and the increasing importance of visual media in modern work and leisure have (amongst other things) prompted the observation that "modern life takes place on screen" (Mirzoeff 1999: 1). Our fractured, post-modern lives, it is argued, are best understood through visual media, just as the nineteenth century is best understood through newspapers and novels (Dikovitskaya in Heywood and Sandywell 2012: 78). The rise of the image as commodity, and of the intermediation of contemporary life through heavily aestheticized marketing (Hand in

Heywood and Sandywell 2012: 516), shapes the lived experience, and this in turn requires alternative ways of interpretation and analysis to traditional methods of approaching the visual. It is in this context, after the so-called 'visual turn' that the significance of infographics to our understanding of news is situated.

In making infographics central to my research, I have sought to understand the shifting meaning and presence of the form, working back to its definitions and the contexts in which those definitions manifest themselves. In so doing I have sought to contextualise the news values at play amongst those specialists whose data visualisations shape the news (data) we consume, in the hope of better understanding what stories are selected for representation in this form, and why. Most rudimentary definitions of infographics take a formalistic approach, predicated on the compounding of the terms involved, for example:

> *Visual representations of data, information, or concepts* (Chandler and Munday 2011: 208)
>
> *A visual representation of information or data, e.g. as a chart or diagram: a good infographic is worth a thousand words.*
> *Origin: 1960s (as adjective): blend of information and graphic.* (Oxford Dictionaries 2013)

Other definitions include an additional purposive dimension:

> *Infographic n. now chiefly Journalism a visual image such as a chart or diagram used to represent information or data in an easily understandable form* (OED Online 2006)
>
> *Charts, graphs, maps, diagrams and tables whose primary function is to consolidate and display information graphically in an organized way so a viewer can readily retrieve the information and make specific and/or overall observations from it. Information graphics may be contrasted with graphics whose primary functions are artistic or for purpose of entertainment, promotion, identification, etc. Such things as engineering and architectural drawings are not included under the classification of information graphics.* (Harris 1999: 198)

It may be said that both in the creation of infographics and in journalism, the process of making information understandable to a mass audience is key to form and function (indeed this argument may represent a continuation of sub-narratives explored in the tabloidization debate above). In my research I have sought to establish common ground between the practitioner discourses that dominate the field of infographics and the theoretical (sometimes deterministic) discourses that dominate thinking in journalism studies (and media studies), an approach that seeks to challenge the ontological certainty of each discourse. This has a direct parallel with developments in the field of visual culture studies where it is argued important ideological studies “tend to be ahistorical and they also tend to ignore the role of the designer and the design process. Design historians interested in graphics, posters and typography are much more likely to pay attention to the latter issues, but also to overlook semiotic and political dimensions.” (Walker and Chaplin 1997: 44)

From an infographic artists' perspective, creative expression in newspapers serves as a function of white space, and this constraint in turn informs the news values of the form (Sullivan 1987). In this practitioner literature, which dominates discourses on the origins, application, and practice of infographic production in UK news, definitions have emerged that separate infographics into those which serve decorative and informational ends; hence 'flavour' or 'fact or information', with the use of informational, or sign systems seen as the key distinction between ‘illustrations’ and ‘graphics’ (Evans and Taylor 1997: 289). Rather than opt for a label of exclusivity, an early UK pioneer of the form (and someone particularly skilled in the production of non-statistical 'information graphics'), Peter Sullivan argued that “anything in a newspaper which is not typographic is in some way an information graphic" (Sullivan 1987: 39). Others use similar definitions to substantiate different expressions; contemporary graphic artist Nigel Holmes, for example, prefers the term 'explanation graphics' (Holmes 1994).

Some definitions are manifest in, alternatively, typological and user-defined approaches to the medium. In terms of the former, a normative list of graphical forms is presented in British Standard 7581 (1992), comprising:

- Table
- Bar graph
- Line graph
- Area graph
- Pie graph
- Isotope graph
- Scatter graph
- Histogram
- Three-dimensional graph
- Superimposed graph
- Thematic map
- Illustrated graph, and;
- Pictorial graph.

This typology is a helpful tool for systematic analysis of the form in so far as British Standards govern data visualisation for a mass (non-specialist) audience, much as newspapers do. This approach can also be useful in terms of facilitating analysis and large-scale comparison; but they do not extend to non-statistical infographics. Alternatively, Bogost et al propose a user-centric approach to classifying and theorising infographics viewing them as either explanatory, exploratory or directed (Bogost et al 2010: 42-3) thus shifting the emphasis from production (or creation) to use (or consumption).

In some manifestations of interactive infographics, the definitional emphasis lies more in the informational aspects, than the graphical. In my research I found that graphical (or rather aesthetic) elements can be less significant than the data yielded through audience interaction; for example, calculators are a popular manifestation in visual data journalism helping the reader to estimate, for example, their post-budget tax burden. The appeal in these forms is in the personalisation they offer, rather than in their visual presentation per se. But of course, this form is not exclusive to interactive journalism; calculators of one kind or another have been represented in non-interactive form from as far back as the 1950s – one example from the sample I covered during my research on ‘The Expressograph’ offers the reader a means of estimating interest on hire purchase items, in order to help the reader avoid paying too much (Daily Express July 18 1959: 4). Some infographics are non-statistical; much of Peter Sullivan's infographics for the Sunday Times are reconstructions, process diagrams or explanatory diagrams. Some infographics are maps; which can be used to convey misleading or erroneous statistical information (Tufte 1983: 20). Non-statistical maps can 'lie' (Monmonier 1996) and have been used to perpetuated propaganda in the past (King 1996: 22). These are areas of infographic design I intend to address in future research, but unfortunately there was no space for such exploration here.

Some have sought to distance technical (and hence esoteric) engineering and architectural drawings from the definition of infographics (Harris 1999: 198), and the form is now mostly associated with the field of journalism and with the process of making concepts easily understandable (OED Online 2006). That infographics are now routinely associated with journalism says much about the popular notion of the form. But it also represents a debt to one particular strand in the history of infographics that may be traced back to Otto Neurath and those in the Vienna Circle, whose ISOTYPE picture language was created as much to visually document the physical world in an efficient form, as it was intended to aid the process of accumulating and retaining knowledge amongst literacy-challenged groups in society (Burke 2009).

But just as they may serve a valued social purpose, we must not lose sight of the fact that infographics can be (and have been) used with the intent of persuasion, and so by implication with propaganda. This runs counter to the positivist ideal of data representation as a means of the accurate and efficient communication of statistical truths (Tufte 1983). Amongst the best known early examples of data visualisation, including works by William Playfair and Florence Nightingale, there are various examples of visual forms that belie the championing of a cause, rather than the objective pursuit of the truth per se (Small 1998). Given the communicative potential in the form it is perhaps surprising that there is little amongst these definitions concerning their potential for bias, and propaganda. This is something I have (through reference to Tufte) brought new knowledge to, in the research I present here.

Much has been written on the historic precedents of the modern infographic, with particular emphasis on the pioneering and excellence of William Playfair (Funkhauser 1937; Fienberg1979; Tufte, 1983; Wainer 1990; Spence 2005; Spence and Wainer 2005). These studies are concerned variously with the origins, classification and relative merit of information graphic forms from a positivist-semiological perspective; they interrogate how soundly the representations and visualisations established by Playfair (and others) convey mathematical and statistical truths. These studies have largely been ignored in journalism studies, where the literature on infographics falls broadly into three categories;

- Studies concerned with the production of visual journalism
- Content analyses of published output
- Studies of how users interact with published output

The second of these categories comprises content-analysis-style case studies of the use of infographics in American (Smith and Hajashm 1988), United Arab Emirates (Bekhit 2009) and Indian (Ghode 2012) newspapers. The third of these categories comprises research concerning; the transgression of standards and rules in newspaper infographics design (Reavy 2003); the accuracy of audience readings of infographics (Prabu 1992); the perceived duality of attractiveness and understanding in news audience comprehension of infographics (Stone and Hall, 1997); and user recall (Ramaprasad 1991; Griffin and Stevenson 1996; Pipps et al 2009), in relation to the use of 'chart-junk' (or heavily aestheticized) forms (Bateman et al 2010). Those limitations that circumscribe Galtung and Ruge's contribution to our understanding of news values (that outputs tells us nothing about journalistic process; interaction and negotiation), may be seen at play in these second and third categories of literature. Moreover, none of the research in these two categories explore how these infographics came to be created or why.

With regards the first categories of literature outlined above, the significance of the role of graphics editors and their outputs in the wider context of news production was established before the ‘global comparative turn’ in journalism studies (Pasternack and Utt 1986). Much of the rest of this field of research is the product of one American academic: Wilson Lowrey. At the time, Lowrey expressed surprise (and concern) that so little research had been undertaken into the professional routines of visual journalism (1999). Later he would apply theories originating in the sociology of work literature to the creation of textual/visual journalism, arguing from the premise that the news we receive is shaped by the conflicting norms held by competing professional sub-groups within the news room (2002).

Looking at the journalism studies field today, those prevailing post-1968 influences of; structuralist semiotics (manifest in the work of Stuart Hall), political economy (manifest in the work of Herman and Chomsky) and more recently cyber (or 'future of news') theory (manifest in the works of contemporaries Jay Rosen and Clay Shirky) has each exerted gravitational pull over the idea of what news values are, and should be. But these approaches all, by degrees, subsume the role of the individual in the process of news production; backgrounding (or eliding altogether) the practice, rituals and routines that those individuals tasked with creating our news engage in. In contrast, Lowrey's approach is predicated on the notion that “Members of each organisational subgroup share unique sets of norms, practices, and values that give meaning to their areas of work” (Lowrey 2002: 412). Lowrey starts (1999) from the structuralist premise that scholarship on the routines of news-making can help inform visual communications scholarship. Routines save time, and expense, and they shape the content of news production, both in its senses as verb or noun - as work or as output. This is as true of non-interactive visual journalism as it is of news interactives. But it also explains the news we receive through the human relations that inform what it is – an approach particularly sympathetic to a practitioner-academic perspective in journalism studies, and so an approach that sits comfortably within my approach to the field.

Some of my findings on interactive news production resonate with Lowrey's earlier work. For example, Lowrey draws our attention to the 'aesthetic routines' found in photographic journalism; the minimisation of elements in shot etc.; similar conventions are found in interactive art too. I also found that journalists do not follow audience research on their output with any great enthusiasm; and indeed, those decisions that inform the selection of stories for treatment, and the selection of interactive form to convey these stories, may be considered a function of journalistic routine and practice. Other findings in my research offer new insight, or bring Lowrey's findings up-to-date. For example, Lowrey found that news specialists may seek to educate non-specialists on design norms (Lowrey 2002: 428), whereas I found that today's interactives specialists even produce bespoke software in order for non-specialists to do their own work; obviating the need for the inculcation of design norms. Some of my findings indicate just how different the offline and online experience of visual journalists is. Lowrey found that "Visual journalists place more importance on pleasing the dominant subgroup than on pursuing accurate representation, which suggests a sort of 'service-department' mentality common to advertising and graphical firms" (Lowrey, 2003: 138). On the contrary, I found that, in the networked newsroom today's interactive professionals are "more self-confident, more creative, and less 'auxiliary' than their non-interactive predecessors" (Dick 2013: 14).

Bertin's (1967) contribution to the theory of infographics takes a structuralist semiological approach. For Bertin, the basic problem inherent to infographics centres upon the issue of efficiency that drives designers' selection of a limited set of visual variables (and combinations of visual variables) towards expressing the realities inherent in the information at hand:

> *Efficiency is defined by the following proposition: If, in order to obtain a correct and complete answer to a given question, all other things become equal, one construction requires a shorter period of perception than another construction, we can say that it is more efficient for this question.* (Bertin 1967: 9)

In keeping with the organisational studies of news production, Bertin argues that efficiency is key to understanding how infographics are produced. But the use of templates and other efficiency-driven norms may simply help to propagate the imperatives of the organisation over the professional; encouraging the shoehorning of news events into pre-existing structures due to limitations of time and editorial will, rather than allowing journalists to explore news events in their own capacity. Indeed, only one of the professionals I interviewed found (or were permitted) time and resources to undertake 'fishing expeditions' in data sets, looking for news stories and treatments at source.

Outside of journalism and communication studies literatures, the relative merit of infographics in news production is subject to two competing, mutually exclusive (but both positivist) discourses. One, which may be described as a mathematical-statistical approach, finds its voice most clearly in the works of Edward Tufte, though others including Huff (1954) and Paulos (1996) have made significant contributions to this field. Tufte's positivist-scientific approach to the infographic form finds its clearest expression in his Principles of Graphical Excellence:

- Graphical excellence is the well-designed presentation of interesting data – a matter of substance, of statistics, and of design.
- Graphical excellence consists of complex ideas communicated with clarity, precision and efficiency.
- Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
- Graphical excellence is nearly always multivariate.

- And graphical excellence requires telling the truth about the data. (Tufte 1983: 51)

Tufte's notions of 'well-designed' and 'interesting data' are an appeal to empirical processes, albeit others have long questioned the lack of evidence behind Tufte's claims, as well as the suitability of this approach to theorising infographics in news production (Prabu 1992). For Tufte graphics necessarily deal in complex ideas (even if news does not). They must explain the world around us clearly and efficiently, with an appeal to 'truth'. Lies in newspaper graphics, Tufte, maintains, are “systematic and quite predictable, nearly always exaggerating the rate of recent change” (Tufte 1983: 76). This critique dovetails with some political economy critiques of news values that criticise the undue emphasis placed on events and 'news hooks' found in much mainstream news coverage (McChesney 2000). Taking aim at Nigel Holmes, Time Magazine's designer of 'explanation graphics' during the late 1970s and 1980s (though not naming him directly), Tufte admonishes newspaper graphics designers in typically positivist terms: “If the statistics are boring, then you've got the wrong numbers” (1983: 80). Tufte decries the rise in ‘chart junk’ in a similar way to those who decry the rise of tabloidization.

3. METHODOLOGY: PARADIGM INTERPLAY

This submission comprises a series of contributory research questions (and one working hypothesis) that have emerged from a range of ontological, epistemological and methodological contexts. The work presented here may be considered to adhere collectively to a meta-theoretical position that has been defined as 'paradigm interplay' (Schultz and Hatch 1996). Representing a post-modernist approach to organisational studies, paradigm interplay maintains that research paradigms (or epistemologies) may be considered to be mutually complementary rather than mutually exclusive. In accommodating *interplay* between paradigmatic approaches, it is argued, it is possible to develop new knowledge in a sequential process (Lee 1991).

In the following sections I will set out a justification for each framework employed. I will explore the operationalisation of each approach; the advantages and disadvantages involved; the nature of evidence gathered; the procedures entailed in doing so; their respective validity and where appropriate, other particular qualities (like legitimacy). I will also evaluate how adequate each approach proved to be in terms of answering my research questions.

The structure to this approach is as follows:

Methodology employed	**Research questions answered**
Action research-as-praxis	**RQ1** What methods and sources can be employed to optimise online newsgather-ing?
	RQ2 What methods can be employed to ensure the veracity of information sourced in online newsgathering?
Ethnographic methodologies	**RQ3** What are the consequences of seamlessly un-publishing online-only news?
	RQ5 What effect is Search Engine Optim-isation having on online news?
	RQ6 What effect are interactive in-fographics having on our online news?
Experimental methodology	**RQ4** What is the optimal way to develop an online news beat?
	RH7 Infographics emerged in the UK press as a means of pushing post-war propaganda at the Daily Express.

3.1 ACTION RESEARCH-AS-PRAXIS

The open-ended nature of my research aim made it particularly well-suited to an action research approach (McNiff and Whitehead 2006). Action research emerged from the field of education, informed by frustrations concerning traditional pedagogic approaches and their inadequacy in accommodating social change (Kemmis and McTaggart 2006: 561), so too it became clear to me that it was a particularly appropriate method for online journalism, where changing relationships between journalist and audience, and new technologies, are rapidly changing what 'journalism' is.

Following my first year at the BBC, my methodological assumptions were developed in such a way that I began to regard myself as an agent working within a methodology that is (in keeping with the research problems they were designed to interrogate) ongoing, open-ended and iteratively developed. My epistemological assumptions were that the knowledge I produced (and continue to produce) while being of practical importance, was uncertain; and as I developed my praxis I evinced a commitment to the journalists I engaged in online search training, as collaborators rather than as subjects. My overarching ontological assumptions in this approach were; an acknowledgement of the value-laden nature of my research and a deep-seated commitment to online journalism practice (which is in turn both validated and legitimised by my continuing engagement in the field as a recognised expert and sought-after trainer).

During my first year at the BBC my approach to pedagogy developed relatively slowly as I accrued a better understanding of the organisational and professional pressures on journalists (and the consequent impact on their online information needs). I steadily built a corpus of knowledge around the seeking and retrieval of information towards solving journalistic research problems online. In this first year, I ran hour-long non-interactive training sessions demonstrating the use of in-house and proprietary sources using a fixed and rigid schedule (established upon the praxis of others who preceded me). The nature of data gathered during this early phase of my working life in online journalism was both primary and secondary. The primary data I gathered included:

- Informal notes taken following training (these were one to one-and-a-half hour-long non-interactive training sessions, pre-organised through my colleagues' contacts, carried out on average once every three weeks);
- Formal notes taken during my attempts at problem-solving following emailed or telephoned exchanges with journalists (one to two hours was spent doing this most working days for the first three months of my time at the BBC)
- Formal minuted meetings; agendas and action points agreed with my peers (these were infrequent; and were of varying length with no structural consistency).

The sample-rate for these methods of data collection varied widely and so the potential in this data towards resolving my research questions was naturally limited. Because these sessions were introductory in nature, and delivered primarily to new recruits to the World Service (not journalists per se), the data generated were rather too broad to be particularly useful. Access to my collaborators was of a relatively formal, non-interactive nature; I gleaned little insight into journalists' working lives and research needs, and there was no real sense of collaboration towards achieving my research goals.

On the other hand I developed a wealth of useful data by means of my attempts to resolve telephoned research queries, and both formal and information communications with my peers generated a good deal of data that contributed significantly towards improving my understanding of the problems at hand, and in establishing search strategies to challenge instances of information overload, if not towards reliable findings and conclusions. My findings from this data, and the conclusions drawn from these findings contributed towards developing basic problem-solving in practical information retrieval; and they informed case studies in search problems (and solutions) that I would go on to use throughout my professional life.

In terms of secondary data gathered during this phase, these included the following:

- Pre-established training scripts;
- Database user manuals (and training sessions organised by database publishers);
- Promotional material;
- Formal written guidance written by experienced staff members, and;
- Data gathered during conversations with colleagues (in my first few months at the BBC, I wrote up notes for an hour or so on most working days; the process became ad hoc thereafter).

All of this material was used to build a base level of understanding of the current state of online search for journalism (though, as above, the potential in this data; my findings from this data, and the conclusions drawn from these findings, were all limited in terms of resolving my top-level research problems).

My second and third years at the BBC involved developing training in a richer range of interactive and reflexive learning contexts, including practical group lab sessions, and self-negotiated one-to-one training in the newsroom. Having started my work amongst World Service and networked news staff, I developed search methods and solutions for journalists working across the network; from factual programming, through networked news, regional news, politics and current affairs and even those working in arts, entertainment and sport. During this period, and in keeping with fundamental action research principles, I moved away from perceiving a hegemonic approach in my pedagogy, towards working with journalists in my praxis; towards negotiation, reciprocity and empowerment (Lather 1986). I would (for example) seek out examples of problems experienced by learners in advance of my sessions, and I would use these examples in order to demonstrate the use of applied search theories (or sources) towards solving research problems in group training.

Between January and March 2005 I interned on BBC Newsnight. During this time I established a bond of trust with the duty editors and producers of the show such that I was asked to conduct information research and brief presenters. For example, I briefed Kirsty Wark on the murder of Robert McCartney, in the pre-transmission editorial conference prior to interview with Martin MacGuinness (broadcast on 10/03/2005). During this time the nature of the evidence I gathered, and the mode of interpretation I undertook, changed markedly, in keeping with my shift to an action research-inspired methodology. During my time at Newsnight, I kept notes on the research needs that emerged during my interactions with journalists. The data gathered here was much richer than that I had previously gathered; as it was generated within the pressures of a broadcast work schedule; I was immersed as an equal in an environment where my collaborators were increasingly open about their thoughts and approaches. The process of undertaking research to production schedules, and negotiating its use during editorial conferences, gave me an insight into the research needs of journalists that would have been impossible as an outsider. Given that the placement was only six weeks long, and that it was undertaken around my day job (three evenings per week); issues of fatigue eventually caught up with me. As such the insights I generated (towards answering my research questions) were rather limited, but I did manage to make a range of contacts during this time, and much of my later training was made possible only as a consequence of this experience.

I continued to gather primary and secondary data during my second and third years at the BBC, in the same ways outlined previously, but the frequency with which I gathered this data increased. I began training every week, sometimes (while training in the nations and regions) over every day of a working week; in group sessions, one-to-one tutorials, and over the telephone. The contexts in which I generated this data varied too; which led to a greater appreciation of the differences in approach to search within the field of journalism. During this period I began to appreciate the different search needs of those working in different journalistic roles and environments. For example, network news desk staff have very limited time, and a heavy reliance on internal contacts, so I framed their training around short-cuts, speed, with an emphasis on sources for finding visual material. On the other hand, investigative journalists are more interested in detailed, intensive, highly-verified material, and are not so tied to broadcast schedules and deadlines; and so their training required more coverage of bespoke informational resources, guidance on approaches to search (such as anonymity) and often with an emphasis on advanced search theory (Dick 2013)

With the addition of some more nuanced data sources during this period (such as an online feedback tool, used to illicit feedback on the perceived utility of my training), in addition to personal development and in-house skills training, I began to reflect upon, audit and verify the quality of my training (and research). This in turn led to primary data with significant potential for addressing particular, recurring problems within my attempts to challenge information overload in the field.

By way of example, one method I developed emerged from the (common) process of finding appropriate background information for broadcast packages. The primary news source used internally at the corporation, Nexis (later Factiva) was, I found, counter-intuitive for many journalists, who short of time, would often use free search engines and less reliable sources instead. Over these two years I developed a methodology to address this problem that I incorporated into my training. In this instance, I developed a search method incorporating the ATLEAST(), LENGTH(), and phrase search functions which, when applied to any particular research problem (and varied iteratively, accordingly) would produce around 15-20 appropriately detailed, depth articles on a particular topic, ideal for the typical 30 minutes production journalists have to *read in*.

By 2007 I had gathered sufficient background material and case studies (including theoretical and practical measures towards verifying online resources) that I was able to present a three-hour practical laboratory session to MA Journalism students at Goldsmiths University (Dick 2007). I would repeat this session at Goldsmiths for four years concurrently, and I have also taught this material at City University London, Brunel University and Westminster University (between 2007 and 2012). I continue to run sessions based loosely on this material for student and journalist alike today – most recently I ran an updated version of this session for PhD students at Oxford University (November 2013).

In early 2008 I struck upon a method for training that led to particularly rich data; a day-long participant observation in the newsroom. In these sessions (which I negotiated personally with duty editors in various networked news sections and in regional newsrooms) I would immerse myself in the rituals and practice of online journalistic work. I was then able to adapt this learning towards generating knowledge for action (Scott and Shore 1979); both reflexively in real-time, and in subsequent training and learning materials. I would often start work during the early morning shift, and would both observe and participate within the newsroom; attending editorial conferences throughout the day, offering proactive and reactive training, advice and guidance. In the process I was able to apply search theory to a wide range of real-world situations, and as my understanding of journalistic information seeking behaviour developed, I was able to develop innovative, iterative (and original) solutions to search problems. The data generated during this phase of my research was mostly primary in nature, and though infrequently sampled, it was rich (sampling was normally organised around two to three concurrent day-long shifts in the newsroom; albeit I undertook these only around once every two months).

The access I negotiated during these sessions was limited in a number of respects; only four networked newsrooms took up my offers, and even this access was largely formalised around newsroom routine. Each interaction had to be carefully negotiated (none were random), and I had to develop capacity to realise when it was appropriate to engage journalists in my training. The journalists, and the contexts in which I gathered this data, were both selective (in that they were gathered from a single, albeit large, news organisation with a public service commitment).

My findings, and the conclusions I drew from these findings became more nuanced and refined, and more practically useful during this period. For example, during my last visit to BBC Radio York, in response to a request for more news diary material, I used an in-house BBC database to identify the forthcoming anniversary of a plane crash in West Yorkshire. I then assisted with researching, sourcing and producing the idea for a radio features package on this topic later that day (24.05.08).

As a result of these findings, and the conclusions I drew, I became particularly adept at contributor finding (or case-study finding) online. An indicative list of examples generated using this means include:

- Using advanced search method in a proprietary database to help two natural history researchers working on a pilot for a documentary on sharks to discover that Lloyd Grossman was a patron of the Shark Trust, and so may be a suitable presenter.
- Using advanced methods in Google to find an academic expert on the role of political spouses for a features journalist (working on the BBC News Magazine).
- Using advanced Google and Nexis search to help a News24 journalist compile a list of addresses of all people who have been the subject of court orders concerning their Christmas lights.
- Helping a World Service editor discover an English-speaking Indian public figure who had previously worked as a farmer for a live 3-way discussion on a political crisis in India.

By 2008 I had developed my teaching in the use of online sources into a series of practical three-hour applied-theory lab sessions, covering; online search, newsgathering online, crowdsourcing and web-writing. In May 2008 I became a sessional lecturer in online journalism at Birkbeck College, University of London. I designed and taught the practical elements of an online journalism module and co-convened this module, as part of Birkbeck's postgraduate diploma in journalism program. These sessions would form the basis of my future sessional work at various London universities and training in NGOs (as well as consultancy work in industry). They would also form the basis of the professional literature that I produced (such as 'Investigative Online Search: CIJ Handbook', 2011 [II]). Eventually this work would inform the applied theory chapters in *Search: Theory and Practice in Journalism Online* (2013).

<u>Legitimacy</u>

Having spent years researching in the field, it became apparent to me that there was very little coverage of search from a professional journalistic perspective in the literature. Certainly there were many texts that proved useful (and influential) in the development of my understanding of search theory (Baylin 2005; Calishain 2004; Henninger 2008; Long 2005; Schlein et al 2000) but these were aimed at a generic researcher, with no reference to the particular research needs and working culture within journalism.

Having observed this gap in the search (and professional) literatures in late 2007, I started to use the knowledge I had acquired over the previous three years to publish my own unstructured thoughts on a public blog (slewfootsnoop.wordpress.com). This approach enabled me to engage with practitioners and academics far removed from my immediate working environment; allowing me, in the process, to establish my legitimacy in the field beyond the narrow confines of BBC culture and practice. I published several posts reflecting upon theory-in-practice, drawn from my own experiences during 2008, many of which offer original insights into the ways in which journalists perform (and can perform) their newsgathering online. These include:

2008. 'Don't stand so close to me: proximity searching the web', slewfootsnoop blog, April 21 (inspired by my work in BBC Sport):
http://slewfootsnoop.wordpress.com/2008/04/21/dont-stand-so-close-to-me-proximity-searching-the-web/

2008 'Training notes from Cambridge', slewfootsnoop blog April 25 (inspired by my work at BBC Cambridgeshire):
http://slewfootsnoop.wordpress.com/2008/04/25/training-notes-from-cambridge/

2008. 'Tips and tricks for developing current affairs ideas', slewfootsnoop blog, July 23 (inspired by my work in BBC Gloucester):
http://slewfootsnoop.wordpress.com/2008/07/23/tips-and-tricks-for-developing-current-affairs-ideas/

Partly as a consequence of this exposure (and partly due to the success of the experimental training methods I had been employing), I was asked by Russel Hayes (News and Current Affairs Lead) in mid-2008 to run a research program (with Rory Cellan-Jones) for all BBC networked news journalists on free-to-use online research tools. I ran a high profile short seminar session, wrote a detailed script, which was published on the BBC News Intranet, and co-produced three instructional videos with Russel during this project. This in turn led to increasing fluidity in my role at the corporation. Having established my expertise in the field, I would become a specialist contributor to the media; helping (amongst others) Rory Cellan-Jones with a technology blog article on search engine Cuill (Cellan-Jones 2008) and Mark Easton with a current affairs blog article on hoaxes surrounding the capture of Radovan Karadzic (Easton 2008). An idea I developed during this project, on the issue of unpublishing in UK news, would become the subject matter for a technology feature (Dick 2008), which became my first experience of professional technology journalism. As I moved from professional (then at the Centre for Investigative Journalism) to academic (at Brunel University), my interest in this issue as 'news' evolved into a systematic interrogation of the phenomenon, towards peer-reviewed academic policy offering, in the form of 'Vanishing Point: On the consequences of lost online news in the UK' (2010 [I]). This example encompasses in microcosm the trajectory of my research epistemology; from trainer of journalists, to reflexive practitioner, to expert contributor to the media, to technology features writer, and finally to research-active journalism studies academic.

This approach also sets the paradigmatic context for my monograph *Search: Theory and Practice in Journalism Online* (2013). In this work I provide a depth of coverage, in terms of detail and authenticity on the issue of journalistic search, verification and newsgathering online, unmatched by competitor critical studies (Ward 2002; Quinn and Lamble 2008; Bradshaw and Rohumaa 2011). The book also offers a breadth of coverage, in terms of context and critical perspective, unmatched by its competitor practitioner texts (Bull 2010; Luckie 2012; Gray et al 2012). It offers a range of new insights into the field of journalism studies, upon which scholars in the wider field may build in future studies.

Today I am recognised as an expert in online search in journalism. For this reason I have been invited to train investigative journalists around the world, including the following workshops:

> 'Investigative Online Search'. A day-long workshop delivered to investigative journalists working on the 101 East strand of Al Jazeera Malaysia, in Kuala Lumpur, October 23/24 2013.
>
> 'Investigative online search'. Three two-hour workshops delivered at the Open Society Institute Investigative Journalism conference, Baku, Azerbaijan, April 14-15 2010. This session was presented to investigative journalists from Ukraine, Georgia, Kazakhstan, and Azerbaijan (translated into Russian in real-time by Azerbaijani dissident Emin Milli).

Validity

In terms of criteria and standards of evaluation (and the essential validity) of the claims to new knowledge I make here, I had no codified performance metrics to rely upon. That is to say, I could not be sure exactly how successfully I had instilled (or at least sought to instil) robust, reliable and replicable methods of addressing information overload and verifying online resources. Indeed, the nearest I came to an audit process in this sense, was a combination of (qualitative and quantitative) feedback from journalists, on my research training; and a very general sense of performance review during annual appraisals.

Instead, I appealed to alternative means of evaluation and validation (McNiff and Whitehead 2006), namely:

- That the work presented here, in its totality, represents originality of mind and critical judgement
- That this work comprises material worthy of publication

The latter is self-evident, but the former requires further explanation. In terms of the establishment of criteria and standards for evaluating my ability to answer the (open-ended) problems of information overload and the problem of online verification during, I have developed (and continue to develop) an informal group of critical friends and ad hoc validation groups that have contributed towards reviewing and evaluating my contribution to new knowledge in journalistic online search; by means of online critique and engagement, and peer-review of my work. This also includes; training feedback (generated in an ad hoc sense), editorial feedback on the videos and professional materials I have created, and public feedback on my blogs. Much of this has been (and continues to be) informed by engaged journalists (including those I have engaged in training in the past) and academics in the wider field of online journalism practice. By means of this process, I have been able to verify my work according to the four validity claims (Kemmis 1998) in Action Research, namely:

- Whether the training I offered was comprehensible
- Whether those involved in my training have true understanding
- Whether this understanding is authentic
- Whether this understanding is morally right

<u>Suitability of this method towards answering RQ1 and RQ2</u>

Action research-as-praxis proved to be a particularly utilitarian method in terms of providing answers to the wide-ranging, ostensibly intractable problems addressed in RQ1 and RQ2; namely, information overload and online verification. This approach yielded a substantial body of theoretical and practical solutions iteratively developed over years. Quantitative methods may have yielded interesting data with respect to journalists' use of methods and tools in the process of online newsgathering and verification, however this thin description' would provide no sense of the human context of the problems identified; of news values emerging from operationalisation in the context of social engagement.

By approaching journalists as collaborators rather than as subjects it was possible to obtain a wider-ranging 'thick description' than would be impossible in an 'objective' study. This approach meant I could establish 'trust' better with my collaborators; and notional barriers between researcher and collaborator were minimised; an important factor in better understanding the micro-level processes involved in relatively new and untried professional routines.

This method afforded a significant degree of flexibility; I was able to gather data from a range of pedagogic methods (from one-to-one training to newsroom consultancy), informed by information sourced from various origins (professional, and organisational information); over a relatively long-term time-frame. This flexibility extended to accommodating my own shifting position with respect to this body of research; from professional trainer, to journalistic source, to freelance journalist, to researching academic. All of these factors contribute substantial, 'thick description' towards answering RQ1 and RQ2 as can be found throughout the many examples provided throughout this section.

On the other hand, the limitations of normativity (especially with respect to the key concepts 'optimisation' and 'veracity', both of which were defined in a relativistic sense) meant that the method chosen was never likely to yield information that is universally relevant, replicable and reliable (in both structural, and temporal sense of the term). The orienting decisions that informed my use of this method also represent a limiting factor on its ability to answer RQ1 and RQ2; namely working in one news organisation. Lacking direct access to the innovation and alternative practices in online newsgathering and verification going on in other newsrooms (both in the UK and around the world), it was not possible to evaluate and validate my own research findings in a wider context, and so give a more robust sense of innovation and practice across the field.

Another limitation to this method relates to the notion of collaboration; for although I tried wherever possible to reject the traditional hierarchy in professional training, it would be misleading to describe this approach as a truly egalitarian 'partnership of equals'; after all, I decided which areas to investigate, and I set the research questions.

3.2 ETHNOGRAPHIC METHODOLOGIES

The approach(s) taken in the research I present here are informed by paradigm interplay; and are bound up with post-modern sensibilities, a position that clearly distinguishes qualitative from quantitative research methods (Denzin and Lincoln 2005: 12). My praxis and all of the professional (and personal) elements have informed (by degrees) all of my subsequent research; in this process I have become a *situated researcher* (Denzin and Lincoln 2005: 21).

Throughout my research I have been concerned primarily with qualitative data. For the most part I was concerned with areas not commonly understood nor discussed in the literature. I sought 'thick description' of these phenomena (Geertz 1973) in order to contextualise the operationalization of routines and practice in online news which don't always have a pre-converged magnitude of consequence, equivalent, or precedent (as is true of unpublishing, Search Engine Optimisation, and interactive infographic Design). Most of the projects I present here were concerned with creating primary data because in each case, I was dealing with nascent, barely known (nor discussed) phenomena.

In defining my approach in these studies my scoping has been shaped by cultural, rather than holistic approaches (Rubin and Rubin 1995: 172). I have purposefully formed a relatively narrow and contextually specific focus, in order to establish key aspects of the shared culture and practice in contemporary, professional online journalism.

In this section I will set out a detailed explanation, and justification for the interviewing and observational methods I undertook in the work I present here.

3.2.1 INTERVIEWING

Interviewing after the critical turn is defined by an historic, political and contextual boundedness (Fontana and Frey 2005: 695). The interview cannot be neutral, and any pretensions towards objectivity or scientific method have long lost credence. Rather this method may be considered a methodological tool to help researchers validate findings in the wider context of their studies. Set free from being a means of pathological diagnosis, the interview may become a means of supporting, and harnessing collaboration (Fontana and Frey 2005: 696); something that speaks to the engaged, nature of my wider praxis-based research.

In terms of sampling, my choice of interviewee was circumscribed by various factors. Those working in web-native specialisms in contemporary newsrooms are relatively few in number; those prepared to provide access, or at least prepared to reply to emails requesting access) were even fewer; and the temporal limitations inherent to project-work designed to be undertaken around wider academic commitments and responsibilities, all formed pressures on the range and nature of my sampling.

Random sampling was therefore deemed inappropriate (bordering on unworkable), and so purposive homogeneous sampling was used to target particular individuals within particular professional groups, in keeping with the exclusive nature of working practice under consideration. By degrees, I also engaged in snowball sampling; conversations and meetings with some collaborators led to names and details for future, prospective interviewees; albeit these were not always taken up (due to time constraints and issues around access). Living in greater London throughout the duration of these projects, I found it logistically possible, at least, to engage professionals at a range of UK national and regional media, many of which are headquartered in the capital.

Interview format

I wanted to establish deep, detailed descriptive data; to understand why participants did what they did (in their own terms), and what they thought about these processes, in terms of wider journalistic functions. It was therefore important that I present myself not only as a sympathetic listener, but also as a skilled practitioner and consultant to industry in this field. This knowledge was offered as a means of assuring my collaborators that I was 'serious' in intent, but also that they may benefit in some way professionally, as well as personally, from engaging with my research. Central to my approach was the principle that "no one perspective can claim exclusive privilege in the representation thereof" (Angrosino 2005: 731). To this end, I sought to establish amongst those engaged a sense of partnership, and collaboration. For example, I offered interviewees a range of means of engaging (phoned, Skyped, face-to-face, and emailed interviewing), depending on their preference, and geared around their commitments. I also offered to provide interviewees with summaries of their contribution prior to publication. This is distinct both from the positivist medical origins of social-scientific interviewing, and from the normal processes involved in journalistic interviewing; but of course, my aims were very different; I sought consensus, and engagement, rather than prognosis, or confrontation.

Body language and other non-verbal signs were not considered as important as the verbal content of interviews; but on the contrary, making it as easy as possible for busy working journalists and other specialists to provide contributions, made it necessary to offer a wide range of interview-type. Secondary materials (such as internet links, files containing audience metrics and analysis, secondary references to practitioner materials and texts, for example) provided a means of authenticating, evaluating and validating findings, as did my own professional understanding of these fields.

Research role

Respondents' answers may differ relative to their perceptions of the interviewer (Rubin and Rubin 1995: 114), so it was important that I established a very clear role and function from the first point of contact with all of my potential collaborators. This allowed me to establish a sense of empathy in terms of discussing their role and working patterns in the news organisation. I used both my specialism in SEO, and my experience of working in the newsroom (and my familiarity with the language and culture of professional news journalism more widely) to put my collaborators at ease, in order to create a 'safe' environment in which we all could talk freely, without fear of consequences.

Interview typology

In the absence of a suitable literature, or pre-existing framework in which to accommodate data about the new, complex phenomena I was concerned with, it became necessary to start interview design on my own terms. Consistency was not essential in these studies; as already stated, I was not (nor am I presently) convinced of the validity of a grand theory in this field, and there were no pre-existing theories arrived at through empirical means to challenge. I did not want to presume anything, given my own knowledge of training in the field; not least because of the variation of engagement with online culture and practice that exist across media organisations, whose informing factors (political, economic, technological) vary widely. Therefore semi-structured interviewing became my preferred approach.

My first foray into the field was a *survey interview*, undertaken during the first published paper I present here (towards answering RQ3). This interview was formal, and was conducted within the interviewee's working office. I employed the 'social scientific prospector' (Holstein and Gubrium: 116) approach here;

towards establishing validity in the data generated. This approach was appropriate to the data sought (on unpublishing policy and practice), but it was in a sense unsatisfying in so far as I didn't acquire a detailed understanding of the 'hows' and 'whys' of operationalization; but rather a theoretical discussion with one or two case study examples.

In terms of answering RQ5 and RQ6, I opted instead for an *active interview* approach. These interviews were conducted in non-formal circumstances; in public spaces (including a cafe in North London on one occasion), or in communal, non-working areas within media buildings (Guardian, BBC etc.). These venues offered a means for my collaborators to speak freely, and reflexively, away from the pressures of daily work, and they served as a useful means of verifying data gathered in the workplace, during participant observation.

In these interviews I envisaged interaction as 'an ongoing interpretive accomplishment' (Holstein and Gubrium: 119), conditioned by the process of interaction. The interpretive conditions that reflected meaning during these interviews included; the research topics being pursued; the working histories (and contexts) of interviewees; and the tensions between professional and organisational norms that shaped collaborators' working practice. Throughout these interviews, I conceived of interviewees as active collaborators rather than passive subjects; according them agency in directing the shape and direction of the interview, as well as in the construction of knowledge generated through these interviews.

Two 'communicative contingencies' (Holstein and Gubrium 1997: 119) were developed in order to achieve this. The first concerned the substantive 'whats' of the interview. Here I encouraged interviewees to consider engaging with my line of questioning from the various perspectives (and according the various roles) they adopt in the newsroom, as; journalists, professionals, employees, SEOs, interaction designers, etc. The second concerned the substantive 'hows' of the interview; encouraging interviewees to consider (and re-evaluate) their interactions according to how they imagine others may react to the questions and themes discussed (i.e. journalistic peers and competitors, others at different levels of seniority at the organisations they work in, members of the audience, etc.). In taking this approach, it became possible for me to gain a clearer understanding of my collaborator's understanding of the topics at hand; and of their thoughts and feelings about tensions in the newsroom (between professional and organisational norms). This approach made it possible to broach topics (such as the incorporation of emotion into interactive design) that some interviewees were either unfamiliar with, or hadn't considered in their work before. By adopting these alternative roles and personae, it was possible to obtain a clearer sense of my collaborators thought processes, and to achieve a more rounded sense of their judgements (in juggling professional and organisational norms).

In terms of the sorts of questions asked, I set out long lists of 'main questions' (Rubin and Rubin 1995: 146) that remained consistent throughout all interviews. In initial interviews (with Paul Roach at the Guardian, and with Ciaran Hughes at Channel 4 News) I made sure all questions were asked (if not answered); and in subsequent interviews and observations, I used feedback from these initial, deeply engaged (and thematically rich) data, in order to focus in on, and make the most of the (by degrees) more limited time constraints imposed by professional time pressures. Probes were necessary in relation to questions about professional routines and practice, as well as in relation to the language and (occasionally confused and competing definitions of) work-relevant concepts. These included questions around the SEO methods attributed most significance in the newsroom. Follow-up questions (which were sometimes arrived at during the longer interviews, but more often arrived at between interviews, as I increased my knowledge and understanding of the terms and concepts used) were used to tease out the specific norms, functions and protocols in each working newsroom. It became clear that, despite a (broadly) common language (with varying degrees of specificity) that organisational norms continue to guide practice in these (emerging) professional and/or technical and esoteric modes of news work.

Validity

It may be argued that bias affects (and hence invalidates) the findings from my interviews, but then it may equally be countered that 'bias' is a danger inherent to all interviews. Because the construction of meaning involves collaboration (Garfinkel 1967), and because all agents involved in the process of interviewing contribute towards the meaning in interviews, in practice all interviews are all tempered with 'bias' in a narrow, positivist sense. Yet this in itself does not invalidate the findings of these interviews, nor does it lessen the value of the 'thick description' generated.

3.2.2 OBSERVATION

Observation, like interviewing, is no longer the exclusive domain of positivist certainty, objectivity and/or neutrality. On the contrary, it is argued that these goals may be neither possible nor desirable, and indeed much debate exists around the validity of observation of people in the natural space in which such activities are situated (Angrosino 2005: 729-730). Nonetheless, eyewitness testimony, though prone to subjectivity, is considered a particularly valuable and compelling form of verification and validation in social science and in law; and so it proved in the research I present here.

In a sense, my observations owe more to the traditions of sociological than to anthropological ethnography, in so far as I used alternative methodologies and data sources (interviewing, documentary analysis, etc.) to help validate my analysis and findings. Nevertheless, I make no claims in the direction of 'scientific' study in these observations. On the other hand, I found it necessary to actually engage with my observees as collaborators, in order to better understand what they were doing, what they thought they were doing, and how these two realities converge around theory-in-practice.

It wasn't necessary for me to engage with the communities I observed long-term as I was familiar with their working environment and with the language and some of the practices involved; but equally, I considered unobtrusiveness to be a sub-optimal way of immersing myself in these new practices. For these reasons, my observations fall somewhere between the traditional three approaches found in the literature, of; participant observation, reactive observation, and unobtrusive observation (Angrosino 2005: 732). I would therefore describe my approach as 'praxis-based observation', modelled on the notion of 'peripheral membership' (Adler and Adler 1987) of the online news communities I engaged with, in accordance with a stated need to engage with informed consent. I made a conscious effort to avoid causing collaborators more harm than necessary in my research; I could find no less harmful method than the one I chose; and the means I chose were selected to avoid undermining my collaborators (Angrosino 2005: 737). This approach differs from Simmel's conception of the stranger (1950) in so far as I engaged with my observational studies from the outset; setting out my background, my experience, and my stated objectives, which were conveyed in a way whereby contributors would benefit in some way from contributing; through access to knowledge of unknown competitors or peers; by being able to compare their own practices, and in some way understand where they stood in a relative sense.

I opted for a narrative appraisal, rather than a more normative checklist or field-guide approach, in part to accommodate inconsistencies, and professional or organisational anomalies (and peculiarities) inevitable in the operationalization of new practices and approaches (RQ4), and in part to accommodate an assessment of people of varying roles and functions in the newsroom; from production journalists to graphic designers, to interaction designers (RQ5).

<u>Analysis</u>

I chose grounded theory as an analytical framework to test the data gathered in the ethnographic methods I have discussed here; an approach informed by the same reflexive pragmatism and empiricism that informed my choice of action research and praxis as research methodologies. That is to say, my epistemological approach to this data is to recognise that it contains real knowledge about the world, but relative only to a very particular context (time, place, organisation), and that this knowledge would only find expression in the context of a very practical engagement with the world, now. I sought to develop, rather than test my research propositions in order to form a basis of knowledge that may be used to test the limitations of grand deductive theories.

This approach I found to be advantageous in two respects. Firstly it offers an alternative way of arriving at and conceptualising knowledge from deductive, 'grand theory' approaches, that are inherently reductive and often lack practical validity. Secondly, the pragmatic nature of the approach represents a bulwark against dogmatism in findings and recommendations; a dangerous notion in a field like contemporary media practice, which comprises complex phenomena.

Here I will outline the conduct of my ethnographic research, using Glaser and Strauss's (1967) theoretical framework, and terminology as a guide. In attempting to analyse research questions (RQ4 and RQ5) I undertook a process of *discovery*, in order to find out more about the operationalization of online practices and protocols in contemporary UK news production.

The online media landscape in the UK is diverse:

- Traditional print and broadcast organisations (with their own traditions, histories and cultures) compete against each other, and with exclusively new media organisations;
- These organisations are funded by a range of different means,
- They are regulated in different ways;
- They include publications with a range of (at times competing and overlapping) news values, and are concerned with reaching (at times) radically different audiences.

It therefore seemed to me that any deductive grand theory may speak past the diversity of journalism as is practised, and so consequently may flatten the validity of my findings.

I employed a process of *constant comparison* during my analysis, realised in a deliberate sequencing of ethnographic engagements, amongst my (albeit limited) choice of collaborators, that in turn allowed me to develop *theoretical sampling* in a chronological, iterative way. In order to design my data gathering, to optimise emerging themes, and to ensure these interviews would encourage depth, detail vividness and nuance (Rubin and Rubin 1995: 76), I coordinated their sequence starting with an organisation who (I knew from professional experience) had a relatively sophisticated operation within the fields I was concerned with (RQ4).

I started with the Guardian; knowing from experience that they have a very public profile in SEO; and knowing too that they take a very public and transparent approach to maintaining the public integrity of their journalism. I did not assume that they would engage fully with all of my questions, merely that it seemed like a good way to establish what a leading news organisation, with a stated commitment to 'digital first' (Sabbagh 2011), was doing. Once theories were developed here (for example, that editorial and news values guide selection of stories in online trends), it was then possible to test these propositions at organisations with less public profile in this field, and with organisations concerned with a very different audience (for example, local news).

This process was further supported by my decision to interview (and observe) across more than one sitting, and in more than one context; taking proven and dis-proven propositions back to the organisations where they were developed. So for example, I interviewed the Guardian's Paul Roach (chief Technical Lead in SEO) in the following contexts, and scenarios:

- Telephoned interview on 8 October 2009 (lasting 40 minutes)

- Face-to-face interview at Kings Place (with Chris Moran) on 20 Nov 2009 (lasting 2 hours)
- Answers to further questions submitted and completed via email during May 2010
- Face-to-face interview at Kings Place (with Chris Moran) on 18 Jun 2010 (lasting 2 hours)

My choice of subjects (RQ4, RQ5) was intended to support the testing of themes during the process of interviewing. I engaged a range of individuals; some experienced long-serving journalists who were very new to the online reporting environment, some highly technical professionals and some relatively new journalists with limited insight into working practice, but with (sometimes) a clearer sense of the potential of social and interactive technologies in journalism. I sought a range of cultures, traditions, and experiences; and in so doing I sought to tease out the diversity of experience and understanding.

The point at which I reached *saturation* in my data gathering (RQ4, and RQ5) was guided as much by the increasing resonance of my findings across the case studies studied, as it was by limitations of access, and project feasibility (organised around the time I had set aside for research in my working life as an academic).

<u>Presentation and organisation of findings</u>

Of the three types of coding commonly associated with grounded theory (Glaser and Strauss 1967), the organisations studies I present here involved an approximation of axial coding. I sought out relationships amongst the categories of data, through both inductive and deductive approaches, emphasising (where possible) causality. An example of this would be in the field of Interactive design (RQ5); where I established (through interview and observation) that the emergence of data journalism as a recognised specialism does not qualitatively change practice in this field, but rather that the strictures of deadlines and finite budgets limit potential. During observation I induced that the data journalists I was concerned with spent most of their time working to commission, rather than in data newsgathering. In formulating this proposition, I then deduced the reasons by means of follow-up interviews with the journalists I had observed.

I did not employ a rigid coding system as my research owes a debt to the interpretivist tradition that privileges the researcher's own experiences, in my case my own journalistic skill, knowledge and experience of journalistic culture.

In order to establish the weighting of themes that arose through interviews (and their presentation in analysis) I undertook a "norming" (Rubin and Rubin 84) approach to some interview questions; as much to illustrate variation, as to indicate coherence. For example, it became apparent that not all media organisations apply methods in SEO in the same way, and indeed that not all practitioners, all of whom were by degrees working primarily in online journalism, took the same approach, or had the same depth of understanding of SEO practices and principles. The following question was asked of all recipients in one study (RQ4), but only half of respondents offered a clear, unambiguous answer (some elided the question):

- What would you say are your key metrics (e.g. conversion rate, percentage referrals), and how do you feel about their ability to demonstrate ROI?

Answers to this and other questions, were therefore organised into a thematic strand concerning the lack of coherence in industry towards SEO practice as a phenomenon.

The assumption that there may likely be a significant variation in the sophistication and seriousness with which SEO (and interactive infographic) practice being undertaken across industry, was borne out in several occasions. So I proceeded to test the reasoning behind approaches used in industry (such as the use of 'kickering', which had yielded positive results at the BBC, RQ4) in subsequent interviews (and observations) to test how reflexive practitioners were in relation to the application of this approach. In the following (emailed) exchange with Paul Roach of the Guardian, it became clear that a duty of care to language, and the reader are seen to circumscribe slavish devotion to some SEO techniques; something all practitioners (by degrees) acknowledged, but which only he was able to express with such clarity:

> *MD: Because search engines prefer documents where each keyword occurs as near as possible to each other, and as far as possible in their original order, must news tend toward the most common language structures?*
>
> *PR: "This approach can actually be counter-productive for SEO. The easiest thing to do is to kicker everything, e.g. put a two or three word search phrase in front of the main headline, such as "Budget 2010:" or "World Cup:". Although this is the easiest way to ensure the targeted keyword appears in the right order in every article it means that the news headlines are homogenised and you have no nuance within headlines or between publications." (Emailed interview with Paul Roach and Chris Moran, 18.06.10).*

My approach in these studies, in keeping with established practice in grounded theory, was not purely inductive. On reading, re-reading, and testing thematic propositions as part of the data gathering process it was possible to establish a series of thematic events, which were then delineated in the form of discrete summary documents, written up to capture the key elements of each interview (in order to form a combined summary of findings). In terms of the presentation of this data, I selected a thematic approach that allowed me to describe and summarise these conversations, while linking this material back to interviewees' experiences concerning the tensions between professional and organisational norms.

In turn this data relies upon the plausibility (and in places corroboration) of facts and experiences divulged in interview with secondary sources of information, including; responses from other interviewees, my own knowledge of some of these research topics, and information already in the public domain (by, for example, analysing the headings and interactive graphics produced in news stories). The validity of the material I present here is predicated upon the narrative authenticity (Ochs and Capps 1997) of my interviewees; all of whom were working professionals in the UK media industry. The process of reading, listening to and re-reading and re-listening to the data generated during interview and observation, constitutes a series of authenticating acts that I believe to be true.

<u>Limitations</u>

The findings from my use of qualitative methods are, in keeping with findings from all qualitative methods, lacking in reliability and generalisability (Kirk and Miller 1986). The subject matter covered here; online-only unpublishing, SEO and interactive graphics, are all hostage to the fortunes of change over time; and indeed, it is by no means clear that those behaviours, attitudes and approaches uncovered, detailed and explained in my research will be easy to 'update' or compare with future studies in these fields. It has been suggested that grounded theory is problematic in a number of respects; that it can lead to an approach that is overly prescriptive; that a level of selectivity can be established which may invalidate the conclusions drawn from data in this way; that its potential to explain (rather than describe) may not be as compelling as its advocates profess (Hodkinson 2008). I acknowledge all of these limitations, and again stress that the works presented here offer a limited, modest, albeit nonetheless important contribution to knowledge in journalism studies.

<u>Ethical considerations</u>

Key components of my ethical approach included the avoidance of deception, the seeking out of permission, honesty with regards the intended use of my research, and (importantly) that my interviewees would not be in some way harmed (whether in terms of professional reputation, or future income) by my research findings (Rubin and Rubin 1995: 94). To this end, I negotiated access to one of my respondents by offering anonymity in return. This approach in turn allowed the respondent to talk freely about material that may normally be viewed as of a (potentially) commercially sensitive nature. It also allowed this respondent to explore the shortcomings of policy in their news organisation, while remaining critical and reflexive with regards to the actual merits of SEO as a policy. I maintained an honest approach to all interviewees; as I progressed through my interviews I offered material garnered from previous interviews and observations, as a means of allowing everyone to compare their own situation, their own attitudes and practices.

<u>Suitability of ethnographic methods towards answering RQ3, RQ5 and RQ6</u>

The methodologies selected towards answering RQ3, RQ5 and RQ6 yielded detailed, 'thick description'. Semi-structured interviews yielded clear insights (in terms of access to esoteric practices and terminology, and in the form of real-world case study material) arising from ethical dilemmas in the newsroom. For example, with respect to the practicalities of unpublishing online-only material (RQ3), it was possible to interrogate the subtleties and nuances of 'invisible mending', and to tease out the wider emerging ethical implications of unpublishing in the context of an increasingly permanent online news archive. Equally, the sharing of case studies where the subjects of news material had sought to have material unpublished (in the form of an anorexic contributor, who contributed to the Guardian several years ago, arguing that this copy is impacting upon their right to privacy, and means of finding employment), yielded crucial insights, validated by narrative authenticity.

Where carried out in conjunction with observation methods, I found that these interviews provided an invaluable means of corroboration. For example, it was possible to verify what had been observed in the newsroom, and importantly, to test my own assumptions about the meaning and significance of these actions; such as use of (or applied emphasis upon) particular metrics and methods in SEO, and what was operationally understood by these concepts (RQ5). This method allowed my contributors to explain practice on their own terms; using their own language, idioms, points of reference, examples, experience and emphasis. I found this to be particularly useful where esoteric languages (originating in the fields of search engine optimisation, computer programming and graphic design) arose in observation. This method proved to be especially useful for gathering data on emerging practice; (such as in SEO and news interactive design) especially where there were inconsistencies in the degree of sophistication, engagement, budget, editorial and management support for these practices across industry, as I found to be the case (RQ5 and RQ6). The effect that SEO and interactive graphics respectively have on our news, is open to interpretation; and is bounded by the factors already mentioned, as well as by more general temporal and cultural contexts. It would be impossible to track these many variables in any meaningful sense; but by situating my research around lived experience, it was possible to shed light on the working realities of these phenomenon.

On the other hand, these methods were in turn limited by a number of structuring factors, not least issues of access and time, that in turn undermine the replicability, representativeness and reliability of my findings. Interview access at only one media organisation may be seen to limit my approach to answering my research questions (RQ3). Other limitations attendant to ethnography also circumscribe these findings; as discussed above, several interviewees (and observees) openly refused to answer some of my questions (on grounds of commercial sensitivity), but equally, the extent to which my interviewees altered other answers in order to provide answers they may have assumed I was expecting, may also represent a limiting factor (RQ3, RQ5, RQ6). That said, I was able to corroborate both interview and observation findings with reference to documentary material, including; private data, in the form of page view statistics, and interactive heat maps of 'click-through' (RQ5), and public data, in the form of best practice guidance in interaction design (RQ6).

Collaboration was perhaps not as equal as could have been in this research. My use of semi-structured interviews gave me an effective 'upper hand' as interviewee; all of my research questions are context-related and in practice, power resided with me as questioner (Andrews 2003: 3). This represents something of a paradox in relation to action research more generally, albeit not one that is unique to the work presented here. The act of observation involved presence in the newsroom that may have had informed unintended consequences in terms of what was observed (RQ5, RQ6).

The findings yielded, and conclusions drawn from these methods is not replicable; none of the material I offer may be compared directly to research carried out elsewhere; a general theory of search engine optimisation, or interactive graphics practice in news, was not possible. Furthermore, the process of positioning myself as expert in search during these observations (RQ5) may have unduly influenced some of my interviewees. On the other hand, the help I gave was, on reflection, relatively limited; and represented merely qualified opinion. It has been suggested that “respondents’ answers may differ relative to their perceptions of the interviewer (Rubin and Rubin 1995: 114), so it was important that I established a very clear role and function from the first point of contact with all of my potential collaborators. Ultimately the process of corroboration, between different methods, but also including supplementary texts and documentary materials, represent a robust approach to yielding findings and conclusions in the research I present here (RQ4 and RQ5).

3.3 EXPERIMENTAL METHODOLOGIES

My attempts to answer RQ4 involved developing a normative news beat a priori, followed by a/b testing. In March 2011, two prototype beats were developed, using a range of RSS feeds with the intention of covering the real world beats of Borehamwood, in Essex, and Ilminster in Somerset. These two locations were chosen as they are urban, and rural beats respectively; so any differences between the feeds would not skew the results toward one 'type' of beat, or the other. The author used the schematic to locate a range of websites and online sources which contained RSS feeds, or which could be used to develop an RSS feed.

These feeds were incorporated into a simple Yahoo Pipes master RSS feed which feature Unique, Sort and Filter modules. Pipes was chosen expressly because of its filtering options; allowing for the removal of job advertisements and other secondary information which may crowd out the feed. Contact was made with journalists in these beats, and over the following months problems arose. Subsequent changes were made to each Pipe in June 2010, to iron out the malfunctioning feeds (Yahoo accounts are required to view these feeds):

Creating beats for each area, town or village in a given beat was streamlined due to time constraints (for example, the Ilminster beat might also include place-name references to Merriott, South Petherton and a number of other small villages, in addition to Chard, Crewkerne, Chardstock and Chard Junction). Moreover full roll-out of each beat, incorporating diary and purpose-created social network accounts was not possible; one of the journalists was unable to access Facebook in the office, at the time of writing. The findings are therefore limited to feed results. Perhaps the most serious limitation to the working model developed here, is the freeware system used to develop these beats Yahoo Pipes, is no longer being maintained.

My attempts to answer **H7** involved a combined content analysis and structural semiotic analysis, a method which was developed iteratively. In order to obtain a reliable sample for this study, I searched The UK Press Online commercial newspaper database, which contains a near-complete collection of archives for The Daily Express from 1900 until the present day. Several generic keywords associated with news infographics were used in the first instance, to obtain an initial sample. These included; an analysis of, analysis, axis, calculated, calculation, diagram, explains, guide, map, number of, our analysis, overview, plan, show, shows, shows that, and working. The infographics and associated stories found using these terms were then analysed for more suitable terminology. The term 'expressograph' was found as were the bylines of several infographic artists (Michael Rand, John Bodle and Derek Dale), and these terms were used to increase the sample. In order to ensure the sample was truly representative of the period chosen, key time-frames established in the initial analysis were re-visited by means of browsing. For example, all published output between August 1956 and April 1958 was manually browsed in order to test an initial hypothesis that no infographics supporting the narrative 'Britain is still a global power' were published during that period.

To establish a representative coding framework for the sample, random examples were analysed for basic, non-optional categories of information. These included: year, date, page number, topic (and sub-topic), type (of infographic), caption (or headline) and summary. As the topicality of news stories can be a slippery concept (Saab 1990), it was decided that an additional, optional narrative 'summary' category was desirable. The classification system used for the infographics found in the sample is based upon BSI 7581:1992 (for statistical infographics). It comprises: tables, bar graphs, line graphs, area graphs, pie graphs, Isotype graphs, scatter graphs, histograms, three-dimensional graphs, superimposed graphs, thematic maps, illustrated graphs, and pictorial graphs.

A number of further optional categories were developed, in order to facilitate analysis of the sample. The recurring narratives found in this study were established by means of a critical reading of histories of The Daily Express (Smith et al 1975; Greenslade 2004), biographies of Lord Beaverbrook (Taylor 1972; Chisholm and Davie 1992), and social, economic and political histories of the 1950s (Montgomery 1965; Peden 1985; Messinger 1992; Pearce and Stuart 1992; Shaw 1996). Categories of numerical bias and semiotic bias were informed by best practice in infographic design and data visualisation. A framework of common bad practice elements (which may in turn be used as a means of delivering misleading propaganda) was sourced from Tufte (1983); Tukey (1990); Holmes (1991); and The BSI (1992).

It is acknowledged that the process of digitising these documents using Optical Character Recognition technology can fail to identify words correctly, and so may impede the full reach of this research, . It is also acknowledged that because some infographics stand alone, and some are neither bylined nor adequately key-worded, that this may also impede the range of the sample identified. Non-statistical graphics (including non-statistical maps) have not been included in this study, as it may be impossible to construct a detailed, normative framework with which to analyse these forms. It is acknowledged that some of the 'best practice' texts identified (such as Tufte 1983) lack empirical basis (David 1992), and indeed that contemporary empirical research shows that, contrary to Tufte's criticisms, people prefer (Inbar et al 2007), are better able to interpret and recall data associated with (Bateman et al 2010) and in any case are not hampered from interpreting (Spence 1990; Blasio and Bisantz 2002; Kulla-Mader 2007) data presented in ‘chartjunk’ format, a style commonly associated with newspaper infographics. It is further acknowledged that some of the models of excellence used are predicated on non-complimentary theoretical approaches. For example, Tukey's identification of impact as central to best practice presents a counterpoint to the Bauhausian minimalism of Tufte's data:pixel ratio theory (Wainer 1990: 341). In short, excellence in data visualisation is circumscribed by the limitations of positivism.

<u>Suitability of experimental methods towards answering research questions RQ4, and RH7</u>

The experimental approach I took (RQ4) accommodated both relativistic concepts (for example, the 'optimal' nature of an online news beat), alongside a means of hypothesis testing (in the form of an a/b test, involving the development of news beat case studies in the form of rural and urban scenarios). This in turn allowed for a utilitarian means of testing my (a priori) normative scheme, leading to iterative improvement a posteriori. It meant that my answers, and the conclusions drawn from my answers, were replicable within cultural and temporal contexts.

Elsewhere, my use of experimental methodology yielded innovative findings and robust conclusions, particularly where extant (for example ideological, structuring) methodologies might otherwise have failed to accommodate the micro-level subtleties of propagandistic method (RH7). On the other hand, in both cases the piecemeal nature of these methods, it could be argued, mean that this research represents an unhappy compromise between truly representative, replicable and reliable conclusions, and a 'thick description' that only goes some way towards dealing with what are otherwise contested, fluid and relativistic concepts, for example propaganda (RH7).

4 FINDINGS, CONCLUSIONS AND CONTRIBUTION OF EACH PIECE OF WORK

4.1 2010. 'Vanishing Point: On the consequences of lost online news in the UK.' *Ethical Space*, 8(1-2): 14-19.

Findings

- Media organisations contribute to public debate and policy in their online-only news, but this influence is not being routinely monitored, nor stored independently for future historians and researchers.
- As a consequence of this, there is a growing body of references, news stories, blogs and multimedia outputs that have disappeared in UK online-only news, the only traces of which remain are maintained by concerned (or motivated) citizens (but for how long?)
- Sometimes complex ethical dilemmas abound in decisions about whether or not online news editors should apply 'invisible mending' (a working manifestation of unpublishing) in online-only news copy.
- The field of corrections policy in online news is opening up a new ethical space in journalism; a consequence of the increasing sense of permanence in online news archives.

Conclusions

- The lack of routine monitoring may be addressed in the formulation of an industry-wide policy in order to balance concerns about the problem of our vanishing history, whilst also accommodating wider ethical issues (and duties of care) media organisations have in relation to unpublishing online-only news.
- A corrections policies in online news media may be formulated into a public Corrections Charter. This charter would acknowledge the rights and responsibilities of publishers and consumers of news, including that:
    - Readers (and future historians) require transparency
    - Newspaper authority is predicated on openness
    - Not everything can (or should) be saved
    - In some circumstances, un-publishing may be justified

  Building upon established excellence in the field, the blueprint for such a code should comprise the following:

    - All non-legal online-only corrections should be annotated
    - Invisible mending should be used only as a last resort
    - All legal removals in relation to online copy should be shared with the British Library
    - Stand-alone corrections should be hyper-linked

Contribution

My contribution to the field in this research is to situate concerns arising in the fields of information science within existing theory and practice in news corrections policy. In the first instance, I have raised the seriousness of the problem of disappearing online-only news; but I have also brought to light the complex ethical dilemmas attendant to unpublishing online-only content. I provide a simple but unifying policy framework that may be used to accommodate the various existing regimes (both in print and broadcast). The legacy of this research will be that it informs future literature in journalism ethics, and inter alia, that it will ensure that the issue of unpublishing online-only content is taken seriously in our increasingly networked (sometimes) permanent news archives.

4.2 2011. 'Search Engine Optimisation in UK News Production.' *Journalism Practice*, 5(4): 462-477.

Findings

- Search engine optimisation is a new and increasingly visible influence in UK online news culture.
- Its practice is informed by varying degrees of sophistication, understanding, budgetary and editorial commitment, and professional engagement, across the UK news industry.
- SEO practice is circumscribed by existing professional norms and organisational routines.
- Some stories are harder to optimise for search than others; especially where they are 'angled' around a concept with various common synonyms.
- SEO creates serious dilemmas for news organisations; in terms of balancing audience expectation and search behaviours against established style conventions, and ethics surrounding certain word-choices.
- SEO is only found to influence editorial decision-making directly in one of the four news organisations studied; but even this is of little value in terms of day-to-day traffic
- SEO has the power to 'stretch' the news agenda temporally; news organisations can be drawn into a battle of forward planning; so the notion of a 'fair fight' on a particular news day becomes less plausible.

Conclusions

- The novelty and inconsistency of practice in this field mean that it is too early to say whether SEO is significantly changing news practice (or news values) in UK news.
- Inconsistencies in practice also indicate that SEO isn't as important (or isn't seen to be as important) to all news organisations uniformly.
- The continued privileging of human-centric decision-making in editorial, a key component in the 'fourth estate' ideal, represents a challenge to the notion that the 'culture of the clickstream' (Curragh 2009) threatens news values.
- Nevertheless, the pressure that SEO may exert over news style, and news values, (on behalf not of readers, or news organisations, but according to the algorithms of third party news aggregators) requires constant revision and re-appraisal.
- SEO has the potential to extend the competitive 'who got their first' spirit in 'news' beyond merely breaking news, towards 'diary', anniversary and current affairs outputs.
- It is entirely feasible that stories that are 'hard to find' may become non-privileged where SEO becomes the prevailing influence on the way online news is sourced, produced and disseminated.

Contribution

This research contributes directly towards a call for further study from within the extant literature; in terms of how the converged nature of the modern newsroom (Harcup and O'neill 2001) and the increasing availability of metrics in converged newsrooms (Wahl-Jorgensen and Hanitzsch 2009: 171) may influence news values.

In the first instance, I present a challenge to concerns arising about a threat to serious news from the 'culture of the clickstream' (Currah 2009). Currah found anecdotal evidence of a "growing tension between editorial values and knowledge of what will actually generate revenue" (Currah 2009: 48). This approach embodies a duality that informs our understanding of SEO in the newsroom; while data may be used to improve coverage (by helping editors see more accurately what their readers' interests are), less popular (but nonetheless important) news stories may suffer exclusion in favour of "populist, click-friendly topics" (Currah 2009: 48). This argument, has clear echoes in past, critical discourses of the phenomenon of 'tabloidization' (Sparks in Dahlgren and Sparks 1992; Franklin 1997; McLachlan and Golding in Sparks and Tulloch 2000). My findings challenge this - instead it is clear that the mythical 'fourth estate' ideal of editorial independence prevails, at least in relation to how journalists perceive they are influenced by SEO. SEO is employed as a linguistic production method, rather than as an editorial tool in the modern UK newsroom. But in turn, this speaks to an interpretation of the media rather more nuanced than the Victorian 'four estate' ideal, in short: “it is likely that to speak of a single unified category of ‘the press’ or ‘newspapers’ or ‘journalism’ conceals very much more than it reveals” (Sparks 1991: 62).

Secondly, my research offers a grounded framework upon which future theories may be tested. Indeed, it has already informed studies formulated around Field Theory (Tandoc 2013; 2014). This is significant in that news values are today being subjected to a range of alternative theoretical and methodological critiques, such as Critical Discourse Analysis (Bednarek and Caple 2014). The emergence of a future semantic web will in turn require future organisational, empirical research to build on the findings I present here.

4.3 2012. ‘The Rebirth of the Beat: A Hyperlocal News-gathering model.’ Journalism Practice 6(5-6): 754-765.

Findings

- It is possible to develop a model for cultivating a robust online news beat using a pre-existing structuring schematic in combination with a range of freeware internet resources.
- The model I present was iteratively developed into two local news beat scenarios, and was tested using two real-world case study news 'beats' (one rural, one urban).
- This model was capable of identifying stories deemed worthy of being published as local news.
- It also has the potential to help journalists identify news stories which may otherwise have been missed, where journalists may otherwise rely solely on their own, 'real world' offline sources.
- However, the success of such a schematic is is in turn informed by the potential of such an approach to assist in the re-configuration of relations between journalist and source in a truly representative, 'multi-perspectival' local networked 'beat'.

Conclusion

- I provide a model that may be used to develop an online local beat that is capable not only of identifying news material deemed worthy of publishing, but also that also this model is capable of finding news stories that might otherwise be missed.
- It is also possible using this model to develop a more democratic, and reciprocal 'beat' than the traditional beat, which is seen to unduly privilege administrative and official voices and perspectives (Foreman, 2010).
- It should be noted that the transient nature of material on the web may require that subsequent to setting up the feed, further maintenance is necessary; which can be time-consuming, and potentially frustrating.

Contribution

In this research I sought to challenge the decline of the news beat (brought about by changes in the political economy of news production) by systematically mapping out hyperlocal online communities and contacts, toward helping local journalism (and journalists) re-engage with audiences. My approach encompasses theoretical issues around pluralism in the range of contributions to and content found within local journalism, while advancing a systematic approach to establishing transparency in communication, and formalising news values as an act. Gans (1979) stressed the need for 'multi-perspectival' news; a call that has in turn informed contemporary critique of converged news-makers as gatewatchers (Bruns 2006). My own contribution to knowledge with respect to this ideal, is to provide a detailed, normative account of how a local UK journalist may ensure their online beat is as pluralistic (or 'multiperspectival'), as possible; leaving behind the limited, bureaucratic bounds (Fishman 1980) of the pre-converged news beat.

4.4 2013. Search: Theory and Practice in Journalism Online. London: Palgrave MacMillan.

Findings

- I offer a critical context for understanding academic studies of journalists' use of the internet as a news-gathering resource
- I offer a nuanced iteration of search theory in the field of journalism practice, developed over several years of practice in the field (conducted in various professional contexts)
- I offer a series of practical and theoretical solutions to two seemingly intractable, major research problems in online journalism; information overload and the problem of verifying online information.
- I interrogate ethical debates that are either new to, or have special resonance in newsgathering online, including; the blurring of public/private space; the ethics of news-gathering in social networks, and the use of subterfuge online (all from a de-ontological ethical perspective; contextualised in a discourse of the 'fourth estate' ideal).

Conclusions

With respect to new or especially resonant ethical debates concerning newsgathering online, my conclusions are two-fold. Firstly, in terms of how UK journalists research their stories online, it is clear that there is an ethical (and regulatory) blind spot. Neither of the dominant professional codes (at the time of writing); the Press Complaints Commission (PCC) Editor’s Code of Practice, and the National Union of Journalists (NUJ) code, mention 'social media', 'search' or 'internet'. In the ethical (if not the legal) purview of journalism, there is an at least theoretical distinction between public and private space, and so what constitutes 'fair game' in terms of reporting (as embodied in regulations with respect to the use of long lens cameras). However research shows that the public has a paradoxical construction of privacy and publicness when it comes to engaging in online communities (Barnes 2006); and so it is clear that the online environment constitutes a new departure in this branch of ethics, if not how newspaper regulation as practised.

Secondly, with respect to the use of fake or misleading social network accounts in the pursuit of news, it is clear that an ethical crisis exists. It is suggested that use of this approach may only be mitigated by means of a public interest defence with an emphasis on 'important' information, the impact of which upon society is significant (Friend and Singer 2007: 81). On the ethics of journalists sourcing news and contributors by means of social networks, I suggest a consideration journalists should use to determine whether or not to even access a social network is to consider the nature and use of the site, and importantly the demographics of its users.

Contribution

My contribution to the literature in this research arises as a consequence of re-situating various theories from information science and the sociology of search within the field of journalism studies, and providing new theorisations and iterations of these approaches within this particular context. As such, this research occupies a unique position, and contributes a corpus of original knowledge situated at the nexus of three strands of extant literature; the first strand concerning search theory and practice (Baylin 2005; Calishain 2004; Henninger 2008; Long 2005; Schlein et al 2000); the second concerning critical studies of online journalism (Ward 2002; Quinn and Lamble 2008; Bradshaw and Rohumaa 2011); and the third concerning non-critical online journalism practice (Bull 2010; Luckie 2012). From this vantage point, I offer new insights in the field of journalism studies, upon which scholars in the wider field may build in future studies.

More specifically, my contribution to theory is situated within theory-as-practice, validated by my several years' accumulated professional collaboration, and expressed in the form of sociologically 'thick description'. In this research I offer unique, and nuanced answers to the seemingly intractable problems of 'information overload', and online verification in online journalism. Sourcing primary data from a range of methodological approaches (from one-to-one training, to newsroom consultancy); and secondary data from various origins (professional, and organisational information); the data I analysed in this research was gathered over a relatively long-term time-frame; and from a range of occupational positions (from professional trainer, to journalistic source, to freelance journalist, to researching academic). Consequently this work expresses a unique and nuanced contribution to the literature, expressed in the form of 'thick description' of search in theory and practice in journalism today.

4.5 2014. 'Interactive Infographics and News Values.' Digital Journalism, 2(4), 490-506.

Findings

- Interactive infographics are created and iteratively subjected to various competing professional discourses (as well as collective, organisational discourses)
- They appear in our news in stand-alone form, in addition to being adjuncts to news stories developed in a range of subject areas
- In the absence of universally accepted, objective standards, news infographics occupy a contested field; informed alternatively by the 'fourth estate' ideal (of the objective reporting and exposition of news data), and 'tabloidization' (expressed in the context of debates about 'chartjunk').
- Some news stories are considered better suited to rendering in online interactive form than others, based upon professional ritual and culture within the networked newsroom
- The creation of interactive infographics is a function of both budgetary constraint and audience expectation; 'big data' does not drive news work in these fields.
- Interactivity is largely informed in this field by the audience as passive end user, (whose actions aren't necessarily tracked for future use) rather than in a genuinely egalitarian balance of interaction; albeit some news organisations undertake ethnographic research in their practice.

Conclusions

Because interactive graphics occupy a valued position in today's networked newsroom, quite removed from the “service department” (Lowrey 2003, 138) function of their pre-converged predecessors, it is apparent that they will contribute increasingly to the form of networked, converged news. This in turn merits serious investigation and engagement in the literature; towards which the research I offer here represents merely a grounded introduction.

While adherence to conventional journalistic values prevail in this new field, the way in which work is undertaken in this field, concerning attitudes to deadlines, and the questioning of conventions and reasoning around coverage or treatment of news stories, arises as a function of alternative professional approaches. This represents a primary tension in the 'quality' spectrum, between the 'fourth estate' ideal, and 'tabloidization' (here manifest in a wide-ranging appeal to a wide-ranging, mass audience online).

Inconsistencies and variation with respect to perceptions of best practice, as embodied in the wider debate concerning ‘chartjunk’ represents a secondary tension. Practitioners see themselves as working somewhere between the minimalism of Tufte (1983) and the experimentalism of McCandless (XXXX), albeit in practice, these ideals are informed just as much by organisational norms and the demands of the news schedule. Interactives are found to embody the ideal of the 'fourth estate' just as they are a manifestation of 'tabloidization'; but this study found no clear tendency towards one side or the other; merely a recognition that a tension exists.

Various aspects of practice in this field represent a continuation of journalistic norms, within the structural realities of news production; especially in relation to budgetary constraints; mitigation of risk and audience engagement. As budgetary constraint circumscribes practice and potential in this field, my findings challenge the utopian notion that sees data journalism (which is in turn intimately interconnected with the development of interactive graphics in the news) as yielding more exploratory potential than traditional forms and manifestations of journalistic practice. Risk aversion, embodied in the use of templates; and expressed in concerns about audience complaint, and in the use of detailed, prescriptive audience research, dampens the experimental potential in this form of news. Moreover, the limited definition of interactivity commonly used in this field, limits the potential for collaboration between journalist and audience as equals (albeit some experts within the 'former audience' are increasingly engaged with as partners). Ultimately the power hegemony of the journalist over the mass media audience prevails.

Interactive infographics are not evenly distributed amongst news topics, or between the teams of journalists that produce news in these fields. They are considered more suited to some news stories than to others, on account of narrative type and the dynamics involved. This represents an intriguing proposition in relation to future news values. Is it possible that in the future, stories that are more difficult to express in interactive form may be de-privileged in favour of stories that may engage audience in a more direct way?

Contribution

As is true of my study into SEO, the research discussed here also contributes directly towards calls for an interrogation of news values in the converged modern newsroom (Harcup and O'neill 2001). In this study I revisit themes raised in classic organisational studies (Tunstall 1971; Tuchman 1978; Gans 1979), contextualised in a later body of organisational studies concerning the production of information graphics, in the newsroom (Lowrey, 1999; 2002). Unlike their pre-networked news predecessors, interactive journalism is today undertaken by specialists who are increasingly valued, and empowered in the newsroom.

As to whether infographics (as illustrations) may represent a manifestation of 'dumbing down' (McChesney 2007) (which may stand as a proxy for 'tabloidization'), my research offers a challenge to this notion. For example, while the tendency towards using circles (at the expense of more traditional, formal graphical forms) might otherwise represent a manifestation of tabloidization (in the form of 'chart junk'), in reality this approach is capable of expressing both serious and trivial news (albeit at the expense of absolute numerical fidelity), with little awareness of the classificatory distinctions between news 'types' often expressed in the literature.

My research compounds the notion that 'tabloidization', as a pre-converged term, somewhat lacks definitional specificity in the world of online news; it struggles to accommodate the realities of modern, networked news production, and the needs (and expectations) of networked, online audiences.

While this study draws upon a structuralist notion of interactivity (certainly in so far as audience use was not explored), and upon a new media conceptualisation (looking as it does at the potential for story telling in interactive infographics, if not concerned with novelty per se), I nonetheless offer groundwork for an organisational studies theory of interactivity, that challenges the three extant branches of literature on interactivity, namely:

- The structuralist (after McLuhan, but more specifically Manovich);
- The symptomatic technology (after Kiousis); and
- The critical after Shultz (2000) and Barry (2001)

The challenge I present here is to re-consider the message, not as medium, nor as user (nor even as a combination of the two) but as a discrete entity in its own right, informed by a complex of human negotiations, norms, and behaviours.

This study offers much-needed empirical data towards informing the discourses that dominate and define our understanding of contemporary online journalism; both the optimistic and the pessimistic. In its theory-building approach it is presented very much as an evidence base for future research and theoretical development in journalism studies.

4.6 2015. 'Just Fancy That: An analysis of infographic propaganda in The Daily Express, 1956-1959.' Journalism Studies, 16(2), 152-174.

Findings

- Over ninety-nine per cent of the statistical infographics found in this study may be attributed to a small handful of news narratives (or values), all of which coincide with Lord Beaverbrook's own personal political philosophy.
- Eight per cent of these infographics accompanied opinion pieces.
- The newspaper that established infographics as a regular feature in UK news, The Daily Express, employed a wide range of bad practices in their visual representation of data
- These bad practice methods are too wide-ranging, and used in too consistent a manner, to have occurred by accident or by random error.

Conclusions

- Infographics emerged as a regular feature in UK newspapers as a form of propaganda, rather than as a means of increasing public engagement with public affairs (the 'Fourth Estate' ideal)
- The clear connection between infographics and opinion indicates that those responsible for the production of the newspaper recognised the power of information design to persuade as well as to inform.
- The infographic work of Michael Rand, who would become a pioneering figure in newspaper design as Art Director at The Sunday Times, contained many of these examples of bad practice; a re-evaluation of his legacy is clearly required.
- Infographics in UK news were borne of a lavishly-financed news organisation whose key decision-makers were deeply concerned with the impact of the visual in news.

Contribution

In this study I offer a new approach to analysing propaganda in visual data journalism, and contribute a new way of thinking about news values (as defined in a structuralist sense); in the design of visual statistics. I also make contributions towards wider debates around tabloidization (as embodied in the form of 'chart junk') and propaganda, in the literature.

In terms of the former, I offer a new critique of the suitability of infographics as an embodiment of tabloidization, as expressed, for example, in McLachlan and Golding's (2000) model which describes 'form' as 'greater use of easily understood illustration' (Sparks 2000: 76/77). It is important to note here a clear distinction between tabloidization as 'dumbing down', and the misuse of standards in infographic design; my contribution is concerned solely with the latter. My research offers an insight into The Daily Express' embrace of 'chart junk' during the mid-1950s, which in turn may be considered to represent the emergence of 'mass culture' (an aesthetically appealing, but nonetheless debased rendering of 'popular culture'). However, it would be false to presume that infographics (whether in 'chart junk' form or otherwise) are incapable of contributing towards the 'Fourth Estate' ideal. On the contrary, it may be argued that Fortune Magazine whose experimentation with (and excellence in) infographic design long pre-dates that of the Daily Express, embodies this legacy.

In terms of the latter, scholars have suggested (Coombs and Nimmo 1993) that propaganda is simply the highest form of palaver; of sleight of hand (Jowett and O'Donnell 2006: 6). This approach is similar to Ellul's (1949) in so far as there is a perception that propaganda may be so pervasive, that it can be difficult to distinguish what is propaganda from what is not. In my own study, I demonstrate that this is not necessarily the case; for it is the absence of universally agreed (or taught, or understood standards – across society at the time) that is the problem, not language per se. If basic standards are adhered to, then many of the anomalies used would immediately become apparent. The originality in my approach is therefore to offer a non-ideological critique of tabloidization as propaganda in infographic form.

5. REFERENCES

Adorno, T. 1970. 'Freudian theory and the pattern of fascist propaganda', in Psyche: Zeitschrift für Psychoanalyse und ihre Anwendungen, Vol 24(7): 486-509.

Allan, Stuart (Ed). 2005. Journalism: Critical Issues. London: McGraw-Hill International.

Allan, Stuart. 2006. Online News: Journalism And The Internet. London: McGraw-Hill International.

Allan, Stuart. 2010. 'Introduction: Re-crafting News and Journalism.' In S. Allan (Ed) The Routledge Companion to News and Journalism. London and New York: Routledge: xxiii-xliv.

Allern, Sigurd. 2002. 'Journalistic and commercial news values: News organizations as patrons of an instution and market actors', Nordcom Review, 1(2): 137-152.

Andrews, Richard. 2003. Research questions. Bloomsbury Publishing.

Angrosino, Michael V. 2005. 'Recontextualizing observation: Ethnography, pedagogy, and the prospects for a progressive political agenda.' The Sage handbook of qualitative research 3: 729-745.

Barnes, Susan. 2006. 'A privacy paradox: Social networking in the United States.' First Monday, 11(9).

Barnett, S. 2011. The rise and fall of television journalism: just wires and lights in a box? A&C Black.

Barnhurst, K. and Nerone, J. 2002. The form of news: A history. Guilford Press.

Barry, A. 2001. Political machines: Governing a technological society. London: Athlone Press.

Bateman, Scott; Regan L. Mandryk; Carl Gutwin; Aaron Genest; David McDine and Christopher Brooks. 2010. 'Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts', CHI 2010 Conference, April 10–15, Atlanta, Georgia, USA.

Bates, Marcia J. 2002. 'Toward an integrated model of information seeking and searching.' The New Review of Information Behaviour Research, 3: 1-15.

Bawden, David; Holtham, Clive and Nigel Courtney. 1999. 'Perspectives on information overload.' Aslib Proceedings, 51(8): 249-255.

Baylin, Edward Norman, and Judith Gill. 2005. Effective Internet Search: E-Searching Made Easy!. Baylin Systems, Inc.

Becker, K. 1992. 'Photojouranlism and the Tabloid Press' in Dahlgren, P. and Sparks, C. (Eds), Journalism and popular culture. Sage: 130-153.

Beckett, Charlie. 2008. SuperMedia – Saving Journalism so it can Save the World. London: WileyBlackwell.

Bednarek, Monika, and Helen Caple. 'Why do news values matter? Towards a new methodological framework for analysing news discourse in Critical Discourse Analysis and beyond.' Discourse & Society (2014): 0957926513516041.

Bekhit, Elsayed. 2009. 'Infographics in the United Arab Emirates newspapers.' Journalism: Theory, Practice and Criticism, 10(4): 492-508.

Bernays, Edward L., and Miller, M. 2005 (1928). Propaganda. Ig Publishing.

Bertin, Jacques. 1967. Semiology of Graphics: Diagrams, Networks, Maps (translation from 1967 edition, 2010). London: ESRI Press.

Blasio, Anthony J. and Ann M. Bisantz. 2002. 'A comparison of the effects of data-ink ratio on performance with dynamic displays in a monitoring task.' Journal of Industrial Ergonomics, 30: 89-101.

Blumler, Jay G., and Stephen Cushion. 2013. 'Normative perspectives on journalism studies: Stock-taking and future directions'.' Journalism: Theory, Practice and Criticism, DOI: 1464884913498689.

Boczkowski, Pablo. 2004. 'The processes of adopting multimedia and interactivity in three online newsrooms.' Journal of Communication, 54: 197-213.

Bogost, Ian; Ferrari, Simon and Bobby Schweizer. 2010. Newsgames : Journalism at Play. MIT Press.

Bourdieu, Pierre. 1998. On Television. New York: The New Press.

Boyce, G. 1978. 'The fourth estate - a reappraisal', in Boyce, G.; Curran, J. & Wingate, P. (Eds), Newspaper History: from the 17th century to the present day. Sage/Constable: 19-40.

Bradshaw, Paul and Lisa Rohumaa. 2011. The Online Journalism Handbook: Skills to Survive and Thrive in the Digital Age. London: Longman.

Breed, W. 1955. 'Social Control in the Newsroom: A Functional Analysis.' Social Forces: 326-335.

Brighton, Paul, and Dennis Foy. 2007. News values. Sage.

Bromley, Michael. 1997. 'The end of journalism? Changes in workplace practices in the press and broadcasting in the 1990s.' In Bromley, Michael (Ed) A Journalism Reader (Communication and Society). London: Routledge: 330-350.

Bruns, Axel. 2005. Gatewatching: Collaborative Online News Production. New York: Peter Lang.

BSI (British Standards Institution). 1992. British Standard 7581:1992 A Guide to the Presentation of Tables & Graphs. London: British Standards Institute.

Bull, Andy. 2010. Multimedia Journalism. London: Routledge.

Burke, Christopher. 2011. 'The linguistic status of Isotype.' Image and Imaging in Philosophy, Science and the Arts 2: 31-57.

Butterworth, Siobhain. 2008. 'Open door: The readers' editor on ... new opportunities to put the record straight online.' Guardian .co.uk, March 17: http://www.guardian.co.uk/commentisfree/2008/mar/17/1

Calcutt, Andrew, and Philip Hammond. 2011. Journalism studies: A critical introduction. London: Taylor & Francis.

Calishain, Tara. 2004. Web search garage. New Jersey: Prentice Hall Press.

Castells, Manuel. 1996. The rise of the network society, the information age: economy, society and culture, Vol. I. Malden, MA: Blackwell Publishers.

Castells, Manuel. 2000. 'Materials for an exploratory theory of the network society.' The British journal of sociology, 51(1): 5-24.

Cellan Jones, Rory (2008) 'How cool is Cuill?' dot life, BBC Technology blog: http://www.bbc.co.uk/blogs/technology/2008/07/how_cool_is_cuill.html

Chandler, Daniel and Rod Munday. 2011. A Dictionary of Media and Communication. Oxford: Oxford University Press.

Chung, Deborah. 2007. 'Profits and Perils Online News Producers' Perceptions of Interactivity and Uses of Interactive Features.' Convergence: The International Journal of Research into New Media Technologies, 13(1): 43-61.

Cohen, L., Mansion, L and Morrison, K. 2000. Research Methods in Education. London: Routledge Falmer (Fifth Edition).

Combs, J. and Nimmo, D. 1993. The new propaganda: The dictatorship of palaver in contemporary politics. Longman Publishing Group.

Conboy, Martin. 2004. Journalism: A Critical History. London: Sage.

Conboy, Martin. 2013. Journalism studies: the basics. London: Routledge.

Corner, J. 2003. 'The Model in Question A Response to Klaehn on Herman and Chomsky.' European Journal of Communication 18.3: 367-375.

Cornwall, Andrea, and Rachel Jewkes. 1995 'What is participatory research?.' Social science & medicine, 41(12): 1667-1676.

Cottle, Simon. 2007. 'Ethnography and news production: New (s) developments in the field*.' Sociology Compass 1.1: 1-16.

Cover, Rob. 2006. 'Audience inter/active Interactive media, narrative control and reconceiving audience history.' New media & society, 8(1): 139-158.

Currah, Andrew. 2009. What's happening to our news. Oxford Institute for the Study of Journalism. Oxford: University of Oxford.

Curran, James. 2006 (I). 'The era of the press barons', in Curran, J. and Seaton, J. (Eds) Power without responsibility. Sixth Edition. Routledge: 38-54.

Curran, James. 2006 (II). 'The liberal theory of press freedom', in Curran, J. and Seaton, J. (Eds) Power without responsibility. Sixth Edition. Routledge: 346-362.

Dahlgren, Peter. 1992. 'Introduction', in Dahlgren, P. and Sparks, C. (Eds), Journalism and popular culture. Sage: 1-23.

Davies, Nick. 2008. Flat earth news: An award-winning reporter exposes falsehood distortion and propaganda in the global media. Chatto & Windus, London.

De Burgh, Hugo. 2008. Investigative journalism. London: Routledge.

De-Liu, Chang. 2006. 'De-skilling Effects on Journalists: ICTs and the Labour Process of Taiwanese Newspaper Reporters.' Canadian Journal of Communication, 31(3): 695-714.

Deuze, Mark. 2004. 'What is multimedia journalism?' Journalism Studies, 5.2: 139-152.

Deuze, Mark. 2008. 'The Changing Context of News Work: liquid journalism and monitorial citizenship.' International Journal of Communication, 2: 848-865.

Dick, Murray. 2007. 'Lecture 3: Faking News: Avoiding Internet Hoaxes': http://static2.docstoccdn.com/docs/24380409/Lecture-3-Faking-News-Avoiding-Internet-Hoaxes

Dick, Murray. 2008. 'Is the future in bits?' BBC News, December 03: http://news.bbc.co.uk/1/hi/7761153.stm

Dick, Murray. 2010. 'Vanishing Point: On the consequences of lost online news in the UK.' Ethical Space, 8(1-2): 14-19.

Dick, Murray. 2011 (I). 'Search Engine Optimisation in UK News Production.' Journalism Practice, 5(4): 462-477.
Dick, Murray. 2011 (II). 'Investigative Online Search: CIJ Handbook.' Centre for Investigative Journalism, City University London, May: http://www.tcij.org/resources/handbooks/investigative-online-search

Dick, Murray. 2012. 'The Rebirth of the Beat: A Hyperlocal News-gathering model.' Journalism Practice 6(5-6): 754-765.

Dick, Murray. 2013. Search: Theory and Practice in Journalism Online. London: Palgrave MacMillan.

Dick, Murray. 2014. 'Interactive infographics and news values'. Digital Journalism 2014, 2(4), 490-506.

Dick, Murray. 2015. 'Just Fancy That: An analysis of infographic propaganda in The Daily Express, 1956-1959.' Journalism Studies, 16(2), 152-174.

Dikovitskaya, Margaret. 2012. 'Major Theoretical Frameworks in Visual Culture.' In Heywood, Ian, and Barry Sandywell, (Eds) The Handbook of Visual Culture. New York: Berg.

Easton, Mark (2008) 'Why Hoax?' Mark Easton's UK Blog, BBC, July 24: http://www.bbc.co.uk/blogs/thereporters/markeaston/2008/07/why_hoax.html

Ellul, Jaques. 1965. Propaganda: The Formation of Men's Attitudes, trans. by Konrad Kellen and Jean Lerner. Knopf.

Epstein, E. J. 1973. 'News form Nowhere: Television and the News.'

Evans, Harold, and Edwin Taylor. 1997. Pictures on a page: photo-journalism, graphics and picture editing. London: Pimlico.

Fenton, Natalie. 2009. 'News in the Digital Age.' In Allan, Stuart (Ed) The Routledge companion to news and journalism. New York: Routledge: 557-567.

Fienberg, Stephen E. 1979. 'Methods in Statistics', The American Statistician, 33(4): 165-178.

Fishman, Mark. 1980. Manufacturing the news. Austin: University of Texas Press.

Fiske, J. 1992. 'Popularity and the Politics of Information, in Dahlgren, P. and Sparks, C. (Eds), Journalism and popular culture. Sage: 45-63.

Franklin, Bob, M. Hamer, M. Hanna, M. Kinsey, & J. E. Richardson. 2005. Key Concepts in Journalism Studies, London: Sage.

Franklin, Bob. 1997. Newszak and news media. London: Arnold.

Friend, Claire and Jane Singer. 2007. Online Journalism Ethics: Traditions and Transitions. Armonk, New York: M.E. Sharpe.

Fuery, Kelli. 2009. New media: Culture and image. London: Palgrave Macmillan.

Funkhauser, Howard Gray. 1937. 'Historical Development of the Graphical Representation of Statistical Data.' Osiris 3: 269-404.

Galtung, Johan, and Ruge, Mari Holomboe. 1965. 'The Structure of Foreign News. The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers.' Journal of Peace Research, 2: 64-91.

Gane, Nicholas, and David Beer. New media: The key concepts. Berg, 2008.
Gans, Herbert. 1979. Deciding what’s news: A study of CBS Evening News, NBC Nightly News, Newsweek and Time. New York: Vintage

Garfinkel, Harold. 1967. Studies in ethnomethodology. Englewood Cliffs, NJ: Prentice-Hall.

Garrison, Bruce. 2000. ‘Journalists' perceptions of online information-gathering problems.’ Journalism & Mass Communication Quarterly, 77.3: 500-514.

Geertz, Clifford. 1973. The interpretation of cultures: Selected essays. Vol. 5019. Basic books.

Ghode, Rajeev. 2012. 'Infographics in News presentation: A Study of its Effective Use in Times of India and Indian Express the Two Leading Newspapers in India.' Journal of Business Management & Social Sciences Research, 1(1): 35-43.

Giardina, Marco and Pablo Medina. 2012. 'Information graphics design challenges and workflow management'. Paper presented at the International Conference on Communication, Media, Technology and Design, (ICCMTD), 09-11 May, 2012 in Istanbul, Turkey.

Gillmor, Dan. 2006. We The Media: Grassroots Journalism by the People for the People. California: O'Reilly Media, Inc.

Glaser, Barney, and Anselm Strauss. 1967. The discovery of grounded theory. London: Weidenfeld and Nicholson.

Golding, Peter, and Phillip Elliot. 1979. Making the News. London: Longman.

Google 2015. Algorithms – Inside Search:
https://www.google.co.uk/insidesearch/howsearchworks/algorithms.html

Gray, Jonathan, Lucy Chambers, and Liliana Bounegru. 2012. The Visual Data Journalism Handbook. London:

O'Reilly Media.

Griffin, Jeffrey L. and Robert L Stevenson. 1996. 'The influence of statistical graphics on newspaper reader recall.' Visual Communication Quarterly, 3(3): 9-11.
Gripsrud, Jostein. 2000. 'Tabloidization, popular journalism and democracy. ' In Sparks, Colin and John Tulloch (Eds) Tabloid tales: Global debates over media standards. London: Rowman & Littlefield Publishers: 285-300.

Habermas, J. 1991. The structural transformation of the public sphere: An inquiry into a category of bourgeois society. MIT press.

Hall, Stuart. 1973. 'The determinations of news photographs.' in Cohen, Stanley and Jock Young (Eds) The manufacture of news: Social problems, deviance and the mass media, Communication and society. Beverly Hills: Sage Publications: 176-190.

Hampton, M. 2004. Visions of the Press in Britain, 1850-1950. University of Illinois Press.

Hampton, M. 2010. 'The fourth estate ideal in journalism history', in The Routledge companion to news and journalism. Routledge: 3-12.

Hand, Martin. 2012. 'Images and Information in Cultures of Consumption', in Heywood, Ian, and Barry Sandywell, (Eds) The Handbook of Visual Culture. New York: Berg.

Hansen, Kathleen, Mark Neuzil, and Jean Ward. 1998. ‘Newsroom topic teams: Journalists' assessments of effects on news routines and newspaper quality.' Journalism & Mass Communication Quarterly 75(4): 803-821.

Harcup, Tony, and Deirdre O'Neill. 2001. 'Galtung and Ruge Revisited.' Journalism studies, 2(2): 261-280.

Harcup, Tony. 2007. The Ethical Journalist : Making Responsible Decisions in the Pursuit of News. London: Sage.

Harcup, Tony. 2014. Oxford Dictionary of Journalism. Oxford University Press.

Harris, Robert L. 1999. Information Graphics: A Comprehensive Illustrated Reference. New York: Oxford University Press.

Harrison, Jackie. 2006. News. Abingdon: Routledge.

Harrison, Jackie. 2009. 'Gatekeeping and News Selection as Symbolic Mediation.' In Allan, Stuart (Ed.) The Routledge Companion to News and journalism studies, London: Routledge: 191-201.

Hartley, John. 1982. Understanding News, London: Methuen.

Hartley, John. 1996. Popular reality: Journalism, modernity, popular culture. London: Arnold.

Hartley, John. 2000. 'Communicative democracy in a redactional society: the future of journalism studies.' Journalism: Theory Practice and Criticism, 1(1): 39-48.

Heinrich, Ansgard. 2011. Network Journalism: journalistic practice in interactive spheres. Abingdon: Routledge.

Henninger, Maureen. 2008. The Hidden Web: Finding Quality Information on the Net (2nd Edition). Sidney: University of New South Wales Press

Herman, Edward S, and Noam Chomsky. 1988. Manufacturing consent: The political economy of the mass media. New York: Pantheon.

Hodkinson, Paul. 2008. 'Grounded theory and inductive research.' In Gilbert, Nigel, ed. Researching social life. Sage: 80-100.

Holmes, Nigel. 1984. Designer's Guide to Creating Charts and Diagrams. New York: Watson-Guptill Publications Inc.

Huff, Darrell. 1954. How to Lie with Statistics. New York: Norton.

Hujanen, Jaana, and Sari Pietikäinen. 2004. 'Interactive uses of journalism: Crossing between technological potential and young people's news-using practices.' New Media & Society, 6(3): 383-401.

Inbar, Ohad, Noam Tractinsky, and Joachim Meyer. 2007. 'Minimalism in Information Visualization - Attitudes towards Maximizing the Data-Ink Ratio.' Proceedings. ECCE, ACM Press: 185-188.

InPublishing Magazine. 2012. 'Whatever Happened To Gob-Smacking Surprise?' May-June: http://inpublishing.co.uk/kb/articles/whatever_happened_to_gobsmacking_surprise.aspx

Jackall, Robert, (Ed). 1995. Propaganda. London: Macmillan.

Jowett, Garth S., and Victoria O'Donnell, (Eds). 2011. Propaganda & persuasion. London: Sage. Fourth edition.

Keeble, Richard. 2008. Ethics for journalists. Abingdon: Routledge.

Kenney, Keith, Alexander Gorelik, and Sam Mwangi. 2000. 'Interactive features of online newspapers.' First Monday, 5(1).

King, Geoff. 1996. Mapping Reality. London: MacMillan Press.

Kiousis, Spiro. 'Interactivity: a concept explication.' new media & society 4.3 (2002): 355-383.

Klaehn, J. 2002. 'A Critical Review and Assessment of Herman and Chomsky's Propaganda Model.' European Journal of Communication 17.2: 147-182.

Krug, Steve. 2000. 'Don't make me think.' A Common Sense Approach to web usability. New Reader Publishing.

Kulla-Mader, Julia. 2007. 'Graphs via Ink: Understanding How the Amount of Non-data-ink in a Graph Affects Perception and Learning.' Master's Thesis, Dept. of Information and Library Science, University of North Carolina.

Langer, John. 1998. Tabloid Television, Popular Journalism and the 'Other News'. London: Routledge.

Larsson, Anders Olof. 2012. 'Understanding non-use of interactivity in online newspapers – Insights from Structuration theory'. The Information Society 28(4): 253-263.

Lasswell, H. 1927. Propaganda technique in the world war. Peter Smith.

Lather, Patti. 1986. ‘Research as praxis’, Harvard educational review, 56(3): 257-278.

Lavie, Aliza, and Sam Lehman-Wilzig. 2003. 'Whose news? Does gender determine the editorial product?.' European Journal of Communication 18.1 (2003): 5-29.

Lee, Allen S. 1991 ‘Integrating positivist and interpretive approaches to organizational research’. Organization science 2(4): 342-365.

Lewis, Justin. 2006. 'News and the empowerment of citizens.' European journal of cultural studies, 9(3): 303-319.

Lindlof, Thomas R., and Milton J. Shatzer. 'Media ethnography in virtual space: Strategies, limits, and possibilities.' Journal of broadcasting & electronic media 42.2 (1998): 170-189.

Lippmann, W. 1922. Public Opinion. New York: Harcourt, Brace and Company.

Long, Johnny. 2005. Google Hacking for Penetration Testers, Syngress Publishing, Inc.

Lowrey, Wilson. 1999. 'Routine news: The power of the organization in visual journalism', Visual Communication Quarterly, 6(2): 10-15.

Lowrey, Wilson. 2002. 'Word People vs. Picture People: Normative Differences and Strategies for Control Over Work Among Newsroom Subgroups', Mass Communication and Society, 5(4): 411-432.

Lowrey, Wilson. 2003. 'Normative Conflict in the Newsroom: The Case of Digital Photo Manipulation,' Journal of Mass Media Ethics, 18(2): 123-142.

Luckie, Mark S. 2011. The Digital Journalist's Handbook. Createspace.

Luft, Oliver. 2008. 'Social networks need to do more to raise awareness about misuse of personal information, says PCC director.' Journalism.co.uk, June 05: http://www.journalism.co.uk/news/social-networks-need-to-do-more-to-raise-awareness-about-misuse-of-personal-information-says-pcc-director/s2/a531706/

Luft, Oliver. 2009. 'PCC targets Sunday Express over Dunblane allegations.' Guardian.co.uk, March 16: http://www.theguardian.com/media/2009/mar/16/pcc-targets-sunday-express-over-dunblane-claims

Lunenfeld, Peter. 'Digital dialectics: A hybrid theory of computer media.' AFTERIMAGE-ROCHESTER NEW YORK- 21 (1993): 5-5.

MacGregor, Phil. 2007. 'Tracking the Online Audience.' Journalism studies, 8(2): 280-298.

Machin, David and Sarah Niblock. 2006. News Production: Theory and Practice, London: Routledge.

Maier, Scott R. .2002. 'Getting It Right? Not in 59 Percent of Stories.' Newspaper Research Journal, 59: 10-24.

Maier, Scott R. 2007. 'Setting the record straight: When the press errs, do corrections follow?' Journalism Practice, 1(1): 33–43.

Manning White, David. 1950. 'The Gatekeeper: A Case Study in the Selection of News,' journalism Quarterly 11: 383-96

Manovich, Lev. 2001. The language of new media. MIT press.

McChesney, Robert. 2000. Rich Media poor democracy: Communications politics in dubious times. New York: The New Press.

McChesney, Robert. 2007. Communication revolution: Critical junctures and the future of media. New York: The New Press.

McLachlan, S., and Golding, P. 2000. 'Tabloidization in the British Press: A Quantitative Investigation into Changes in British Newspapers, 1952 - 1997', in Sparks, C and Tulloch, J. (Eds) Tabloid tales: Global debates over media standards. Rowman and Littlefield Publishers.

McLuhan, Marshall. Understanding media: The extensions of man. MIT press, 1994 (1964)

McMillan, Sally J. 2002. 'A Four-part model of cyber-interactivity: Some Cyber-places are more interactive than others.' New Media and Society, 4(2): 271-291.

Middlestadt, Susan E., and Kevin G. Barnhurst. 1999. 'The influence of layout on the perceived tone of news articles.' Journalism and Mass Communication Quarterly, 76(2): 264–276.

Mirzoeff, Nicholas. 1999. An introduction to visual culture. London: Psychology Press.

Monmonier, Mark. 1996. How to lie with maps. Second edition. London: University of Chicago Press.

Morris, Merrill, and Christine Ogan. 'The Internet as mass medium.' Journal of Computer-Mediated Communication 1.4 (1996): 0-0.

Morrison, David E. and Tumber, Howard. 1988. Journalists at war: the dynamics of news reporting during the Falklands conflict. Sage Publications.

Morrison, James. 2010. 'Spin, smoke-filled rooms, and the decline of council reporting by local newspapers: the slow demise of town hall transparency.' Presented at the 61st Political Studies Association Annual Conference:Transforming Politics: New Synergies: http://www.psa.ac.uk/journals/pdf/5/2010/612_322.pdf

Nadarajan , Belinda and Penghwa Ang. 1999. 'Credibility and Journalism on the Internet How: Online Newspapers Handle Errors and Corrections.' Paper submitted to the Communication Technology and AEJMC for the 1999 AEJMC Convention: http://list.msu.edu/cgi-bin/wa?A2=ind9909B&L=aejmc&P=R15182

Neilsen, Jakob. 1997. 'How Users Read on the Web.' Nielsen Norman Group, October 1: http://www.nngroup.com/articles/how-users-read-on-the-web/

Neilsen, Jakob. 1999. Designing Web Usability: The Practice of Simplicity. San Francisco: New Riders Publishing.

Nichani, Maish and Venkat Rajamanickam. 2003. 'Interactive Visual Explainers — A Simple Classification.' eLearningPost, September 01: http://www.elearnspace.org/blog/2003/09/02/interactive-visual-explainers-a-simple-classification/

O'Maley, T. 1997. 'Labour and the 1947-9 Royal Commission on the press', in Bromley, M. and O'Malley, T. (Eds) A journalism reader. Psychology Press: 126-158.

Örnebring, Henrik. 2010. 'Reassessing journalism as a profession.' In Allan, Stuart. News culture. Third edition. Maidenhead and New York: Open University Press: 568-577.

Outing, Steve. 2005. 'Sophisticated Web Stats Give Editors Better Idea of Reader Interests.' Convenience Store News, July 26.

Oxford Dictionaries. 2013. Oxford: Oxford University Press: http://www.oxforddictionaries.com/definition/english/infographic

(OED) Oxford English Dictionary. 2006. OED online. Oxford: Oxford University Press: http://www. oed. com

Palmer, Jeremy. 2000. Spinning into Control: News Values and Source Strategies, Studies in Communication & Society. London: Continuum International Publishing Group Ltd.

Pasternack, Steve, and Sandra H. Utt. 1986. 'Subject perception of newspaper characteristics based on front page design.' Newspaper Research Journal, 7: 29–35.

Paulos, John Allen. 1996. A mathematician reads the newspaper. New York: Anchor Books.

Pavlik, John. 2000. 'The impact of technology on journalism.' Journalism Studies 1(2): 229-237.

Picard, Robert G. 1998. 'Media concentration, economics, and regulation.' In Graber, Doris, Denis McQuail and Pippa Norris (Eds) The politics of news and the news of politics. Washington DC: CQ Press: 193-217.

Pipps, Val; Walter, Heather; Endres, Kathleen; Tabatcher, Patrick. 2009. 'Information Recall of Internet News: Does Design Make a Difference?' Journal of Magazine and New Media Research, 11(1): 1-20.

Prabu David. 1992. 'Accuracy of Visual Perception of Quantitative Graphics: An Exploratory Study.' Journalism Quarterly, Summer: 273-292.

Qualter, T. 1962. Propaganda and psychological warfare. Random House.

Quandt, Thorsten. 2008. '(No) News on the world wide web?' Journalism Studies, 9(5): 717-738.

Quinn, Stephen, and Stephen Lamble. 2008. Online newsgathering: research and reporting for journalism. London: Focal Press.

Ramaprasad, Jyotika. 1991. 'Informational graphics in newspapers: attention, information retrieval, understanding and recall.' Newspaper Research Journal, 12(3): 92-103.

Reavy, Matthew. 2003. 'Rules and the real world an examination of information graphics in time and Newsweek.' Visual Communication Quarterly, 10(4): 4-10.

Redden, Joanna and Tamara Witschge. 2010. 'A new news order? Online news content examined.' In Fenton, Natalie (Ed.) New Media, Old News: Journalism and Democracy in the Digital Age. London: Sage: 171-186.

Robinson, Sue. 2007. 'Someone's gotta be in control here: The institutionalization of online news and the creation of a shared journalistic authorit.y, Journalism Practice 1(3): 305-321.

Rosenberry, Jack. 2005. 'Few papers use online techniques to improve public communication.' Newspaper Research Journal, 26(4): 61.

Rowe, David. 2009. 'Tabloidization of News.' In Allan, Stuart (Ed), The Routledge Companion to News and Journalism Opens in a new window. Oxon, UK: Routledge: 350-361.

Royal Pingdom. 2010. 'Study: Ages of social network users.' RoyalPingdom.com, February 16: http://royal.pingdom.com/2010/02/16/study-ages-of-social-network-users/

Rubin, Herbert J., and Irene S. Rubin. 1995. Qualitative interviewing: The art of hearing data. Sage.

Sabbagh, Dan. 2011. ‘Guardian and Observer to adopt 'digital-first' strategy’, Guardian.co.uk, Thursday 16th June:
http://www.theguardian.com/media/2011/jun/16/guardian-observer-digital-first-strategy

Schlein, Alan M., James R. Flowers, and Shirley K. Kisaichi. 2000. Find it online: The complete guide to online research. New York: BPR Publishers.

Schlesinger, Peter. 1987. Putting 'reality' together: BBC news. Taylor & Francis.

Schlesinger, Peter. 1992. 'From production to propaganda?' in Scannell, P., Schlesinger, Peter. and Sparks, Colin. (Eds). Culture and power: a media, culture & society reader. Sage: 293-316

Schroeder, Roland. 2004. 'Interactive Info Graphics in Europe--added value to online mass media: a preliminary survey.' Journalism Studies, 5(4): 563-570.

Schudson, Michael. 1997. 'The Sociology of News Production.' In Berkowitz, Dan (Ed) Social Meanings of News: A Text Reader. New York: Sage: 7-22.

Schultz, Majken and Mary Jo Hatch. 1996. ‘Living with Multiple Paradigms: The Case of Paradigm Interplay in Organizational Culture Studies’, The Academy of Management Review, 21(2): 529-557.

Schultz, Tanjev. 2000. 'Mass media and the concept of interactivity: an exploratory study of online forums and reader email.' Media, culture & society 22.2: 205-221.

Schultz, Tanjev. 1999. 'Interactive options in online journalism: A content analysis of 100 US newspapers.' Journal of Computer‐Mediated Communication, 5(1): http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1999.tb00331.x/full

Schulz, Winfried Friedrich. 1982. ‘News structure and people's awareness of political events.’ International Communication Gazette. 30(3): 139-153.

Scott, Robert A., and Arnold Shore. 1979. Why sociology does not apply: A study of the use of sociology in public policy. New York: Elsevier.

Seale, Clive. 2004. Social research methods: A reader. Psychology Press.

Shepard, A.C. 1998. 'To Err Is Human, To Correct Divine.' American Journalism Review, June: http://www.ajr.org/article.asp?id=1612

Shirky, Clay. 2009. Here Comes Everybody: the power of organizing without organizations. London: Penguin.

Simmel, Georg. 'The stranger.' The Sociology of Georg Simmel (1950): 402-408.

Skinner, David, Mike J. Gasher, and James Compton. 2001. 'Putting theory to practice A critical approach to journalism studies.' Journalism, 2(3): 341-360.

Small, Hugh. 1998. 'Florence Nightingale's Statistical Diagrams.' Paper presented at Stats & Lamps Research Conference, organised by the Florence Nightingale Museum at St. Thomas' Hospital, March 18.

Smith, Edward J. and Donna Hajash. 1988. 'Informational Graphics in 30 Daily Newspapers.' Journalism & Mass Communication Quarterly, 65(3): 714-718.

Spark, David. 1999. Investigative reporting: A study in technique. Taylor & Francis US.

Sparks, Colin. 1991. 'Goodbye, Hildy Johnson: the vanishing 'serious press'', in Communication and Citizenship: Journalism and the Public Sphere in the New Media Age. London: Routledg: 58-74.

Sparks, Colin. 1992. 'Popular journalism: theories and practice', in Dahlgren, P. and Sparks, C. (Eds), Journalism and popular culture. Sage: 24-44.

Sparks, Colin. 2000. 'Introduction: the panic over tabloid news', in Sparks, C. and Tulloch, J. (Eds) Tabloid tales: Global debates over media standards. Rowman and Littlefield Publishers.

Spence, Ian and Howard Wainer. 2005. 'William Playfair and His Graphical Inventions: An Excerpt from the Introduction to the Republication of His 'Atlas' and 'Statistical Breviary'.' The American Statistician, 59(3): 224-229.

Spence, Ian. 1990. 'Visual Psychophysics of Simple Graphical Elements.' JEP: Human Perception and Performance, 16(4): 683-692.

Spence, Ian. 2005. 'No humble pie: The origins and usage of a statistical chart.' Journal of Educational and Behavioral Statistics, 30(4): 353-368.

Staab, Joachim. 1990. 'The role of news factors in news selection: a theoretical reconsideration.' European Journal of Communication, 5(4): 423-443.

Stevenson, Nick. 2000. 'THE FUTURE OF PUBLIC MEDIA CULTURES.' Information, Communication & Society, 3:2, 192-214,

Stone, Gerald and Patti Hall. 1997. 'Do newspapers graphics have two dimensions?' Visual Communication Quarterly, 4(4): 4-10.

Sturken, Marita, and Lisa Cartwright. 2001. Practices of looking: An introduction to visual culture. Oxford: Oxford University Press.

Sullivan, Peter. 1987. Newspaper Graphics. Darmstadt: IFRA.

Tandoc Jr, Edson C. 2013 Web Analytics, Social Media, and the Journalistic Doxa: The Impact of Audience Feedback on the Evolving Gatekeeping Process. Diss. University of Missouri--Columbia (Cites my work)

Tandoc Jr, Edson C. 2014. 'Journalism is twerking? How web analytics is changing the process of gatekeeping.' new media & society: 1461444814530541.

Tankard, James W., and Hyun Ban. 1998. 'Online newspapers: Living up to their potential.' Mid-winter conference of the Association for Education in Journalism and Mass Communication, Dallas, Texas.

Tedlock, Barbara. 2005. 'The observation of participation and the emergence of public ethnography.' The Sage handbook of qualitative research: 467-481.

Thomson, Oliver. 1999. Easily led: A history of propaganda. London: Sutton Publishing Limited.

Thrift, Nigel. 2005. Knowing capitalism. Sage.

Toffler, Alvin. 1984. Future Shock. New York: Bantam.

Tuchman, Gaye. 1973. 'Making news by doing work: Routinizing the unexpected.' American journal of Sociology: 110-131.

Tuchman, Gaye. 1978. Making news. New York: Free Press.

Tufte, Edward, R. 1983. The Visual Display of Quantitative Information. Cheshire, Conn: Graphics Press.

Tukey, John W. 1990. 'Data-based graphics: visual display in the decades to come.' Statistical Science 5(3): 327-339.

Tunstall, Jeremy. 1971. Journalists at work: Specialist correspondents: Their news organizations, news sources, and competitor-colleagues. London: Constable.

Turner, Graeme. 1999. 'Tabloidization, journalism and the possibility of critique,' International Journal of Cultural Studies 2(1): 59-76.

Usher, Nikki, 2014. Making News at The New York Times. University of Michigan Press.

Volkmer, Ingrid. 2003. 'The global network society and the global public sphere,' Development 46(1): 9-16.

Wahl-Jorgensen, Karin, and Thomas Hanitzsch. 2009. ‘Introduction: On Why and How We Should Do Journalism Studies.’ In Wahl-Jorgensen, Karin, and Thomas Hanitzsch (Eds) Handbook of journalism studies. London: Routledge: 3-16.

Wainer, Howard. 1990. 'Graphical Visions from William Playfair to John Tukey.' Statistical Science, 5(3): 340-346.

Walker, John A and Sarah Chaplin. 1997. Visual culture: An introduction. Manchester: Manchester University Press.

Ward, Mike. 2002. Journalism Online. London: Focal Press.

Wells, Matt and Randeep Ramesh. 2004 ‘BBC reputation hit by Bhopal interview hoax.’ Guardian.co.uk, December 04:
http://www.theguardian.com/media/2004/dec/04/india.broadcasting

Wickström, Gustav, and Tom Bendix. 'The' Hawthorne effect'—what did the original Hawthorne studies actually show?.' Scandinavian journal of work, environment & health (2000): 363-367.

Wiener, Joel. 1988. Papers for the millions: The new journalism in Britain, 1850s to 1914, No. 13. Westport, Connecticut : Greenwood Publishing Group.

Witschge, Tamara, and Gunnar Nygren. 2009. 'Journalism: a profession under pressure?' Journal of Media Business Studies, 6(1): 37-59.

Zelizer, B. 2004. 'When facts, truth, and reality are God-terms: on journalism's uneasy place in cultural studies.' Communication and Critical/Cultural Studies, 1.1: 100-119.

APPENDIX 1: VANISHING POINT: ON THE CONSEQUENCES OF LOST ONLINE NEWS IN THE UK

As far back as 1662, the principal of legal deposit has ensured that publications are stored in national libraries, so that researchers and historians might gain a detailed insight into our life and times. Yet fifteen years since UK newspapers started publishing online, there is still no collective, publicly held archive of online newspapers. This paper demonstrates that the lack of a centralised, public archive for online news represents more than a theoretical loss to posterity. Drawing on several examples, this paper shows how our living history online rests on shifting sands. Chris Elliott, Readers' Editor of The Guardian, is interviewed on issues arising out of online corrections policy. A voluntary code on the removal of, and/or editing of online content for UK online media is proposed.

Keywords: legal deposit, online news, corrections policy, British Library, Newspaper Publishers Association, provenance

Introduction

In 2008, The Guardian's then Readers' Editor, Siobhain Butterworth, stated that some 40% of content created for guardian.co.uk never appears in the Guardian newspaper (Butterworth, 2008). Little data exists with regard to other UK news titles, but there can be no doubt online and off-line versions of the UK national press are diverging. Aside from web-only news stories, and updates, other web-only content published routinely by the UK media include in-house blogs and community pages, charts, graphics, polls, multimedia, and other interactive media, not to mention article-copy re-written for search engines (including alterations to headlines, supporting text and other page furniture).

There are many reasons why online-only content may be removed, altered, or invisibly mended; whether temporarily or permanently. Libel and contempt of court rulings may be resolved by means of deleting an article, and/or publication of a correction. Resolutions negotiated via the Press Complaints Commission, where it is often in the interests of wronged parties to raise little fuss, need not be declared online (the PCC having no stated duty to set the public record straight) (Cain, 2010). Third party intellectual property rights and the correction of mistakes also take a significant toll.

UK online media often block third party search engines and archives from caching and archiving their content (Dick, 2008), because such archives may compromise future commercial exploitation. While the British Library has been collecting material by means of the UK Web Archive, this content represents a tiny fraction of UK online news output.

News is diverging by medium, but the regulatory regime which exists to ensure history is not lost is still catching up. Online-only news is not being stored for posterity under current UK legal deposit legislation. More worryingly still, it is disappearing.

The crisis in legal deposit

Legal deposit has a significant legacy in English law, dating back to 1662. The principal that written publications (from books, through to pamphlets and newspapers) should be systematically stored by national libraries as a living record of our life and times was codified in the Copyright Act 1911. The power, privileged access and influence the press wields as our *fourth estate* is in turn held in check by means of this independent archiving. The British Library's collections of newspapers comprise a significant part of our future histories, but with the development of mass communications online, the legislative framework is in need of overhaul.

The Legal Deposit Libraries Act 2003 was framed (as enabling legislation) in order to regulate materials for which the 1911 Act was never intended. As part of a long-scale strategy, the Legal Deposit Advisory Panel (LDAP) ran from 2005 until August 2010. Its remit, as an independent committee which included those with vested interests in this field, was to advise and take into consideration the problems inherent to legal deposit of non-print materials.

In December 2009, the Department for Culture Media and Sport (DCMS) launched a public consultation in relation to free-of-charge online publications, with the expectation of further proposals for commercial and protected online publications later. The LDAP's recommendations include a preferred option of statutory regulation (as opposed to self-regulatory, unregulated or fully privatised archiving regimes) (DCMS, 2010).

In its response, the British Library made the case for 24/7, 'without walls' access to such online archives (British Library, 2010). The Newspaper Publishers' Association (NPA) has been extremely critical of this position, and circumspect on the extension of legal deposit to their digital assets. In their response, the NPA raised fears about the ambitions of the British Library whose "objectives vis-à-vis the exploitation of legal deposit items would impact adversely on publishers' ability to derive full value from their copyright material" (NPA, 2010). Legal deposit "should not be allowed to become an alternative free distribution medium for digital content" (NPA, 2010) at the risk of harming the UK's still-emerging e-publishing market. While this stand-off continues, our online-only news is vanishing, with only localised media corrections, clarifications and amendments policies preserving our past.

Corrections policies in news media

The removal and/or amendment of newspaper copy often falls under the rubric of corrections policy. Newspaper corrections have been categorised into 'minor' (eg. spelling, typographical errors, homophones), 'moderate' (eg. misspelled proper nouns) and 'egregious' (eg. mistakes that malign, or cause significant offence) (Butterworth, 2007). Others have distinguished between "objective-type errors" (eg. typographical and minor factual errors) and "subjective-type errors" (eg. omissions and misquotes) (Nadarajan and Ang, 1999). However they are categorised, corrections policies designed to deal with serious errors, do not always work well for minor mistakes in practice (Butterworth, 2007), and do not account for the many other reasons archived online copy may be altered.

Much evidence of public dissatisfaction with newspaper corrections policy abounds. The "corrections box" culture in US newspapers has been found to be lacking in the eyes of many news consumers (Maier, 2007). Other studies have elaborated on shortcomings in policy; a perception of inconsistency (Lester, 2005) and of insufficient rigour in application (Shepard, 1998), have been cited as public concerns. Nevertheless, where news organisations are seen to have a systematic approach to corrections, it has been argued they accrue a dividend of trust (a thesis based on limited research) (Mayes, 2004).

The Guardian's corrections column, in place since 1997, was a direct challenge to the 'culture of resistance' in acknowledging mistakes within UK journalism (Mayes, 1999), a culture whose origins owe as much to professional arrogance (Diamond, 1998), as they do to the parlous state of UK libel law (Rusbridger, 1997) (albeit, Observer Readers' Editor Stephen Pritchard maintains that far from increasing the likelihood of court action, a robust and transparent corrections policy has seen legal problems recede at his newspaper) (Pritchard, 2008).

It has been suggested that stand-alone corrections columns in newspapers can ghettoise the process, especially in the case of significant and/or prominent errors (Diamond, 1998). This can be a problem online too. At the Guardian only around 30% of users access the column from a link on the home page, which in turn limits the potential for fully-distributed transparency (Butterworth, 2008).

Frost argues that corrections policy is framed by ethical codes of conduct (2000), albeit in practice this is not necessarily the case. Article one of the Press Complaints Commission Editors' Code of Practice, on accuracy, applies only to 'significant' inaccuracies, while the National Union of Journalists' Code of Conduct makes reference to the correction of 'harmful' inaccuracies. In both cases, the importance of the public record is not accounted for; these codes do not make up for shortcomings in law.

Harcourt suggests a range of best practice measures for dealing with corrections in newspapers (2007), but none are considered in regard to online-only news. In the US, online news corrections policies come in different guises; some newspapers delete content without leaving a note, while others leave old versions intact with notes appended (Friend, 2007). Some in the industry are fundamentally opposed to foot-noting corrections, suggesting instead that errors should be corrected where they appear (Agrawal, 2007). Moreover, appending corrections at the bottom of stories online can lead to further errors, if readers (and journalists) to do not continue to the bottom (Thompson, 2004), a recurring theme found in the web usability literature, with regard to the way we consume information online (Neilsen, 2000).

It has been suggested that online journalists may benefit from those tools and conventions used by bloggers (Friend, 2007). The use of strike-through to correct mistakes, and the culture of openness amongst bloggers are lauded as something news publishers can learn from (Friend, 2007), though it may be countered that some errors (which might include insults, implied or inferred) do not lend themselves well to such remedy. Friend draws upon the work of Blood’s *Weblog Handbook* (2002) for further guidance. Blood advocates that bloggers should “write each entry as if it could not be changed; add to, but do not rewrite or delete, any entry” (Blood, 2002). Moreover, deleting blog posts should be avoided at all costs, as this “destroys the integrity of the network” (Blood, 2002). But in news some errors are so profound, they are impossible to redress without doing violence to the story (Thompson, 2004), so may require complete removal.

In the US, many online corrections policies carry directly over from print (Nadarajan and Ang, 1999). But online news has its own unique properties. For example, in some cases it can be impossible to un-publish content due to the 'viral' nature of dissemination online (Alexander, 2010), albeit with no centralised archive, this can be difficult to navigate. It has been suggested that a grounded corrections policy, composed as native to the web, might improve the reliability of information published online (Nadarajan and Ang, 1999).

Vanishing point: going, going gone...

Online, a news story can change beyond all recognition at the stroke of a key. On April 25 2010, following Nick Clegg's widely lauded performance in the televised political leadership debates in the run-up to the 2010 general election, The Daily Mail published a story titled 'The story of privilege behind Nick Clegg’s wife’s lingerie shopping trip'. This story subsequently changed entirely into a Jan Moir opinion piece, the original story is no longer available (Hundal, 2010).

Newspaper blogs are another area where online-only content can disappear without trace. On January 24 2010, the Telegraph's James Delingpole blogged about a letter sent to prospective Conservative candidates at the 2010 election asking about their position on the environment. In subsequent comments on this post, the author's personal information and contact details were published, following which the blog post was removed (Monbiot, 2010).

Elsewhere on telegraph.co.uk, on June 04, 2010, Strategic Events Editor Tom Chivers wrote an excoriating piece on climate 'sceptic' Viscount Monckton, describing him as “a fantasist, a blethering popinjay useful only for amusement”. In an update to the post, Chivers mentioned that Monckton had contacted him, but that he “refused at least for now to take the blog down, until I have spoken to my editor” (Littlemore, 2010) The post subsequently vanished with no explanation.

Interactive features can also vanish. On June 19, 2009, the Daily Mail published an interactive poll featuring the question “Should the NHS allow gipsies to jump the queue?” (Ekklesia, 2009). The poll led to a backlash from those who questioned the the ethics and framing of the poll, leading to its removal from the Mail website.

Online hoaxes are a common source of corrections. On June 27, 2010, The Daily Mail ran a story claiming that CEO of Apple Corp, Steve Jobs, had used Twitter to give notice of a recall of their Apple iPhone 4 product – the journalist in question had inadvertently quoted a notorious hoax Twitter account ('Zee', 2010). The story has since disappeared, and there is no other reference to it on dailymail.co.uk.

In all of these case, and in many others across the web, so-called citizen journalists have been responsible for raising awareness about vanished content, ensuring it is not completely lost. Blogs and websites including: The Sun – Tabloid Lies, Express Watch, Mail Watch, Enemies of Reason, Bad Science, Liberal Conspiracy, Desmoblog, The Media Blog, and Regret the Error all express the concerns of those who feel strongly enough about the power of the media to give up their time to track and monitor what they say.

But there is no guarantee that the handiwork of these individuals will exist forever. Blogs (indeed blogging platforms) are not famed for their longevity. The means by which people might find such references are subject to the whims of search engines and other aggregators, not to mention the sophistication with which bloggers make their copy 'findable' online. It is therefore necessary that a consistent approach to the correction of online news is established.

Interview with Guardian Readers' Editor Chris Elliott

In 2007, research by the International Center for Media and the Public Agenda found guardian.co.uk to be the most transparent online media company in the world, based on a number of criteria, including the availability of corrections information (Moore, 2007). As such, Chris Elliott, current Readers' Editor of The Guardian, is in a particularly strong position to offer insight into how online corrections policy can be operationalised. Elliott was interviewed for this research on October 15, 2010. A series of open-ended questions were asked during an hour-long interview. The interview was recorded, then summarised in around 1000 words.

The Guardian approach serves as a robust exemplar to wider industry. There is an online corrections column, linking to stories which have been corrected, wherein footnotes explain the nature of the error, while at the top of stories, an Article History interactive contains a cross-reference to the hard-copy paper story (where available), a publication date stamp for online, and the last modified date-stamp for the online story.

Elliot confirmed that since 2008, the amount of web-only content produced for guardian.co.uk has risen from 40%, saying he "would be very surprised" (Chris Elliot, interview, 2010) if web-only content was less than 50% of all content published, though he did not have a precise figure.

Live reporting, where breaking and/or ongoing stories which inevitably render previous facts obsolete, is a vexing topic for Elliott. He suggests that neither reader nor publisher would benefit from an exhaustive itinerary of mistakes appended on such stories – this would make copy very difficult to read or make sense of; not just for consumers of news, but for future researchers. This issue, he contends, highlights the danger of taking a principal to an absolute extreme.

The Guardian has been using *invisible mending*, or *scrubbing* in online content since before 2006, (Mayes, 2006; Butterworth, 2007). In the past it has been acknowledged that "amendments to web-only items don't appear in the corrections column and some errors are fixed but not acknowledged" (Butterworth, 2008). But Elliott explains that this is only permitted in certain circumstance. "Our rules are: if a story is not open to comments, if the change is to be made on the same day and if it is a straight-forward misspelling of a name (or something like that) then you will invisibly mend" (Elliot, interview, 2010). The reasoning behind restricting scrubbing to stories not open to comment is because "if you invisibly mend, then you risk making your commenting readers who have pointed out the discrepancy look stupid" (Elliot, interview, 2010). Elliot states that they cannot invisibly mend if the mistake in question is subject to legal action, as this would require visible redress.

The Guardian are currently revising their approach in the wider field of online ethical conventions, extending far beyond simple corrections procedure. This is broadly in line with concerns explored by both of Elliott's predecessors: that the permanency of online archives "raises for some the possibility that history might very easily be amended to their advantage" (Mayes, 2006), and that "the archive needs to be controlled if we are going to resist pressure from readers and journalists to reshape the past" (Butterworth, 2007).

The issue of un-publishing is taking up much more time than it used to. A number of (confidential) examples were shared, ranging from what may be considered the trivial (the updating of biographical details for an individual who appeared in a feature) through to the very serious (an anorexic contributor, who contributed to the Guardian several years ago, arguing that this copy is impacting upon their right to privacy, and means of finding employment). Invisible mending has its dissenters – it can be construed, according to Craig Silverman, as unethical (Butterworth, 2008). Elliott acknowledges that this is a difficult area, adding "There is something intrinsically unfair which I can't put my finger on" (Elliot, interview, 2010) when making a decision not to un-publish.

The preliminary stages in securing opinion and actuality in the Guardian's journalism are also currently subject to ongoing review: "We need to develop better mechanisms for assigning consent. In other words, making people understand that when they speak to us, there is a very high likelihood that this will be there forever" (Elliot, interview, 2010).

Conclusion and findings

Corrections policy in online news is opening up a new ethical space in journalism. Newspapers continue to influence public debate and policy, but this influence is not being routinely monitored nor stored independently for future historians and researchers.

Corrections policies in online news media should be formulated into a public Corrections Charter. This charter would acknowledge the rights and responsibilities of publishers and consumers of news, including that:

Readers (and future historians) require transparency
Newspaper authority is predicated on openness
Not everything can (or should) be saved
In some circumstances, un-publishing can be justified

Building upon established excellence in the field, the blueprint for such a code should comprise the following:

All non-legal online-only corrections should be annotated
Invisible mending should be used only as a last resort
All legal removals in relation to online copy should be shared with the British Library
Stand-alone corrections should be hyper-linked

If we are to build upon the centuries of archiving modern day researchers have at their fingertips, we must acknowledge that online news corrections are more than merely footnotes in history.

Sources for tracking missing online news

Bad Science:
http://www.badscience.net/

Desmogblog:
http://www.desmogblog.com/

Enemies of reason:
http://enemiesofreason.co.uk/

Express Watch:
http://expresswatch.co.uk/

Liberal Conspiracy:
http://liberalconspiracy.org/

Mail Watch:
http://www.mailwatch.co.uk/

Regret the Error:
http://www.regrettheerror.com/

The Media Blog:
http://themediablog.typepad.com/the-media-blog/

The Sun – Tabloid Lies:

http://the-sun-lies.blogspot.com/

References

Agrawal, Rocky (2007) 'Googling all the news that's fit to correct', Re-Design Blog, August 29th
http://blog.agrawals.org/2007/08/29/googling-all-the-news-thats-fit-to-correct/
(Accessed 30.11.10)

Alexander, Andy, (2010) 'A need for "Best Practices" on "unpublishing" Web content', Washington Post, August 9th:
http://voices.washingtonpost.com/ombudsman-blog/2010/08/a_need_for_best_practices_on_u.html
(Accessed 30.11.10)

Blood, Rebecca (2002) 'Weblog Ethics', from the Weblog Handbook: Practical Advice on Creating and Maintaining Your Blog. (Accessed 22.10.10):
http://www.rebeccablood.net/handbook/excerpts/weblog_ethics.html
(Accessed 30.11.10)

British Library (2010) British Library response to the Government's consultation on the Collection and Preservation of UK Offline and Microform Publications and UK Online Publications (Available free of charge and without access restrictions), March:
http://www.culture.gov.uk/images/consultation_responses/British_Library_-_Consultation_response_Digital_Legal_Deposit.pdf
(Accessed 30.11.10)

Butterworth, Siobhain, (2007) 'Open door: The readers' editor on ... the invisible mending that happens every day, The Guardian, May 7th:
http://www.guardian.co.uk/commentisfree/2007/may/07/comment.pressandpublishing
(Accessed 30.11.10)

Butterworth, Siobhain (2008) 'Open door: The readers' editor on ... new opportunities to put the record straight online', Guardian, March 17th:
http://www.guardian.co.uk/commentisfree/2008/mar/17/1
(Accessed 30.11.10)

Cain, Matthew, (2010) 'Why the PCC should record all breaches of the code', PCC Watch. November 3rd:
http://pccwatch.co.uk/pcc-record-breaches-of-the-code/
(Accessed 30.11.10)

Department for Culture Media and Sport(DCMS) (2010) Digital Legal Deposit – Non-print Off Line Publications Impact Assessment, July 1st:
http://www.google.co.uk/url?sa=t&source=web&cd=2&ved=0CBkQFjAB&url=http%3A%2F%2Fwww.culture.gov.uk%2Fimages%2Fpublications%2FImpact_assessments-digitallegaldeposit2011.pdf&rct=j&q=LDAP%20recommendations%20%22Statutory%20Regulation%22&ei=TmfYTImhNdGGhQf719z1Aw&usg=AFQjCNFJcMqxQ-liGYmpntyEpxU-pMlktg&sig2=aZNOA_-U27bjhWDgT873EQ
(Accessed 30.11.10)

Diamond, John (1998) 'Sorry state; There is no point having a "Readers' Editor", says JOHN DIAMOND,if all he does is apologise for misprints', The Evening Standard, November 11th.

Dick, Murray (2008) 'Is the future in bits?', BBC News Online, December 3rd:
http://news.bbc.co.uk/1/hi/technology/7761153.stm
(Accessed 30.11.10)

Ekklesia, (2009) 'Daily Mail removes poll backing priority for gypsies in NHS', Ekklesia, June 20th:
http://www.ekklesia.co.uk/news/daily_mail_poll_gypsies
(Accessed 30.11.10)

Friend, Cecilia, and Singer, Jane (2007) Online journalism ethics : traditions and transitions. New York: M.E. Sharpe

Frost, Chris (2000) Media Ethics and Self-Regulation. London: Longman
(Accessed 30.11.10)

Harcup, Tony (2007) The ethical journalist : making responsible decisions in the pursuit of news. London: Sage.

Hundal, Sunny (2010) 'Jan Moir wages class-war on Clegg family as Daily Mail removes articles', Liberal Conspiracy, April 26th:
http://liberalconspiracy.org/media/2010/04/26/jan-moir-wages-class-war-on-clegg-family-as-daily-mail-removes-articles/
(Accessed 30.11.10)

Lester, Will (2005) “Survey: many people think journalists ignore mistakes”, Register-Guard (Associated Press Report), 25 May, p A-4.

Littlemore, Richard (2010) 'Monckton bashing "left" and "right"', desmogblog. June 9th:
http://www.desmogblog.com/monkton-bashing-left-and-right

Maier, Scott R. (2002) 'Getting It Right? Not in 59 Percent of Stories', Newspaper Research Journal.

Maier, Scott R., (2007) 'Setting the record straight: When the press errs, do corrections follow?', Journalism Practice, 1751-2794, Volume 1, Issue 1, Pages 33 – 43.

Mayes, Ian (2004) 'Trust me — I'm an ombudsman', British Journalism Review Vol. 15, No. 2, pages 65-70.

Mayes, Ian (2006) 'Open door: The readers' editor on ... why requests to alter the archive are resisted', The Guardian, February 27th.

Mayes, Ian (1999) 'Open door: No getting away with it', The Guardian, January 16th.

Monbiot, George (2010) 'James Delingpole leads Telegraph into vicious climate over email', The Guardian, January 27th:
http://www.guardian.co.uk/environment/georgemonbiot/2010/jan/27/james-delingpole-climate-change-denial
(Accessed 30.11.10)

Moore, Martin (2007) “Transparency and news organisations”, Media Standards Trust, June 15th:
http://martinjemoore.com/transparency-and-news-organisations/
(Accessed 30.11.10)

Nadarajan , Belinda and Ang, Penghwa (1999) Credibility and Journalism on the Internet How: Online Newspapers Handle Errors and Corrections. Paper submitted to the Communication Technology and AEJMC for the 1999 AEJMC Convention:
http://list.msu.edu/cgi-bin/wa?A2=ind9909B&L=aejmc&P=R15182
(Accessed 30.11.10)

Newspaper Publishers' Association (2010) 'PROPOSAL ON THE COLLECTION AND PRESERVATION OF UK OFFLINE AND MICROFORM PUBLICATIONS AND UK ONLINE PUBLICATIONS AVAILABLE FREE OF CHARGE AND WITHOUT ACCESS RESTRICTIONS, March:
http://www.culture.gov.uk/images/consultation_responses/NPAPA_response_Digital_Legal_Deposit.rtf
(Accessed 30.11.10)

Neilsen, Jakob (1999) Designing Web Usability: The Practice of Simplicity, New Riders Publishing: Indianapolis.

Pritchard, Stephen, (2008) 'Holding ourselves accountable', British Journalism Review, December; vol. 19, 4: pp. 63-67.

Rusbridger, Alan (1997) 'BUILDING A RELATIONSHIP BASED ON TRUST; Editor ALAN RUSBRIDGER announces the appointment of a new ombudsman as part of a radical plan to strengthen the paper's relationship with its readers', The Guardian, October 3rd.

Shepard, A.C. (1998) 'To Err Is Human, To Correct Divine', American Journalism Review, June:
http://www.ajr.org/article.asp?id=1612
(Accessed 30.11.10)

Thompson, Mark (2004) 'To Fix or Not to Fix: Online Corrections Policies Vary Widely', Online Journalism Review, July 28th.

'Zee' (2010) ' DailyFail: Newspaper quotes (Fake) Steve Jobs tweet in story about iPhone 4 recall DailyFail: Newspaper quotes (Fake) Steve Jobs tweet ...', The Next Web, June 27th:
http://thenextweb.com/uk/2010/06/27/dailyfail-newspaper-quotes-fake-steve-jobs-tweet-in-headline-about-iphone-4-recall/
(Accessed 30.11.10)

APPENDIX II: SEARCH ENGINE OPTIMISATION IN UK NEWS PRODUCTION

This paper represents an exploratory study into an emerging culture in UK online newsrooms – the practice of Search Engine Optimisation (SEO), and its impact on news production. Comprising a short-term participant observational case study at a national online news publisher, and a series of semi-structured, depth interviews with SEO professionals at three further UK media organisations, the author sets out to establish how SEO is operationalised in the newsroom, and what consequences these practices have for online news production. SEO practice is found to be varied, and application is not universal. Not all in the UK media are making the most of SEO, even though some publishers take a highly sophisticated approach. Efforts are constrained by time, resources and management support, as well as off-page technical issues. SEO policy is found, in some cases, to inform editorial policy, but there is resistance to the principal of SEO driving decision making. Several themes are established which call for further research.

KEYWORDS Search Engine Optimisation; SEO; news production; online news; search; editorial standards

Introduction

The UK media have been active online for around 15 years, but it is only in the last few years that newsrooms have seen SEO practice having a direct impact on journalistic work-flow. The incorporation of techniques designed to ensure high ranking in search engine results pages (SERPs) means that journalistic output is increasingly written not for the benefit of the audience per se, but for those popular search algorithms which determine 'relevance'. This raises a number of important questions with regard to traditional news values as applied in the selection, creation, presentation and distribution of news output. As competition for readership moves away from the news-stand and towards the browser, this research aims to trace how SEO, which seeks to drive organisational behaviour (Moran and Hunt, 2006) is informing UK journalism, and what consequences SEO is having for our news and wider culture.

Literature

This research overlaps by degrees with MacGregor's (2007) study concerning the influence of server tracking data in the newsroom. MacGregor sets out three strands in the literature relevant to his study; journalists' attitudes to their audience, market pressures in the web era, and news values. While the author acknowledges the relevance of these themes to the present study, he chooses instead to focus on scholarship in SEO, and an emerging 'rise of the clickstream' literature which addresses those areas where market pressures impact on journalistic values and routines.

SEO: an introduction

SEO is the blanket term for a range of practices whose end-purpose is to increase organic search traffic referrals to web sites. It comprises four elements; keyword research, search engine indexing, on-page optimisation and off-page optimisation (Malaga, 2010). In theory therefore, the processes involved in SEO, from establishing appropriate keywords with which to frame (or inspire) stories, to developing relationships with external sources with whom to share links (link-bulding) can be significant to newsgathering, news production, and news distribution – the full online news production chain.

SEO is predicated on the assumption that people search with intent, rather than allowing serendipity to guide their interests, and that by tracking keyword search referrals, web publishers can gain an insight into the psychology of the online audience (Enge et al, 2010). The process offers a means of tracking the language and idioms used by search engine users, which can then be mirrored in the language and idioms used by web publishers in order to maximise search traffic.

Search engines are an important source of traffic to many news websites today. In May 2009, telegraph.co.uk received around a third of its traffic from search (Coles, May 2009), down from July 2008, when the figure was said to be around 50% (Charlton, 24/07/08), but nevertheless still significant.

High ranking in SERPs can have a significant impact on search-driven traffic. Around 62% of search engine users click on a result in the first page of results, and 90% within the first three pages (iProspect and Jupiter Research, 2006). A previous study of AOL server logs found that 42% of users tend to click on the first placed result in SERPs, while only 12% click on the second placed result, and 9% on the third. The first ten results receiving 90% of all click-through traffic, and the second page only just over 4% (Enge et al, 2010).

Ranking factors

Google is the pre-eminent search engine in the UK. In July 2010, Google domains accounted for more than 90% of the UK search market (Experian Hitwise Data Centre, 2010). Google claims to use over 200 'signals' in addition to their patented PageRank algorithm, in order to determine the ranking of web pages. But beyond providing general guidance and support on how to optimise for search, Google is evasive both on what these 'signals' are, and on how they are weighted in ranking. The lack of transparency around this issue stems from a fear of giving web publishers free reign to 'game' the search engine ranking system (Moran and Hunt, 2006).

In the absence of official guidance, the biennial SEOMoz search engine ranking factors survey, which aggregates the views of 72 of the world's leading SEO practitioners (SEOMoz, 2009), offers practice-based estimations. The key ranking factors in this study include:

10. Keyword used in HTML title tags (online news headlines are normally placed in title tags)
11. Anchor text used in inbound links: where external sites use rich, descriptive words when embedding hyper-links to a site, this site will experience a higher ranking.
12. Global link authority of site: the authority (in ranking) of any given page in a site is influenced by the site's overall link authority (derived at via PageRank). The PageRank for a news domain is therefore a significant factor in the ranking of each of its individual stories.
13. Age of site: sites which have developed their link authority over a longer period of time generally have a higher PageRank.
14. Link popularity within the site's internal link structure: well-structured collections of links within websites are used by search engines to gain a

sense of what publishers themselves consider their 'most important' content. Relevant 'related links' sections, and well-structured topic (subject) pages contribute towards improved ranking.

15. Topical relevance of inbound links: Reciprocal (and 'relevant') links between online sources who are concerned with the same or similar topics will boost the search ranking of both sources.
16. Keyword use in body text: keywords and synonyms carried over from Title content into body text will boost ranking. Keyword use in file names, and in other fields (such as picture captions) are also ranking factors, albeit less significant.

The speed at which users navigate back to search results (the less time, the less 'relevant' the link clicked on is considered to be in terms of the search terms used) can be used in ranking. User data (or 'contextual relevance') including a surfer's location, search history, profile data (where applicable), and language preference are also weighted. There is disagreement on the degree to which these factors impact on relevance. Only the top three factors in this survey were agreed upon by over 60% of respondents, and indeed the degree of consensus about each of the other factors varies considerably.

Some factors have either declined in significance over the years, or are no longer considered significant at all. For example, keywords placed in the the HTML keywords tag no longer play any part in Google's ranking algorithm (Cutts, 2009), and keyword density, which used to be considered significant (Moran and Hunt, 2006), is no longer considered relevant.

Many of the UK's leading online news organisations now employ SEO specialists – The Telegraph have an SEO executive (Press Gazette, 2010), while the Daily Mail appointed an SEO manager in July 2008 (Oliver, 2008). SEO can be controversial – The Mirror's Associate Editor Matt Kelly has publicly criticised both its principles and practitioners (Andrews, 2009), and indeed, accusations have circulated concerning unethical (or 'black hat') SEO practices in the UK media (McGaffin, 2005; Cook, 2010).

On the other hand, the keyword research and trend-spotting elements of SEO can also be used to inspire original journalism (Broomhall, 2009), as well as offering insights into wider social trends (Tancer, 2009). For Richmond (2008), the rules of SEO should be as important as the 'Five Ws' and house style in the construction of online journalism, while for others it is considered more important yet (Niles, 2010). The online medium, Richmond argues, merits its own conventions, rather than re-packaged conventions carried over from print media. Basic on-page SEO techniques are in his analysis, beyond debate in ethical terms, and the rules of SEO are value-neutral (Richmond, 2008). But this belies the fact that 'relevance' in search is a quantitative assumption, weighted by a trademarked algorithm owned by a powerful corporate entity: Google.

The rise of the 'click-stream' in newsrooms

The interpretation of search data in newsrooms can be problematic. Outing (2005) demonstrates that traffic statistics do not exist in a vacuum, and that a surge in traffic to news sites can result from a link from a popular site, or a prominent ranking in a search index. Outing does not consider the influence SEO techniques can have in establishing such links, and in ranking.

Both Outing and MacGregor (2007) found that online news editors in their respective studies were exerting caution as to the use of (and conclusions drawn from) readership metrics. MacGregor shows that traditional news (and brand) values represent a counter-weight to commercial imperatives brought about by the increasing access to (and prominence of) click-stream data in newsrooms – the potential dangers of which were first articulated by Cohen (2002), in a re-consideration of McManus (1994). Nevertheless, while the 'nose for news' is ceding power to reason, ratings and reaction, otherwise termed a 'culture of the click' (Anderson, 2008), in practice server data is not found to be unbalancing the relationship between an essentially passive audience, and the 'gatekeeping' editor.

Dissenting in part from this view, Lee-Wright (2009) has established that page hits influence story priority at BBC News Online, even though the lack of supporting data are acknowledged to skew assessment. Lee-Wright noted a disjuncture between “audience predilection and the News editors' perception of news priority” (Lee-Wright, in Fenton, Ed, 2009, p77), whereby 'hard' news stories continue to be prioritised over 'soft', more popular stories in terms of placement on topic pages.

This more nuanced interpretation is taken further by Currah (2009), whose research bears witness to a "growing tension between editorial values and knowledge of what will actually generate revenue" (Currah, 2009, p48), with the rise of what he terms 'the clickstream'. Currah's research gives voice to a central concern in SEO. While some feel search and traffic data can improve the quality of journalistic output by providing editors and journalists with a means of knowing what their readers want, others fear less popular public-interest stories will cede ground to "populist, click-friendly topics" (Currah, 2009, p48) in news-gathering and editorial decision making.

Currah maintains that the increasing visibility of click-stream data in the newsroom exerts pressure on journalists "to justify the value and relevance of news stories to the brand and the bottom line" (Currah, 2009, p78). As people increasingly go online for their news, the click-stream will "assume an even more important role in the shaping of the news agenda" (Currah, 2009, p88). Those media who opt to pursue the large volumes of traffic associated with soft, populist content, should be careful not to "erode the distinctiveness of the brand and its connection to a specific audience" (Currah, 2009, p89). Currah's analysis does not, however, accommodate the possibilities of diverging off- and on-line news brands.

Methodology

This research employs two methodologies; a short-term participant observation case study, and three series of semi-structured, depth interviews. All subjects have some grounding in the theoretical principles of SEO, though by varying degrees of technique and application. They are all SEO practitioners working in the UK news media, whose labours are exclusively online.

In total 10 hours of interactive observations were undertaken with BBC News Online SEO executive Martin Asser over three separate days: 02, 16 and 17 June 2010. These were followed by a post-observation interview conducted via email, sent on 27 July. During the period of observation, Martin provided several files and data, which allowed the author to check for bias in observed behaviour. The observation was designed to coincide with a discrete SEO campaign at BBC News Online – coverage and treatment of the group stages of the World Cup 2010.

Three semi-structured, depth interviews were undertaken with two SEO staff at The Guardian; Paul Roach (Technical Lead for SEO) and Chris Moran (SEO editorial executive). One telephone-based interview with Paul Roach took place on October 08, 2009, and two face-to-face interviews with Paul and Chris took place on November 20, 2009, and June 18, 2010. Two semi-structured interviews were undertaken with Northcliffe Digital Director Robert Hardie (on June 23 and July 07, 2010) with added input from Northcliffe's Head of Audience Response and Optimisation via email. A further semi-structured, depth interview was undertaken with an anonymous online journalist at an undisclosed national news site on July 01, 2010. In total, these interviews comprised 10 hours. Short-hand notes were taken during observations and interviews, and findings were summarised from all notes in around 11,000 words.

The subjects chosen for this research were not selected at random – the field of SEO professionals in UK news is small, and the author was guided primarily by access.

Questions asked during observation and interviews covered many areas, including:

- How long an SEO policy has been in place,
- How SEO policy is operationalised and applied
- What (if any) impact SEO has had on traditional journalistic work
- The extent to which SEO informs editorial
- How search traffic (and search terms) are measured and audited
- What (if any) resources are used to track 'buzz' online
- Whether or not stories have been inspired by search data
- Whether there is a linking and/or link-building strategy in place
- What (if any) are the consequences of SEO on news language, and on news values.

These research themes were developed iteratively, and framed according to subjects' experience of terminology and process. They were informed by the literature in SEO, as well as by the author's knowledge as a consultant to industry. A mixture of open- and closed-questions was used. The weighting of responses is broadly proportionate to the range of participants, but reflects individual concerns.

Participant observation case study findings

Background: BBC SEO policy, and the SEO Specialist

In April 2009, representatives from the BBC's Future Media and Technology Journalism (FM&T J) and Editorial Development (News) divisions worked on a trial SEO project for implementation across the World News team, focussing on the Middle East index page. Following a successful up-turn in traffic during this period, the chief implementer (Martin Asser - the subject for this research) was appointed SEO Specialist, on a short-term, rolling contract. Martin reports to the Head of Editorial Standards in BBC News, though his working relationship with those in FM&T J has been complicated by organisational change. Between October 2009 and March 2010, around 600 BBC journalists were trained in basic, on-page SEO theory and practice. Martin's job is, as he sees it, to monitor output and cascade best practice down to sub editors and journalists.

Day-to-day: keeping track of referrals, terms and competition

Martin uses various tools in his SEO work. The BBC has an account with Experian Hitwise, which provides details of performance on downstream clicks from Google and Google News (the two greatest sources of traffic to BBC News Online by a significant margin), against competitors (Guardian.co.uk and Telegraph.co.uk). Other competitors (such as the Independent.co.uk, theSun.co.uk, and DailyMail.co.uk) are also factored into analysis. Bespoke graphs are used for specific SEO projects – factoring in other competitors in niche fields.

Search referral data from an in-house system called Sage Analyst is used to inform Martin's keyword research. This data is neither provided in a user-friendly way, nor is it exhaustive for SEO needs. There is no breakdown of search terms by UK region (or indeed by country). The system does not appear to provide analysis of IP addresses, so it is not possible for Martin to monitor the proportion of visits that are new and returning. All BBC journalists have access to audience metrics via the "Livestats" service. A dashboard provides interactive heat maps of click-through for the top 10 pages on the BBC news site, which shows Martin how well headlines on index pages are performing in real-time.

Martin undertakes sporadic competitor analysis, tracking the development of a particular story in terms of ranking performance in Google and Google News, but he does not have time to do this in a more systematic way. Tools for keyword research and monitoring search and social web trends are not generally used: "the philosophy is not to follow trends... it is to continue following our news agenda as we have always seen it, but to make sure we are availing ourselves of current SEO techniques in order to get the maximum amount of traffic from search for the agenda that we are following" (Martin Asser, personal communication, July 28 2010).

An instinctive knowledge of BBC news style drives SEO policy at the BBC, and preferred spellings and transliterations from BBC Monitoring take precedence. The BBC do not have a uniform, written SEO styleguide which journalists can consult. Martin will never alter headlines directly – this is editorially forbidden.

BBC News has no active presence on YouTube, though its videos are available via some other third party aggregators. Martin is fully aware that his SEO activity, in particular his ability to grow search referrals, is curtailed by this lack of presence, but the BBC has significant rights issues to consider. Martin has occasional meetings with BBC News' head of social media but they do not work closely. This compromises his ability to fully grasp the potential in SEO for planning, adapting to and developing story ideas around online 'buzz'. But having monitored search and social web trends in the past, Martin believes that many of the keyword topics which trend tend to be of little value to BBC News audiences.

SEO in practice

Martin spends most of his time working at the sub-editors desk in the BBC News hub. His day-to-day work on SEO is compromised by other project work which is considered more pressing. Throughout the observations, Martin was involved in dealing with problems with the latest release of the BBC News content management system, which he had been involved in for several weeks. Martin explained that this causes him frustration, because it means SEO is neglected. Martin does not attend editorial meetings, as he doesn't have the time, but he is well aware of any given day's news agenda.

When not working on particular SEO projects, Martin is concerned with fine-tuning headlines on stories linked 'above the fold' (prominently) on index (topical) pages. He checks that headlines are adequately keyword-rich and well-structured, and that links in 'related links' sections are up-to-date. This work is very much about influencing practice, rather than ranking.

Martin's project work takes a news diary approach. The pre-election budget in March, the general election in May, and the World Cup in June, all formed the basis for discrete SEO projects in the Business, Politics and Sports online teams respectively, throughout 2010. By choosing forward-planned stories around which to concentrate efforts, Martin is working around what could be natural increases in the volume of searchers seeking out BBC information, which could make it difficult to attribute search performance to SEO actions taken.

BBC News SEO policy aims are to increase the percentage of visits from search, and the percentage of total visits more generally. Without any conscious effort in SEO (prior to current arrangements), BBC News Online has achieved a PageRank matched only by the Guardian in UK media – reflecting the age of this site, and its wider online authority. Martin sees SEO as part of the public service approach the BBC must take. SEO is undertaken in the interests of the license fee payer, and in the wider public interest.

All of Martin's communications on SEO are by way of suggestion and promotion, though he does have the authority to communicate his concerns to senior news management. When Martin sends out emails about headlines, he will often send them to the senior journalist sub-editors group, carbon-copying in the author of the piece. Martin has rarely experienced resistance or criticism, and being a journalist by trade, he is sensitive to his colleagues' working environment.

SEO and technical issues

The first sentence of any news story is automatically inserted into Meta 'description' tags by the in-house Content Management System (CMS), but the keywords tag is not used. Keywords are not used in naming conventions for the thumbnail pictures which accompany stories, even though the Google Images vertical is one of the top five referrers for BBC News content. Stories cannot be filed in the BBC News CMS without alternative descriptive text (in keyword form) for all images.

Since November 2009, BBC News has run dual headlines for every story – a short headline (31-33 characters) which features on index pages and Ceefax, and a longer headline (up to 55 characters) on story pages. This approach was taken in order to accommodate more keywords in headline fields. Index (short) headlines don't have to include keywords so long as they give what Martin calls a "whiff" (Martin Asser, observation, July 2, 2010) of the news. Through experience, Martin believes these headlines have little bearing on ranking. Some off-page SEO issues are felt to constrain Martin's efforts. The use of unique numbers rather than keywords in BBC News Online URLs (web addresses) is a legacy issue over which he has no control. Likewise, at the time of writing there is no site map (a file used by high-output publishers to improve SEO by automatically notifying search engines about new content), though this will soon change.

SEO project – World Cup 2010

Before the World Cup started, Martin had meetings with key editorial figures in BBC Sport to discuss SEO issues. A week into the World Cup, Martin put together a 'dashboard' of metrics on how the site would be monitored on a weekly basis throughout the campaign – made available to Sports staff. Martin used June 01 (10 days before the tournament started) as the benchmark for comparisons – he acknowledged that this is “arbitrary” (Martin Asser, observation, June 17, 2010) but felt he had to start somewhere.

After the first week of the tournament, Martin sent a round-up email to subs and other senior journalists outlining performance. The proportion of search (as a total of page views) peaked at 6.1%, averaging at around 4.5% throughout the week, around 2% higher than pre-tournament referral traffic. Martin was pleased that the second and third highest referring keywords on the June 14 were BBC "branded searches" (Asser, observation, June 17, 2010) – navigational searches including the term 'bbc' (these were 'bbc world cup' and 'BBC world cup 2010'). These navigational searches are of value, as they imply searchers are actively seeking out BBC News Online.

During the second day of observation, Martin sorted through the top 250 search terms, for the June 12 and 13. Those keywords yielding 30 searches or less were discounted, and the rest integrated into a pre-existing keyword list. Martin passed over those terms which have stopped trending since his monitoring began, so there is a danger he misses out terms whose popularity ebbs and flows over a longer period of time.

Martin's second weekly email to Sport news staff (sent on June 17) included a potential newsgathering issue:

'brazil fans – 221 searches on the 14th'

'brazil fans – 763 searches on the 15th'

Martin drew this issue to the attention of the Sports team, but in a very informal way

– with good reason. A search of Google Images showed that this search term returns images of near-naked female Brazil fans. Martin does not feel it is his job to suggest editorial treatments of data like this – he is happy just to supply the information, and let journalists and editors interpret it as they would any other source.

Kickers

Martin has found that inserting trending keywords into kickers (a print convention, whereby headlines are prefixed with a two- or three-word clause followed by a colon) across major on-going news stories can have a beneficial effect on traffic. He says he feels that this approach has been useful in instilling discipline in the process of headline writing. On June 14, the third most popular search term was 'BBC World Cup 2010'. The kicker 'World Cup 2010:' was applied to much World Cup coverage in BBC Sport, and proved to be popular across much of the UK's online media for several weeks (subsequent research showed some UK media had been using 'World Cup 2010:' in headlines since the first week of May 2010). Martin acknowledges the negative impact this approach has on SERPs for world cup content. He agrees that these Google News results are difficult to interpret – they are "a sea of information" (Asser, observation, June 17, 2010) – but he feels he is locked into an arms race. He sees that he has a public duty to compete, and that his competitors have a commercial need to compete, but can see no way around this problem, other than to vary their use from time to time, and so potentially lose traffic.

SEO and 'hard to find news'

On the third day of observation, Martin worked on a story about CCTV cameras in Birmingham that were being covered over with plastic bags to render them ineffective. Martin spent much time trying to re-write the original headline which started with the sub-optimal keyword 'Plastic bags'. It proved to be very difficult to re-write – all of the keyword elements were conceptually difficult. The cameras could be referred to as 'terror cameras', 'cctv cameras', 'surveillance cameras', or even 'spy cameras' – as was found in past BBC news output, and amongst the BBC's competitors.

Martin concluded that the headline was far from ideal, but also that it was a very difficult story to bring out in headline keywords – there were so many options it would have taken half an hour to consult keyword tools, and the competition, to check them all. Some stories, he acknowledged, are hard to optimise for, even relatively 'big' stories (this story was the second most prioritised story on the UK News index at the time).

SEO and the politics of news language

I ask Martin for his views on a story from late May 2010 concerning the murder of three prostitutes in West Yorkshire. I suggest that the prominent use of the keyword 'prostitute' in headlines might imply SEO values are impinging on the ethics of formal style across the media. Martin says he has no desire to be "prurient" (Asser, observation, June 17, 2010) in his pursuit of increased traffic. This remains essentially an editorial issue.

Internal linking, Link-building and linking out

Martin has input into the naming conventions used within the BBC News Online information architecture. During the third day of observation, the BBC news topic page concerning BP's Deepwater Horizon oil spill was called simply 'Oil and water' (albeit, the title for this page was 'Oil Disaster'). The visual heading for this page was inspired by the branding for a World Service radio series on the disaster by this name. However, accommodating a much wider pool of BBC multi-media content, the page subsequently reverted to the far more SEO-friendly title 'BP oil disaster'. Martin is not involved in any formal link-building (establishing contact with other sites in order to share links, to boost ranking). On the BBC's general policy on linking out, Martin said that they are at "a crossroads" (Asser, observation, June 17, 2010). The public service brief informs a policy of increasing in-line (body text) and deep (direct) linking, but there has been some concern that this could lose the BBC traffic.

Interview findings

Background

The Guardian's in-house SEO team was formed in 2008, and comprises three members including; Paul Roach (Technical Lead for SEO, and SEO by trade) who is based in Technical, and Chris Moran (SEO Editorial Executive, and sub-editor), who is based in Editorial. At Northcliffe, an SEO specialist has been in place since 2006, working on the 'thisis' sites to Digital Director Robert Hardie. An undisclosed media group's SEO history is not shared (hereafter referred to as Y, and the SEO journalist subject X).

SEO culture in the news room

All Northcliffe and Guardian journalists are trained in SEO, but most SEO efforts are undertaken by Guardian sub-editors, Northcliffe Digital Editors and Y's sub editors respectively. There is no blanket SEO policy nor catch-all guide for writers at any of these organisations – and for Guardian and Northcliffe copy, optimisation is prioritised for those stories featuring on the home page, or prominently on topic (index) pages. At the Guardian, Reviews and Sport have particular SEO requirements.

Not everyone sees the importance of SEO to the same degree within these organisations. Guardian sub-editors prioritise headlines, standfirsts, and picture captions, though journalists select suitable (internal and external) related links. Body text is not a priority for SEO treatment at any of the three organisations.   Both Paul and Chris are wary of impacting the reading quality of Guardian copy by forcing (and overly repeating) keywords. Topical in-body links to internal keyword pages are automatically generated on all three sites. At the Guardian, the SEO team aim to look beyond isolated stories towards creating topically-consistent hyper-linked portmanteaus, to improve the site's wider optimisation. Topic (subject) pages are generated automatically at the Guardian, and at Northcliffe.

Northcliffe's general approach is to lead with real, intuitive language. The natural use of synonyms is encouraged, albeit top-ranking keywords will feature most prominently. Northcliffe Digital Publishers manually re-work automatically generated topics pages, as well as on-page SEO and link-building for key stories.

SEO is undertaken at Y by sub-editors, not all of whom are enthusiasts. There are no consistent rules on key-wording in HTML meta tags or picture captions.   A significant amount of news (including picture-led news) on Y is wires copy, which is not re-worked for SEO purposes. X estimates that one person will have around an hour to optimise the day's content – with a heavy emphasis on checking for errors in headlines.

Monitoring search traffic

The Guardian's SEO team have access to a Hitwise account, and all three organisations have in-house traffic referral monitoring consoles. At the Guardian, search traffic analysis is ephemeral, though monthly analysis is also carried out. Key metrics include "Unique users, page views, visits to various areas of the site from specific referrer types and level of engagement of each" (Paul Roach, interview, June 18, 2010). Engagement is measured in several ways; from dwell time on site, through to sign-up, and sign-in to online community accounts, and even conversion to products bought online. Most literature conceives the end-goal of SEO to be a conversion or transaction (Kent, 2008, Thurow and Musica, 2009, Enge et al, 2010), but for the Guardian, the equivalent is a unique visit, followed by click-through to other pages, followed by a subsequent visit from this IP address to guardian.co.uk, via a navigational search (ie 'guardian football'). The difference in click-through between first and second place in search results is considered to be hugely significant, especially when it comes to the 'one-box' Google News vertical results found in Google search results for trending terms. The Guardian SEO team is therefore very proactive in its competitor analysis.

At Northcliffe, search traffic reports are analysed monthly, but there is a daily headline report too. Northcliffe use a combination of Hitwise and their own in-house Audience Dynamics research, data from which is shared with all Digital Publishers.

X and his team are aware of internal statistics throughout the day. Day-to-day search and third-party referral key words are not considered important – responding quickly to web trends and topics is not something Y has the resources, inclination or time to do. Neither is there time for competitor analysis. News carried in Y is published on the website at midnight. These are print stories without SEO fine-tuning which are retrospectively checked, optimised, placed and adjusted on the home page, or landing pages in the hours following publication.

Editorial policy and social web and search trends

Chris acts as SEO representative at Guardian digital editors meetings. At these meetings, and in his communications, he is very careful to only impart trending topics where they have (as he perceives) a resonance with Guardian readers. He is conscious that if he were to be seen to be pushing irrelevant topics, he may loose interest and hence influence within the organisation. This selective approach "leaves traffic on the table" (Chris Moran, interview, June 18, 2010) for competitors, but this is considered a price worth paying to maintain the Guardian brand. The SEO team's contribution to editorial comprises three strands: ensuring staff understand the culture of online news consumption, helping them optimise their content for search, and helping to identify opportunities in treatment, news-gathering and commissioning.

Chris distributes daily social and search trend updates at The Guardian using several freely available online tools. General web trends are considered useful not just to establish blank-spots in coverage in any given point in the news cycle, but also for helping determine when topics are ephemeral or long-lasting among the searching public.

There are no central web trend communications at Y. X dislikes the idea of being driven by trends, but he is aware of the pressure the promise of such traffic drives in industry. Y's current editor is adamant that heavy-trending stories which are "not right" for Y will not be published (X, interview, July, 2010).

SEO does influence editorial process at Northcliffe directly, but web trends are of minimal use in day-to-day work. Search and social web trends are considered to be too nationally-focused for a local news online publisher. There is no formal daily trends email sent round journalists at Northcliffe – but the Audience Response and Optimisation (ARO) team deal with this on a case-by-case basis, contacting particular digital publishers on specific topics.

SEO and news language

At The Guardian, SEO principles always defer to the in-house styleguide, and so Paul and Chris are not immediately concerned about the rise of jargon in keyword selection. From time to time this is to the detriment of organic traffic. Traditional style conventions in headlines such as 'Formula 1' (as opposed to the more frequently searched 'F1') make optimising for SEO difficult.

Robert, as a former sub-editor, is concerned that search might "reduce the scope of language" (Robert Hardie, interview, July 07, 2010) in news. Any fears about clichés and jargon should be considered in the context of how readers use language, and how they refer to the world around them, rather than blindly following subbing advice from off-line. During the period of interviews, Northcliffe were starting to consider the incorporation of SEO into their styleguide, and Robert was aware that a potential conflict between SEO and conventional writing style is something he will have to confront.

X said he would not use a keyword he didn't think shared the values of Y just to maximise traffic. Use of the SEO-friendly term 'Crossbow cannibal' in referring to the alleged murderer Stephen Griffiths (who used this term to describe himself in Court, in May 2010), led to some editorial soul-searching at Y in terms of showing respect for the victim's families. They opted to run with it, though not from an expressly SEO-driven perspective – this was a decision shared by most UK media. X relies on his news instincts and editorial judgements (via the Y styleguide) when it comes to choosing keywords to research.

When copy-editing for online, X is wary of doing violence to big impact, visual front-page splashes which often use simple, intuitive, but striking headlines. He often ends up re-writing them completely, incorporating keyword-driven pull quotes, sub-headers, and other page furniture, into what he feels can be "a poor imitation" (X, interview, July, 2010). The trick, he suggests, is to "embody the editorially elegance" (X, interview, July, 2010) within such news stories online but this he concedes can be difficult.

Generating news ideas

None of the sources use trend data to inspire news stories or features regularly, and no examples were given of past use.

Kickers

Paul and Chris are opposed to the slavish use of kickers in pursuit of traffic, but acknowledge they have to do it sometimes. Kickers are "the easiest way to ensure the targeted keyword appears in the right order in every article... [but] ...it means that the news headlines are homogenised and you have no nuance within headlines or between publications" (Roach, interview, June 18, 2010). To obviate the problem, they try to "target search terms within the headline and page title in a way that makes sense, is in context and reads well" (Moran, interview, June 18, 2010). Living Stories, a Google experiment in creating interactive pages for big, rolling news events, was mentioned as an example of a possible design-led solution. At present it is felt Google News rewards those publishers who output lots of discrete stories and link well around a key-word relevant topic, rather than bringing ongoing stories around a topic into a rolling narrative on one page. Paul believes this is a crucial issue for the future of SEO in news.

X is aware that kickers are not ideal, and he is averse to the use of phrases in kickers. He feels search engines have a responsibility to ensure that when online publishers follow their terms and conditions (and guidance) to the letter of the law, that the end-user isn't presented with “a sea of near-indistinguishable headlines” (X, interview, July, 2010).

Robert is also aware of a problem with kickers but rather than consider this is an ethical dilemma for publishers, he believes it will be resolved by the market. The public will have to become more attuned to what they want, and so will change their behaviour, or indeed the search engines will have to work a path around this problem, or fail competitively.

Calendar news stories

Paul suggests that SEO has the power to stretch the news agenda around big calendar stories. He is aware that there is a danger of media groups setting out to pre-prepare topic pages for news events months in advance of the actual event taking place, preparing this content more in the pursuit of ranking in future weeks and months, than in informing public awareness.

X regrets this approach, feeling that “a fair fight” (X, interview, July, 2010) on reporting, insight, and opinion at the actual time of the event in question, should be the primary influence on ranking. X believes this is contributing to information overload.

Hard-to-find news

Paul acknowledges that some stories are harder to find than others, but points out that the Guardian's internal key-wording taxonomy "means that an article will always be linked from several keyword pages as well as edited fronts" (Roach, interview, June 18, 2010). X also acknowledges some stories may be harder to find than others.

Linking

Paul says "Linking out to the relevant sources wherever possible is actively encouraged throughout guardian.co.uk" (Roach, interview, June 18, 2010), including competitor sites. The media "have a duty to link to relevant sources wherever appropriate", but they do not pass link authority on to commercial links or user generated content (Roach, interview, June 18, 2010). Robert sees external linking as an important way of harnessing the full power of the internet, though with similar caveats. When it comes to link-building, Northcliffe publish on average two to three stories per week, which are submitted to Fark.com (an aggregator of 'quirky' news). Northcliffe use a range of third party sites in order to build traffic, establish brand, and develop reciprocal relationships in a number of niche topical areas. Robert mentions a recent example of a local story concerning Cliff Richard, whereby one of his Digital Publishers spent around two hours researching and distributing this story around relevant fan sites and forums, to boost traffic to this page.

Analysis and conclusions

SEO is still very much in its infancy in the UK media, with some organisations still focusing on influencing culture, rather than applying policy. SEO activities are constrained by time, resources and support. Even when a specialist in place, SEO can defer to other tasks and projects. Some technical, rights-based and off-page SEO factors effectively tie the hands of SEO practitioners in terms of maximising traffic. In short, SEO is some way short of being institutionalised across the UK media.

The culture of SEO in these organisations is visible, but not dogmatic. This research found only one example of a media organisation where SEO directs editorial decision-making, for all others SEO is acknowledged to inform (by varying degrees), but never lead editorial, with significant resistance to this idea in some quarters. It is acknowledged that in yielding to editorial (and style) conventions these organisations are losing out on traffic. But this is broadly considered a price worth paying in order to preserve the brand, and maintain the news values of the profession.

Search traffic measurement tools are used across all organisations studied, but some organisations have a clearer sense of how they are performing in SEO terms than others. Search and social web trends are not being routinely used to identify original areas for reporting, but they do inform news-gathering. Other SEO tools, including keyword research tools, and search or social web trend measuring tools, are not uniformly used in industry. On the contrary, language, style and ethical conventions take absolute precedence over the language and idioms people use to search.

The perceived bias towards 'soft' (including celebrity, entertainment and 'quirky') topics in search and social web trends is considered unfit for sharing with some journalists, and by extension with news audiences. Linking out is considered important amongst all the subjects I spoke to, but with caveats. The process of link-building is not a priority in national news organisations, but a local news publisher was proactive in attempting to build links with external sources towards driving traffic.

This research shows that approaches to SEO go beyond mere 'common sense'. Standards in journalistic writing reflect best practice and ethical concerns, and help publishers pull together a consistent news brand. But this research demonstrates that SEO is applying pressure to these standards not in the perceived interest of the reader or because of publication constraints (as once may have been the case), but in the interests of a third-party commercial arbiter in online distribution: Google.

A new conformity in the language and 'aboutness' of news online is emerging, mediated by Google. Google stands between those terms used by the reading public and journalist alike, implicitly regulating the 'market place of ideas' as it goes. Relevance as a quantitative measure is challenging what people feel qualitative 'relevance' to be. But for now at least, some of the UK's leading news brands are refusing to bow entirely to the demand for more and more search traffic.

This research uncovered various themes which merit further investigation.

The first relates to the use of trending keywords in kickers to prefix ongoing news headlines online. Kickers are an old media convention given new life and significance online. News organisations use this blunt tool because Google is perceived to reward their use, leading to an arms race whereby the surfing public suffers. There is a danger that news organisations will damage both their own brand, and the brands of leading search engines, if they do not reign in indiscriminate use. It has been suggested that were Google to reward news organisations who publish big news themes on one page (as opposed to a myriad of small, discrete stories), that this may obviate the problem. It is prescient that some in industry are calling for an end to the journalistic convention of the 'story' in online journalism (Marsh, in Miller Ed, 2009) at a time when such an approach might help lead to a solution to this problem.

Another theme relates to the increasingly early development of pages for big calender news events, designed with the intent of improving ranking nearer the time the news event takes place, as much as informing the public. This represents another potential arms race in news SEO. This approach has the potential to stretch the definition of 'news' to breaking point, while contributing yet further to the problem of information overload.

Thirdly, SEO professionals accept that some news stories are harder to find online than others by their very nature. Some stories do not lend themselves to simple keyword analysis, and without a firm editorial steer, these stories could be relegated in priority, in the same way that less visual stories are perceived (by some) to have less priority in broadcast news.

Present day SEO practices may be looked back upon in the decades ahead as a relic of early developments in search, as we move towards a 'semantic web'. But SEO will be with us for some time yet, and this research serves merely as a starting point for further study in this field.

References

Andrews, Robert (2009) “Daily Mirror's Matt Kelly: put SEO in its place”, **The Guardian**, 2nd December:
http://www.guardian.co.uk/media/pda/2009/dec/02/mike-kelly-seo-journalism-world-newspaper-congress-keynote accessed 17 December 2010

Anderson, Chris (2009) “Web Production, News Judgement, and Emerging Categories of Online Newswork in Metropolitan Journalism”, presented at the Online Journalism Symposium, University of Texas, Austin, 18 April.

Brooker, Charlie (2008) “Online POKER marketing could spell the NAKED end of VIAGRA journalism as we LOHAN know it”, *The Guardian*, 21 July: http://www.guardian.co.uk/commentisfree/2008/jul/21/charliebrooker.pressandpublishing accessed 17 December 2010

Broomhall, Drew (2009) “Building content and traffic using search trends”, presentation for the Association of Online Publishers SEO Forum, London, 25 March.

Charlton, Graham (2008) “Site review: Telegraph.co.uk revamp”, *Econsultancy*, 24 July: http://econsultancy.com/blog/2628-site-review-telegraph-co-uk-revamp accessed 17 December 2010

Cohen, Elisa (2002) “Online Journalism as Market-Driven Journalism”, *Journal of Broadcasting & Electronic Media*, Volume 46, Issue 4 December, pp 532 – 548.

Coles, Malcolm (2009) “Find and Share Information on the Internet : Part 2”, *Fumsi*, May: http://web.fumsi.com/go/article/share/3907 accessed 17 December 2010

Cutts, Matt (2009) “Google doesn’t use the keywords meta tag in web search”, *Matt Cutts Blog*, 21 September: http://www.mattcutts.com/blog/keywords-meta-tag-in-web-search/ accessed 17 December 2010

Enge, Eric; Spencer, Stephen; Fishkin, Rand; and Stricchiola, Jessie (2010) *The Art of SEO: Mastering Search Engine Optimization*, Sebastapol: O"Reilly.

Cook, Mark (2010) “Express Group contacts SEOs to sell links”, *Further.co.uk*, 22 April: http://www.further.co.uk/blog/Express-Group-contacts-SEOs-to-sell-links-269 accessed 17 December 2010

Lee-Wright, Peter (2010) “Culture Shock: New Media and Organizational Change in the BBC” in: Fenton, Natalie (Ed) *New Media, Old News: Journalism & Democracy in the Digital Age*, London: Sage.

Experian Hitwise Data Center (2010), “Top Sites & Engines”: http://www.hitwise.com/uk/datacentre/main/dashboard-7323.html accessed 17 December 2010

iProspect and Jupiter Research (2006) “Search Engine User Behaviour Study”, April: http://www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior.pdf accessed 17 December 2010

Kent, Peter (2008) *Search Engine Optimization for Dummies*, Indianapolis: Wiley.

MacGregor, Phil (2007) “Tracking the Online Audience”, *Journalism Studies*, 8:2, pp 280-298.

Malaga, Ross, (2008) “Worst Practices in Search Engine Optimization”, *Communications*

of the ACM, December, Vol 51. No.2.
Marsh, Kevin (2009) “Death of the Story”, in: Miller, Charles (Ed), *The Future of Journalism: Papers from a conference organised by the BBC College of Journalism in London*, 2008:
http://www.bbc.co.uk/blogs/theeditors/future_of_journalism.pdf
accessed 17 December 2010
McGaffin , Ken (2005) “The Financial Times Website and Hidden Links, *Linking Matters*, 11 June:: http://www.linkingmatters.com/2005/06/financial-times-website-and-hidden
accessed 17 December 2010
McManus, John (1994) *Market Driven Journalism: Let the Citizen Beware?*, London: Sage.
Moran, Mike, and Hunt, Bill (2006) *Search Engine Marketing* (2nd Edition), London: IBM Press.
Niles, Robert (2010) “Student journalists need to learn SEO more than they need AP style”, *Online Journalism Review*, 21 April:
http://www.ojr.org/ojr/people/robert/201004/1843/
accessed 17 December 2010
Oliver, Laura (2008) “Mail Online names Simon Schnieders as search strategy head”, *Journalism.co.uk*, 11 July: http://www.journalism.co.uk/2/articles/531917.php
accessed 17 December 2010
Outing, Steve (2005) “Sophisticated Web Stats Give Editors Better Idea of Reader Interests”, *Convenience Store News*, 26 July.
Richmond, Shane (2008) “How SEO is changing journalism”, *British Journalism Review*, 19;51.
SEOMoZ Ranking Factors (2009):
http://www.seomoz.org/article/search-ranking-factors#ranking-factors
accessed 17 December 2010
Press Gazette (2010) “Telegraph creates 'digital operations unit'”, 20 April:
http://www.pressgazette.co.uk/story.asp?storycode=45325
accessed 17 December 2010
Tancer, Bill (2009) Click: What We Do Online and Why It Matters. London: HarperCollins.
Thurow, Shari and Musica, Nick (2009) *When Search Meets Web Usability*, Berkeley: New Riders.

APPENDIX III: THE RE-BIRTH OF THE 'BEAT': A HYPERLOCAL ONLINE NEWSGATHERING MODEL

Murray Dick

Scholars have long lamented the death of the 'beat' in news journalism. Today's journalists generate more copy than they used to, a deluge of PR releases often keeping them in the office, and away from their communities. Consolidation in industry has dislodged some journalists from their local sources. Yet hyperlocal online activity is thriving if journalists have the time and inclination to engage with it. This paper proposes an exploratory, normative schema intended to help local journalists systematically map and monitor their own hyperlocal online communities and contacts, with the aim of re-establishing local news beats online as networks. This model is, in part, technologically-independent. It encompasses proactive and reactive news-gathering and forward planning approaches. A schema is proposed, developed upon suggested news-gathering frameworks from the literature. These experiences were distilled into an iterative, replicable schema for local journalism. This model was then used to map out two real-world 'beats' for local news-gathering. Journalists working within these local beats were invited to trial the models created. It is hoped that this research will empower journalists by improving their information auditing, and could help re-define journalists' relationship with their online audiences.

KEYWORDS: beats; hyperlocal; news-gathering; network society; online journalism; theory

INTRODUCTION: A DEFINITION OF THE NEWS 'BEAT'

The 'beat' has a rich lexical history. It has been defined in the sense we understand it today, as a 'course habitually traversed by any one'; in the literary canons of two heavyweight journalistic authors; Charles Dickens (in Sketches of Boz, 1836), and later Mark Twain (in Roughing It – 1872) (Oxford English Dictionary, 1989). The term has been applied in natural history to the routines of wild bears (Matthews, 1951), and where man comes into contact with nature, such as “a stretch of water fished by an angler” (Concise Oxford Dictionary, 1999). The concept is redolent of an attempt to impose order, sometimes to a masochistic degree; the expression 'beating the bounds', owes its meaning to the proce of tracing “...boundaries in a perambulation, certain objects in the line of journey being formally struck, and sometimes also boys whipped to make them remember” (MacDonald, 1977, p113).

Later definitions move away from geographical space, toward something most journalists can relate to in professional life (albeit not a turn of phrase used often in current times); a piece of news published in advance of a journalists' competitors (Oxford English Dictionary, 1989), (Random House, 1987). But the beat is more than a scoop or a geographical locus; it is a social construct, which is under pressure from the economic, political and social forces which have historically shaped it.

A SOCIOLOGY OF THE NEWS 'BEAT'

Fishman (1980), whose research concerned news-gathering and sourcing in crime journalism, observed that the demands of the news schedule (a regular, and consistent flow of stories) had led journalists toward a dependence on bureaucratic organisations to pad out the boundaries of their beats, including; councils, police forces, hospitals and other 'official' local sources. The routines of news-gathering around such sources help journalists plan out their agenda (Harrison, 2006, p141), and lend newspaper copy authority in the form of a proxy for 'objectivity' in such sources (Tuchman, 1972). But this comes at a cost (Foreman, 2010), for cynicism can spread from such rhetorical methods, leading to disillusionment with professional news reporting, as the public become ever more distanced from the 'news' taking place in their communities (Carey, 1999, p58). .

Schlesinger (1987) found that BBC news was dependent on routine sources for up to 80 per cent of its output, while later Franklin and Murphy (1991) established that local press relied upon local government, courts, police, business and voluntary organisations for 67 per cent of their stories (in Keeble, 2009: p114-15).

Behind the ritualisation of sourcing lies a series of consensual and mutually affirming processes, whereby both journalist and source have much to gain – this sustenance of mutual dependence is something which marks early literature in this field (Zelizer, 2004). In his analysis, McManus (1994) argued that this is a market in which the media are an active arbiter; bartering “access to the public to news sources in return for information needed to fill the paper or newscast.' (McManus, 1994 p5). But non-institutional organisations have been found to be taking advantage of the diminution of editorial resources found in many modern newsrooms too (Davis, 2000); which would suggest that this is a market in which social capital is also a recognised currency.

Gans (1979), arguing that efficiency is the mitigating factor which drives all time-poor journalists; formulated six source considerations which may speed this process up:

Past suitability
Productivity
Reliability
Trustworthiness
Authoritativeness
Articulateness (as summarised in Harcup, 2009: p59)

Others have found that such routine-based criteria may not be so important in establishing which voices in a beat are heard most frequently and which are missed; that it is in fact ideological motivations which trump all other factors in the newsroom (Cottle, 2000).

The beat has faced stern challenges it modern times. In the US, the mid-1990s saw 'topic teams' emerge, supplanting beat reporters, and loosening the connection between local audiences and the journalists who serve them. These practices were considered to have a negative effect on news routines, according to the journalists working within them (Hansen et al, 1998). On the other side of the Atlantic, the 1990s saw the rise of new technologies and marketing, which re-enforced the dominance of official voices in news production, but which also brought rapid consolidation (and job cuts) in industry, putting strain on the beat. Fewer journalists were covering the same beats; the consolidation of routines around regional hubs and away from local communities emerged; the profession suffered a brain-drain, and former journalists began boosting the ranks of public relations organisations who, in turn, were increasingly spoon-feeding the remaining time-poor journalists easy-to-use copy (Machin and Niblock, 2006, p70).

Today local beats are very much under sustained pressure. In local government, although performance data is available on a hitherto unimaginable scale, the new decision making processes (post-Local Government Act 2000) have become even more opaque (Morrison, 2010). The steady rise of the council newspaper 'propaganda sheets' at the expense of local newspapers (Gilligan, 2009) has abated, but plans to curb their influence have subsequently been revised (Department for Communities and Local Government, 2011). Court reporting in the UK is considered so vulnerable, that police forces are beginning to publicise court proceedings themselves (BBC News, 2011).

PUBLIC JOURNALISM AND THE NETWORK SOCIETY

The rise of 'public journalism' (or 'civic journalism') in early 1990s America represented an attempt to de-institutionalise the profession. The beat system, with its rituals and practices around private (and protected) relationships, were seen to be out of step with popular opinion, a disaffection stretching back to the Pentagon Papers. Those cosy practices which tie the beat journalist to her sources were viewed as a conspiracy against the audience (Carey, 1999). In seeking to address the dual crises of plummeting trust between citizens and their governments; and plummeting trust between news audiences and news organisations, industry began experimenting with new journalistic forms, and with new methods of engagement with the audience, on a more level-footing (Haas, 2003; Haas and Steiner, 2006).

Today we may view these experiments in the context of the rise of the network society. Castells (2000) argues that today's society is networked and predicated on three interwoven elements; the modern world is informational, global and networked. Where once political power was embodied in hierarchical organisations and institutions (such as journalism), today that power is spread across networks. While some nodes in networks may be 'more important' than others, power relations are flattened into a binary of inclusion or exclusion from within and out-with the network. (Castells, 2000, p15).

The binary nature of the network is problematic; especially where public interest journalism is concerned. Social networks are self-selecting; they do not proportionally reflect wider society in terms of use – only one in 20 of those over 65 engage with social media frequently, and less than a third of 35-44 year-olds are active online (Ofcom, 2010). Journalists should be careful how they develop their networks, as exclusivity within an online network bodes ill for media pluralism.

Nevertheless, the process of re-calibrating relationships between journalist and local audiences is already under way online; journalists have been found to be reaching out beyond their institution, sharing their talents for storytelling as part of a network – a new form of journalism (Robinson, 2007). Elsewhere, Davis Mercy (2009), sought to establish whether or not the techno-utopian predictions of the late 1990s, that technology would kill geography (Cairncross, 2007; Bulkely, 2008), were borne out in the relationships between people and their local newspapers. On the contrary, survey findings showed readers of beat-blogs feel more attached to their geographic communities than their online communities (Davis Mercy, 2009).

The rise of the network, evidenced in everything from user-engagement via Twitter, to the processing of User Generated Content, offers a means of extending this approach; and of re-invigorating the 'beat'. It permits the re-constitution of journalism's traditional power-base, re-connecting journalists with their audience online within a wider social network.

SOURCE SCHEMATIC FOR LOCAL NEWS

The author sought out a schematic which may help in mapping out potential sources of news within a local news geography. One such schematic is offered in Smith (2007), who lists the following as places journalists routinely ‘get all the news’:

TV, radio & the web
Wire services
Publications

Your own newspaper
Other newspapers
Magazines and journals
Diary jobs
The calls
Handouts and press releases
Submitted articles
Readers' letters
Tip-offs
Leaks
Public opinion
In-house activities
Anniversaries
And you… (Smith, 2007)

This is problematic (in terms of application) on many levels; not least that these are non-exclusive categories, and include few real-world examples which may be mapped online.

Randall (2007) frames where 'good stories' come from around 'the habits of successful reporters'; several of which ("Getting out and about", "Hanging around") are at odds with the contemporary journalistic experience as is found in the journalism studies literature. As an alternative to stories derived via social means, Randall suggest a series of 'non-obvious' sources; Universities and research institutes, specialist and underground press, books, esoteric magazines, international institutions, blogs, and classified advertisements; a (very limited) mix of official and unofficial sources, but nevertheless, a starting point of sorts.

In The Journalists' Guide to the Internet, Callahan (1999) outlines three proposed types of online beat; The Geographic Beat, The Issue Beat, and The Institution Beat. While this is a useful way of conceiving of local beats, these categories fall squarely into 'old' journalistic beat routines. They are so general, it is not difficult to see how some sources may appear in all three columns, and it offer little in the way of specificity which may help map them. Harcup offers a more normative, and far lengthier list of 'common sources of news stories', which include Academic Journals, Eyes and ear, Leaks, People and Pressure Groups (Harcup, 2009: p62). This is a comprehensive list but it is lacking in classification, and contains some conceptual cross-over. This approach does not differentiate between the importance of sources (in terms of weighting), nor is there an attempt to distinguish between people and institutions. A telling aspect of the problems inherent to constructing such lists can be found when comparing this list in first edition (2004, p46) and second edition (2009: p62) of this book. Within the space of five years, blogs and social networking sites rise from obscurity to become valued sources of news.

Keeble attempts to differentiate between primary and secondary sources, a practical convention drawn from the literature. The former include "...councils, MPs, and Euro MPs, courts, police, fire brigade, ambulance service, hospitals, local industries and their representative bodies, the local football, cricket and rugby clubs. Schools and colleges, churches, army, naval and air force bases, local branches of national pressure groups and charities are secondary sources. In rural areas, other contacts in this category will include village post office workers, publicans and hotel keepers, agricultural merchants, livestock auctioneers, countryside rangers or wardens. In coastal areas they might include coastguards, harbourmasters and lifeboat stations" (Keeble, 2009: p113). While such a list is instructive (if not helpful in this paper), such a distinction is, it could be argued, relative; after all, if a foot and mouth epidemic were to occur, agricultural merchants would soon become 'primary' sources (albeit perhaps for a limited time).

A NEW SCHEMATIC FOR LOCAL 'BEAT' NEWSGATHERING

The following model looks to sources (of all kinds) of information in a local community. This approach is not presented as an alternative to conventional off-line news-gathering methods, nor as a replacement for them; there aren't enough people or information sources online (and there may never be) for this to be a realistic goal. This model is presented as a tool which may help journalists adapt from offline to online news-gathering, systematise news-gathering online, eke out stories online which might otherwise be missed completely, audit an unfamiliar or new beat, or reinvigorate relationships with online audiences.

It is helpful to choose a taxonomy-model which reflects the administrative bounds of the beat. In the case of a local reporter, the structure of the local authority is a good fit; but for a subject area (like health, for example), a thesaurus taxonomy of that field would be better suited.

Once a model is established, keywords must be selected, which will be used to find sources and set up Really Simple Syndication (RSS) feeds. For a geographical beat this will necessarily include a full run-down of all significant local place-names, but it will also (at different points in the news calendar) feature other terms too (including local people, organisations, companies, events, etc.). Some sources will not require keywords as such, merely pre-existing automated feeds (local blogs, for example).

The proposed structure must be kept up-to-date, as it is subject to change over time; it needs to reflect the shifting nature of peoples' lives, and communities. Some online sources have no RSS feeds, and 'scraping' feeds for these sources can be problematic. Once relationships are established with sources, a social network begins to develop; such that journalists may find news more efficiently than via official sources. For all of these reasons, this news-gathering model comprises three parts:

1. RSS feeds (aggregated in an Feed Reader such as Google Reader, or filtered in Yahoo Pipes, etc.)
2. A diary of non-RSS sources which require systematic re-visit (ideally kept in a spreadsheet) and;
3. Membership of various social networks (including Twitter, Facebook, LinkedIn, Delicious etc.)

The following are not ordered in terms of importance. Sources are not presented with URLs due to space constraints.

1. NEWS (META) SOURCES

These are competitor news sources. What Bordieu referred to dismissively as the 'circular circulation of information' is a crucial element in exhaustive news-gathering (something most news organisations, if they are to claim to truly involve themselves in 'news', must aspire to). Any beat will require several news search feeds – covering references to the 'beat' at local, regional, national and international news levels, and possibly even including references to the beat in the trade press. RSS search feeds should be set up in the following aggregators:

Google News
Yahoo News
Bing News

Daylife

There will be some duplicate copy, but this can be filtered out later – ultimately these different services have different indexes, so it is important to cover as much as possible. Manual searches may be necessary (in NewsNow, for example) – these should go in the news diary. If there is a subscription sources budget, then further feeds will be available (in Nexis, FT.com, Times Online, etc).

Social news sites are of less importance to local journalism, but are still worth monitoring:

Reddit
Stumbleupon
Digg

2. ADMINISTRATIVE SOURCES

Many 'official' news sources are available online; from local authorities, to local single-issue charity groups. These can be organised within three domains:

GOVERNING SOURCES:

Council website RSS feeds.
Council newspaper (may sometimes require monitoring if online).
Council committee minutes – via OpenlyLocal (if available).
Online Forward Plan (will require to be found, and pages monitored).
Hansard (available via TheyWorkForYou, require place-name and politician name searches).
European Parliament (no feeds on site, advanced search necessary).
Freedom of Information (feeds in WhatDoTheyKnow for all local public authorities, including NHS trusts, schools, council etc.).

POLITICAL SOURCES:

Parish council (websites/social media groups and accounts of parish councillors).
Local politics (committee SM groups, officers, councillors, mayor, opposition).
National politics (MPs, websites/SM presence party websites/SM presence).
International politics (MEPs websites/SM presence).

STAKEHOLDER SOURCES:

Community: council social services, animal welfare groups, asylum groups, carer groups, children groups, women's groups, minority groups, Surestart etc.
Leisure: local arts groups (theatres, arts centres, galleries, etc.), community centres, conservation groups, sports groups, pro- and amateur teams, societies, etc.
Health and social care: Primary Care Trust management, carers groups, doctors General Practitioners and hospitals, mental health groups, health awareness groups, patients groups.
Education and learning: Schools (primary, secondary, etc.), colleges, universities, council information, educational groups, teachers, parents associations, and online sources such as Schooloscope, Ofsted etc.
Transport: transport committees and lobbies (motorists, cyclists, etc.), as well as online sources like FixMyStreet.

Environment & planning: Planning committees, NIMBY groups, environment groups etc.
Occupation/industry: employment agencies, business groups, businesses, local trade union representatives, and online sources such as OpenCorporates, WhosLobbying etc.
Law and order: council advice, local courts, local police services, neighbourhood watch groups, local lawyers/solicitors, Citizens Advice Bureau, local arbitration, etc.

3. SOCIAL SOURCES

For all social sources, it is important to consider who (i.e. individuals, groups) in conjunction with either where (geography of beat) or what (topic of beat), whichever is apt.

SOCIAL NETWORKS

Facebook (account required):

Search for and keep up-to-date with people, pages, groups, and events.
For people concentrate on searching by location, education or workplace .
Speculative geographical search can be done using Advanced Search 2.0 Beta plug-in.
Public updates and conversations are available in IceRocket, YourOpenBook, and Google.

Twitter (account not required for search feeds):

Set up place-name (and other) search feeds.
Set up Geo-location feeds by place-name (watch for limitations).
Follow key/popular figures in beat, measured by followers, but including famous sons and daughters (footballers, public figures, other local celebrities – see Knowhere and Wikipedia pages for suggestions).
For generic Twitter search feeds it may be necessary to filter for content that has been re-tweeted or which contain links, especially for breaking news.
Find followers in beat using geo-twitter and Tweepz searches (which provide feeds for new people).

Delicious:

Sign up for tag-based RSS feeds.
May be necessary to create other social bookmarking accounts too.

LinkedIn:

Keep up with companies within a beat
Search for and friend individuals working within beat.

Online forums. Sometimes anonymity can be a strength, not just in terms of finding whistle-blowers, but also in contentious/competitive fields – the main weakness with forums can be lack of RSS). Try place-name terms in:

Omgili (payment may be necessary for some feeds)
Google Groups

Yahoo groups
Boardtracker
Boardreader
Possibly some other bespoke forum sites (like Mumsnet, or local team boards).

BLOGS

Blogs:

Search for posts featuring local place-names via Google Blog search, Icerocket,or Social Mention.
Search for bloggers whose About includes mentions of a local place-name.

WEBSITES

Conventional keyword search in Google, Bing, Yahoo, Ask, Delicious etc. for local place-names.

MULTIMEDIA

Video:

Youtube
Blinkx
Google Videos
Truveo

Images:

Flickr
Panoramia
Twitpic (included in Twitter feeds). May be necessary to follow other image sharing sites too.

REFERENCE

Wikipedia: Set up searches to monitor all pages of interest in beat (places, people, subject matter etc.)
Google maps: need to do this manually now, RSS on Google maps is no longer.

DEVELOPMENT OF PROTOTYPES

In March 2011, two prototype beats were developed, using a range of RSS feeds with the intention of covering the real world beats of Borehamwood, in Essex, and Ilminster in Somerset. These two locations were chosen as they are urban, and rural beats respectively; so any differences between the feeds would not skew the results toward one 'type' of beat, or the other. The author used the schematic to locate a range of websites and online sources which contained RSS feeds, or which could be used to develop an RSS feed.

These feeds were incorporated into a simple Yahoo Pipes master RSS feed (http://pipes.yahoo.com/pipes/) which feature Unique, Sort and Filter modules. Pipes was chosen expressly because of its filtering options; allowing for the removal of job advertisements and other secondary information which may crowd out the feed. Contact was made with journalists in these beats, and over the following months problems arose. Subsequent changes were made to each Pipe in June 2010, to iron out the malfunctioning feeds (Yahoo accounts are required to view these feeds):

Borehamwood Wire: http://bit.ly/kcrHJH
Ilminster Wire: bit.ly/lufRod

LIMITATIONS

Creating beats for each area, town or village in a given beat was streamlined due to time constraints (for example, the Ilminster beat might also include place-name references to Merriott, South Petherton and a number of other small villages, in addition to Chard, Crewkerne, Chardstock and Chard Junction). Moreover full roll-out of each beat, incorporating diary and purpose-created social network accounts was not possible; one of the journalists was unable to access Facebook in the office, at the time of writing. The findings are therefore limited to feed results.

Full working for each beat can be found at the following addresses:

Borehamwood beat working: http://bit.ly/lguHgh
Ilminster beat working: http://bit.ly/iToxFN

FINDINGS

The journalist for whom the Ilminster beat was developed is a relatively active agent in his geographical beat, while at the same time also being relatively active online in seeking out stories. He has set up alerts in Twitter around mentions of several beat place-names (including Chard, Ilminster, Crewkerne, and South Petherton), which have yielded a number (and a broad range) of news stories in the past; from features on local shops, to emergency services stories. He acknowledges that there is likely less activity online in this beat than there may be in urban beats. He felt that the Pipe did, occasionally, yield some irrelevant results – particularly relating to duplicate place-names (i.e. references to Ilminster Intermediate School in Gisborne, New Zealand), and references to Lord Armstrong of Ilminster in Hansard (a politician who doesn't have any direct relationship with Ilminster as a news beat). But, he added, the unified, and thorough nature of the feed make up for this weakness; it is, he summarised “a good supplementary tool, which is worth using on a daily basis” (Jamie Brooks, interview, July 11, 2011)

When asked whether the Wire had led to the finding of any stories which may otherwise not have been found, he replied that it had: "a couple of community groups in Ilminster were planning events which were shown on updates in the Wire, before they had contacted us or the paper had found them” (Brooks, interview, July 11, 2011). Moreover, he added, “the most useful discovery though was the updates on the Ilminster Midsummer Experience website. The Wire found me one of these – when information was released on the site about the Ilminster Forum’s green event and on contacting them about this I found they were looking for more coverage for this year’s festival, with more events planned than ever before. The result of making contact with festival organisers so early was a real positive – and meant we had later exclusives on other events planned such as the Pimp My Scooter competition and led to our advertising department getting in contact and running a pull-out for the week of the Ilminster Experience. I would not have found these website updates otherwise – so the Wire acted as a great stating point to both get further stories and obtain some feedback from the community" (Brooks, interview, July 11, 2011).

CONCLUSION

This paper sought to establish an exploratory schematic for online local beats, which can be used by journalists to develop an online 'beat'. While constrained by the limitations of the technology used (and more generally, present-state web infrastructure), this approach was found to help in constructing an online beat which, when put to real-world application, generated information deemed worthy of developing into, and publishing as, 'news'. It should be noted that the transient nature of material on the web may require that subsequent to setting up the feed, further maintenance is necessary; which can be time-consuming, and potentially frustrating.

The success of such a schematic is hostage to more than what is considered to be optimum coverage. Taking after Metcaffe's Law, the value in such a chain, and the transformative nature of this approach, is dependent on the engagement of a local community online. This in turn informs the potential of such an approach to assist in the re-configuration of relations between journalist and source in a local networked 'beat'.

REFERENCES

BBC News (2011) Court results put on Twitter by West Midlands Police, bbc.co.uk/news, April 19:
http://www.bbc.co.uk/news/uk-england-birmingham-13127533

Bulkeley, William (1998) 'Corporate seers: who knows better what the future holds than those who make a living thinking about it?', The Wall Street Journal, 16 November, p37.

Cairncross, Frances (1997) The Death of Distance: How the Communications Revolution Will Change Our Lives. Harvard Business Press: New York.

Callahan, Christopher (1999) A Journalists' Guide to the Internet: The Net as a Reporting Tool. Allyn and Bacon, New York.

Castells, Manuel (2000) Material for an exploratory theory of the network society, British Journal of Sociology Vol No. 51, Issue No. 1 (January/March 2000) pp. 5-24.

Cottle, Simon (Ed) (2003) News, Public Relations and Power. Sage: London.

Concise Oxford dictionary (1999). Tenth Edition. Oxford University Press: London.

Carey, James (1999) 'In Defense of Public Journalism', in Glasser, Theodore L (Ed) The Idea of Public Journalism. The Guildford Press: London.

Davis, Aeron (2000) 'Public relations, news production and changing patterns of source access in the British national press', Media, Culture and Society, Volume 22, pp39-59.

Davis Mercy, Rachel (2009) Online News Users' Sense of Community, Journalism Practice, 3:3, pp347-360.

DCLG (Department for Communities and Local Government (2011) Code of Recommended Practice on Local Authority Publicity (The Publicity Code), March 31:
http://www.communities.gov.uk/publications/localgovernment/publicitycode2011

Fishman, Mark (1980) Manufacturing the News. University of Texas Press: Austin

Foreman, Gene (2010) The Ethical Journalist: Making Responsible Decisions in the Pursuit of News. Wiley-Blackwell: Chichester

Franklin, Bob and Murphy, David (1991) Making the Local News: Local Journalism in Context. Routledge: London.

Gans, Herbert (1979) Deciding What's News: A Study of CBS Evening News, NBC Nightly News. Random House: New York.

Gilligan, Andrew (2009) The propaganda newspapers, The Evening Standard, July 27:
http://www.thisislondon.co.uk/standard/article-23724285-the-propaganda-newspapers.do
Haas, Tanni (2003) 'Importing Journalistic Ideals and Practices? The Case of Public Journalism in Denmark', The International Journal of Press/Politics 8, pp90-97.
Haas, Tanni and Steiner, Linda (2006) 'Public journalism: a reply to critics', Journalism, Vol. 7, No. 2, pp238-254.
Hansen, Kathleen; Neuzil, Mark and Ward, Jean (1998) 'Newsroom Topic Teams: Journalists' assessments of effects on news routines and newspaper quality', Journalism and Mass Communication Quarterly, Winter; 75, 4; pp803-822.
Harcup, Tony (2009) Journalism Principles and Practice, 2nd edition. Sage: London.
Harcup, Tony (2004) Journalism Principles and Practice. Sage: London.
Harrison, Jackie (2006) News. Routledge: London.
Keeble, Richard (2009) Ethics for Journalists, 2nd Edition. Routledge: London.
Matthews, Eric (1951) A Dictionary of Americanisms on Historical Principles. University of Chicago Press: Chicago.
MacDonald, Alisdair (1977) Chambers Twentieth Century Dictionary New Edition. W & R Chambers: London.
McManus, John (1994) Market-Driven Journalism: Let the Citizen Beware? Sage: California.
Machin, David and Niblock, Sarah (2006) News Production: Theory and Practice. Routledge: London.
Morrison, James (2010) Spin, smoke-filled rooms, and the decline of council reporting by local newspapers: the slow demise of town hall transparency. Presented at the 61st Political Studies Association Annual Conference:Transforming Politics: New Synergies:
http://www.psa.ac.uk/journals/pdf/5/2010/612_322.pdf
Ofcom (2010) UK Adults' Media Literacy Report, May 17:
http://stakeholders.ofcom.org.uk/market-data-research/media-literacy/medlitpub/medlitpubrss/adultmedialitreport/
Oxford English Dictionary (1989) Second Edition, prepared by J.A. Simpson and E.S.C. Weiner. Clarendon: London.
Partridge, Eric (1973) Routledge Dictionary of Historical Slang. Routledge and Kegan Paul: London.
Randall, David (2007) The Universal Journalist, third edition. Pluto Press: London.
Random House Dictionary of the English Language (1987). Unabridged version. Random House: London.
Robinson, Sue (2007) 'Someone's Gotta be in Control Here', Journalism Practice, 1:3, pp305-321.
Schlesinger, Philip (1987) Putting 'Reality' Together: BBC News. Taylor & Francis: London.
Smith, Jon (2007) Essential Reporting: The NCTJ Guide for Trainee Journalists. Sage: London.
Tuchman, Gaye (1972) 'Objectivity as Strategic Ritual: An Examination of Newsmen's Notions of Objectivity', American Journal of Sociology, Vol. 77, No. 4 (Jan), pp. 660-679.
Zelizer, Barbara (2004) Taking Journalism Seriously: News and the Academy. Sage: Thousand Oaks.

APPENDIX IV: SEARCH: THEORY AND PRACTICE IN JOURNALISM ONLINE

The context to this book

It is not often that the process of news-gathering becomes front-page news, yet this was the story that defined late 2011, when David Cameron was compelled to set up the Leveson Inquiry. It is a well-worn cliché that woe betide any journalist who becomes part of the story. But when the way journalists do their job becomes a front-page splash, it seems all too inevitable that serious consequences must follow.

Since 2006, and the publication of the Information Commissioner's report *What Price Privacy?* we have known that British newspapers used illicit means and ways to eke out stories. Journalists, and private investigators working on their behalf, have indulged in unethical means of securing evidence such as payment of witnesses, and illegal methods of obtaining private information, such as phone-hacking and blagging. And yet in a way, it is not these indiscretions and crimes individually which warrant pause for reflection. It is rather a matter of scale.

So frequently were these methods used, they must have seemed like second nature in the newsroom, little different to keeping an eye on the wires, and certainly not a last resort for use in investigations which materially affect society. But just as more on this industrialisation of unethical and criminal practice comes into the public domain, journalists must continue to bring us the news. How do we ensure news-gathering remains ethically and legally sound, without putting a brake on the way news is made altogether? There seems no better time than now to re-establish the importance of ethical news-gathering as an essential tool in journalism. This book will, I hope, make a useful contribution in this direction.

Many thorny questions arise from news-gathering online, amongst the wealth of social networks, and other spaces people frequent. To what extent is it fair to use information and files gleaned from social network profiles? Where do we draw a line between private and public information here? What if the user doesn't fully understand the privacy settings on their social network account?

To further confuse these already thorny issues, we must not forget that social norms are prone to change over time. When W.T. Stead popularised the 'universal interview' as a newsgathering method at the turn of the 20th century, it caused great consternation amongst the patrician elite of late-Victorian Britain. Henry James, whose distaste for what he termed 'newspaperism', gave voice to this fear: in consenting to this new journalistic device, the masses may do more violence to the dignity of social order than any muckraker ever could (Rubery, 2009). How quaint and stuffy such views seems now, and yet this experience prompts us to consider what future privacy has in the information age, given that the dividing line between private and public space is measured in bits.

The debate about the extent to which online technologies are shaping our notions of privacy rages on. But there is no doubt that the massive (and growing) quantity of information about us online, when aligned with instant access, are changing many facets of modern life. This can be seen in legal areas including libel and contempt of court, in the regulation of use of evidence obtained online, and even in basic professional (and personal) online etiquette.

Can anyone say for certain that the public even care what methods journalists use to bring them their daily read? Given how far-reaching the criminal and extra-judicial investigations around phone-hacking have been, it is perhaps surprising that no large-scale survey has (at the time of writing) been conducted asking the UK public for their thoughts on this, and other related issues. But whether or not the public has a say, there remains the not inconsiderable matter of professional propriety to consider.

Anyone who has been engaged in news-gathering professionally knows that it can be a difficult, and conflicting business. But it can also be massively rewarding. Those of us who have only ever undertaken legal methods in our news-gathering, and those of us who may have used illicit practices but only as a very last resort when investigating matters of grave public concern, have an obligation to place on public record just how important it is not to abuse the public trust.

But neither must we lose sight of the fact that we live in a democratic, and open society. While there are undoubtedly moral issues to be addressed concerning the gathering of information (even information firmly in the public domain), a balance must be struck between privacy and free expression, in the public interest. Open societies require that we know our neighbour, or at least have a means of finding out who they are. The free flow of information is key to an open and transparent society, and in many ways the internet is a particularly apt medium to help us realise transparency in our public and social affairs.

What this book is not

This book does not start from the premise that everyone has something to hide, nor that “privacy is for peados” (Williams, 2011).

This book is not a muckraker's manual – it does not assume that there is something intrinsically 'fishy' about those who happen to be in the news. On the contrary, ethical questions permeate every aspect of a journalist's working life, and so by extension, ethical criteria must govern the acquisition and use of online resources. From the gathering, to the management, to the distribution of information found online, there is more to questioning how journalists should do their work, than merely how quickly.

It has been argued that declining trust in newspapers is a major factor in declining circulations, a position which has been strengthened by the sort of truths the Leveson Inquiry has laid bare. Others feel that the audience is, whether willingly or unwillingly, complicit in the conspiracy, and that trust has no bearing on why punters part with their cash. Others still argue that the media habits of the public are governed more by the possibilities of choice wrought by new technologies and mediums, than by loyalty to a particular source of news. Wherever the truth lies, and whatever the state of the industry, it behoves any journalist to act with dignity and decorum; whether because this is how we would wish to be treated ourselves, or because such behaviour brings its own virtue.

That said, this book is not a philosophical treatise. While an analysis of the Categorical Imperative may tell us much about life and the universe, it tells us little specific about the practice of journalism, which is central to this book. Ethical issues are therefore framed very much around analysing the practices, guidelines and lessons learned in the field, and in other related or similar professional fields.

Theory is explored in this book as a means of anchoring online resources and tools to the context in which they were developed. The technologies which define this field are prone to change, and obsolescence, often in a very short period of time. But while many of the resources covered in the following pages will be mothballed in the months and years ahead, the thinking and use which make them such invaluable resources in the first place, will remain.

Use of this book

This book offers a practical overview of the journalistic potential in various new (and developing) online tools, informed by various related theoretical fields.

The internet is a fast-changing environment, in which journalists must play catch-up in response to emerging online resources. But modern journalism is also a high-pressure environment. Converging industries and media, and increasing competition from non-traditional media in breaking news, make it more and more important that journalists are skilled in effective and efficient search methods.

But the emergence of these technologies also requires a re-appraisal of key theoretical issues, that inform, and that are informed by practice.

Whether uncovering breaking stories, finding reliable background on on-going stories, or finding witnesses and contributors, there is today a wealth of information freely available at journalists' fingertips. But these tools arrive in a messy and unstructured way, and it is not always easy to know where to find them, let alone learn how to use them.

The new social web offers journalists a means of collaboration with fellow professionals, experts and sources (not to mention the 'former audience'), which can revolutionise the research process, but this in turn requires an informed debate around several theoretical issues.

References

Rubery, Matthew (2009) The novelty of newspapers. OUP: Oxford.

Williams, Olivia (2011) 'Privacy Is For Paedos' Paul McMullan Shocks Leveson Inquiry Over Hacking Culture, Huffington Post, November 30:
http://www.huffingtonpost.co.uk/2011/11/29/notw-journalist-paul-mcmu_n_1118764.html

Chapter 1: Search in theory

From information overload to filter failure

Information overload is a modern concept, though not a new one. Long pre-dating the rise of the internet, the term was popularised by Alvin Toffler. His book, Future Shock (1984), conveyed the *nausea* of the information age, and the sense of dislocation individuals and groups feel as a result of excessive change in too short a period of time. The concept has been studied and analysed in a range of information professions; from accounting to marketing and consumer research. In most cases the common denominator is that individual performance (in terms of decision-making) varies relative to the volume of information available, up to a point. Beyond that relative point, further information results in rapidly declining performance (Eppler and Mengis, 2004).

As far back as 1997, information overload was found to interfere with journalists' ability to gain traction, or 'grip' over the news (Nicholas, 1997). Journalists have complained of information overload caused by modern working life, and not least excessive PR communications, or 'information subsidy' (Curtin, 1999). More recently, scholars have predicted that information overload will continue to be a 'key challenge' for journalist and citizen alike in today's networked world (Servaes, 2009).

The internet makes it possible for anyone to publish information; which has led to rapid growth in the production of information online. Indeed, Google indexed its landmark trillionth web page as far back as 2008. This sea of information can be disorientating without adequate support; and the risk of misinformation to journalists poses a serious danger to professional credibility.

Information overload continues to inspire new literature, albeit at a declining rate since around 2004. Some critics have argued that the volume of information on the Internet is making us more 'stupid' (Carr, 2010), others disagree (Battelle, 2008). Striking a pragmatic note, Shirky (2008) turned the concept of information overload on its head, arguing that the true malaise of the modern era is rather 'filter failure'. This is conceived as the collapse of those systems we use to help us tell good from bad (the likes of which we make use of everyday in our off-line lives). While on the surface this may seem a mere semantic twist, Shirky's term moves us away from the individual (not 'user') as a helpless, information 'junkie'. It moves us instead towards a place where, if we can build filters online to help determine the useful from the useless more efficiently, then we may be able to plot a course through this sea of information.

This book will offer an overview of the current state of information filters available to journalists working online. It is intended to help journalists address the filter failure that plagues contemporary, networked journalism.

How Journalists use the internet

Throughout the last century, most large media organisations employed teams of researchers and librarians whose job it was to provide journalists with research systems (such as newspaper clippings) and a full reference library service. A combination of economic factors (including declining circulations and industry consolidation) and technological developments (including the emergence of powerful and affordable database technologies) have changed this. As with many other professions throughout the global economy, media research is increasingly becoming the domain of the all singing, all dancing all-rounder: the journalist.

Today most journalists do their own research from the comfort of a desk. But navigating the internet as a researcher or journalist is a different proposition from using it to book holidays and listen to music. One difference is that a journalist can't just give up on a story if she is struggling to find the information needed, another is that the news schedule will not slow down to accommodate journalists who cannot find information in time.

Various studies have considered how journalists use the internet to help them source news and contributors. Nicholas (1996) found that Guardian journalists (who demonstrated degrees of capability in online search) weren't as ignorant of technique as some librarians (and librarianship literature) assumed them to be.

The mid-1990s saw a tipping point in US journalists' usage of the web in news-gathering. One study found that daily use of the internet had risen from 25% of respondents in 1994, to 92.4% in 1998 (Garrison, 1999) . This rise in usage has fuelled concerns about the content found online, especially concerning the verifiability and reliability of this information.

By 2001 web search was out-stripping the use of commercial online research tools in US journalists' news-gathering routines (Garrison, 2001), and by 2005, it was found that almost two-thirds of journalists were using competitor news found on the web in their researching and reporting (EURO RSCG Magnet & Columbia University, 2005). However, just because journalists today have access to many unofficial sources online does not mean they use them. On the contrary the old, long-established 'conventional' means of sourcing stories remains (Jha, 2007).

More recent studies have moved away from the basic measurement of internet use, towards trying to understand the contexts within which the internet is accessed by journalists. This includes issues arising out of professional context; such as time constraints, and means of access, in terms of the range of tools available to journalists, and those which are actually used (Hermans et al, 2009). While it is important not to confuse professional approach with medium in journalism, online journalists have been found to be more trusting of news they find online than print journalists are (Cassidy, 2007), suggesting a relationship between the two.

In perhaps the most comprehensive research study to date into journalists' use of online search tools, Machill and Beiler (2009) found that although Google plays a decisive role in most German journalists' news-gathering research today, it is nevertheless one of relatively few online tools commonly used in the newsroom. They found that most journalists achieve only moderate levels of success in online search. Those who apply most thought to search problems were found to perform best overall. It was also found that journalists' concerns about the reliability and verifiability of material found on the web have led to an increase in the cannibalistic approach to news-gathering; that journalists will tend to reference their own and other media at the expense of reflecting the whole web. Further research has shown that just over half of journalists are oblivious to blogs and social media in terms of sourcing the news (Oriella PR Network, 2011).

Search theory

Morville (2005), explains that the central issue in search (or 'information retrieval' as it was originally known) has traditionally been the concept of *relevance*. Traditionally, developers in this field have conceptualised search in terms of the binary (and inversely related) concepts of 'precision' and 'recall' as measures of *relevance*. These two concepts have very precise meanings in search engine development, thus:

Precision = Number of relevant and retrieved/Total number retrieved

Recall = Number of relevant and retrieved/Total number relevant

(Morville, 2005, p51)

Morville goes on to explain these concepts in the following terms: “precision measures how well a system retrieves only the relevant documents... recall measures how well a system retrieves all the relevant documents” (Morville, 2005, p49). How important these two concepts are (and how search strategy should be amended to accommodate them) depends on the type of search undertaken. For searches which require a certain (manageable) number of search results (a situation common to many busy, time-poor news journalists), precision is key. But for exhaustive searches, where unearthing a fact may require hours of painstaking search (a situation most investigative journalists will be able to relate to) recall is key.

But this approach is fundamentally compromised by the imprecision, ambiguity and vagueness of language as it is used by the one variable which no laboratory conditions can impose order upon; the searcher. When *relevance* is defined by search engines, it is an aggregate, quantitative measure; whereas we humans think of relevance in a rather more fluid, ambiguous, qualitative way. This is why search requires additional *aboutness,* and additional keywords for content, not to mention Boolean operators, field-specific functions, and other advanced options.

In the mid-1990s, the focus of work in information retrieval moved away from the hard science of precision and recall, and towards the study of how humans interact with information (Bates, 2002). Putting the searcher at the centre of the process changes the landscape. Acknowledging that information needs to evolve as searchers interact with the tools at their disposal (and change their search needs) changes the game.

Search developers are aware of the importance of iteration in search; that what is considered to be the right result is an (often) internatilsed process of negotiation. Expert search is akin to *berrypicking* in that it is discriminating (Bates, 2002). Users *satisfice* their search needs. As such, those search options which are available to us can change the nature not only of what can be found, but of what we seek (Halavais, 2009, p87). Post-modernity, it could be argued, came late to search, but information literacy can help journalists clear a path through the search wilderness

How search engines work

The term 'search engine' is often used to refer to two entirely different types of search resource; human-powered directories (which will be covered in more detail later in this book) and crawler-based search engines. The first form of these, human-powered directories, such as the Open Directory Project, DMOZ (http://www.dmoz.org/), are designed to aid browsing for information within a collection of materials organised along the lines of human expertise, and presented in an intuitive way.

The second form of these, crawler-based search engines like Google, Bing, and Yahoo, have three components. First is the 'spider', the algorithms which pass through the Internet looking for new content and for changes to existing content. This process involves analysing all information on a page, parsing that information, and then storing it in the second component: the search engine's 'index'. The 'spider' then exits via the links found on that page, so if a web page has been spidered but not indexed, it will not be found via the search engine. The third component (the search engine) is a program which helps the searcher interact with the index; and which ranks content returned in terms of 'relevance'.

Major search engines differ only by degrees, in particular with regard to the metrics (and the weighting of these metrics) used to determine 'relevance'.

Google is said to use over 200 'signals' to help determine the relevance ranking of web pages, including their patented PageRank algorithm (which uses co-citation between web pages to inform ranking). However, apart from some general technical (and editorial) advice, the company is deliberately vague on how these signals are weighted in ranking. This absence of transparency is deliberate; it stems from not wishing to give unscrupulous web publishers ammunition with which to 'game' the system (Moran and Hunt, 2006). This is a fast changing situation. Some of those methods which have been significant in search relevance in the past, such as the use of metadata tags to add meaning to online content, have been exploited by 'spammers' and so their significance in ranking has been muted. For example, the <Keywords> HTML meta tag is no longer factored into ranking (Cutts, 2009).

More recently, Google has been forced to acknowledge the need to move toward qualitative notions of quality and relevance. The 'Panda' algorithm (Google Blog, 2011), has been developed in the midst of wide-ranging concern about the rise of 'content farms'; websites whose content is optimized for search, but which owe their online authority more to effective exploitation of Google's Adsense marketing platform, than to reputation for providing quality information (Roth, 2009).

Political economy of search

Jurgen Habermas conceived of the public sphere as a social arena which exists between the private sphere of enterprise and the government. It is somewhere people can get together to talk openly about the issues of the day, to debate social problems, and organise action accordingly. This neo-Enlightenment concept, it is argued, found its apogee in the coffee houses of 17th and 18th century London. These were discursive spaces; places where deliberative democracy could flourish (Habermas, 1991). But does the economic structure and technology comprising today's search industry contribute to, or detract from this ideal?

Monopolies and oligopolies, in effect at least, abound within the search industry. In July 2010, Google domains accounted for more than 90% of the UK search market (Experian Hitwise Data Centre, 2010) – an effective monopoly unrivalled in virtually any other industry. But Google has a major influence upon the information we consume due to search behaviour too. Research on AOL server logs has found that 42% of users click on the first placed result in search engine results, while only 12% click on the second placed result, and 9% on the third. The first ten results receive 90% of all click-through traffic, and the second page only just over 4% (Enge et al, 2010). People do not have the time or resources to check every result, and as the Internet grows, we are becoming more dependent upon Google as an arbiter of relevance (as a proxy for *truth*) in our lives.

Google's corporate literature used to described the PageRank algorithm as 'uniquely democratic'; a claim which has led some in academe to question just what kind of democratic system the company have developed. In a theoretical consideration of search indexing method, Introna and Nissenbaum (2000) argued that far from being an egalitarian domain, a small number of elite online sources dominate on the web. Elsewhere, those concerned with deliberative democracy have argued that Google's PageRank algorithm has a less democratic sense of relevance than the company may like us to believe. Commercial motivation drives content creation on search engines according to Hargittai (2004); who argues that there is an association between the economic might of corporate firms and the influencing of search rankings (albeit no empirical data was provided to corroborate this view). On the other hand, it has been argued that the imperfection of current state search engines means that searchers can actually benefit from serendipity (Lev-On, 2006).

Challenging those who claim that the Internet (as represented by the PageRank method of link citation) is 'inherently democratic', Diaz draws upon empirical research by Barabási (2002), which found that the Web's structure is far from egalitarian in nature, but is rather organised around a small number of elite 'hubs' – a process described as the 'rich get richer' phenomenon (Diaz, 2009, p14). Well-linked web-pages benefit from more and more links (as they are easy to find), and so entrenchment bias or 'Googlearchy' prevails. What started with the professionalisation of information has now led to information industrialisation and it is argued, to a crude utilitarianism expressed in the “digital version of the vox populi” (Hinman, 2008, p67).

One rather obvious manifestation of bias online is censorship; and Google's attempts to censor material in countries around the world, in order to accord with law (and in some cases government whim). While it could be argued that the censorship of anti-semitic material in France and Germany is justified, it is harder to excuse the fact that references to Taiwan independence and Tianenmen Square have been censored in the Chinese version of Google. Indeed, it has been argued further that the local versions of 'search' Google has established around the world, make it easier to enforce censorship locally (Halavais, 2009). So relevance, as established by factoring in country-specific factors into search ranking, can work counter to the truth.

Diaz conceives of search engines as “general interest intermediaries” (2009, p15) which should be scrutinised in the same way as any other mass media organisation. But Hinman goes further still, arguing that search engines “construct knowledge through control of access” (Hinman, 2008, p73). Our public trust rests in the hands of a commercial body whose imperative to thwart industrial espionage by keeping the ranking recipe secret, mean we can never judge for ourselves whether or not bias exists: “...the political possibilities of the Web are constrained by its architecture. The end-to-end design of the Web might not limit the political sites that citizens visit, but the link structure of the Web certainly does” (Hindman, 2009, p57). Others, of course, would counter that it is unfair (and indeed patronising) to play down the agency of searchers (or web authors) in the process of search.

Search and privacy

Amongst the 200 signals Google uses to rank search results, are elements defined as 'contextual relevance'. These relate to personal information about a searcher and their search habits, and may include location, search history, profile data, and language preferences. It may even include the *bounce rate* for links clicked on (a measure of user time spent on a page before returning to search results - the less time spent dwelling on a page, the less 'relevant' the link clicked on is considered to be). There is disagreement within industry upon the degree to which these factors impact on relevance, but personal information is undoubtedly a factor in relevance.

In imagining the 'perfect search engine', Zimmer (2008) put forth a theoretical axis of relevance driven by two mutually inclusive concepts; 'perfect reach' and 'perfect recall'. The former necessitates access to all available information; the latter comprises personally-relevant, bespoke search results. But achieving 'perfect recall' would require search engines to harvest more and more of our personal information (such as contextual relevance), which in turn represents a threat to our personal privacy. Eric Goldman (2006) re-framed this issue, claiming that the subjective, changing and secretive approach to search engine relevance, which he termed 'search bias', is both necessary and desirable as a public good, in so far as it reduces information overload. It is inevitable, he further argues, that this bias will be consigned to the dustbin of history by the rise of personalised search results.

More recent critiques argue that the rise of personalisation (whether that be via contextual factors, search history, or friend's recommendations in search) can dampen the effects of search bias. This means of determining 'relevance', it is argued, has the potential to cause the wider public to become intellectually complacent, and quiescent. In this view, personalisation represents a pernicious cultural relativism that cocoons us all from competing or opposing truths, trapping us in our own beliefs and prejudices. This, it is argued, impoverishes us all in so far as we are instantly removed from opinions and facts which may challenge our most basic assumptions (Pariser, 2011).

In turn Pariser's assumptions have been questioned. Morozov (2011, I) critiques Pariser's utopian belief that internet companies should do anything other than provide a means of finding things out. On the contrary, he argues “algorythms do not 'think' - they compute. And while computing the 'is' (i.e. relevance) is something they can accomplish, computing the 'ought' (i.e. our information duties as citizens) is a much more contentious and value-laden process.” (Morozov, 2011, I). But on the other hand, it may equally be countered that the weighting of the 200 'signals' Google uses to determine relevance owe more to editorial judgement (as found in any newsroom), than to the lightening-quick processing of zeros and ones. Morozov takes after Goldman in arguing that regulatory intervention may create more problems than it solves; for what makes one bias better than any other? (Morozov, 2011, II). And yet left to their own devices search engine companies only have to do the bare minimum in order to protect our privacy.

In 2009 Tim Berners Lee voiced concern about 'behavioural targeting' in search, arguing that our quality of life may be adversely affected where our search history is factored into our ability to buy life assurance, or work (Poulter, 2009). For journalists freedom of expression is a direct concern too. In the US, The Patriot Act (2001) has led to state intrusion into search habits. Internet search histories have been used in the prosecution of alleged terrorists (Graham, 2009) while in the UK, the Prevention of Terrorism Act (2006), includes a positive obligation to report information which may materially assist the authorities in preventing terrorism (something any journalist may stumble across when researching sensitive issues). In other circumstance, the loss of privacy in search can lead to the identification of dissidents and proscribed groups. In the wrong hands this can pose a challenge to freedom of expression, and even a risk to life. For this reason, it is important that journalists know how to protect their search privacy should they need to work on a sensitive topic.

For journalists who are concerned about intrusions into their search (especially with regard to investigative journalism in the public interest), the best option is to use an encrypted (proxy) Internet Connection, ideally via a reputable proxy system such as Tor (https://www.torproject.org/), and to use an anonymising, private search engine in conjunction with this software, like DuckDuckGo (http://duckduckgo.com/).

But it is not just a journalists' own search history that can raise ethical problems. The search term prompts offered by Google Suggest are based on volumes of past search queries; and they represent an interesting quandry for journalists. On the one hand, this function can speed up search, and can even form the basis of original research and journalism. Search trends have been used to anticipate crises as they develop, such as the Mexican swine flu outbreak in 2009 (Madrigal, 2009), and they have been used by the Bank of England to tap into indicators of economic growth (McLaren, 2011). But equally they can lead to a certain circularity of thought – where the dominant terms and language used can unduly influence individual search at an atomic, pervasive level. Knowing how and when to switch this option off, is just as important as is knowing when to step back from journalistic cliché in the process of writing.

References

Barabási, Albert-László (2002) Linked: The New Science of Networks. Perseus Publishing: Cambridge.

Bates, Marcia (2002) 'Toward an Integrated Model of Information Seeking and Searching' (Keynote Address, Fourth international Conference on Information Needs, Seeking and Use in Different Contexts, Lisbon, Portugal, September 11, 2002.), New Review of Information Behaviour Research, 3, 2002, pp1-15.

Battelle, John (2008) Google: Making Nick Carr Stupid, But It's Made This Guy Smarter, Searchblog, June 10th:
http://battellemedia.com/archives/2008/06/google_making_nick_carr_stupid_but_its_made_this_guy_smarter#ixzz1O1maEKdZ

Carr, Nicholas (2008) 'Is Google making us stupid?', Atlantic Magazine, July/August:
http://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/6868/

Cassidy, William P. (2007) 'Online News Credibility: An Examination of the Perceptions of Newspaper Journalists', Journal of Computer-Mediated Communication 12 (2007) 478–498

Curtin, Paul (1999) 'Re-evaluating public relations information subsidies: market-driven journalism and agenda-building theory and practice, Journal of Public Relations Research, 11 (1), pp53-90.

Cutts, Matt (2009) “Google doesn’t use the keywords meta tag in web search”, Matt Cutts

Blog, 21 September: http://www.mattcutts.com/blog/keywords-meta-tag-in-web-search/

Diaz, A (2008) 'Through the Google Googles: Sociopolitical Bias in Search Engine Design', in Spink, Amanda and Zimmer, Michael (eds) (2008) Web Search: Multidisciplinary Perspectives. Springer: Berlin.

Enge, Eric; Spencer, Stephen; Fishkin, Rand; and Stricchiola, Jessie (2010) *The Art of SEO: Mastering Search Engine Optimization*, Sebastapol: O''Reilly.

Eppler, Martin J., and Mengis, Jeanne (2004) 'The Concept of Information Overload: A review of Literature from Organization Science, Accounting, Marketing, MIS and Related Disciplines', The Information Society, 20: pp 325-344.

EURO RSCG Magnet & Columbia University Survey of the Media. (2005). Rebuilding trust: Can credibility be rebuilt in the newsroom and the boardroom. Retrieved May 31, 2006 from http://magnet.mediaroom.com/file.php/binaries/31/RebuildingTrust.pdf

Experian Hitwise Data Center (2010), “Top Sites & Engines”: http://www.hitwise.com/uk/datacentre/main/dashboard-7323.html

Garrison, Bruce (1999) 'JOURNALISTS’ PERCEPTIONSOF ONLINE INFORMATION-GATHERING PROBLEMS' A paper presented to the Newspaper Division of the Association for Education in Journalism and Mass Communication, Southeast Colloquium, Lexington, March 5-6, 1999

Garrison, Bruce (2001) Diffusion of online information technologies in newspaper newsrooms Journalism August 2001 vol. 2 no. 2 221-239.

Goldman, Eric (2005) 'Search Engine Bias and the Demise of Search Engine Utopianism', in Spink, Amanda and Zimmer, Michael (eds) (2008) Web Search: Multidisciplinary Perspectives. Springer: Berlin.

Google (2011) Company Philosophy - Ten Things:
http://www.google.com/about/corporate/company/tenthings.html

Google Blog (2011), Finding more high-quality sites in search, February 24:
http://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html

Graham, Kevin (2009) TESTIMONY CENTERS ON 'USEF' INTERNET SEARCHES, St Petersberg Times Florida, August 20.

Habermas, Jurgen (1991) The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society. The MIT Press: New York.

Halavais, Alexander (2009) Search Engine Society. Polity Press: London.

Hargittai, Eszter, (2004). 'Do you "google"? Understanding Search Engine Popularity Beyond the Hype', First Monday, 9(3).

Hermans, Liesbeth, Vergeer, Maurice, Pleijter, Alexander (2009) 'Internet adoption in the newsroom: Journalists' use of the Internet explained by attitudes and perceived functions' Communications. Volume 34, Issue 1, Pages 55–71,

Hinman, LM (2008) 'Searching Ethics: The Role of Search Engines in the Construction and Distribution of Knowledge', in Spink, Amanda and Zimmer, Michael (eds) (2008) Web Search: Multidisciplinary Perspectives. Springer: Berlin.

Hindman, Matthew (2009) The Myth of Digital Democracy. Princeton University Press: Oxford.

Introna, Lucas and Nissenbaum, Helen (2000) Shaping the Web: Why the politics of search engines matters:
www.nyu.edu/projects/nissenbaum/papers/searchengines.pdfSimilar

Jha, Sonora (2007) EXPLORING INTERNET INFLUENCE ON THE COVERAGE OF SOCIAL PROTEST: CONTENT ANALYSIS, Journalism and Mass Communication Quarterly; Spring 2007; 84, 1; ABI/INFORM Research
pg. 40-57

Levi-On, Azi (2008) 'The Democratizing Effects of Serach Engine Use: On Chance Exposure and Organizational Hubs', in Spink, Amanda and Zimmer, Michael (eds) Web Search: Multidisciplinary Perspectives. Springer: Berlin.

Machill, Marcel and Beiler, Markus (2009) 'THE IMPORTANCE OF THE INTERNET FOR JOURNALISTIC RESEARCH', Journalism Studies, 10: 2, 178 — 203

Madrigal, Alex (2009) Google Could Have Caught Swine Flu Early, Wired.com, April 29:
http://www.wired.com/wiredscience/2009/04/google-could-have-caught-swine-flu-early/

McLaren, Nick (2011) Using Internet Search Data as economic indicators, Bank of England Quarterly Bulletin, Q2:
http://www.bankofengland.co.uk/publications/quarterlybulletin/qb110206.pdf

Moran, Mike, and Hunt, Bill (2006) *Search Engine Marketing* (2nd Edition), London: IBM Press.

Morozov, Evgeny (2011) I: Your Own Facts (book review of The Filter Bubble, by Eli Pariser), New York Times, June 10:
http://www.nytimes.com/2011/06/12/books/review/book-review-the-filter-bubble-by-eli-pariser.html?_r=1

Morozov, Evgeny (2011) II: The Net Delusion: How not to liberate the world, Allen Lane: London.

Morville, Peter (2005) Ambient Findability. O'Reilly: Sebastopol.

Nicholas, Dave (1996) 'An assessment of the online searching behaviour of practioner end users', Journal of Documentation, Vol. 52 Iss: 3, pp.227 – 251

Nicholas, Dave and Martin, Helen (1997) 'Assessing information needs: a case study of journalists', Aslib Proceedings, Vol. 49 Iss: 2, pp.43 – 52.

Oriella PR Network (2011) Click, Community and Conversations: The State of Journalism in 2011, Digital Jouranlism Study:
http://www.orielladigitaljournalism.com/view-report.html#/4/

Pariser, Eli (2011) The Filter Bubble: What The Internet Is Hiding From You. Penguin Press: London.

Poulter, Sean (2009) Internet ad tracking system will put a 'spy camera' in the homes of millions, warns founder of the web , Daily Mail, March 12:
http://www.dailymail.co.uk/news/article-1161215/Internet-ad-tracking-spy-camera-homes-millions-warns-founder-web.html#ixzz1K44kJYur

Roth, Daniel (2009) The Answer Factory: Demand Media and the Fast, Disposable, and Profitable as Hell Media Model, Wired Magazine, October 19:
http://www.wired.com/magazine/2009/10/ff_demandmedia/

Shirky, Clay 'It's Not Information Overload, It's Filter Failure', talk at Web 2.0 Expo New York City, September 2008:
http://blip.tv/web2expo/web-2-0-expo-ny-clay-shirky-shirky-com-it-s-not-information-overload-it-s-filter-failure-1283699

Servaes, Jan (2009) 'We are all journalists now!', Journalism, June vol.10 no.3 pp371-374.

Toffler, Alvin (1984) Future Shock. Bantam: New York.

Zimmer, Michael (2008) 'The Externalities of Search 2.0: The Emerging Privacy Threats when the Drive for the Perfect Search Engine meets Web 2.0', First Monday, Volume 13, Number 3 - 3 March 2008

Chapter 2: Search in practice

Approaches to search

In the previous chapter we considered how search engines work, and how 'search engine' is an umbrella term for different types of search resource (specifically between those sources such as human-powered directories, or structured collections of links to materials which encourage browsing, and crawler-based search programs which allow you to key-word search an index). These different type of resource encourage different approaches to search, and the different types of searches we undertake in order to satisfy search needs. If what you are searching for exists within a hierarchy for example, it may be better to browse for the answer using a directory, rather than using a search engine. Of course there are many different types of search engine, some of which index material which is more appropriate for some searches than for others. Where searching general search engines fails to yield positive results, you may want to opt for a specialist search engine or site.

Browsing, or searching?

A very basic search strategy involves weighing up (in a non-scientific way) the concepts of precision and recall introduced in the previous chapter, and establishing which is more important in the search need. It may be said that most day-to-day search queries will have more emphasis on precision than on recall. The time constraints and transient nature of news journalism means that the accuracy of search results relative to search query, is more important than their exhaustiveness. However this is not uniformly the case – anyone who has ever worked in negative checks (checking the names of fictitious characters against real people), or anyone who has undertaken investigative online search, will more likely than not require results weighted more towards recall than precision. In turn, knowing which concept drives *relevance* in your search need can inform appropriate use of advanced operators and functions, whether the searcher should opt for those which restrict, or those which expand the flow of results.

Online search expert Nora Paul suggests that surfers should let their strategy be guided by what most journalists will recognise as Kipling's *Six honest serving men*: Who, What, When, Where, Why, How (Schlein, 2004). This approach can be used to focus in on those key words and phrases to use in search.

Alternatively, visualising the (hypothetical) information you want, in the words you would expect to find them on the page (or document), can be useful. This will help with establishing keywords, though may involve a good deal of imaginative leg-work, depending on the search query. It is helpful to think about this from a document analysis perspective:

Who (or what kind of person) may be the author of the words you are looking for? Is it someone whose educational/social/demographic profile explicitly impacts the words and language that they use (and that you will need to search for)?

Is it published in a source whose formal style can be mined for clues as to which words to search for (Manchester United footballer Wayne Rooney is routinely referred to as *Roo* or *Wazza* in UK tabloid newspapers, but referred to as Rooney in the broadsheet press). This should also help inform the keywords used.

Is it written in the first, second, or third person? Although Google (and other search engines) compensate for grammatical variation in search terms (such as tense), it is possible that some keywords are conceptualised differently according to who is saying them.

Is there more than one answer to any of the above questions? If so, this may also impact choice of search terms.

In terms of choosing search terms, it is always useful to make use of a thesaurus, or even a subject directory (depending on your subject and approach). Trying to avoid words with multiple meanings will help with precision. If that is not possible, try incorporating those terms into a phrase, to avoid passing references to homonyms.

Search engine marketing professionals categorise queries into three distinct categories, which they use to design the information architecture of web pages more effectively. The three categories of search intent are:

Transactional: search queries which suggest the surfer is using the search engine to try to do something. For example, the search string *train ticket wigan* would suggest the searcher wants to buy a train ticket to Wigan.

Informational: search queries which suggest the surfer is trying to find something out. For example, *where is Osama Bin Laden buried* would suggest the searcher is looking for information.

Navigational: search queries which suggest the surfer is trying to get to a particular website. For example a search for *bbc news* would suggest that the searcher is looking for the BBC news website. (Thurow and Musica, 2009)

It may therefore be instructive for searchers to consider whether their searches are informational or transactional before searching. This may in turn inform the sort of approach to search taken.

Alternatively, there are some things not directly related to online research, but which nevertheless have an impact on search success. Calishain (2005) identifies an approach to search in this vein as ‘the principle of every scrap’. A methodical and meticulous approach to online search, it is argued, is just as important for difficult search needs, as is the process of searching itself. Keeping a solid account of what you have done is key. This might include:

Saving or bookmarking relevant related material.
Recording the keywords used to find these resources.
Analysing the types of website returned for your searches (by domain, or other

criteria).

Without these records, you stand doomed either to miss out on things you have forgotten to search for, or to go round and round in circles doing the same searches again and again.

Henninger (2008) offers some useful general advice on searching online, suggesting that searchers should:

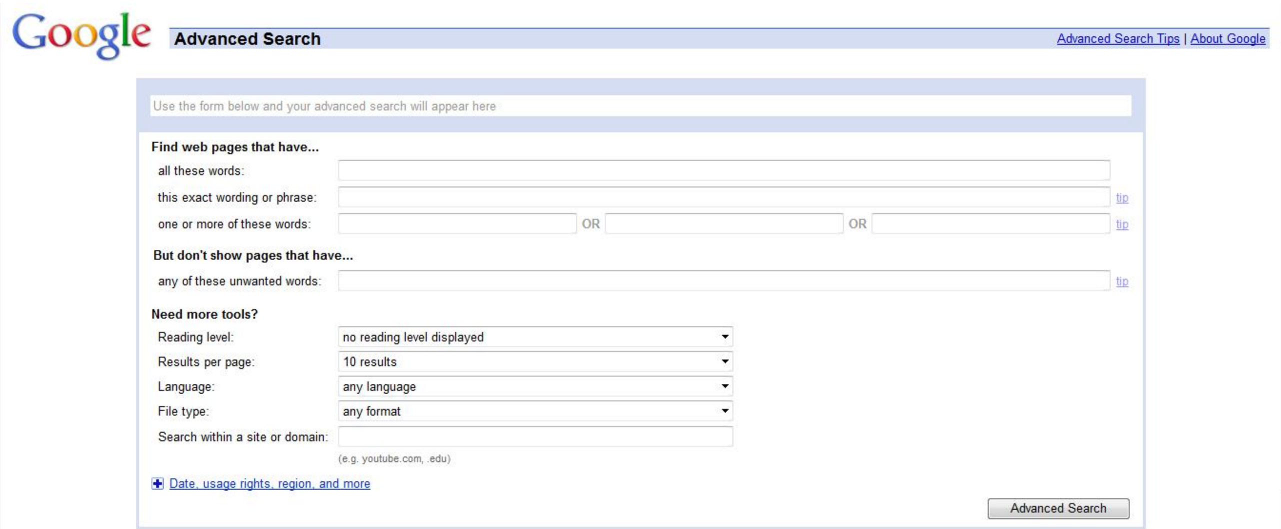
17. Develop a firm concept of what you are looking for – write it down.
18. List synonyms or phrases that reflect each concept.
19. Select the appropriate tool – Internet, known web server, database vendor, or reference tool.
20. No matter which searching tool, use advanced searching techniques and take advantage of its special features.
21. Examine the titles, summaries and/or contents of documents found to see if they are relevant to your request.
22. If they are not relevant then refine the search by modifying the strategy to get a list of documents that are more relevant.
23. If not enough documents are found, modify the strategy to increase the number of relevant documents retrieved.
24. If you do not find anything relevant, try a different research tool (return to step 6).

(Henninger, 2008, p91)

Screen breaks

One last piece of advice which is very important – it is wise to take regular screen breaks between searching sessions. Unless you keep a clear and focussed mind, it is easy to become distracted by irrelevant material, which drag you further from your search goal, and which eats away valuable time. Screen breaks give the opportunity to pause and reflect – which can be essential when undertaking large-scale, or frequent online search tasks.

Advanced searching Google



Google's advanced search interface

The simplest way to refine a search in Google (or any search engine) is simply to use more terms, because the more terms you add, the less results you will get back. This is equivalent to the AND operator in Boolean Logic; a means of formal logic, or search grammar, which can help to shape the relationship between words in a search query. In Google the AND function is implicit in the space between two keywords. But adding more words to your search can be taxing, and overly restrictive. For this reason it is a good idea to avail yourself of the many other operators and functions available in Google, which can help either refine or expand a search. Here is an overview of the key operators and functions, and how you can use them.

OR (|): When you are searching for something which can be expressed in more than one way, it can be useful to run an 'OR' search using a range of terms. For example, using either the pipe symbol (**|**) or the upper-case word OR here:

blair wmd OR weapons.

...will return pages containing the word blair and either weapons, or the acronym WMD.

NOT (-): Just as you can add keywords to your search string, so too can you remove them if they are distorting your results. For example, were you to search for *rangers*, with the aim of finding pages on Glasgow Rangers, you might want to remove references to other sporting teams with Rangers in their title, such as Texas Rangers, Queens Park Rangers (QPR) etc. Hence:

rangers -texas -qpr

It is worth noting, however, that using the NOT function too often can, in some cases, inadvertently filter out relevant content from your search. It is perfectly possible, for example, that the information you seek is available on a long, narrative, textual page which references both Glasgow *Rangers* and *QPR*, but paragraphs apart (say, essays on British football).

Phrase search (" "): When searching for terms which are likely to appear contiguously, in a particular order, it is wise to use phrase searching. Searching for *Putin Berlusconi* for example, is not as precise nor as restrictive as:

"putin and berlusconi"

Phrase searching can be a little restrictive however – so in this case, a safer bet to find articles which are expressly about the relationship between Putin and Berlusconi, the search would be better put as:

"putin and berlusconi" OR "berlusconi and putin"

Phrase searching is also important when the expression you are concerned with contains a stop word (common words such as; *is*, *that*, *and*, and *to*) which Google (and other search engines) tend to ignore when you search.

Phrase searching also supports a means of applying one aspect of search theory covered in the last chapter. It is possible to use phrase searching to answer hypothetical questions you may have in your search results, by searching for a fragment of language you would expect to find in answer to your question. So for example if you wanted to find reference to the UK’s wealthiest banker, try searching for…

"UK's wealthiest banker"

...which may appear in the answer, where it is phrased “the UK's wealthiest banker is...”.

Phrase searching also allows you to force keywords. Google tends to ignore any common words used in search because in many circumstances they don't help refine a search query. In addition, because Google's PageRank algorithm is a principal factor in Google ranking, sometimes pages which don't actually contain all your search terms, but most of them, and that have other pages linking to them containing your remaining words, will be included in your results. This is one reason why you may not always find the keywords you searched for on the pages you have returned.

To ensure Google includes such words in your search results, place in double quotes (“”) any word you insist should feature in the search results. This approach also blocks Google from automatically stemming your words, that is finding your keyword with different conjugations or word endings. This is useful if you are interested in a word in a particular case, or tense. Though you may be able to force Google to show you any particular word, Google ignores all but a tiny handful of characters, and is case insensitive (Long, 2005).

Synonyms (~): The tilde operator can be used to take advantage of Google's internal thesaurus, if you seek other terms related to those you are searching for. For example, *~marriage* will return related concepts which include references to *love*, *weddings* etc. This can be useful in many searches, especially where the terminology you would expect to find used, is fluid (and non-specialist language). At the time of writing this function is not supported in all languages (for example, there is no Russian equivalent), but is available by default in some personalised search results. The synonym operator can be thought of as being diametrically opposite to the + operator in terms of results returned – it improves recall, but in an intuitive way, by exploiting the conceptual similarity between words.

Wildcard (*): Many search engines allow you to employ a wildcard operator which can be used as substitute for a particular letter (or a number of letters) where you are unsure how a certain word may be spelled by the sources you are searching for. Google doesn’t support the wildcard used in this way because it uses an internal thesaurus to provide automatic stemming to find alternate word-endings and spellings for your terms. In Google you can use a * when phrase-searching in order to serve as a place-holder for a word which could be one of many alternatives. This can be useful in some very specific areas of journalistic search. For example, the following query:

"putin and berlusconi" ~expert " said"*

...will return results where an expert (or specialist) on the relationship between “putin and berlusconi” has been quoted in a news source (or elsewhere). The * serves as a placeholder for the hypothetical surname of the expert in question (which could be anything), and takes advantage of the journalistic convention of ending quotations with the expression 'so-and-so said'.

Numrange (..): It is possible to force Google to bring back numerical results from a range, using the *numrange* function (two dots placed, without spaces, between two numbers). So:

terrorism 1972..1977

...should (in theory) bring back results about terrorism during the mid-1970s. However, this operator tends to conflate numbers from different areas of the same page which can badly skew results. This function does not appear to recognise symbols (i.e. £), so is best used on raw numbers, rather than values. It may be said that this function is better suited to commercial rather than informational searches.

Domain function (site:): The domain function can be used to refine your search to a particular domain, a type of domain, or a sub-domain in a website. It has many uses – here follow just some.

This function can be used to find expert academics, in a three-stage process, requiring:

Subject terms to cover expertise
A term connecting the subject to his/her profession (i.e. *expert, department, professor* etc.), and
The domain function: *site:.edu* (filtering results to pages from US universities)

Using this method to find a US expert in solvent abuse might result in a search query like this:

"solvent abuse" professor site:.edu

Where American universities have .edu domains, UK universities tend to have *.ac.uk ones*. For other countries around the world, conventions vary – New Zealand uses *.ac.nz*, while Australia uses *.edu.au*. Moreover, in some countries there are no conventions at all in this sector. Top-level domains must be approved by the Internet Assigned Numbers Authority (IANA), and Wikipedia contains a full list of top-level domains (http://en.wikipedia.org/wiki/List_of_Internet_top-level_domains).

Example: advanced searching experts

While training in BBC online features, I once asked a features journalist to give an example of a story where finding contributors was particularly difficult and time-consuming. She suggested a story about the influence of political spouses, saying it had taken her hours to find an expert on this rather obscure field. I talked through the above method, and found the same expert she had, in the first result (in less than three minutes), by searching for:

"political spouses" professor site:.edu

Another possible use for the domain function is for finding non-profit groups, or NGOs. These groups often have a *.org* (or in the UK, *.org.uk*) upper-level domain. So the following search:

aberdeen drugs site:.org.uk

...should help unearth Aberdeen-based drug charities, by means of which you might get in touch with experts and professionals (while avoiding references to Aberdeen in

Washington, in the US).

But the domain function can also be used to exploit naming convention in web design, towards finding certain types of contributor. For example, WordPress blog accounts (and many websites) feature an *about* folder by default, for profiles, while Blogger accounts feature a *profile* folder. So, for example, by searching for...

"climate scientist" site:.com/about

...you may more accurately find experts in climate science who blog, or who have websites. In addition, finding conversations on Facebook can be done by exploiting the /topic folder used within the forum section of the site. So:

CNN site:facebook.com/topic

...will return conversations about CNN, which can be useful in terms of reputation management, or newsgathering. This cannot be done using Facebook's search interface.

This function can also be used negatively in order to find content about a topic, but from anywhere other than a particular online source. So if you are a media correspondent, interested in news about The Australian, but not published on The Australian's website, you could try:

“the australian” -site:theaustralian.com.au/

Searching within title (intitle:/allintitle:): This function lets you refine by terms found in the title of a web page. In Search Engine Optimisation, it is widely held that the words entered into a title field in a web page are weighted as highly significant when it comes to determining the 'relevance' of this page in search. By taking advantage of this widely acknowledged convention it is possible to refine your search. Most online news organisations provide backgrounders on big news stories and themes, and they are usually presented (or branded) in a consistent way. So if you want to find background (analysis, not news) on the Somali war in 2007, by taking advantage of the conventions used by some online media, you could try:

somalia war intitle:Q&A

Alternatively, try any of the following keywords: *depth/comment/analysis//brief/background*

Searching within URL (inurl:/allinurl:): This function can be used to find content according to words which appear in the address (i.e. URL, or server file path) of a particular web page. So if any type of website may contain your terms, but you do want to find keywords considered ‘important’ to a particular topic, this can be useful. This function can be used as an alternative to the domain function where search is concerned more with words in the folder structure than words in the domain name. If you are searching for news, then *inurl:* can in practice serve as a very similar function to *intitle:*, in so far as modern news content management systems often generate unique URLs by using keywords from the headline and/or story in question. But for finding content on other, less changeable sites, it serves a unique purpose.

Searching by file (filetype:): There are many reasons why it can be useful to filter your search by the type of file you imagine may contain the information you are looking for. For example, government and corporate reports are often published in PDF format, for ease of printing (these documents are most often read and acted upon not online, but offline). Alternatively, if you are concerned with finding a specialist or expert, it makes sense to confine your search to Powerpoint slides, in so far as experts often present to conferences, and subsequently make these presentations available online. Statistics are often published in Excel (.xls) format, so if you are interested in finding out what were house prices like in Toronto in 2007, the following search will filter out all of those estate agent websites:

“house prices” toronto 2007 filetype:xls

Google will only index the 'rendered' (visible) view of a file, so key metadata which might help identify an author of a file is not necessarily indexed – you have to download the file to check it for yourself. As a good alternative to searching by file-type, there are a range of search verticals which deal exclusively in certain types of media, such as Slideshare (www.slideshare.net), which is concerned with presentations, and Docstoc (www.docstoc.com), which is concerned with sharing text documents.

The language filter: For some types of search it is necessary to omit search results in foreign languages – however those factors used in contextual reference (including Internet Protocol recognition, and language) often render this issue much less of a problem than once it was.

Search in anchor text (inanchor:): depending on the naming conventions and care with which a website has been constructed, it may be useful to refine a search by concentrating on those words used to embed hyperlinks. Web developers looking to maximise their ranking on Google will use rich, descriptive keywords in their links rather than generic ones – but of course, the person who uploaded the information you are looking for won't always necessarily be an expert in the field, which is why this function may be of less value than others.

Proximity search (AROUND:): Google has a little-known proximity search function, which allows you to specify, where you are searching for two words, that they appear within a certain number of words of each other. It exploits the principle that words which appear closer together in text are more likely to be connected to each other – and where you are searching for names, this can be a useful way to weed out passing references. It also gives more control over search than the wildcard option. So for example, if you were doing some research on former French President Nicolas Sarkozy (about his time as president, rather than what he is up to now), you might try:

president AROUND(2) “sarkozy”

Cached pages (cache:): When search engines index web pages, they often take a snap shot image of the page (if the creator of the page permits them to). These 'cached' results often appear within search results next to the web address of the page. They can be essential in seeking out relevant results, as pages can be removed, altered or otherwise compromised in the time between indexing and being returned in your search, and the cache is the only way to get to the information you searched for. Cached pages can be particularly useful in breaking news situations, where online content is removed suddenly. It is not so useful for investigating what sites looked like months, or even years ago - The Internet Archive (http://www.archive.org/index.php), covered in Chapter 6, is a better option in this instance.

Similar pages (related:): If you find a website that is particularly useful to you, you may want to check to see if there are any other sites out there which are similar. If you employ the *related:* function on a web address:

related:irishtimes.com/

...Google will use co-citation to present you with as many as 30 'similar' sites to the one you've found. It should be noted again, though, that this is a quantitative measure of similarity, not a qualitative one.

Combining operators

While each of these operators has its own particular use, the real value in advanced searching comes when combinations of operators are used on a particular search problem. There are far too many combinations to list here, but here follows a few examples of what is possible.

If you wanted to find out what guidance is being given to help combat sectarianism (sectarian habits, or sectarian behaviour, for example) in New Zealand schools, then try:

*site:.ac.nz OR site:.school.nz "sectarian * schools" ~guide*

If you needed to find a list of (current or former) advisors to Northern Ireland's Ulster Unionist party, try:

"ulster unionist party" advisor site:linkedin.com intitle:United Kingdom

If you were investigating the oil industry in Azerbaijan, and you were looking to find potential whistleblowers, try looking at different companies:

bp OR shell azerbaijan site:linkedin.com/

If you were looking to track details of US arms contracts worth between 20 and 25 million pounds, try:

site:.gov missiles 20M..25M

If you needed to find a contributor who has previously worked at Ikea try:

"ikea" intitle:"curriculum vitae" OR intitle:cv

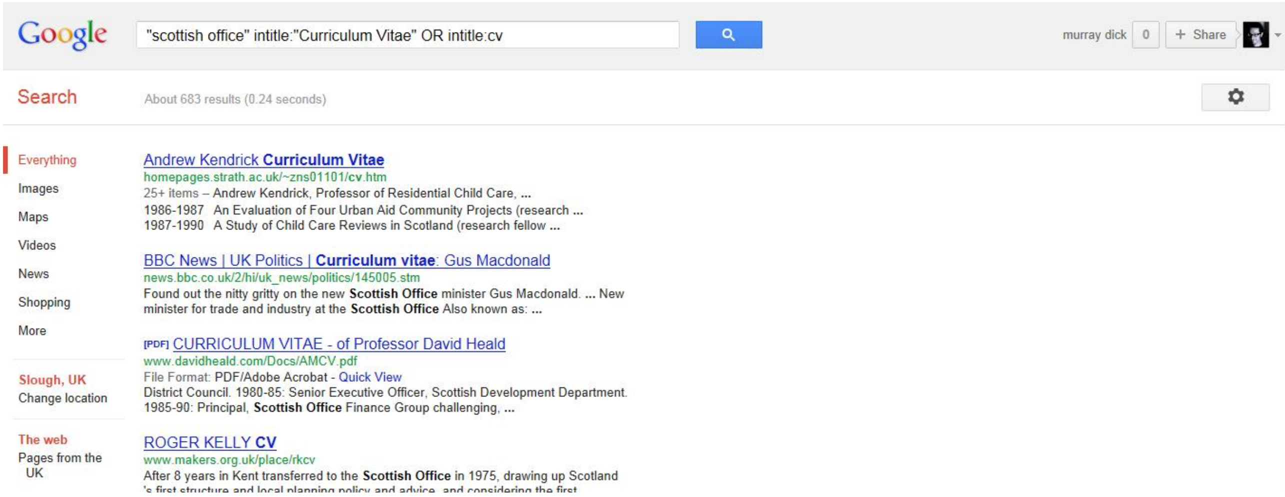
In each case you should bear in mind that Google indexes lots of different types of information in lots of different formats – and for this reason distortion is often a problem. But by refining you should be able to get closer to your search need. Bear in mind too that some advanced functions can be used negatively.

Remember that all Google advanced operators share the same basic syntax, namely:

operator:search_term

There should be no spaces between operator, colon or search term, failing which your operators will be treated as a search term. Operators can generally be used in conjunction with each other, and with boolean operators and special characters, but the ALL operators can be problematic.

Search result filters



Google's advance search filters (left-hand column)

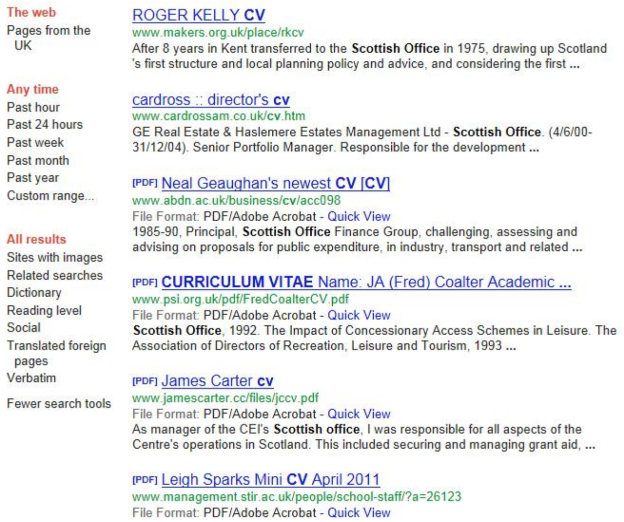
The process of refining your search is made easier because Google provides a range of filters on the left-hand side of search results – allowing you to filter by type of source (whether it be images, or forum discussions), by date-range (particularly useful for when time scale is crucial to your research), and by place (you can over-ride the automatic location setting Google selects for you, based on your IP address).

Google's predictive search (Google Instant) can be useful in so far as it can shed light on those areas of the web, prompting keywords and areas of study you may not have thought of before starting your research. Alternatively it can get irritating, and can be switched off within the settings section (wherein, so long as you disable Google Instant, you can specify how many results are returned per page).

In addition to offering refinement by media-type, Google offers a range of other interesting filters which can be useful, and which are often hidden under *More search tools*. It is possible to filter by *Language level*, whereby the words used on the documents returned are analysed, with those pages which contain rarer words are differentiated from those documents which contain more commonly found words. Google's *Timeline* option can also help with some types of search, but only where date of web publication is significant to research. The *Nearby* and *Custom location* options can also very helpful in ironing out differences in meaning across location. A search for 'Ospreys' (the name of a Rugby team) in Swansea will return very different results to the same search done in New York – take advantage of 'local knowledge' as expressed in the words of a locality by changing your location.

Using the Find option to browse a long web page

When it comes to web search – perhaps the most useful utility you will find on your browser is the *Find* utility, available from the Edit menu on your browser, or by pressing the <Control> and F keys at the same time. This lets you cut a dash through reams of information on long, textual pages – and can save you a great deal of time in the process.



Google's advanced filters continued (left-hand column)

Advanced searching Bing

While Bing is Google's main rival for search traffic in the UK, it is, and has been for some time, a long way behind in terms of market share. Nevertheless, and despite accusations that Bing have been guilty of copying Google (Sullivan, 2011), Bing's index is sufficiently different to make it an important addition to any journalists search tool-kit. Several advanced operators are worthy of note, so here follow some highlights.

Pages containing files (contains:): This operator can be used to find web pages which contain particular file types, rather than the just files. This is a significant improvement on Google's *filetype:* operator in so far as web authors can be remiss in terms of how they name and label their files. Web pages give authors far more scope with which to define, describe and contextualise the files they are sharing. So...

apiculture contains:ppt

... will find pages that contain the word "apiculture" and have links to files with the .ppt extension. Some of the other file types you can search for using this method include *doc, pdf, ppt, rtf, txt* and *xls*.

Proximity search (near:): Bing has a proximity operator which can be used in a similar way to Google's AROUND function:

expert near:3 "chinese politics"

...will find you a range of experts (it is worth noting that Bing will alternate the order of these two expressions).

Co-citation (LinkFromDomain:): this operator identifies the outgoing links from a specific domain, which in turn lets you build a profile of any site you wish to investigate. This operator does not work on all sites.

Availability of Really Simple Synidcation (RSS) feeds (HasFeed:): this operator specifies that any page retrieved which contains your search terms must include a link to an RSS feed. It doesn't search within the content of RSS feeds; it only returns web pages which contain them. Advanced operators in Bing are employed with the same syntax as is found in Google.

Advanced searching – some other major search engines

Yahoo (http://uk.yahoo.com/) may be some way behind even Bing in UK search share, but it is nevertheless a major, international search engine. However, while once Yahoo boasted a number of highly useful advanced operators and filters (such as the *region:* operator which allowed searching for content by geography) these are mostly gone now, leaving only a handful of key metasearch options (http://help.yahoo.com/l/us/yahoo/search/basics/basics-04.html).

Ask (http://uk.ask.com/) is another alternative, but like Yahoo it lacks much in the way of innovation, and indeed it has outsourced much of its web crawling to Google, seeking to become a question and answer specialist (Sullivan, 2010). Nevertheless it can be important to use all of these search engines, especially for hard-to-find material. The domain function in both Yahoo and Ask is the same as is found in Google – *site:*.

Exalead (http://www.exalead.com/search/) is a French-based engine, incorporating a number of advanced options which are unavailable in most conventional web search engines, such as phonetic searching, word adjacency and word frequency searching, all available within a user-friendly drop-down on clicking the Advanced search button. This last option is particularly useful for finding background on issues and people.

Metasearch engines

Metasearch engines (or metacrawlers, or federated search engines) are devices which can be used to scour the indexes of several search engines simultaneously. They are often found in commerce (where comparisons are crucial to getting best value), but they are available in generalist fields too. Some search the most popular engines in a given field, others the most obscure; some present results by source (engine), others present all results cumulatively, using an internal relevance ranking. Advanced search in these tools is determined by the extent to which all of the engines searched share advanced search syntax.

Dogpile (http://dogpile.co.uk/) allows you to find results from Google, Yahoo and Bing, and is a great one-stop-shop for a very broad overview of search results. The engine provides its own form of relevance, and lists clearly which engines each result was returned from. There are some limitations to using this engine; there are no cached options, advanced search is limited to Boolean operators, language and domain options, and the sources available from within the news vertical are weighted towards US sources and content. Similar to Dogpile are Webcrawler (http://www.webcrawler.com/) and Metacrawler (http://www.metacrawler.com/) (the caveats about searching in Dogpile apply equally to both of these).

Yippy (http://search.yippy.com/) (which used to called Clusty) represents a genuine alternative to all of the above. Its preferences allow the user to set up bespoke tabs from a menu of options (covering web, news and blog verticals), though with the News options spanning CNN, Reuters, Yahoo! News and NY Times, there is a major US-bias in this engine. Results are easier to interpret thanks to the clustering option, which allows you to refine your search by 'sites' (at the top level this includes domain type, but further down you can refine by particular site). Domain type includes *.co.uk* which can be used as a loose proxy for UK sites.

Kartoo (http://www.kartoo.com/) is another metasearch engine that presents results using its own relevancy criteria. The designers of this engine opted for clustered results rather than offering advanced operators – and those clusters of results include several which hint at the consumer-driven nature of the engine. Each page in the Kartoo results has its own ranking profile, albeit it may be worth mentioning that Kartoo has no information on a sizable number of sites. This engine is relatively weak when it comes to finding UK people and names, with mis-ordered US name results often out-ranking UK results.

search.com (http://www.search.com/) provides a federated search across Google, Ask, MSN and public directory DMOZ results. Its advanced search features appear, at first glance, to offer a good range of options which all other metasearch engines fail to match, but the language, file-type, domain, link and related options all seem to be drawn exclusively from Google results (and not the others). The Images vertical draws its results form Webshots and Flickr, but results can be limited to just these two sources (it is rare indeed for searches in Google images to result in no results, but not necessarily so in search.com).

One metasearch engine which offers minimalistic search experience with greater reach than virtually all of the above, is Forelook (http://www.forelook.com/), albeit this engine offers less reliability. Zuula (http://www.zuula.com/) doesn't blend search results from different engines, but rather lets you select which engines to search from a tabbed menu – though there is no way to specify location, and results are US-led. Lastly, ixquick UK (http://ixquick.com/uk/), which claims not to track your IP address, represents a UK-biased alternative to all of the above.

Building your own search engine

Google custom search (www.google.co.uk/cse/) allows you to build customised search engines using its index; which for some specialisms can be a useful way to save time and effort. Simply by collecting a range of domains together, it is possible to search more precisely, and avoid information overload. Alternatively, Rollyo (http://www.rollyo.com/) lets you create your own customised search from Yahoo's index. It used to be possible to pull results together from both engines together into a build-your-own metasearch service (via an online tool called Agent55), but this is no longer available. Nevertheless, just because you can't view results from both in one place doesn't negate the value in searching each individually.

Semantic search

Most traditional search engines used to take a rudimentary, some may say ‘dumb’ approach to the content they index. Large swathes of information would be indexed, and surfers would search through it, but the engines wouldn’t ‘understand’ what the searcher was looking for, any more than it understood words as strings containing a series of characters. Today all major search engines make great efforts to try to determine what searchers ‘mean’ by the terms they search for, by interpreting the search terms searchers use, and by differentiating indexed words according to context (like a thesaurus does). This process has been referred to as *semantic search*.

While vestiges of *semantic search* are incorporated into mainstream search, it is equally possible to obtain a purer form of semantic search, by using any of the following engines.

Hakia (http://www.hakia.com/) would more fairly be described as a search portal than a search engine. Semantic search engines often involve natural language queries, rather than keyword search, so advanced operators are not always necessary. If you search Hakia for the expression:

> *what was 'The People's Charter'?*

...the engine should return results which recognise the tense of the question, and can differentiate ‘The People’s Charter’ from other political tracts described in this way. Search results for complex searches like these in Hakia are often comparable with, if not better than Google results.

DuckDuckGo (http://duckduckgo.com/) is another engine which places a high premium on quality in search results. While it has its own self-generated content pages, its main focus is on crowdsourced areas of the web. The site has its own unique and detailed syntax for performing functions (like working out numerical conversions) and advanced one-click searching in certain popular domains (for example !bbc "prince william" will search the BBC index for stories about Prince William).

Self-styled 'Web 3.0 search engine' Kngine (http://www.kngine.com/) offers an alternative, albeit rather US-centric, option. They also offer a statistics vertical, drawing exclusively upon UN, World Bank, CIA Factbook and other official material. If you needed to find out adult literacy levels in Belgium, then it might make sense to start here rather than with a generic search engine.

Visual search

Most search engine developers recognised long ago that the 10-links-on-a-page format of search results is an unsustainable way of dealing with the ever-growing volume of information on the Web. For this reason, there is today a wide range of visual search engines, whose results are presented in a more intuitive way. Sometimes this is an aesthetic measure, aimed at muting the volume of results, but for other engines the visual nature of search results makes for a different search experience, whereby the surfer can gain new insights into search results

Visual search engine NewsMap, which exploits the Google News API

NewsMap (http://newsmap.jp) is probably the best place to go on the net for a visual overview of key stories of the day, as drawn from the index of Google News. News stories are presented as a treemap visualization. This site offers some versatility when it comes to coverage selection; you can include or leave out world news, and other topical areas. The size of each cell reflects the volume of related articles found in Google News story clusters around these themes, so in essence it gives you an insight into the dominant news values of different groups of media. This site is as close as it is possible to be to a round-up in one place, from a good selection of news sources. (and it is reasonably up-to-date, albeit not strictly speaking real-time).

Quintura (http://www.quintura.com/) is a long-established, and much-praised engine in this field. However, for some searches this engine will return borderline-irrelevant content – no doubt due to the fact that it lacks any sense of understanding the meaning in the words which crop up. Nevertheless, the basic technology (and the building-blocks approach to search refinement) serves as a good introduction to the potential in visual search.

Spezify (http://www.spezify.com/) can be useful for visualising information across consumer, social media and multimedia sites (from Amazon, to Twitter, to SoundCloud), providing a mixed palate of information across fields and media-types.

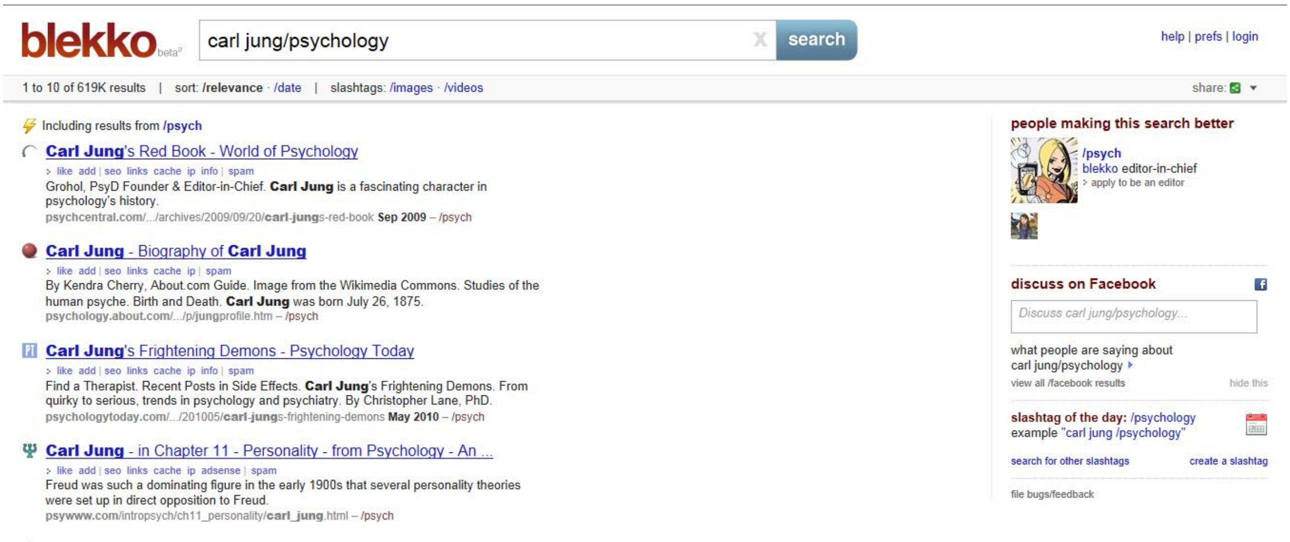
Search cloud (http://www.searchcloud.net/) can be used to weight the words you want, in terms of their relative frequency, across a set of search results. So for example, if you were interested in researching the UK cash for peerages scandal of 2006/7, and you were interested in finding reports and material which featured Tony Blair, Lord Levy, Dr Chai Patel and John McTernan, it is possible using Search cloud to 'weight' those terms individually – such that if John McTernan were the key focus of the research, by increasing the size of font in the 'cloud' (relative to the font size for the other search terms), more material should come back which references him more than the others. Unfortunately the site doesn't give much away about how its underlying technology works, about how re-scaling the size of each term impacts on relevance, nor about where its index comes from, but it nevertheless offers a very different, and innovative alternative to search.

Social search

The term “social search”, or even the application of the term 'social' to the concept of search is conceptually redundant in so far as the ranking methods used in leading engines represent a form of endorsement, and so are inherently social (Halavais, 2009, p160). But those developing and using the online tools which fall into this category do not seem to mind.

It has been argued, as an alternative to conventional information retrieval, that human intercession in relevance may go beyond an atomistic search “need”, towards enabling the potential in social capital to solve community problems; otherwise conceptualised as “sociable media” (Donath and Boyd, 2004). But whether this field is conceived as 'sociable media', 'collaborative filtering', 'social search' or 'crowdsourcing', the principal is simple; rather than rely exclusively (or mostly) upon metrics when it comes to determining what is 'relevant' in search, social search engines harness the wisdom (and efforts) of the crowd (or social graph), and communities of interest, to bring relevance to search.

For the purposes of structuring this book, I have differentiated between the process of social search as defined above, and both people-finders (many of which are social search engines), social bookmarking sites, and social news sites, all of which are dealt with separately, in later chapters.



blekko: a social search engine on the rise

blekko (http://blekko.com/) represents a radical, social alternative to Google in search – and its developer have come into conflict with big UK online web companies (such as CompareTheMarket.com) over the interpretation, and exclusion of 'spam' from their index (Arthur, 2010). This engine's index is structured around human activity, rather than the spidering of an algorithm. It's approach to search is described as 'slashtag search', which hints at the hierarchical, taxonomical, human-centric nature of organisation used. At present some searches are a little light on content by comparison with its long-established competition. But the social nature of this engine (and its wider recognition at least within the search community) mean that if any product is likely to challenge received wisdom in search, then this may be it.

Scour (http://www.scour.com/) is a composite social/federated search engine. It takes the principal of social news sites (where relevance is determined by people voting up and down the articles they most enjoy or trust), and applies it to results generated by Google, Yahoo!, Bing and OneRiot (as well as using its own ranking). The idea is to develop a community of trust, so taking part in voting (and also commenting on web search results – which often contain partisan posturing rather than insightful analysis) is dependent upon registering for an account with the service. The company clearly hope that surfers will encourage friends to do likewise, and so harness 'relevance' based on real-world communities.

This approach brings into question the efficacy of voting as a measure of *relevance,* in terms of journalistic enquiry. It is unclear, for example, whether people are voting for or against material based on its accuracy, or if they are simply voting content up or down because they *like* (or dislike) particular documents (or authors). Moreover, just because consensus can be reached on a topic does not mean conclusions drawn from this consensus are correct, or accurate. Eurekster's Community Powered Web Search (http://community-powered-web-search-swicki.eurekster.com/) provides an alternative to Scour.

ChaCha (http://www.chacha.com/) attempts to harness the wisdom of its users to answer common questions. The result is something most journalists should be able to appreciate – search for any concept (say presidential election), and results are returned in a who, what, where, when-type format. A more intuitive service, and one based on the very public expertise of its community, is provided by Quora (http://www.quora.com/)

Mobile search

At the time of writing there are few significant alternatives to the major search brands who have carved out a significantly alternative niche in mobile search (not withstanding Siri for Apple iOS, which is more of a personal assistant than a search vertical per se). In future, developments in mobile search will likely be driven by geo-location, predictive search and voice-recognition.

Some sources in search

There are many sources online which offer updates on developments in search; here follows a list of the most useful:

Bing Search Blog:
http://www.bing.com/community/blogs/search/

Phil Bradley's Weblog:
http://www.philbradley.typepad.com/

Google Operating Systems (unofficial):
http://googlesystem.blogspot.com/

(Official) Google Blog:
http://googleblog.blogspot.com/

Karen Blakeman's Blog:
http://www.rba.co.uk/wordpress/

LifeHacker:
http://lifehacker.com/

Mashable:
http://mashable.com/

Pandia:
http://www.pandia.com/

Read Write Web:
http://www.readwriteweb.com/

Search Engine Land:
http://searchengineland.com/

Search Engine Watch:
http://searchenginewatch.com/

Tech Crunch:
www.techcrunch.com/

The Next Web:
http://thenextweb.com/

Browser extensions for search

Some highly useful search tools come in the form of browser add-ons (or extensions), which are small programs run in conjunction with Internet browsers.

Surf Canyon's browser extension (http://www.surfcanyon.com/extension.jsp), used in conjunction with Firefox, will personalise results when searching in particular search engines. Likewise, Google's Similar Pages beta (https://chrome.google.com/extensions), when used in conjunction with the Chrome browser, will suggest similar pages to the ones you find.

There are many other extensions available, so it is worthwhile keeping on top of these over time. Here follow the 'big four' browser application directories (the Firefox directory is probably the most active – some add-ons created for this browser will be re-visited in subsequent chapters in this book).

Chrome extensions directory:
https://chrome.google.com/extensions/?hl=en

Firefox extensions directory:
https://addons.mozilla.org/en-US/firefox/

Microsoft Internet Explorer extensions directory:
http://www.ieaddons.com/gb/

Safari extensions directory:

http://extensions.apple.com/

References

Arthur, Charles (2010) New search engine Blekko labels MoneySavingExpert and CompareTheMarket 'spam', The Guardian, November 02:
http://www.guardian.co.uk/technology/blog/2010/nov/02/blekko-search-engine-labels-money-sites-spam

BBC (2006) Google Censors itself for China, bbc.co.uk/news, January 25:
http://news.bbc.co.uk/1/hi/technology/4645596.stm

Calishain, Tara (2005) Web Search Garage. New Jersey: Prentice Hall PTR (Pearson Education, Inc).

Donath and Boyd (2004) 'Public displays of connection', BT Technology Journal Vol 22 No 4 October 2004, pp 71-82:
http://smg.media.mit.edu/papers/Donath/PublicDisplays.pdf

Henninger, Maureen (2008) The Hidden Web: Finding Quality Information on the Net (2nd Edition). UNSW Press: Sidney

Long, Johnny (2005) Google Hacking for Penetration Testers, Syngress Publishing, Inc.

Schlein, Alan M (2004) Find it Online, The Complete Guide to Online Research (Fourth Edition). Arizona: Facts on Demand Press.

Sullivan, Dan (2011) Google: Bing Is Cheating, Copying Our Search Results, Search Engine Land, February 01:
http://searchengineland.com/google-bing-is-cheating-copying-our-search-results-62914

Thurow, Shari and Musica, Nick (2009) *When Search Meets Web Usability*, Berkeley: New Riders.

Chapter 3: The invisible web

The 'deep', 'invisible', 'hidden' web...

Search engines are not the only means of finding information online. Indeed some of the most useful online tools in the journalists' armoury are expressly not available via search engines, because this content does not exist on the *surface web*, where web crawlers can find it. Some research tasks are far better resolved by going straight to source; by finding interactive mediums, or sources of expertise beyond the wisdom of the crowd. This is especially true where search tasks are complex, hard to express in simple terms, or unlikely to exist in a convenient place online.

The *deep*, *invisible*, or *hidden* web, as it has alternatively been called, is a concept which has challenged the web development community since the mid-1990s. Broadly speaking, it is conceptualised as those areas of online material which are non-accessible to general search engines; World Wide Web content which lurks beneath the areas routinely indexed by major search companies. More detailed definitions of this concept have been prone to change over time.

Sherman and Price (2001) offer a thorough overview of what the 'hidden' web contains, and why this content is there:

Type of invisible web content	**Why it's invisible**
Disconnected page	No links for crawlers to find the page
Page consisting primarily of images, audio or video	Insufficient text for the search engine to “understand” what the page is about
Pages consisting primarily of PDF or Postscript, Flash, Shockwave, Executables (programs) or Compressed files (.zip, .tar, etc.)	Technically indexable, but usually ignored, primarily for business or policy reasons
Content in relational databases (often created using SQL)	Crawlers can't fill out required fields in interactive forms
Real-time content	Ephemeral data; huge quantities; rapidly changing information
Dynamically generated content	Customized content is irrelevant for most searchers; fear of “spider traps”

(Sherman and Price, 2001, p61)

In addition to this typology of hidden content, the authors went on to explain the existence of the invisible web by dividing it into four sub-divisions of information. The invisible web is therefore conceived as an umbrella term for:

The Opaque web (files which can be, but are not indexed)
The Private web (deliberately excluded; passwords/robots.txt/noindex)
The Proprietary web (hidden behind paywalls)
Truly Invisible web (the edge of search engine development) (2001, p70)

But even back in 2001, it was acknowledged that “what may be invisible today may become visible tomorrow” (Sherman and Price, 2001, p56), due in part to indexing policy, and in part to advances in search engine technology.

In 2008, Henninger refined these early definitions to exclude password-protected content, and content which uses HTML tags to prevent indexing and caching. This definition of the invisible web encapsulated “publicly accessible, non-proprietary pages that are not 'seen' by the spiders of general search engines” (Henninger, 2008, p162). Broken down into its constituent forms, this includes: *grey literature;* material not commercially published, or hard to find commercially, such as committee reports, and other official documentation (Henninger, 2008, p169).

More recently still, some have questioned the usefulness of the 'invisible web' as a concept (given that a number of research findings used to define it are becoming evermore obsolete, with the passage of time). It has been questioned just how useful ephemeral online information within the invisible web really is (Ryan, 2008). Equally, content which is now widely accessible in general search engines (indexing policies aside), but which was not when the concept first came to pass, include:

Pages in non-HTML formats (pdf, Word, Excel, PowerPoint), now converted into HTML.
Script-based pages, whose URLs contain a *?* or other script coding.
Pages generated dynamically by other types of database software (e.g., Active Server Pages, Cold Fusion). These can be indexed if there is a stable URL somewhere that search engine crawlers can find.
(University of California, Berkeley, 2010)

Today Google et al are trying to solve the invisible web problem by developing technologies which analyse and query databases on the web, whose content has until now been off-bounds, but this is a major project, with the breadth and range of database types available (Wright, 2009).

But whatever current or future definitions of the 'invisible web' we may use, general search may never provide the answers to many research questions. Today it is often far more efficient to go directly to directories, portals (community-driven subject directories), vortals (a portal focussed on one industry), gateway services (subscription resources where specialists help meet research needs) and other alternative search *verticals*.

People finders: subscription and free

Most media organisations have subscriptions to directories for finding people. These are proprietary online databases which combine official public data, such as (in the UK) the edited electoral roll, or directory enquiries, and which often include other sources of people-finding information too.

One of the better known premium sources in this field is 192.com (www.192.com/). This service offers a range of useful data sets; from current (and historic) electoral roll data, through information on directorships, births marriages and deaths, house prices (current and historic) business listings, maps and even company credit reports. Some of this information can be obtained elsewhere, and not all content is premium, such as directory enquires, which can be found freely elsewhere (http://www.thephonebook.bt.com/). However, the availability of all this information in one place makes for a particularly powerful (if relatively expensive) people-finding tool.

The ability to sort results by age-range for some (not all) people makes for a hugely useful filter (especially for common names). An inbuilt thesaurus which looks for contractions of common names, a visualisation option for checking when people have been registered at a particular property, and the option to view by house price, all contribute to improve search strategy. Director reports can also be mined in order to undertake social network analysis on individuals; showing who company directors know, and how far their public influence extends.

Example: follow the money...

While working in a regional newsroom in South Yorkshire, I was asked if it were possible to find out the associates of a local businessman convicted of fraud online. Using 192.com, we found director reports for his companies, all of which listed co-directors and other affiliates, which the journalists then used to pursue further leads.

As an alternative to 192.com, Tracesmart (www.tracesmart.co.uk/)(also a subscription) offer a different range of specialist tools, in addition to public directories. This service includes a (limited) mobile phone directory and a range of other public records directories (including individual small share holdings). Further premium alternatives to 192.com and Tracesmart, include:

Cameo (http://cameo.bvdep.com/) (electoral roll information only)
GB Accelerator e-Trace (http://www.gb.co.uk/gbgroup/products/gb-accelerator-e-trace-debtor-tracing) (which includes Royal Mail Postcode Address File data).

But it is important to consider one caveat when it comes to these official sources of public information. Since 2003, as a consequence of Data Protection legislation, people have had the right to opt out of many such databases. Moreover, the increasing number of people making their landline numbers ex-directory and opting for mobile telephones as an alternative to landlines contrive to make these *official* methods of people finding less and less useful as time goes on. It has been predicted that landline telephones have less than fifteen years of mass use left (Economist, 2009), which may serve as a benchmark for how much longer official people-finders along the lines discussed above may remain useful.

Yet ironically, just as people are asserting their right to privacy by having their names struck off public directories, so too are they voluntarily sharing aspects of their personal lives across the web. Most of the leading search engines do not offer people-finder verticals, so a gap in the market has opened up into which several free people-finding search engine brands have stepped.

123people (http://www.123people.co.uk/), is a free search engine that focusses on indexing those areas of the web most likely to yield information about people, including areas of the hidden web. It then sorts results by *type* of web content (and so images, email addresses and phone numbers are presented separately). Web content on 123people is a composite of Yahoo, MSN and Google results, and media content is organised by format, (so Word documents, PDFs, and other formats are all distinguished from each other). The tag cloud viewer offers more than merely an interesting representation of results – this way of representing information goes beyond the linear strictures of 10 search results, and lets you see which words are most associated with the name you are searching for.

One significant weakness with 123people (as with all free people finders), is the use of 'best match' (as opposed to 'exact match') in search results. The web is not a neatly constructed database with matching content, and easy-to-access, consistent data. So when searching for images, pages containing images of other people (but textual references to the person you searched for) will be returned. When it comes to emails, it is perhaps inevitable, given the various conventions people and companies use in configuring email accounts, that inconsistency abounds. These can include:

firstname.surname@domain
firstinitialsurname@domain
surname.initial@domain

On the other hand, such free people finders do offer something conventional search engines cannot – Google, for example (officially at least) does not index @ signs (Long, 2005).

Pipl (http://pipl.com/) is a popular alternative to 123people. However, Pipl is arguably less helpful when it comes to identifying UK domains (either by registration, or upper level domain) than it is when searching for US content. Much content in this database comes from social media sites, with a heavy emphasis on Youtube and Myspace content; all archives (geaneological), and news articles are from US sources (at the time of writing). Pipl encourages you to select a location, but doesn't let you include keywords in your search, which may help refine results where you are searching for a common name.

Yasni (http://www.yasni.com/) is a German competitor to both 123people and Pipl, which (unlike the other two) offers an email alert service (similar to Google News alerts) for new mentions of a name across the Web over time. Social Register (http://www.socialregister.co.uk/) is another option for people finding online.

In addition to these web-based engines, today there are many mobile application people-finders available, which can be used in breaking news situations to track what is happening in a given place. These include:

BrightKite (http://brightkite.com/)
Buzzd (http://www.buzzd.com/)
Foursquare (account required) (https://foursquare.com/)
Gowalla (http://gowalla.com/)

As an alternative to these free people finders, there are today a range of file-specific search engines, that is to say, engines which let you search for particular types of file, which can in turn help you to find particular types of contributor (such as experts who have produced publications on their specialism). They are not quite equivalent to doing a filetype: search in Google (covered in the previous chapter), in so far as they have their own indexes. They include docstoc (www.docstoc.com/) and Scribd (www.scribd.com/).

If you are searching for US citizens, there are many many more people finders which can help including:

iSearch (http://www.isearch.com/)
Intelius (http://www.intelius.com/)
Spokeo: (www.spokeo.com/)
PeekYou (www.peekyou.com/)
Wink (www.wink.com/)
Yahoo People Search (http://people.yahoo.com/)

Because of the current unstructured nature of web content, the subscription model of 'official' people finder will continue to be necessary to marketers, policy makers, credit-checkers, law enforcers and any number of other professions besides journalists. Until the internet contains a higher quotient of meaningful (semantic) information, the current generation of people finders will need to be supplemented.

Example: hidden needles...

Free people finders may lack precision, but they can be useful when combined with other, official sources. While consulting for an NGO, I was asked to find the contact details of a popular (and controversial) public figure. Electoral roll results were misleading – the results returned did not match the individual's social class profile. However, searching for the name in 123people returned an Amazon wish list which indicated that the person in question is known by his substituted middle names. Armed with this hidden knowledge, his identity became clear with subsequent searches of the electoral roll.

Finding local knowledge

When it comes to finding local people, and where they spend their time (as opposed to where they live), then local knowledge is key. So when seeking out vox pops in an unfamiliar area, The Knowhere Guide (http://www.knowhere.co.uk/) can be useful. The site was set up by skateboarders around the UK, sharing information on places to skate and places to avoid. Today most towns in the UK have a profile on the site. Here you can find out (via the *Hookup Spots* pages) places to go and places not to go to meet people, while the 'Cringing Cult of Celebrity' section tells you of famous sons and daughters in the area (much like Wikipedia's entries for towns do, albeit these are probably less reliable). There is no Wiki system of authority on this site, so every fact found here will require verification (as should any other piece of information you find online).

Finding experts

Perhaps the easiest means of finding experts is to search for expertise within search verticals covering books and literature; which would include Amazon and Google Books. Amazon permits browsing by topic, and results filtering by both popularity (one way of establishing authority in some – though not all, fields) and date of publication (which can be used to find relevant forthcoming releases – some releases in Amazon are years ahead of publication).

As an alternative to such mainstream book verticals, Abe Books may help with more obscure specialisms, albeit (obviously) the longer a book has been out of print, the less likely the author will be around to field your call. Other options in literature verticals include:

Biblio (http://www.biblio.com/)
Google Books (http://books.google.com/)
Library thing (http://www.librarything.com/)
What Should I Read Next (http://www.whatshouldireadnext.com/)

For a more tailored experience, Profnet (https://profnet.prnewswire.com/) has been used by UK journalists as a means of sourcing PR, expertise and case studies since 1992. Free to journalists (but not the 'experts' whose details are indexed), it helps with most sourcing problems, though of course ultimately, there is no guarantee that the experts are comfortable in front of a camera or with a microphone under their nose. It could equally be argued that this is altogether too convenient a way to source expertise, and that it risks limiting public opinion in news coverage to the contents or a relatively tiny sample of the population. When used in conjunction with the search strategies outlined in the previous chapter, however, it makes a valuable contribution towards news-gathering.

For legal expertise, Sweet and Maxwell's Legal Hub (http://www.legalhub.co.uk/) contains a Bar Directory for sourcing lawyers and an Expert Witness directory. For an alternative to the Bar Directory, Lawyer Locator (http://www.lawyerlocator.co.uk/) can be helpful.

Finding academics

There are a number of ways to find scientific expertise online. For newsgathering purposes, specialist research outlets like Alphagalileo (http://www.alphagalileo.org/) and Eurekalert (http://www.eurekalert.org/) can prove fruitful. Alternatively, Google Scholar is near equivalent to the expensive academic resources universities subscribe to. Google Scholar search results are ordered by volume of citations by default, which is one way to measure authority – the peer review process in academia is (it could be argued) a more reliable way of measuring this than volume of sales. Google Scholar offers email alerts.

Finding contributors via charities

Finding contributors for some social issues can be fraught with ethical dilemmas. In the first instance, it is probably best to approach people with sensitive conditions through the organisations they come into regular contact with, rather than contacting them directly. Asking a charity, care group or other NGO to circulate a request for contributions is more likely to yield positive results, and less likely to end with potential interviewees feeling harassed.

For ‘issues-based’ stories; for example health or social issues, poverty or social mobility, AskCharity (http://www.askcharity.org.uk/) contains a detailed directory of charities likely to yield both experts (such as practitioners and professionals in certain fields) and those who have experience of the issue at hand (who have come into contact with practitioners). The site offers journalists a case study mass mail-out to all relevant organisations in its index (though this is expressly not for the use of student journalists). For an alternative, browsable index of charities and activist groups on several alternative issues, see the contacts and links database in Schnews (http://www.schnews.org.uk/links/index.php).

While for the most part the Charity Commission website (http://www.charity-commission.gov.uk/) was established to monitor NGOs' financial details, it can also be used to keyword search groups working in fields which are unfamiliar to you. For example, Policy Library's site (http://www.policylibrary.com/) contains many think tanks, research centres and policy workers, papers, research and press releases from around the world. Equally, the National Council for Voluntary Organisations (http://www.charity-commission.gov.uk/), have significant news section, where campaigns at national and local levels are publicised. Internationally, the Worldwide NGO directory is a more fitting source (http://www.wango.org/resources.aspx?section=ngodir).

Finding celebrities

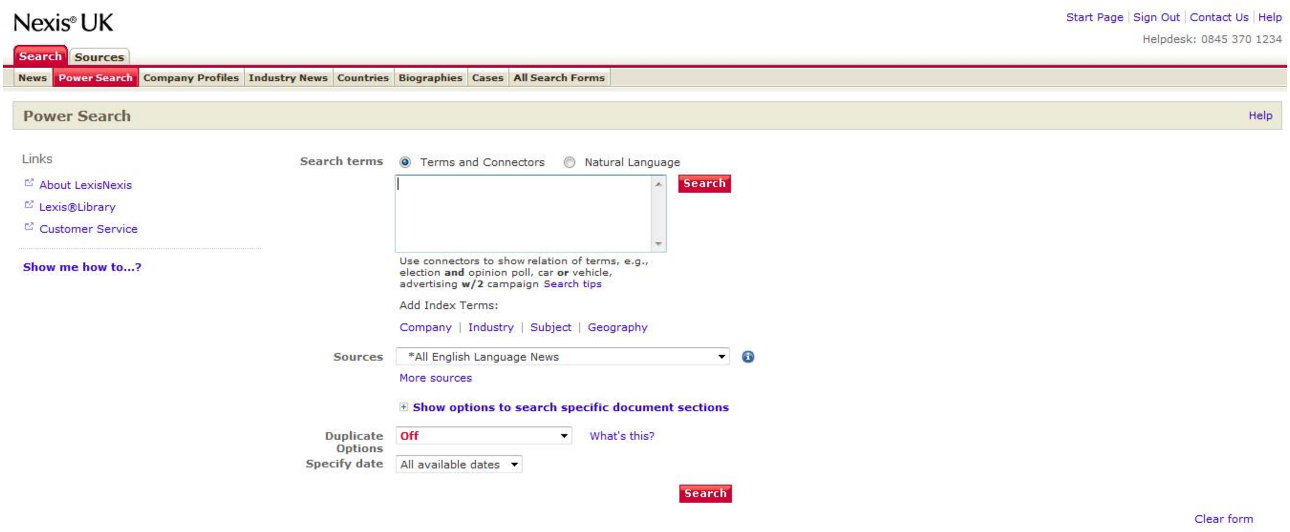
Spotlight (http://www.spotlight.com/) is the industry standard subscription database for finding actors and actresses (and child actors); albeit not all are by any means 'celebrities'. There is an element of quality control in this long-standing subscription-based brand; everyone herein must be professionally trained or experienced performers. A free-to-use alternative to this model is available in the form of Casting Call Pro (www.uk.castingcallpro.com/).

If you can't afford subscriptions to Celebrities Worldwide (www.celebritiesworldwide.com/), or Who's Who (www.ukwhoswho.com/), there are some free (albeit less reliable) alternatives. Most of the bigger acting agents have their own websites, and indexed directories.

Finding people via news archives

While some in academe have raised concerns about the lack of pluralism involved in the way journalists return again and again to the same sources, using the media to source contributors for some stories can nevertheless be essential. For example, if FIFA were to issue a new safety procedure covering football stadia, you may want to speak to someone who has lived through a disaster at a football stadium to substantiate your piece. But it you cannot find anything in your own newspaper's archives, you may dig through a newspapers archive, and find the names (and most likely ages – journalistic convention should see to that) of those who have been affected. Once you have these, you can take your details to a people-finding directory like 192.com, and check (using their age-profiling option).

Nexis and subscription newspaper archives



Nexis: newspaper archives are a fairly reliable way to source information online

Nexis is a newspaper cuttings database whose archives go (for the most part) back to the early-mid 1990s, and in some cases further still. It is an enormous repository of news archives, designed and developed long before the user-friendly and intuitive search experience we (mostly) associate with current-state search engines. As such, the unfamiliar must come to terms with a different kind of search experience when trying to get the best out of Nexis. Though studies suggest that the service may be lacking in the archival of wires (press agency) copy (Weaver and Bimber, 2008), it is nevertheless one of the most popular (and regularly used) tools available to journalists in their research.

An easy way into Nexis is to go for the *Power Search* option (see the left-hand column). At the top of the page you'll see a Simple Search box. Note the radio buttons for *Terms and connectors* and *Natural Language*. The default is set to the former, which allows you to add search functions, which will give you more control over your searching than merely using natural language.

The sheer volume of content in Nexis puts a premium on refining; and because all information is in a common format (i.e. news articles, with common fields), it is possible to provide some very specific field operators, which can significantly speed up search. Here are a selection of some of the most useful ones:

AND: you have to put this between all terms and connectors (Google allows spaces, Nexis doesn't).

OR: if you want to return results containing one or other term, use this.

AND NOT: use this to remove results containing words you definitely don't want (though be wary of false positives).

ATLEAST8(): this function allows you to set a threshold (i.e. 8) for the frequency with which your terms appear in results. If you are looking for an interview with an individual, or for a backgrounder for an issue, use this to refine your results, and miss out all the passing references. Use the function on the person's surname, then include a phrase search for their full name – i.e. *ATLEAST5(cameron) AND "david cameron"*.

LENGTH>(): This operator allows you to return results either above (>) or below (<) a certain word-length. This is useful for finding depth articles – few analyses of any major issue would be less than 800-900 words, so you can miss out all those news in briefs.

BYLINE(): If you work in a particular beat, and want to find out if one of your competitors has written something on an issue before, use this function to pull back all their work. Alternatively, use the **SECTION()** function.

HEADLINE(): this operator provides a very restrictive search, when you are getting too many results, and other filters aren't helping you.

Once you've decided what terms and operators to use (usually an iterative process), you can then determine which papers (or groups of papers) to search, and the time-span you want to search within. Here follows a series of examples showing how to use these operators together, where you want to find some detailed background information about the **G20**, with sources set to *UK Broadsheets*, and time-span set to *previous five years*.

> A search for: *G20* ...returned more than >3000 results.

This search returns too many results; even when changing the ranking to relevance. We can filter out passing references to G20, and increase the relevance of these results by using the ATLEAST function.

> A search for: *ATLEAST7(G20)* ...returned <200 results.

This search is getting much more accurate, but there are still too many stories to read in 10 minutes, to get a backgrounder on this organisation. It is possible to filter out those short stories which clog up the 'relevance' results ranking, and so focus on the lengthy pieces which feature the term G20 at least 7 times.

> A search for: *ATLEAST7(G20)* AND LENGTH(>1000) ...returned <100 results.

This is another improvement, but not enough. If references to Gordon Brown were important to this search, the following string would suffice:

> A search for: *ATLEAST7(G20) AND "gordon brown" AND LENGTH(>1000)* ... returned <60 results.

The results are refined, but it is possible to specify another very restrictive filter on this search – that all results coming back must feature G20 in the headline (this includes standfirsts):

> A search for: *ATLEAST7(G20) AND "gordon brown" AND LENGTH(>1000) AND HEADLINE(G20)* ... returned <50 results.

Working through these advanced options allows search refinement to the point where a manageable number of results is returned. But in addition, when results are returned, they can also be sorted by *relevance* (see the drop down on the top-left) which can help speed the process. Finally, all of these articles can be saved as a text or Word file. Nexis can also be used to generate original features ideas, depending on timing. By searching for at least 5 mentions of the keyword 'alcohol', within the *Society, Social Welfare and Lifestyle* subject, for articles written more than ten years ago, for example, it is possible to harvest some features ideas which may merit re-visiting, ten years later.

While archive search is often associated with the past, it can also be put to good use for finding upcoming events too, by searching the past few days' or weeks' content in the archive. Try using something like:

“home office” AND “next week” OR “next month”

As an alternative to Nexis, Factiva (http://www.dowjones.com/factiva/index.asp) offer a similar (subscription) service, and similarly useful advanced operators, much like those available in Nexis. These include *same* for finding two terms in the same paragraph, *w/n* for finding one word within a certain number of words (n) of another word, or *near/n* for finding one word within a certain number of words (n) of another word, where they appear in any order in the text.

Example: don't Sam so close to me...

Proximity search can be particularly useful when seeking out connections between people. When training a Sports journalist, I was asked to find context on the relationship between former Newcastle Football Club manager Sam Allardyce, and Manager of Manchester United, Alex Ferguson. A search in Factiva (subscription) for:

“sam allardyce” near “alex ferguson”

…brought back a court report in a local paper about a lawyer who stole from a paralysed man. Hidden at the bottom of the article it was mentioned in passing, that the accused “was a member of the prestgious (sic) Mere Golf and Country Club where regulars include football managers Sir Alex Ferguson, Sam Allardyce”. While this information might be useless in and of itself, it could be useful for anyone trying to contextualise comments made by either manager about the other in the press (i.e. supportive remarks). It highlights what interesting pieces of information can be teased out of online search, which would be massively time-consuming using basic free-text.

As for web-based alternatives to Nexis and Factiva, which anyone can access, there are further alternatives:

The British Newspaper Archive (http://www.britishnewspaperarchive.co.uk/)
Google News Archive (http://news.google.co.uk/archivesearch)
HighBeam (http://www.highbeam.com/)

Legal resources

Whether you need to consult legislation, or to keep track of who is appearing in court in the next few days, there are a handful of essential tools available in the UK, including:

Bailli (http://www.bailii.org/rss/): free updates on UK and Irish case law and legislation (includes RSS feeds).
Courtserve (http://www.courtserve.net/): subscription for up-to-date cases at Royal, Crown and County courts, as well as Scottish and Welsh courts, and UK Tribunals.

HM Courts (http://www.hmcourts-service.gov.uk/): for forthcoming court lists and judgements.
Legislation for all UK Acts of Parliament, by jurisdiction: (http://www.legislation.gov.uk/)
Outlaw (http://www.out-law.com/page-0): free legal guidance on technological issues (by Pinsent and Mason).
Scottish Courts (http://www.scotcourts.gov.uk/): for forthcoming court lists and judgements.
Westlaw (http://www.westlaw.co.uk/): subscription for finding cases, amongst other things.

Financial search tools

Sometimes it is helpful to look up an individuals' business records to give insight into their character, or something you have heard about them. Here follows several UK financial sources which help shed light on the business goings on of individuals, and corporations:

Companies House (http://www.companieshouse.gov.uk/): subscription for all official public company documentation, which includes disallowed director information free.
Corporate Data (www.corporateinformation.com): subscription which includes a decade of earnings information.
Credit Safe (http://creditsafeuk.com/): subscription for credit checks: you need permission of the person you are searching, according to the Data Protection Act 1998.
Director Check (http://company-director-check.co.uk/): find companies by director.
Individual Insolvency Register Search (http://www.insolvency.gov.uk/eiir/).
Level Business (http://www.levelbusiness.com/): free company report and accounts.
London Stock Exchange (https://www.londonstockexchange.com/): offers email alerts for breaking news on company matter.
Land Registry (www.landregistryservice.co.uk/Ownership): subscription useful for finding out who owns a property (by title, plan or lease).
Orbis (Bureau Van Dijk) (https://orbis.bvdep.com/ip/): subscription covering key data for companies operating around the world. Includes corporate structures, which can be useful when investigating international corporate relationships.
Red Flag Alert (http://www.redflagalert.com/): subscription which, in addition to company information, includes a unique 'alert' measurement of company 'health'.
Registry Trust (http://www.trustonline.org.uk/): details of County Court judgements.
Scoot (http://www.scoot.co.uk/): business directory, like Yellow Pages.
SCoRe (Search Company Reports) (http://www.score.ac.uk/): integrated search covering various UK collections of historical company reports. Useful when searching for past connections between companies.
Search Systems UK (http://publicrecords.searchsystems.net/): exhaustive list of subscription services covering everything from Aircraft registrations to World War II warships – includes many rare databases.
UK Data: company credit reports (http://www.ukdata.com)

Finding people by their family history

When it was discovered in 2010 that Ed Miliband is not named on his eldest child's birth certificate, this became front-page news for the Daily Mail. This story is evidence of the potential in ancestry databases for sourcing news. Here are some key UK sources in this field:

> Ancestry.co.uk (http://www.ancestry.co.uk/): subscription, for all family tree, genealogy and census records; includes data sets on births marriages and deaths, census, migration, military records.
> Find My Past (http://www.findmypast.co.uk/): competitor service to Ancestry.co.uk.
> National Archives (http://www.nationalarchives.gov.uk/documentsonline/): provides databases from many sources of public and state information, from air force records to wills. You will need to visit the archives in Kew to access some online tools.

Your local public library

Many of us are used to borrowing books, CDs, DVDs, and even drinking frothy coffee in local public libraries. But some UK local library authorities provide other services which are especially useful in journalistic research. Some of Greater London's borough libraries, amongst others across the country, provide access to highly useful online subscription resources. To access these sources from home, you need a library card number and (sometimes) an electronic pin number for the library authority that subscribes.

NewsUK and Newsbank

For those who cannot afford subscription news archives, there are free alternatives available at your local library. Although the search functions are pretty crude (and limiting) by comparison with Nexis and Factiva, NewsUK (www.newsuk.co.uk/) and Newsbank (www.newsbank.com/) offer phrase searching, date spans, a broad range (and depth) of regional and national titles, and the option to sort results by relevance (whereby those articles which feature your terms most frequently appear uppermost in your search results).

KnowUK

KnowUK (www.knowuk.com/) is a subscription ‘meta-search engine’, allowing users to search across more than 100 reference sources, many of which contain information not freely available over the net. For ease and convenience, the databases in KnowUK are organized into categories on the left-hand navigation. Because its uses are so many and so diffuse, here follows just some of the more useful categories for journalists.

Biography offers 20 biographical databases, which include generic sources (Debrette’s and Who’s Who for finding 'the great and good'), political sources (a database for each governmental body in UK politics), through clerical, civil service, sporting and even obituaries sources. Most of these sources include information such as work and life history, as well as hobbies and even sometimes contact details. As such they can be put to a number of uses.

Example: shark sandwich...

While training two natural history researchers, I was asked if it were possible to find a celebrity (for a pilot) who has a genuine interest in sharks. Take such a request to a castings agency, you may likely be sent anyone; but if you search through the biographical databases found in KnowUK you will find that Lloyd Grossman is a patron of the Shark Trust, and a PADI-certified divemaster. You'll also learn that Francis Rossi lists Coy Carp as a pass-time, and that Elaine Paige is a keen clay pigeon shooter.

As for the parliamentary databases, these include lots of useful information about MPs, such as their majority at the last election, their career history, membership of committees, their political interest, their commercial interests, and contact details. These are therefore useful contacts databases for any budding political journalist looking for a political contributor. House of Commons Biographies (published by Dods), European Parliamentary Biographies, House of Lords Biographies, Scottish Parliamentary Biographies and Welsh Assembly Biographies cover each legislature by degrees.

If you are working on a story about a local school, search to see if any MPs or other public figures attended it, they may provide some colourful actuality. House of Commons Biographies even includes former professions. Alternatively, if your MP is active on Twitter, you will find them on Tweetminster (http://tweetminster.co.uk/mps/search), but if not, there are alternatives such as the Urban75 list (http://www.urban75.com/Action/politicians.html).

Events contains forward planning resources. It can be used to search by keyword or location for events coming up for the year, and perhaps most usefully it can be searched by category.

General Knowledge contains some useful statistical sources in the form of Regional and Social Trends, in an easily searchable format (unlike the ONS site, which can be difficult at times). See too Titles and Forms of Address, which helps to avoid faux pas when addressing the landed gentry (be it for interviewing, or for rights to film on property).

Organisations, Associations and Charities is a self-explanatory area of the site. For the purposes of contributor-finding there are broadly two types of charity or NGO: policy groups, and practitioners (albeit larger organisations will sometimes fall into both camps). When using this section of the site, you don't have to rely on the ability of these organisations to promote themselves on the web, as reflected in web search results – this is a dedicated source. Professional associations (searchable within this section) are also a good way to find specialist contributors. If you want a local historian on Humberside – you may find the professional association (i.e. search local history).

Dictionary of National Biography

The Dictionary of National Biography (www.oxforddnb.com/) is an extensive obituaries database, whose content is generally written and maintained by academics and experts. While it may be of limited use to time-deprived national news journalists, it can be handy should you venture into features writing, and is especially useful in regional and local journalism. You can approach the site in two ways – searching for an individual by name, or key-word searching the database for places, events, landmarks and movements (or groups). The *People Search* in *Advanced search options* allows you to search by birthplace – an invaluable way for local and regional journalists to establish forgotten local heroes and notables. This source can also help further develop features. For example, searching for *Clifton suspension bridge* in Free-text will (obviously) bring back Isambard Kingdom Brunel's entry, but it will also return other major figures involved in the bridge's construction (such as civil engineer William Henry Barlow), as well as those influenced by the bridge (such as sculptor Kenneth Armitage).

Grove Dictionaries of Art and Music

As with the Dictionary of National Biography, Grove Art (www.oxfordartonline.com/) and Grove Music (www.oxfordmusiconline.com/) are both niche sources which can nevertheless be used effectively within regional and local journalism. Use the *Biographies* link (in either database) within *Advanced Search* to find artists and musicians by place of birth.

> **Example: small town Saturday night...**
>
> When training in regional news rooms up and down the country, I have found that searching the Grove databases for births and deaths by local place-name will often generate potentially useful subject matter for developing features. These bespoke databases are more thorough than free alternatives when it comes to biographical detail on obscure artists and musicians.

The Times Digital Archive

The Times Digital Archive (archive.timesonline.co.uk/tol/archive/) is an excellent primary newspaper source- providing access to The Times archive going back to the late 18th century, in the original format in which it was printed. While sometimes used as a source for making astons in broadcasting, it is more often the source of primary research in its own right; *The Thunderer* was, for much of the 19th and 20th centuries, the official organ of state. Several other UK newspaper archives are also available online, including the Guardian and Observer (http://archive.guardian.co.uk), the Financial Times (http://gale.cengage.co.uk/financial-times-historical-archive.aspx), The Economist (http://gale.cengage.co.uk/the-economist-historical-archive-18432003.aspx), and collections of The Daily Mirror, The Daily and Sunday Express and The Daily Star (http://www.ukpressonline.co.uk/ukpressonline/open/index.jsp). There are also nineteenth century newspaper archives (http://www.mediauk.com/article/32686/newspaper-archives-to-be-found-online). All are necessary in deep, historical journalistic research, because Nexis, Factiva and other archives mainly deal in post-1990s content. But they can also be really useful for inspiring a return to a feature topic, or to inspire a topic in a new context.

Encyclopaedia Britannica

It is always worthwhile cross-referencing the factual information you find with Encyclopaedia Britannica (http://www.britannica.co.uk/). This is the standard, industry recognized reference tool of choice, albeit research has shown that it contains equivalent numbers of serious errors as some of its free counterparts (Giles, 2005).

Events and anniversaries

Significant anniversaries and events (whether past, present or future) are core to a good deal of features, soft-news, and news-diary copy. Thinking ahead towards forthcoming events (such as the release of government reports or votes on certain issues), allows journalists to plan in advance their angle, background and copy. Most media organisations have subscriptions to databases which help, including:

Entertainment News: (http://www.entnews.co.uk/)
Foresight News (http://www.foresightnews.co.uk/)
Year Ahead (http://www.yearahead.co.uk/)

Some of the larger media groups have databases put together in-house, but there are plenty of free alternatives on the web.

A great deal of factual and current affairs programming today is generated around the anniversaries of significant historical events (usually in factors of ten years, but sometimes in factors of five). Some editors find these so-called *handles* editorially lazy and even desperate, but it is nevertheless possible to derive some value out of historical facts when woven into the narrative of a story. While isolated historical facts are not in themselves worthy of news or features coverage, their significance (such as in the form of a chronology of events) could most certainly be valid:

BBC On This Day (http://news.bbc.co.uk/onthisday/)

The Wikipedia home page also features an 'on this day' section, but their advanced search options (http://en.wikipedia.org/wiki/Wikipedia:Searching) allow for more thorough searching of past events. Here follow some speculative searches, all of which are angling at a 100 year anniversary of one kind or another (for birth years of famous composers, authors and English poets, respectively):

1910 intitle:composer

1910 -list intitle:author

1910 incategory:English_poets

Those who spend more time looking ahead (as opposed to looking to history) might instead want to use Whats On When, a database used to sell tickets for events, but which contains lots of useful local, regional, national and international event information:

What's on When (http://www.whatsonwhen.com/)

There are other tools for helping to source national and international events including events (elections, publication of reports etc.), such as:

Eventful (http://eventful.com/)
FT Week Ahead (http://www.ft.com/world/weekahead)

Some fairly obvious places to find out what's coming up the arts would include Time Out (http://www.timeout.com), and IMDb Upcoming Releases (http://www.imdb.com/calendar/?region=gb). There are many subject-specific options in hardcopy too (Prospect magazine, for example, has a monthly calendar of talks and other events which is essential forward planning for anyone who takes their philosophy/politics seriously).

Finding PR copy

This is for some the most contentious areas of sourcing news; while for others it is an essential part of any journalist's daily routine. There are many online agencies more than willing to distribute tsunamis of copy (and case studies) from those organisations who crave exposure. Here are just a few. Subscriptions, in some cases, may apply – and most of these services require sign-up:

FeaturesExec (http://www.featuresexec.com/)
Gorkana Media Database (http://www.gorkanadatabase.com/)
Pressbox (http://www.pressbox.co.uk/)
Prfire (http://www.prfire.co.uk/)
PR Web (http://ukservice.prweb.com/)
Response Source (http://www.responsesource.com/)
SourceWire (http://www.sourcewire.com/) (Tech and business)

General reference

Wikipedia (http://en.wikipedia.org/wiki/Main_Page) is the leading light in reference material online – but it does not come without its problems, not least the potential to mislead. If you are looking for authoritative, reliable reference website content, then using an open directory, such as DMOZ (http://www.dmoz.org/), or Yahoo Directory (http://dir.yahoo.com/) can be a more effective means of research. Such public directories, put together by dedicated human beings rather than ethereal algorithms, allow you to either browse down or search through. This can be very useful where searching for establishment sources, whilst trying to avoid the many hoaxes which exist in hyperspace (which are dealt with later in this book).

Henninger (2008) outlines four reasons why researchers should opt for these bespoke, hand-crafted public directories, namely:

> Manual classification generally assures the relevancy of the documents within the subject category.
> The initial focusing of the search process has been done for you.
> By browsing in a broad subject area you should arrive at a more specific aspect of the subject automatically.
> Documents on similar subjects are grouped together so all items are potentially relevant. (Henninger, 2008, p61)

However, one major disadvantage of using these tools is the vast difference in volume between automated indexes and manually-managed ones. For many searches, directories will simply not have the capacity of content to cope. Further problems with reliance upon open directory content have been explored in the literature. Sherman and Price cite unseen (and so unknown) editorial policies, timeliness, 'lopsided coverage' and the fact that some directories charge website owners for inclusion as mitigating factors (Sherman and Price, 2001, p24-26).

General research aside, there are some sections within the open directory which are especially useful to journalists; not least the following writing aides:

Dictionaries of slang:
http://www.dmoz.org/Reference/Dictionaries/World_Languages/E/English/Slang/
Etymological dictionaries:
http://www.dmoz.org/Reference/Dictionaries/Etymology/
Homonyms:
http://www.dmoz.org/Kids_and_Teens/School_Time/English/Grammar/Homonyms/
Literary dictionaries:
http://www.dmoz.org/Reference/Dictionaries/By_Subject/Humanities/
Rhyming Dictionaries:
http://www.dmoz.org/Reference/Dictionaries/Rhyming/
Style guides:
http://www.dmoz.org/Arts/Writers_Resources/Style_Guides/Grammar/
Thesauri:
http://www.dmoz.org/Reference/Thesauri/

Some other useful language tools not connected to the Open Data Project include:

Graphical dictionary:
http://www.visuwords.com/
Reverse dictionary:
http://onelook.com/reverse-dictionary.shtml

There are many other reference tools on the Web which can be helpful in journalistic research, including search engines, portals, vortals, and gateway services. Here follows just a handful of the most useful, from a UK perspective:

InfoPlease (http://www.infoplease.com/) : an extensive, world-wide Almanac – features an 'on this day' section for anniversaries.
Intute (http://www.intute.ac.uk/): the largest single collection of academic subject guides and resources, maintained by UK universities.
Nation Master (http://www.nationmaster.com/index.php): useful visual tool containing key economic, policical and social indicators from from the CIA World Factbook, OECD, World Bank etc.
UK National Statistics' Publication Hub (http://www.statistics.gov.uk/hub/index.html): self-explanatory collection of official UK data.
Wolfram Alpha (http://www.wolframalpha.com/): semantic search engine whose index is comprised of a selection of key data sets, whose purpose geared toward solving scientific and factual problems, and queries).
The World Factbook (https://www.cia.gov/library/publications/the-world-factbook/): the CIA's best annual attempt to publicly analyse countries from around the world.

If you need to contact fellow journalists at other media organisations, ABYZ News (http://www.abyznewslinks.com/) and PaperBoy (http://www.thepaperboy.com/) both offer extensive lists of news media online, and act as free alternative to Willings Press Guide (which is available via some local libraries online collections, in KnowUK).

References

British Library virtual reference library:
http://www.bl.uk/reshelp/findhelprestype/webres/humanitiesresourcesforresearchvirtualreferencelibrary/humanitiesvrl.html

Economist (2009) The decline of the landline, Unwired, August 13:
http://www.economist.com/node/14213965

Giles, Jim (2005) Special Report Internet encyclopaedias go head to head, Nature, 438, 900-901 December 15:
http://www.nature.com/nature/journal/v438/n7070/full/438900a.html

Henninger, Maureen (2008) The Hidden Web: Finding Quality Information on the Net (2nd Edition). UNSW Press: Sidney

Long, Johnny (2005) Google Hacking for Penetration Testers, Syngress Publishing, Inc.

Ryan, Kevin (2008) Must We Unlock the Deep Web?, Search Engine Watch, 12 November:
http://searchenginewatch.com/3631665

Sherman, Chris and Price, Gary (2001) The Invisible Web: Uncovering Information Sources Search Engines Can't See. Information Today, Inc: New Jersey.

University of California, Berkeley (2010) Invisible or Deep Web: What it is, How to find it, and Its inherent ambiguity, UC Berkeley - Teaching Library Internet Workshops, January 08:
http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html

Weaver, David and Bimber, Bruce (2008) Finding News Stories: A Comparison of Searches Using Lexis Nexis and Google, Journalism and Mass Communication Quarterly; Autumn 2008; 85, 3; ABI/INFORM Research pg. 515-530

Wright, Alex (2009) 'Exploring a ‘Deep Web’ That Google Can’t Grasp, New York Times, February 22:
http://www.nytimes.com/2009/02/23/technology/internet/23search.html?th&emc=th

Chapter Four: Social Media Theory

Crowdsourcing the news

> ...under the right circumstances, groups are remarkably intelligent and are often smarter than the smartest people in them. Groups do not need to be dominated by exceptionally intelligent people in order to be smart. Even if most of the people within a group are not especially well-informed or rational, it can still reach a collectively wise decision.
>
> (Surowieki, 2004)

Most journalists are generalists, and as such rely upon the opinions of experts and sources in a range of ways; from only trusting 'news' when they see it published by a news agencies, to forming close (Tuchman, 1973), some suggest too close, friendships with politicians, public relations professionals, industry leaders and NGO activists.

These working practices often take place in an environment of private information, which has served journalists well over the years, for good or ill. But in today's many-to-many communications platforms, these cosy, private relationships are coming unstuck. It is now possible for people who have hands-on experience or specialist knowledge concerning news events to broadcast their own 'news', on their own medium, dis-intermediating news media from the story, and so wrecking the exclusivity which drove newspaper journalism. It is now possible for experts and opinion formers, to establish themselves as publishers or broadcasters of their own news, and so connect directly to the public. Just as the advent of the electric telegraph helped to establish news as being synonymous with newspapers, so internet technologies now threaten to unpick this semantic connection.

But it would be misleading to think of journalists as perpetual victims of this new state of possibilities. On the contrary, journalists may take advantage of this new publishing landscape, to source actuality and expertise from the great sway of readers (and indeed non-readers) on news stories of the day – and many news media are now deeply engaged in involving their readership in uncovering stories. The Guardian's appeals for their reader's help during the MPs expenses (Rogers, 2009) and Trafigura (Leigh, 2009) scandals, and their Open Newslists (http://www.guardian.co.uk/news/series/open-newslist) experiment, are leading lights in a wider industry trend. While newspapers have often made appeals for help in the past, it is now possible for the public to get involved in a very visible and direct way, and for readers to engage in making news in real-time.

This reality has inspired fear in some for the future of journalism, for if journalists lose their status as gatekeepers; as breakers and makers of 'news', then (some feel) they lose their sense of purpose. But these social online tools offer journalists a means of sourcing opinion, expertise and even first-hand evidence about news events from a far wider range of sources than was ever possible before. And while most journalists may be generalists, this does not mean that journalistic skills, insight and ethics are not still crucial to the developing news ecology. Indeed, the importance of these skills today, and for the future, has been expressed convincingly by several authors, conceptualising the rise of Networked Journalism (Beckett, 2008), Grassroots Journalism (Gillmor, 2004), and the shift in journalism from gatekeeping to gatewatching (Bruns, 2004).

In short, journalists need to think of ways to apply their skills and knowledge to events as they break in this online environment. This requires, it is argued, a shift in the focus of journalistic convention – where transparency and openness will become every bit as important as fact-checking, and ethical discretion.

Social networking in the network society

Years before the rise of online social networks, sociologist Manuel Castells had argued that we live in a new epoch; a new technological paradigm where economies are primarily organised around the production and management of information, rather than organised around the production of energy, as they had been throughout the 20th century. It is not that information and knowledge are uniquely crucial in the current age (they arguably always have been in human societies), it is that our new technologies affect our economies, and so our societies, in a more significant way than past information technologies have affected bygone eras. Castells' thesis is predicated on three interwoven elements. The present age, he contends, is informational, global and networked.

The rise of the network is central to this new paradigm. Where once political power was embodied in hierarchical organisations and institutions, it is argued, today that power is spread across networks. Some nodes may be more important to the network than others, but all are essential to the project. Networks are democratic after a fashion; they decentralise power and decision making; they convert power relations into a basic logic of inclusion or exclusion. This is a contentious issue so far as public engagement in journalism is concerned. Online social networks are a self-selecting means of newsgathering – they do not fairly reflect wider society – only one in 20 over 65s engage with social media regularly, and less than a third of 35-44 year-olds are active online (Ofcom, 2010). Journalists should be careful how they develop their networks online, as exclusivity within an online network does not serve a genuinely pluralist press.

While social networking is not new, its electronic manifestations today is often criticised, even ridiculed, in the mainstream media because the output created is considered to be trivial, (Heffer, 2010) or because the process is considered to be an elitists fad (Street-Porter, 2009). However, these criticisms misunderstand fundamentally that social networks are not a cultural medium, so much as a platform. They are not an alternative to journalism, nor are they in and of themselves journalism. Twitter, Facebook and other online social tools are a communicative platform, not an end product. They are social media, with the emphasis on social (and all that it connotes). To compare Twitter or Facebook to the conventional media in terms of quality or standards, is to misunderstand the essential difference between the two.

Social networking harnesses a many-to-many communicative model, as opposed to the conventional one-to-many model upon which mainstream media was based. As such, it opens the door to anyone with internet access, living out-with oppressive regimes, to join in a global conversation, or broadcast their thoughts or experiences to the world. And while we must tread cautiously within such a medium, there are nevertheless opportunities for journalists to get involved in this conversation, and help eke out stories which might otherwise have been ignored.

Newspaper groups make great use of Facebook and other social networks, in the distribution of their stories (Jaffe, 2010), but their use in newsgathering and contributor finding is equally as significant. Social networks have been a staple source for US journalists for some time (Cison, 2010) and they are an increasingly popular source in UK online newsrooms too – research has shown that political reporting has changed irrevocably because of Facebook and Twitter (Newman, 2010). Facebook even provide a guide to using the platform in journalism:

https://www.facebook.com/journalists?sk=app_201416986567309

Being culturally and socially relative, social media have different functions and different groups of users, and this is changing all the time. Today much of Europe, North Africa, North America and India spend time on Facebook, but outliers persist; Poland's fascination with Nasza-klase (http://nk.pl/), Russia's fixation with Kontakte (http://vkontakte.ru/), Brazil's obsession with Orkut (http://www.orkut.com/) and China's love affair with QQ (http://www.qq.com/) remain out-with international trends. There remain many communities wedded to using a social network of choice – and it serves any journalist well to understand who they will likely encounter when seeking out stories, tip-offs and contributors via social networks.

Social networking and online newsgathering – a new ethical space?

As far back as the mid 1970s, the consequences of applying computer technology to medical procedures compelled some, chief amongst them Walter Maner, to argue that computers were opening up a new ethical space. Some question whether this can be true; arguing that computers do not materially change society, but merely speed up those processes to which they are applied; that we mistake quantitative change for qualitative change. Nonetheless, advances in technology do have consequences for media regulation; and not just with regard to broadcasting bandwidth. The UK Press Complaints Commissions' (PCC) Editorial Code acknowledged the potential in new technologies to blur the old certainties of public and private space, when in 1993 the definition of private property was tightened, with regard to commercially available advanced long-lens photography. Nevertheless, there are dangers inherent to altering legislation (and regulation) in order to accommodate (and deal with) problems arising out of technological innovation. Whichever philosophical school of thought you use to justify action or agency, it is important to recognise that online scenarios throw up some ethical issues which have no direct equivalent in the physical world.

Respect for online public space

Privacy is an area of ethical concern for developers and users of online services which depend on real-world social structures for their meaning; this is as true of the personalisation of search, as it is of the rise of social networks. But social networks are also virtual spaces, which can be used to harness a range of different types of relationship. While some argue that privacy is the enemy of free speech, others can point to the European Convention on Human Rights for evidence that privacy is important in Western societies.

Social networks can be private or public spaces, or a hybrid of both. The purpose of an online space is essential in determining how open to outside intrusion it is. For example, commercial networking sites, which exist to help in self-promotion for employment opportunities, are more obviously a public-facing domain than are re-uniting sites, which exist to re-connect individuals who have lost touch with each other, and who want to re-establish personal friendships. Another factor is the nature of site membership. Social networks used predominantly by real-world friends to keep in touch, can be considered to be a more private environment than subject-specific groups, which are open to anyone from anywhere. Use and uptake conventions are also important factors; privacy measures in Facebook are more nuanced than those associated with Twitter (which essentially has a binary privacy status). So for those users who understand these privacy measures (and not all Facebook users do – which is a separate ethical issue) the locus of any users' privacy on Facebook can be plotted on a spectrum, where Twitter users either have private or public accounts. Journalists must take an informed, reasoned, and case-specific path through online space; where (ethical matters aside) the risk to reputation for transgression can be significant.

At the more extreme end of privacy intrusion, 'ambulance chasing' is a phenomenon which long pre-dates the advent of the internet, and which has long inspired public opprobrium. Reporters who rushed onto Facebook in order to seek out students caught up in the Virginia Tech shooting atrocity of April 2007, badgered victims and witnesses alike with legitimate accounts, created fake accounts to badger them further, and even created memorials in order to ensnare grieving students. This led many to question the fundamental ethics of journalism in the age of the social network (Hermida, 2007).

In the UK two years later, two Sunday Express journalist invaded the privacy of survivors of the Dunblane shooting atrocity who, now teenagers, were found to be doing what most teenagers do on social networks (namely posting details and images of drinking, and swearing). The resulting controversy led to an uncharacteristically robust rebuke from the Press Complaints Commission (PCC, 2009) for the Express.

The platform which was so egregiously invaded by these Express journalists, Bebo (http://www.bebo.com/), is the UK's fourth most popular social network after Facebook, Youtube and Twitter as of June 2010 (Goad, 2010). It is an online community that all journalists should tread very warily around. Almost half of the users on this network are under 17, and the average age of its users is 28 (Royal Pingdom, 2010), significantly younger than other social networks. Given the special consideration journalists must take around minors, it is wise to weight intrusion into such networks against the very strongest of public interest defences; for most issues it is hard to conceive of a justification, beyond the most traumatic areas of investigative journalism.

Using material found on social networks

The terms 'social media', 'search' and 'Internet' do not feature in the National Union of Journalists Code of Conduct (http://media.gn.apc.org/nujcode.html), nor in the PCC Editorial Code of Practice (http://www.pcc.org.uk/cop/practice.html). Indeed, in 2008, following public disquiet about the exploitation of social network information, Tim Toulmin, then Director of the PCC, insisted that guidance on journalist's re-use of social network material need not be written expressly into the PCC code, as it is catered for already (Luft, 2008).

The BBC, which is regulated in a far more intrusive way than the UK press (at the time of writing), suggests this, in its editorial guide:

> Although material, especially pictures and videos, on third party social media and other websites where the public have ready access may be considered to have been placed in the public domain, re-use by the BBC will usually bring it to a much wider audience. We should consider the impact of our re-use, particularly when in

> connection with tragic or distressing event. (BBC, 2010, II)

Some have argued further that there is even a private sanctuary in public social network accounts; that some users of these services assume a certain amount of privacy due to a 'needle in a haystack' principal which governs how likely information shared is to end up being seen by significant numbers of people (Zimmer, 2010). On the other hand, it may be argued that this approach makes a mockery of the distinction between public and private space. Indeed, a number of cases brought before the PCC, including a policeman who made flippant remarks on his private Facebook account after the unlawful killing of Ian Tomlinson (PCC, 2009), and a civil servant who complained about the republication of her Tweets (PCC, 2011), have been rejected outright. There is no uniform position on this theme across all UK media; which points to a collective failing in these regulatory regimes.

Using subterfuge online to source news

The PCC has previously ruled that the use of a bogus online identity represents merely 'mild subterfuge', and so is legitimate in the presence of a public interest test (PCC, 2010) (though disclosure in and of itself is construed as a public interest defence, according to the PCC's Editorial Code). So far as the police are concerned, this form of entrapment is considered legitimate at the highest levels, albeit with caveats (Evans and Lewis, 2011). But while such behaviour may be legitimized in regulatory and even legal domains, as an ethical issue, there is a wider debate to be had. Where journalists undertake to use false personas to access closed, private online groups, without disclosing who they are or what their interests are, then they risk straying into deeply unethical territory. This can only fairly be mitigated by a public interest defence with the emphasis on 'important' information, whose impact on society is significant (and not transitory). (Friend, 2007, p81).

This is acknowledged within some media organisation's ethical guidance, including those rules employed at Reuters, which state:

> Reporters must never misrepresent themselves, including in chat rooms and other online discussion forums. They do not “pick locks” in pursuit of information, nor do they otherwise obtain information illegally. Discovering information publicly available on the web is fair game. Defeating passwords or other security methods is going too far. (Reuters, 2010)

Jeff Jarvis (2008) has argued that source transparency is replacing objectivity (or 'balance') as a new guiding ethic in journalism online – this notion is predicated upon the central, functional importance of the hyperlink in online communications as the glue which holds the Web together. Others maintain that such technologically determinist notions should not simply be allowed to trample over basic, universal human needs (such as privacy). Elsewhere, and more subtly, it is argued that attitudes to privacy are being challenged by the new ways people share information about themselves, and re-construct their lives online (Solove, 2008).

Some question the legitimacy of the 'off the record' principle on entirely separate grounds, arguing that journalists have every right to employ subterfuge in terms of identity (Niles, 2008), because there is a deeper problem with the struggle between transparency and privacy online. Such perspectives factor without the sanctity of personal privacy (whether on the part of the journalist, sources, or members of the public). The psychological importance for journalist and public alike to be able to communicate without fear or favour, to reason through complicated issues, to voice the unthinkable 'offline', where otherwise the consequences of sharing such information would be massively damaging to careers, cannot be ignored. While the primacy of the First Amendment informs some US thinkers, in the UK human rights legislation and conventions (and media regulations which pre-date them) recognise the importance and sanctity of personal privacy (in principal at least).

Some media organisations have sought to obviate the issue of 'objectivity' in their reporting by laying down rules and regulations about how staff conduct themselves in public (and private) capacities online. The BBC's social media guidelines can be seen to be merging professional and public space online, certainly as far as editorial journalists are concerned:

> Editorial staff and staff in politically sensitive areas should never indicate a political allegiance on social networking sites, either through profile information or through joining political groups. This is particularly important for all staff in News and Current Affairs, Nations and Regions and factual programming and applies regardless of whether they indicate that they are employed by the BBC or not. (BBC, 2010, I)

In 2009 the Washington Post's social media guidelines were leaked, showing that some media companies were going a step further than the BBC. The document encouraged Post journalists to engage with political individuals, groups and movements online, but to be seen to be doing so in a 'balanced' way:

> Our online data trails reflect on our professional reputations and those of The Washington Post. Be sure that your pattern of use does not suggest, for example, that you are interested only in people with one particular view of a topic or issue. (Kramer, 2009)

This attempt at 'keeping up appearances' represents for some an absurd last ditch, shoring up a convention whose time has (at long last) passed. The principal of 'objectivity' in news (which has its critics), owes its origins not just to the potential in new technologies from the late 19th century, but also to the advertisers who, demanding access to mass-market audiences, provided the cash-flow which kept print news in business from the mid-19th century onwards (at least according to a classical liberal reading of the history of the press). While some may question the significance of who we follow, or in what we re-publish online in terms of our own views, networked space clearly has serious implications for the scrutiny of 'objectivity' in news, and for news publishers and journalists.

Rupert Murdoch says sorry: legally and illegally sourced news

On Saturday July 16, 2011, the UK's national newspapers all ran with an extraordinary advertisement; an unavoidably humiliating apology from the most powerful and influential news baron on the planet: Rupert Murdoch. Published the day after Murdoch lost two of his most senior executives in Les Hinton and Rebekah Brooks, this was a long-overdue apology for the long-denied illegal news-gathering techniques employed by his journalists at The News of the World. Aside from bugging the phones of celebrities and senior politicians, journalists at this top-selling Sunday newspaper had hacked the phones of murder victims, victims of terrorist atrocities, and even those soldiers killed on duty whose rights the newspaper sought to campaign on behalf of.

Thirteen days before, the final edition of The News of the World rolled off the Waping presses. That day's copy bore contrasting emotions. On one hand, there was contrition for the illegality that had led to the newspapers' demise, but on the other a familiar self-congratulatory braggadocio concerning the papers' 'proud' record of journalism. Shameless and conflicted to the end, The News of the World died because their brand of storming, salacious, scandalous journalism relied too heavily on illegal methods of obtaining information.

But the News of the World were not alone in sourcing illicitly obtained information from private investigators in this way; on the contrary, a report by the Information Commissioner in 2006, *What Price Our Privacy*, showed that such practice was widespread throughout the UK news media. Operation Motorman traced illicitly obtained information which included "details of criminal records, registered keepers of vehicles, driving licence details, ex-directory telephone numbers, itemised telephone billing and mobile phone records, and details of 'Friends & Family' telephone numbers" (Information Commissioner's Office, 2006), to 305 named journalists from across the UK newspaper industry.

For the most part these illegal practices relate to offline news-gathering routines (phone-hacking, illegally obtaining personal information, 'blagging' etc.), and so fall out-with the remit of this book. But it is also alleged that the illicit use of malaware (or Trojans; programs sent often via email attachment to infect computers, which then allow the sender to log keystrokes, and monitor use of the computer) also abounded within this arm of Rupert Murdoch's News Corp empire.

While there is today a critical mass of information in the public domain about the scale of phone-hacking at News International, it is as yet uncertain how widespread the exploitation of computer hacking may have been. We have the anecdotal evidence of a former Prime Minister made in parliament (Brewster, 2011), anecdotal evidence from two celebrity bloggers (Leyden, 2011), and a Panorama documentary which proved that emails belonging to a former British Intelligence operative were hacked by a private detective, and obtained by then News of the World editor Alex Marunchak (BBC News, 2011). It remains to be seen just what lessons can be learned here; and two UK parliamentary enquiries (not to mention an FBI enquiry) will likely shed more light over the coming years.

It is the industrialisation of these practices at News International that makes their illegality and immorality all the harder to justify. When such practices become routine (rather than a last resort, when all legal methods have been exhausted), then this threatens to undermine the very public interest defence that investigative journalism has always relied upon to hold the corrupt to account in the UK. In a country whose legal system permits the rich and powerful (both individuals and corporations) to super-injunct legitimate investigation into, and criticism of their behaviour, the phone-hacking scandal may yet poses a serious danger to the proper functioning of British democracy.

But while leading politicians and public figures, and even some journalists (Bould, 2011), have framed these events as a crisis for British journalism, others might contend that this could not be further from the truth. After all, it was the relentless and fearless investigating undertaken by Nick Davies (and others) at The Guardian over several years, during which time threats, smears and casual dismissals were thrown at them by those implicated (amongst them some of the most powerful figures in contemporary British life). Today we can say Rupert Murdoch has experienced his Watergate (Bernstein, 2011), though the consequences for News Corp remain to be seen. Nevertheless, it is a credit to British journalism, in the pursuit of the public interest which drove Davies that we now know what we know; and that we can move on.

References

BBC (2010, I) Editorial Guidelines: Social Networking, Microblogs and other Third Party Websites: Personal Use:
http://www.bbc.co.uk/guidelines/editorialguidelines/page/guidance-blogs-personal-full

BBC (2010, II) Editorial Guidelines: Privacy and consent:
http://www.bbc.co.uk/guidelines/editorialguidelines/page/guidelines-privacy-privacy-consent/#material-from-social-media

BBC News (2011) News of the World executive obtained hacked e-mails, BBC News, March 14:
http://www.bbc.co.uk/news/uk-12712400

Beckett, Charlie (2008) SuperMedia – Saving Journalism so it can Save the World. WileyBlackwell: London.

Bernstein, Carl (2011) Murdoch's Watergate?, Newsweek, July 11:
http://www.newsweek.com/2011/07/10/murdoch-s-watergate.html

Bould, Sarah (2011) NoW scandal has 'sullied the name of journalism', Hold The Front Page, July 13:
http://www.holdthefrontpage.co.uk/2011/news/scandal-has-sullied-the-name-of-journalism-says-men-editor/

Brewster, Tom (2011) The former Prime Minister says Trojans were used by the now deceased tabloid as the hacking scandal rages on, IT PRO, July 14:
http://www.itpro.co.uk/634912/gordon-brown-claims-malware-used-in-notw-hacking

Bruns, Axel, (2005) Gatewatching: Collaborative Online News Production. Peter Lang Publishing: New York.

Castells, Manuel (2000) Material for an exploratory theory of the network society, British Journal of Sociology Vol No. 51, Issue No. 1 (January/March 2000) pp. 5-24

Cison survey (2010) Journalists' use of social media for research increases, but traditional media are still seen as more reliable:
http://insight.cision.com/content/GWU-request

Evans, Rob and Lewis, Paul (2011) Met chief says officers' use of fake identities in court was not illegal, The Guardian, October 27:
http://www.guardian.co.uk/uk/2011/oct/27/met-police-activists-fake-identities

Friend, Cecelia and Singer, Jane (2007) Online Journalism Ethics, Traditions and Transistions. ME Sharpe: New York.

Gillmor, Dan, (2006) We The Media: Grassroots Journalism by the People for the People. O'Reilly Media, Inc: New York.

Goad, Robin (2010) Social networks now more popular than search engines in the UK, June 08:
http://weblogs.hitwise.com/robin-goad/2010/06/social_networks_overtake_search_engines.html

Heffer, Simon (2009) Facebook: Everything you never wanted to know..., Telegraph, June 13:
http://www.telegraph.co.uk/comment/columnists/simonheffer/5524402/Facebook-Everything-you-never-wanted-to-know....html

Hermida, Alfred (2007) Social media poses digital dilemmas for journalists, Journalism Ethics, June 08:
http://www.journalismethics.info/feature_articles/social_media_poses_digital_dilemmas.htm

Howe, Jeff (2008) Crowdsourcing: How the Power of the Crowd is Driving the Future of Business.

Information Commissioner's Office (2006) What Price Privacy? The unlawful trade in confidential personal information:
http://www.ico.gov.uk/upload/documents/library/corporate/research_and_reports/what_price_privacy.pdf

Jaffe, Alexandra (2010) Facebook Social Plugins and Open Graph: What they mean for online newspapers, Editors' Weblog, April 22:
http://www.editorsweblog.org/analysis/2010/04/facebook_social_plugins_and_open_graph_w.php

Jarvis, Jeff (2008) On Transparency versus Objectivity, Big Think (video):
http://bigthink.com/ideas/1260

Kramer, Staci (2009) WaPo's Social Media Guidelines Paint Staff Into Virtual Corner; Full Text of Guidelines, paidContent, September 27:
http://paidcontent.org/article/419-wapos-social-media-guidelines-paint-staff-into-virtual-corner/

Leigh, David (2009) Trafigura drops bid to gag Guardian over MP's question, The Guardian, October 13:
http://www.guardian.co.uk/media/2009/oct/13/trafigura-drops-gag-guardian-oil

Leyden, John (2011) 'I caught a virus from Murdoch's organ' – famous secret hooker, Malware de Jour also hit Girl With One-Track Mind, The Register, July 12:
http://www.theregister.co.uk/2011/07/12/malware_de_jour/

Luft, Oliver (2008) Social networks need to do more to raise awareness about misuse of personal information, says PCC director, Journalism.co.uk, June 05:
http://www.journalism.co.uk/news/social-networks-need-to-do-more-to-raise-awareness-about-misuse-of-personal-information-says-pcc-director/s2/a531706/

Newman, Nic (2010) #UKelection2010, mainstream media and the role of the internet: how social and digital media affected the business of politics and journalism. Reuters Institute for the Study of Journalism: Oxford University, July 6.

Niles, Robert (2008) There's no such thing as 'off the record' anymore, Online Journalism Review, April 16:
http://www.ojr.org/ojr/stories/080416niles-on-the-record/

Ofcom (2010) UK Adults' Media Literacy Report, May 17:
http://stakeholders.ofcom.org.uk/market-data-research/media-literacy/medlitpub/medlitpubrss/adultmedialitreport/

Press Complaints Commission (PCC) (2011) Ms Sarah Baskerville, February 08:
http://www.pcc.org.uk/news/index.html?article=NjkzNA==

Press Complaints Commission (PCC) (2010) Mr Paul Smith vs Hull Daily Mail adjudication, July 27:
http://www.pcc.org.uk/news/index.html?article=NjU0Mg==

Press Complains Commission (PCC) (2009) Ms Mullan, Mr Weir & Ms Campbell vs Scottish Sunday Express adjudication, June 22:
http://www.pcc.org.uk/news/index.html?article=NTc5Mw==?oxid=a8ea2047270292dc8f98ecbcb23a6d0c

Press Complaints Commission (PCC) (2009, II) PCC rejects privacy complaint about Facebook comments, December 02:
http://www.pcc.org.uk/news/index.html?article=NjA4Mg==

Reuters (2010) Social Media Guidelines:
http://handbook.reuters.com/index.php/Reporting_from_the_internet#Social_media_guidelines

Rogers, Simon (2009) How to crowdsource MPs' expenses, The Guardian, June 18:
http://www.guardian.co.uk/news/datablog/2009/jun/18/mps-expenses-houseofcommons

Royal Pingdom (2010) Study: Ages of social network users, Royal Pingdom, February 16:
http://royal.pingdom.com/2010/02/16/study-ages-of-social-network-users/

Solove, Daniel J. (2008) Do Social Networks Bring the End of Privacy? Scientific American Magazine, September.

Street-Porter, Janet (2009) Editor-At-Large: Twitter ye not, for it will not change the world, August 16:
http://www.independent.co.uk/opinion/columnists/janet-street-porter/editoratlarge-twitter-ye-not-for-it-will-not-change-the-world-1772833.html

Surowiecki, James (2005) The Wisdom of Crowds. Abacus: London.

Tuchman, G (1973) 'Making news by doing work: Routinizing the unexpected', American Journal of Sociology, 79(1), pp110-131.

Zimmer, Michael (2010) Is it Ethical to Harvest Public Twitter Accounts without Consent?, MichaelZimmer.org, December 02:
http://michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent/

Chapter 5: Social networks and newsgathering

Introduction

There are many social networks available to journalists who need to make contacts, and source news. Here follows a close look at how to use these sources, including an overview of their relative potential as people-finders.

Facebook

Since 2009, Facebook (http://www.facebook.com/) users have rapidly been getting older (Schroder, 2009) – those aged between 35-54 are joining at a higher rate than all other age-groups – though all age-groups are (at the time of writing) increasing on Facebook, as reflected in the landmark half-a-billion users milestone achieved in July 2010 (Myslewski, 2010) In the US, research has shown that more affluent, and more urban users will likely be heavier social networkers, and that Facebook users are on average more affluent than, for example, MySpace users (NielsenWire, 2009) The average Facebook user is 38 (Royal Pingdom, 2010), but in the UK the greatest number of users is aged between 18-25 (Su, 2010). Facebook users are, on average, more female than male (Royal Pingdom, 2009). But these figures have changed significantly over time (Social Media Optimisation, 2008), and will no doubt continue to do so.

It has been argued that long-term Facebook could become a genuine rival in search to Google in search, as we move toward seeking out information filtered through the interests of our friends and contacts online (Goodman, 2009). In any case, social networks are, as of June 2010, more popular than search in the UK (Goad, 2010). Many of Facebook's users are interested in international social and political issues, and some are experts in their field – the site contains groups based around themes and issues from around the world.

Facebook requires you to sign up in order to make any meaningful use of it. At first glance it doesn't seem a difficult site to search, but when results come back, it is worth bearing in mind the various filters which comprise this particular social network (listed as *navigation* on the left hand side of search results). The *People* option lets you search for people by name – allowing refined results by location, school or workplace (or a combination of these) once your initial results are returned. It is possible to put a place name into people search, but these results can be inconsistent. There will be groups for place-names which are more inclusive, but they may contain many people with a connection to the place in question, rather than current residents. On the other hand, those people found by place-name search may have moved elsewhere without updating their profile, not to mention the many people who haven't bothered to populate this field in their profile. You can search for profiles by email address if you have one – this is one of the few mass-market online tools which allow this.

Pages can be useful for padding out a geographical beat. Try searching for the place-names in your beat, and become a fan of these pages to keep up on issues both trivial and serious. There are, at the time of writing, over 500 pages featuring the term 'peterborough'

– many (though not all) of them will be relevant to a journalist whose beat includes this town. The same could be said for *Groups* – which differ from Pages in so far as they denote a sense of activity in their members, rather than a loose association, or general interest in something (say a brand, a celebrity or an institution). Groups can be especially useful in tracking future events – membership is controlled, and the privacy settings and responsibilities involved can make for a more personal, as opposed to corporate or general network experience.

If you have developed a healthy following, *Posts by Friends* can be a good place to search for interesting news. *Posts by Everyone* lets you find publicly shared status updates from all of Facebook – be aware though, that many people choose not to make their status updates public. Facebook encountered severe criticism when, in 2009, it made all status updates public by default – inspiring some concerned individuals to shame the company into changing its policy, by making all of these updates easily searchable in Your Open book (site now defunct). Otherwise, you can browse for contributors by a number of criteria including place and employer (http://www.facebook.com/find-friends/browser/).

Applications: there have been, during Facebook's history, some very useful Facebook applications for journalists, which are now defunct (including Advanced Search 2.2 Beta for general search, and Truescoop for finding US citizens with criminal records). One app which continues to prove its worth however is the Facebook Chat History Manager add-on (https://addons.mozilla.org/en-us/firefox/addon/facebook-chat-history-manager/) for Firefox (or Chrome). This tool ensures that your Facebook conversations are stored for later (just in case you delete your inbox).

It has long been rumoured (Van Grove, 2010) that Bing will incorporate public Facebook status updates into its index, however at the time of writing Bing's social vertical (http://www.bing.com/social) doesn't yet appear to provide them. Neither are they available in Google, for now. But that is not to say that conventional search engines can't help you make sense of Facebook content in ways which the site's own search doesn't allow. Facebook has many public topic pages; forums which bring people together around different issues – but the content of these conversations is not searchable within Facebook. As an alternative, try the following domain search in Google, in conjunction with a search term you are interested in:

site:facebook.com/topic bbc

Because Google indexes the whole web page, keywords buried in text (and not present in topic titles or names) can be found – this can be a hugely useful means of digging out obscure information, for the purposes of search, or for reputation management.

Twitter

One of the principal advantages of using Twitter (https://twitter.com/) over Facebook and many other social networks, is that you don't have to have an account to make use of its research and newsgathering potential. Though of course, in terms of actually communicating with others, and building both your reputation and your online contacts, it certainly helps.

Sixty four per cent of Twitter users are over 35 (Royal Pingdom, 2010). By November 2009, the UK had 5.5 million 'Tweepz'. According to a YouGov survey in 2009, Twitter users are younger than average, more likely to vote Labour, more liberal, more London-centric (and less Northern), and more likely members of lower social classes (Sparrow, 2009). It is not, however, as popular with users in their teens, according to research undertaken by Morgan Stanley a few months earlier (Kollewe, 2009). The conventions used on Twitter have traditionally been driven by its users, but more recently the group have published a guide to Twitter use in newsrooms (http://media.twitter.com/newsrooms).

Anatomy of a Tweet

Though it is not essential to get involved in Twitter in order to find news, it is nevertheless useful to understand some of the basic shorthand conventions of the medium. Here is a basic primer:

Tweet by environmental campaigner Joss Garman aimed at climate 'sceptic' thinktank The Global Warming Policy Foundation

Re-tweet: Tweet has been re-tweeted (ie re-published) to the followers of six different users (you can see the avatars of five of them below the message). The process of re-tweeting alerts a users' followers that they are sharing a Tweet they've found. This can be done manually (by typing out RT before a re-published message), but is automated in most Twitter client platforms. Alternatively, **MT** means that someone has modified the tweet they are sharing.

.@TheGWPF: This is the intended recipient of this tweet (albeit, everyone else is effectively copied in, by means of the full-stop just before the @ sign). Users can send private Tweets (or direct messages) which will not appear publicly too.

#BENNYQA: The # symbol is used to develop keyword-driven topics, to focus everyone's conversation around a particular theme. It can be quite a messy process finding trending topics, but their distribution is very much bound up in the social nature of the medium. That is to say, someone starts up a trending topic, which their friends can see and contribute to, and which then in turn their friends' friends can contribute to, and so on. If you want to find out if a trending topic for a subject exists, just try out some speculative efforts in the Twitter search box, making sure to prefix with #. This particular topic concerns a question and answer session by Benny Peiser, of the GWPF. Trending topics in Twitter are discussed in more detail later in this chapter.

Shared links: Given the 140-character limit on tweets, it is important to be able to shorten any link you want to share with one of many url shortening services – something Twitter does automatically. There are alternatives (i.e. bit.ly, or tinyurl.com), and all are worthy information sources in their own right. You can search the bit.ly index for content around a theme, which will show you relevance by social discovery, rather than search engine algorithm.

For those who don't want a Twitter account, there is a separate advanced search (https://twitter.com/#!/search-advanced). Full boolean options are available, as well as a hashtag search option (for searching out trending topics – prefixed with a #). The *People* filters are fairly limiting, and the *Attitudes* filter extends sentiment analysis only so far as the use of emoticons in posts (which is not particularly helpful for seeking out hard news stories). The *Containing links* option can be used to dig out conversations around links or even multimedia. The full list operators (which can be used in simple search) includes the following:

To find tweets by a user, use *from:<username>*
To find tweets to a user, use *to:<username>*
To find a tweet from someone near a place, use *near:“aberdeen”*
To specify the approximate distance of the tweeter from a place use within:25k (where k = Kilometres. Can also use m for Miles)
To find tweets posted since a particular date use *since:2012-12-09* (this uses US date format)
To find tweets posted before a particular date use *until:2012-12-09*
To find tweets containing URLs use *filter:links*

The real value in this search tool is in the *Places* option – whether used on its own, or in conjunction with other search terms and options. It will return people based either on the location they have typed into their profile (which may or may not be reliable), or if they have enabled the *Tweet with your location* option when using Twitter on a mobile device.

Once you have a keyword search, you may want to create a Really Simple Syndication (RSS) feed for it, to help you keep on top of your beat - more detail on RSS can be found later in this book, in Chapter 7. This is not a simple process in the current web interface (it is worth remembering that many users access the service via third-party applications and sites). To get an RSS feed by keyword, use the following URL, replacing <keyword> with your own search term:

http://search.twitter.com/search.rss?q=<insert keyword>

For setting up RSS Twitter feeds with more elements, here follow a number of templates which can be altered using your own keywords and other search parameters:

To keep a user's timeline: http://twitter.com/statuses/user_timeline/<insert user ID>.rss

(you will require the user's unique identifying number for this, which can be found via www.idfromuser.com/)

Trending topic feed: http://search.twitter.com/search.atom?q=%<insert topic>

User mention search feed: http://search.twitter.com/search.atom?q=%40<insert username>

Multiple keyword feed: http://search.twitter.com/search.atom?q=<insert keyword 1>+<insert keyword 2>

Either/or keyword feed: http://search.twitter.com/search.atom?q=<insert keyword 1>+OR+<insert keyword 2>

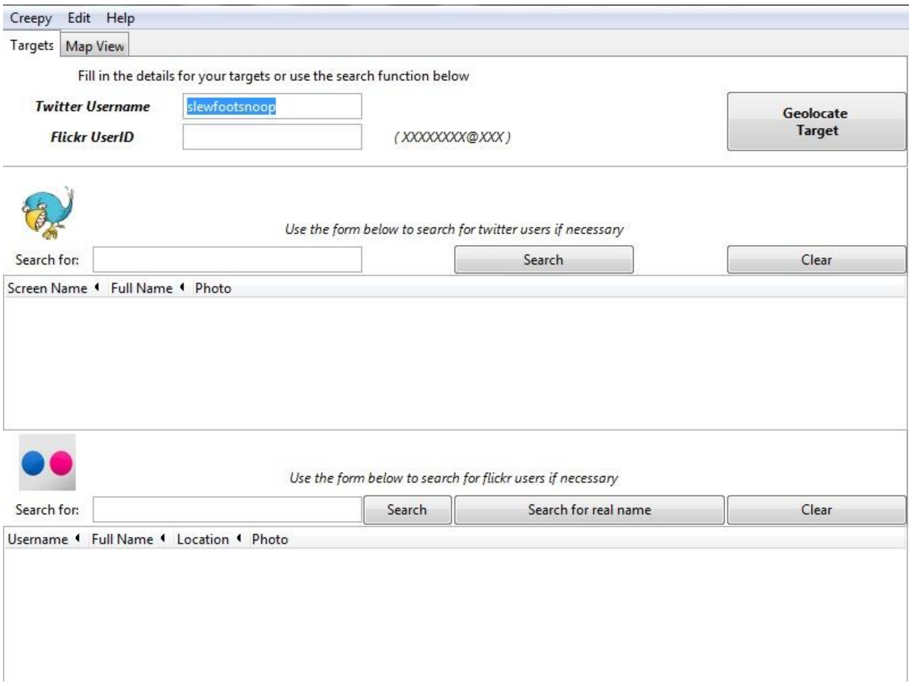
Geo-location keyword feed: http://search.twitter.com/search.atom?geocode=<insert longitude>%2C<insert latitutde>%2C<insert radius>mi&q=+<insert keyword 1>+<insert keyword 2>

To find the GPS co-ordinates for physical locations, run a place search in Google Maps and then copy the *Embed code*, found in the S*hare* option – look through this code for a number directly preceded by an equals sign – this is the longitude for your location, and immediately after this (separated by a comma) is the latitude.

```
Hacked URL:

http://search.twitter.com/search.atom?geocode=51.501166%2C-0.14227%2C0.2mi

Code:

<iframe width="425" height="350" frameborder="0" scrolling="no" marginheight="0"
marginwidth="0" src="http://maps.google.co.uk/maps?
f=q&amp;source=s_q&amp;hl=en&amp;geocode=&amp;q=buckingham+palace&amp;aq=
&amp;sll=51.479902,-
0.565436&amp;sspn=0.006268,0.021136&amp;g=sl39hl&amp;ie=UTF8&amp;hq=bucking
ham+palace&amp;t=h&amp;ll=51.501166,-
0.14227&amp;spn=0.006295,0.006295&amp;output=embed"></iframe><br /><small><a
href="http://maps.google.co.uk/maps?
f=q&amp;source=embed&amp;hl=en&amp;geocode=&amp;q=buckingham+palace&amp;a
q=&amp;sll=51.479902,-
0.565436&amp;sspn=0.006268,0.021136&amp;g=sl39hl&amp;ie=UTF8&amp;hq=bucking
ham+palace&amp;t=h&amp;ll=51.501166,-0.14227&amp;spn=0.006295,0.006295"
style="color:#0000FF;text-align:left">View Larger Map</a></small>
```

Embedding code from Google maps: note GPS co-ordinates in bold

Be wary of using conventional place-names in search – duplicate place-names from around the world will cause problems, and US users are still far more prevalent on Twitter than are UK users. Twitter no longer supports geo-spatial coordinate searching, nor post code searching, but you can track down the co-ordinates of individual Tweepz using the freeware cree.py (http://ilektrojohn.github.com/creepy/).

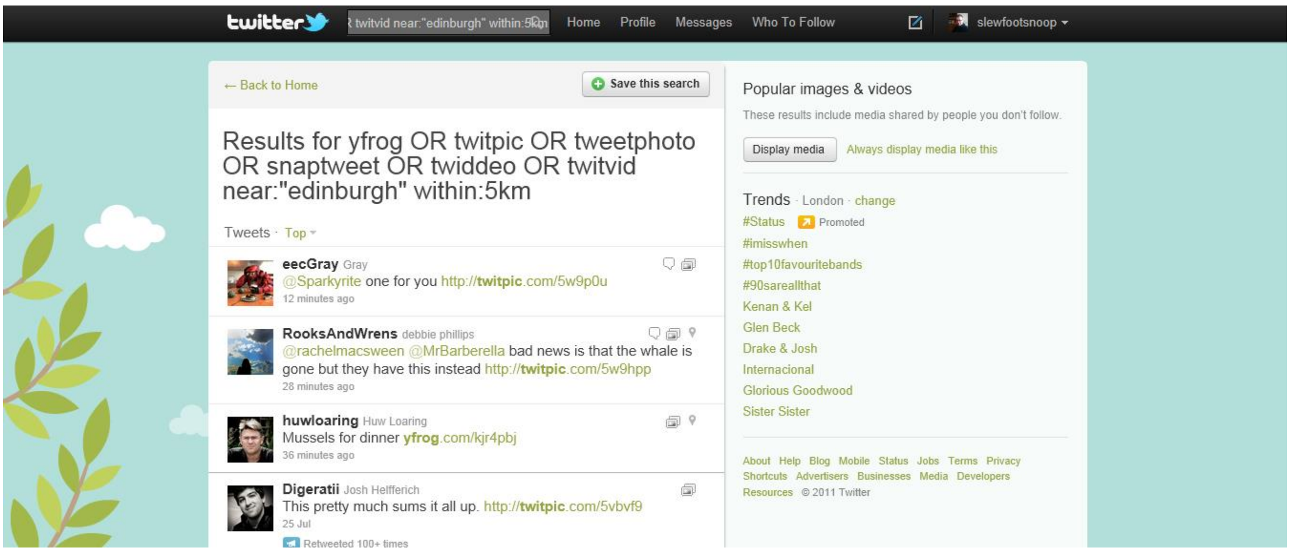


Creepy's interface: this tool was developed by Infosecurity enthusiast Yiannis Kakavas

To filter results so that you see only those linking to pictures, videos and other media shared on popular networks, try the following search:

> *yfrog OR twitpic OR tweetphoto OR snaptweet OR twiddeo OR twitvid near:edinburgh within:10k*

This will return tweets posted within 10 kilometres of Edinburgh which contain links to some of the more popular media-sharing twitter services. This can be particularly useful for seeking out images from near a breaking news event, though it will inevitably include irrelevant images if the place in question is a big city. If you find this approach returns results from other Edinburghs, try searching in Trendsmap (http://trendsmap.com/), which disambiguates place-names by country.



Advanced Twitter search for multimedia posted by Tweepz in Edinburgh

Searching older tweets is a serious problem in Twitter – Twitter's own search only returns roughly the last 10 days' Tweets (Perez, 2009). US searchers may be able to access Twitter's archives via an Updates option in search results (which draws selectively from the Library of Congress archive of all Tweets). As of late 2009, many Tweets have been indexed on Google (Bunz, 2009). so Google's domain function (covered earlier in this book) can be instructive here. Try the following search:

budget site:http://twitter.com//status*

This should find archived tweets mentioning the budget from any user (note the use of the wildcard for the folder preceding */status*).

Those who syndicate their Tweets via Friendfeed have a far deeper searchable archive. However, is worth bearing in mind that even those services with access to the Twitter 'fire hose' (direct API) do not have anything even approaching a complete Twitter archive, so it is important to try different services. In the absence of the now defunct Google Realtime search, which took advantage of the 'firehose', the most comprehensive archive of Tweets can be accessed via Topsy (http://topsy.com/). Some other tools which are useful for searching Twitter's archive include:

Searchtastic (http://www.searchtastic.com/)
The Archivist (archive your own and selected tweets) (http://archivist.visitmix.com)
Tweetboard (Alpha invite at time of writing) (http://tweetboard.com/alpha/)
Tweetscan (backup option may help) (https://www.tweetscan.com/index.php)

The problem of archiving in Twitter may be allayed more permanently (and thoroughly) once the Library of Congress, who have acquired all Twitter data, make it publicly available (albeit some archive material is available to access for businesses,via marketing company Datasift). There are more tools for digging out archived tweets which can be found later in this chapter, under *Real-time search*.

Twitter works especially well as a newsgathering source when you have a large pool of people to follow, and in turn, a large number of followers whose tweets can inform your newsgathering. But how do you acquire a large, and meaningful following? At a very basic level you can go through the process of finding all your email contacts on sign-up to Twitter. You can also perform a basic search for names you know, then once you've found them, go through their followers (and their Twitter lists) to see who else may be worth following. Browsing through hashtags, trending topics, individual posts (and their associated profiles, and lists) is a good (if laborious) way to dig out potential contributors around certain topics. But there are a number of dedicated directories which are worth bearing in mind when it comes to expertise or experience.

Follower Wonk (http://followerwonk.com/) provides a search twitter profiles, which can be essential if you are searching for people by profession, location or interests; and especially a combination of these. The results show you how many updates each person has made – which can show you how proactive a user they are, which in turn may suggest not only how well-connected they are in their field online, but also how likely they will be to get back to you within a short deadline, via Twitter.

Users of Twitter might assume that url-shortening tools are all about distributing content, but they can have a research and contributor-finding value too. The leading url shortening service, bit.ly (https://bitly.com/) provides analysis which lets you track the flow of your links shared via the service, not to mention clicks on those links over time (see the Manage tab

- login to Twitter account required). Searching for urls using Twitter Advanced search is not reliable, and won't include shortened links. But it can be a useful way to find people motivated by certain issues. This can in turn feed into developing your own Twitter following, and your own pool of contacts.

Conversations between individual accounts can be hard to track on Twitter, but can be the source of intriguing information; Aaron Swartz has created a tool which can help with this (http://twitter.theinfo.org/). You can obtain a Tweet's individual ID by clicking on its time stamp. There are many other Twitter-related research tools, which will be introduced later in this chapter. For help on using Twitter, take a look at their Help Center, (http://support.twitter.com/) and Get Satisfaction (http://getsatisfaction.com/twitter) forums.

Other social networks

Google+ (https://plus.google.com/): with more than 90 million users as of January 2012 (Wasserman, 2012) this platform has been feted as a long-term challenger to both Facebook and Twitter. This platform offers real potential in newsgathering, not least from the 'hang out' option, a mass-video conferencing facility, which can be used to group-chat with news sources. When launched, its terms of service insisted upon real-name disclosure, a policy which may serve the purposes of transparent, objective news coverage, but which militates against whistleblowing. It has been stated though, that this policy is due for overhaul (Galperin and York, 2011) At the time of writing, there is a major US bias in users and in content, and a lack of any real activity around anything other than the most popular UK news topics. Another late arrival to social networking, Pinterest (http://pinterest.com/) is a life-streaming platform which boasts a significant, and overwealmingly female userbase (Constine, 2012). Again however, it lacks a UK-focus at the time of writing.

FriendFeed (http://friendfeed.com/) has far fewer users than Twitter (and indeed some Twitter users duplicate their feeds via FriendFeed, and visa versa). Its advanced search (http://friendfeed.com/search/advanced) offers a range of options to keep up on issues. Tumblr search (http://www.tumblr.com/search) can be inconsistent, and there is no formalised profile element to search within in this platform.

Myspace (http://www.myspace.com/) the UK's fifth most popular social network, has a significant under 17-year-old userbase - 33% according to the latest research (Royal Pingdom, 2010). That said, many older people use Myspace too. Myspace doesn't have an advanced search, but it does offer various means of filtering your search results, whether by channel (People, Music, Videos and Images) or by demographic factors (such as gender, or age). MySpace offers groups, and as with Facebook, Myspace has public forums which can be searched in Google using the domain function:

london site:forums.myspace.com

Hi5 (http://hi5.com/) and Netlog (http://en.netlog.com/) have significant UK users, though far fewer than Facebook and Twitter.

If you are looking for professional communities then, aside from Facebook's networks, LinkedIn (http://uk.linkedin.com/) is probably the best resource. LinkedIn is a social networking site aimed at the professional community – for those seeking contacts and jobs in their industries. Although there’s a heavy US presence here, LinkedIn contains a sizeable range of UK-based professionals. Registration is not necessary to search, albeit it will allow you to make contact, and seek out introductions to people. If you want to send mails, request more than five introductions at a time or see expanded results, you will need to buy a Pro account (there are different levels of access, at different tariffs). LinkedIn can be especially useful for seeking out people by employer – present or past. This can be a good means of approaching a sensitive issue, or an issue where employees are being discouraged from talking to the press.

Example: follow the money (even when it has disappeared)...

While training in a London newsroom on the day Lehman Brothers filed for bankruptcy, I was told that journalists prowling the bars around Canary Wharf were having little success in sourcing information (due to the legal agreements employees had signed). But a search of LinkedIn highlighted the names of those who used to work at the stricken investment bank, which proved to be an alternative, indirect means of accessing opinion and experience of events.

LinkedIn is best used when a work-related research problem arises – people will not necessarily treat their LinkedIn account in the same social way as they would with a Facebook account – they may be well-established in their field, and don't need to share this information with the wider world. As an alternative to LinkedIn, there is Xing (http://www.xing.com/) The average user on this network is aged 44 (Royal Pingdom, 2010).

If you need to find older contributors, Sagazone (http://www.sagazone.co.uk/) is a good alternative. Likewise, if you are looking for people according to the school they once attended, Friends Reunited (http://www.friendsreunited.com/) and My Old Mate (http://www.myoldmate.net/) often offer a more nuanced means of finding people than do Facebook's school pages.

Perhaps the best place to start looking for specialist groups online is amongst those services which allow online communities to create their own social networks. Try searching Ning (http://www.ning.com/) or SocialGo (http://www.socialgo.com). Google Profiles search (http://www.google.com/profiles) should not be discounted; they may not include contact details, but can give an insight into potential contributors.

If you've set up accounts for some of these sources, you may want to be able to follow them all from one source – there are a few services which can help here, such as CollectedIn (http://collectedin.com/) and AOL Lifestream (http://lifestream.aol.com/) both of which allow you to log in to each of your accounts, and keep on top of them all from the same page.

Mobile social networks

The rise of the smart phone has made possible some very intuitive and innovative ways to find people, and engage with communities. Mobile social networks are often considered in terms of game-play; where users are rewarded for performing a task (something which lends itself well to commercial and PR professionals). But in marrying social networking and geo-location, mobile social networks can provide an alternative means of finding sources; establishing impromptu meetings, or even tracking the development of a story as

it is unfolding in real-time, geographically.
UK data journalist Nicola Hughes has offered insight into the innovative use of the geolocation function available via these social media platforms. When using Foursquare in conjunction with Google Maps, Nicola was able to find, and then verify the reliability of a potential witness to a volcanic eruption in Indonesia, despite the scale of 'background noise' created in trending topics from this region, at the time (Hughes, 2011).

There are several options out there; the most popular being:

BrightKite (http://brightkite.com/)
Foursquare (https://foursquare.com/)
Gowalla (http://gowalla.com/)
Plazes (http://plazes.com/)

Online forums

Google's *Discussions* vertical contains Google Groups (https://groups.google.com/?hl=en), which in turn incorporates Usenet newsgroups dating back as far as 1981. This is arguably the most comprehensive forum to seek out contributors, and to keep up to date on your beat or topic area on the web. A good way into Google groups is to search for general, topical keywords within the 'search for a group' search bar. The menu of results allows you to filter by region, topic, post frequency, volume of members, language (which differentiates between US and UK English) and the number of days since last post. The last filter is perhaps the most useful in news and current affairs contributor-finding and newsgathering, as those groups which have had little activity over the past month may be less likely to yield a response to short deadlines. Groups can be generalist or specialist, serious or trivial – there are many groups comprised exclusively of academics, professionals, and experts of all kinds, and indeed some of the most popular groups require (free) membership.

Given the depth of archive, it is always a good idea to sort results by date rather than relevance – people may have given up their accounts in the 10 years since they posted the messages you've just found. The advanced search in Google Groups (http://groups.google.com/advanced_search?q=&safe=on&) provides one of the few means of online search by email domain. A search for @bbc.co.uk for example, returns many messages posted by BBC journalists and researchers seeking out contributors. Be wary though – Newsgroups are closed communities, often organised around issues and hobbies taken seriously. It is unwise to barge into online forums assuming everyone there will be awed by your profession, and assuming they will respond in kind. Treat newsgroup contributors with tact and civility – don't belittle hobbies, or come across as taking but not giving back, if you want to make the most of these forums.

Yahoo Groups (http://groups.yahoo.com/) is a worthy alternative, with conversational content reaching back to 1988. This is a less utilitarian tool for day-to-day work – their advanced search (http://groups.yahoo.com/group/flyfishermensretreat/msearch_adv) is limiting. Whether public or private, access to these groups often requires sign-up, but some of the bigger groups are updated very frequently – it's a good idea to include a location in your searching to avoid the US-bias in membership.

Omgili (http://omgili.com/) is a crawler-based service which can be thought of as a metasearch engine for forums dotted around the web – use this tool to find general chatter across forums, discussion groups and mailing lists. It can be a great source of unofficial information, but given the often anonymous nature of forums, it isn't the most reliable of sources . There is an advanced search option on results pages, but the sliding scale options for time scale, numbers of replies and numbers of commenters, give more than enough scope to filter out passing references and irrelevant pages. At the time of writing, they offer two RSS options; a (paid) premium option which provides near-realtime content, sorted-by-date, and a free one which doesn't.

Boardtracker (http://www.boardtracker.com/) is another forum meta-search engine, but one that draws its results from a different range of sources to those above. Its results include a helpful pie chart showing proportionally where your results have come from, around the globe. Slightly less helpfully, there is no option to filter by country of origin. However, Boardtracker does offer free RSS for search results, and an alerts service if you are happy to divulge your email address.

Such general ways of scouring forums are helpful in conventional news sourcing, but many journalists will have her own particular spread of interests and specialisms – so it is always a good idea to find and then routinely access forums in your beat.

For example, if you write on education, you will want to follow Teacherstalk (http://www.teacherstalk.co.uk/) If you are a home affairs correspondent, you will want to keep an eye on UKpoliceonline (http://www.teacherstalk.co.uk/). If you are a generalist reporter and a parenting issue arises one morning, Mumsnet (http://www.mumsnet.com/) will be worth scouring for potential contributors. There may be many other forums and blogs which will be relevant to you, but which may not be easy to find. One good way to find more relevant sources is to take advantage of Google's *related:* function. For every relevant forum you can find, perform a *related:* search on the domain to eke out more relevant sources, like this:

related:ukpoliceonline.co.uk

Google Blog search (http://blogsearch.google.com/) is also a useful place to dig out forum discussions – include the term *forum* in these searches.

Blogs

As with most online sources, it is best to build up and maintain a list of individuals whose interests (or area) is relevant to your beat. Additionally, it is important to track references to your beat from across the blogosphere.

Several blog search verticals are available, though some in the search community have long decried a general malaise across this field (Smarty, 2008). Technorati (http://technorati.com/) is perhaps one of the longest established means of searching the blogosphere. Their advanced search (http://technorati.com/search?advanced) offers some useful options; such as searching by category, and filtering by blog authority (Technorati's own measurement). In reality though (and at the time of writing), minimal UK content is available via this engine.

Google blog search (http://blogsearch.google.co.uk/) is a high profile alternative. Though it has far more UK content available, the advanced search is in some ways, less utilitarian than Technorati's (especially with regard to ranking results by *authority*). It can be flaky –

searching within the UK index won't prevent sites referring to non-UK place-names being returned, so it may be necessary to include negative references to related place-names from outside the UK. In addition, the index is more generally fully of forums and listings sites (especially property and cars) which can crowd out local or topical news sources. To remove forums posts and listings, it is best to include the term 'blog' in your search – somewhat counter-intuitive given that this is a blog vertical. There is no way to filter out non-UK content and references, so the best hope is to include the term UK in searches. A search for:

cambridge UK -MA

...will filter out most references to Cambridge Massachusetts (often abbreviated to *MA*), though perhaps at the expense of material which is not tagged with the term 'UK'.

Unfortunately there are no 'similar' options in Google Blog search results pages. When you find an appropriate blog, and want to find others like it, Google blog search is probably not the best answer. It is probably easiest to browse down the related links options on the blog in question, but for a more robust approach, try searching for the blog in Open Site Explorer (account required) to unearth other blogs linking to it, which may be related.

Google's main index also has potential in blog search. For subject or beat-specific blogs, run a search including the term blog (i.e. *middlesbrough blog*), then refine using the Pages from the UK filter from the search options panel on the left of the results page. (you may need to click on the *More search options* link to drop this option down). For more blogs like the one you have discovered, seek out the *similar* option under the search result.

Twingly (http://www.twingly.com/search) provide another alternative blog (and social media) search. Their search results pages offer a spam-free filter under the *Any Source* option at the top of results pages, but this will not help you deal with the avalanche of listings and jobs pages which are returned – neither will incorporating the term *blog* into your search.

Of all the blog search verticals available, perhaps Blog Pulse (http://blogpulse.com/) offers the most. In this engine, an advanced search (see option on top right of screen) for:

cambridge NOT MA

...returns page after page of UK-specific, fresh blog results, and an RSS feed for every search too. Their conversation tracker (http://blogpulse.com/conversation) vertical also permits the tracing of debate around your keywords across several popular online news sources and blogs.

Provenance is an ongoing issue in blog search, which from time to time inspires authors to provide tips and source which may help (Schofield, 2008). Finding new blogs local to your interests (especially in the UK) can be difficult and time-consuming. It doesn't help that most blogging platforms go with a .com upper-level domain (rather than .co.uk), nor that some blogs don't include detailed nor standardised biographical information.

Social bookmarking

Most web users are used to bookmarking or saving useful sites and sources once they've found them – it is a crucial part of keeping a grip on research, of making sense of the mass of online content available at our finger tips – and it is also a process which connects all stages of newsgathering into an interconnected chain; from finding sources, to publishing them. Saving these resources in the cloud online (making them accessible anywhere, at any time), and sharing them in a social context (where other users can join together in knowledge networks) can help connect journalist, story and audience by exploiting what Halavais calls 'social remembering' (Halavais, 2009, p167).

The most popular site which allows this kind of knowledge sharing is Delicious (http://www.delicious.com/) (formerly del.icio.us). This social bookmarking site, which allows its users to create its structure, represents the most efficient way to get a grip of the content you've found online. Because it is platform independent, it allows you to access your saved (or favourite) websites from anywhere – which is useful where you are hot-desking, using one or more home computer, or using a mobile in addition to desktop computers, without the hassle of uploading your bookmarks to all the machines you use in the routine of your work.

It is a good idea to use bookmarklets (http://www.delicious.com/help/bookmarklets) to help speed up the process of saving, describing and tagging bookmarks. If you are concerned about privacy, you can mark your saves 'private' using a tickbox on the 'save a bookmark' page. Being a social network, you can also share bookmarks with colleagues and friends using the *network* option. The site can help you both audit a topic or beat, and keep on top of it using RSS.

But you don't need a delicious account to make good use of this source. It is also a great place for newsgathering, and for finding valued sources and material online, because it harnesses the power of social discovery. You can surf those sites people actually value; you can measure this value by the volume of people who have saved a particular site, and you can isolate every individual who has saved this site, establish what you have in common, and effectively browse their saves and follow future saves (or particular tags, or search terms) to keep on top of discoveries in your field as time goes on.

You can either search all of delicious using the search box, focus on particular tags *Explore a tag* (http://www.delicious.com/tag/), or even find a particular web address using the 'Look up a url' (http://www.delicious.com/url/) option. Search results indicate the number of people who have saved each link, and the tags used to describe them. Click on the number of users, and you can access the full list – however, Delicious user profiles are often less populated than profiles in other sites, and usernames rarely give away who the saver in question is. RSS feeds are available for all types of search results.

Although a timeline is provided in search results, there is no way to sort results by the date they were saved – the default format is by relevance (based on keywords found in titles, notes, and tags. There are many add-ons, plugins, extensions and other Delicious-based tools available, some of which offer benefits in search. But Delicious is no cure-all for search – people don't always tag content in a useful or consistent way (if at all), and indeed some have questioned the value of links people save, arguing that an arms-race informs behaviour, with people pursuing personal popularity and 'authority' online. Since it was sold by Yahoo in April 2011, some users have migrated elsewhere. But for now, it still contains a significant reach of curated material.

Delicious isn't the only social bookmarking tool available for use in online research. Since March 2010, Google Bookmarks have been experimenting with public lists (https://www.google.com/bookmarks/). Though less popular than delicious, Google Bookmarks (which requires a Google account to access) does index the entire page of your bookmarks; which gives more control and more options in searching those bookmarks which have been shared publicly, while lessening the impact of inaccurate or inconsistent tagging.

Blinklist (http://blinklist.com/) offer an alternative search option, albeit one with relatively little UK content, and lacking much of the functionality found in delicious. Faves (http://faves.com/SampleProfile.aspx) contains a good deal of content, but doesn't have a big user baser. Likewise, social annotation tools like Diigo (http://www.diigo.com/) (which incorporates FURL - account required) may be useful alternatives. CiteULike (http://www.citeulike.org/home) and Connotea (http://www.connotea.org/) offer an academic take on social bookmarking, and can both be useful accompaniments to Google Scholar for digging out expertise, or esoteric research.

Realtime news aggregators

Traditional search engines are good for finding structured, fixed, archived information, but they have tended to have a blind-spot when it comes to indexing and presenting breaking news. Shrewd independent developers spotted a potentially lucrative niche in the market, and so the major search providers have had to play catch up.

Real-time aggregators are a particularly useful place to watch news as it breaks – though there is some disagreement about a working definition of the term. For the purposes of news-gathering (and the tools introduced here) it is useful to think of real-time news as online media, whether status update, tweet, breaking news story, or even links to recently published material, which is published, and indexed (and so available to view on appropriate platforms) near-instantly.

Social Mention (http://socialmention.com/) is one such platform. Styling itself as a ' social media search engine' more than a real-time search tool, its primary focus is on many of those sources which fuel some of the other engines mentioned here. It is perhaps one of the most far-reaching of all the real-time engines available – indeed it goes far beyond mere breaking news to offer a depth of perspective via many social sites. Though it provides a smattering of trending topics on the homepage, the real value in this service lies in the various means provided to filter and isolate information needs. Once results are returned, it is possible to filter by adding more (commonly found) keywords, or by isolating the usernames of those who have contributed most using your keywords. Social Mention draws from a range of multimedia sources, listed conveniently on the left-hand side navigation in search results. There is real depth to this service, and an archive which few competitors can equal.

IceRocket's (http://www.icerocket.com/) Big Buzz vertical (selected from the top-right of the search bar) provides timely results organised around different social media strands; blogs, Twitter, Facebook status updates, video, news and images. Twitter results show how many followers a poster has, though Facebook updates do not. All text-based results tell you how recently they were posted, and it is worth noting that news results do not discriminate between local, national, international, old- and new-media sources. A down-side to using this tool is that it is very search-dependent – the *Buzzing searches* on IceRocket's home page are fairly inappropriate for hard news journalism.

There are several means of keeping on top of Twitter in real-time. Twingly live (http://live.twingly.com/) offers a means of setting up rolling real-time search results, incorporating as many (or as few) keywords as is appropriate to your subject area. Results offer a preview of links so you can detect spam, which, at the time of writing is a pervasive problem in real-time search.

Twittorati (http://twittorati.com/) brings the worlds of blogging and microblogging together – allowing users to establish followers on Twitter by how highly ranked their respective blogs are. Though it should be noted, those 'bloggers' who comprise the top 100 (as presented on the left-hand column of Twittorati's home page) represent a who's who of the usual suspects in news media and technology. At the time of writing all the top-ranked political blogs are US-based, leaving little scope for those concerned with UK politics.

Topsy (http://topsy.com/) provide yet another alternative. Their experts directory (http://topsy.com/experts) offers a means of seeking out experts by mentions across Twitter, and its depth of archive (dating back to May 2008, at the time of writing) makes it an especially useful tool. Tweetmeme (http://tweetmeme.com/) offers a range of channels, one of which lets you browse through the top news linked stories by most recent, Top in 24 hours and Top in 7 days as shared across Twitter, though technology stories (and by all accounts self-promotion) tend to crowd out results. Twitturly (http://twitturly.com/) lets you track those links which have prompted so much debate across Twitter, with helpful verticals for news, images and video.

To track big topics and your own bespoke interests in realtime, Twitterfall (http://www.twitterfall.com/) provides an elegant and robust solution – so much so, it has been used in industry (Oliver, 2009). Follow general trends, or set up your own rolling feeds via Twitter lists; filter by search, by geolocation or by search logic to construct your own news platform. Both Twitter advanced search (http://search.twitter.com/advanced) and Geochirp (http://www.geochirp.com/), allow you to limit your search to within 1KM and 1 mile of a postcode or place, respectively (the former even supports search by fractions of of a mile). As some of these sources develop, and provide a geographically determined means of presenting news by popularity, their utility will surely improve for UK researchers and journalists.

Some other real-time search options include: WhosTalkin (http://whostalkin.com/) and Kurrently (http://www.kurrently.com/).

Finding social search trends

Trends in search have been used in real-time to track the spread of disease (http://www.google.org/flutrends/), and historic search trends have even been explored by economists as a means of predicting future economic activity (McLaren, 2011). The use of web trends in newsgathering can be controversial. Some warn that using search trend data as a means of deciding which stories to cover, or chasing traffic associated with some popular trends at the expense of traditional newsgathering methods, can put pressures on journalistic standards, and risks freezing out genuine, public interest journalism as a result (Currah, 2009). On the other hand, it may fairly be argued that web trends represent the purist form of demand online – and after all, a journalists' role is to give their readers the content they want to read.

By keeping on top of web trends in the newsroom it is possible to discern what the public are interested in. But the following resources for establishing what is trending can also be used pro-actively in the newsroom; to inspire original stories, to help journalists make the most of their coverage of a popular issue, and to get a sense for how your content matches with global (or local) search demand, out-with the confines of your own in-house analytics tools.

Google Trends UK (http://www.google.co.uk/trends) offers a modest top 10 'Hot Topics' at any given time, which you can compare and contrast with US hot searches (for which, at the time of writing, there are some twenty trending terms listed. You can also search for trends, but these are not UK-specific.

Google Insights (http://www.google.com/insights/search/?hl=en-GB#) takes the Google Trends concept further, providing various options for depth analysis of internet memes. You can isolate particular verticals (such as News, or Images), and regions – a heatmap shows the popularity of terms geographically. Search results include those terms most associated with your terms; both 'top' and 'rising' trends. Only a portion of web traffic across Google's domains is used, which is presented relative to the total number of searches undertaken in Google over time. In terms of analysing your results, it is important to bear in mind that a declining line doesn't mean absolute traffic for your term is decreasing – just that the overall popularity of this term is decreasing. Google Insights can be a source for news stories, nationally (Warman, 2010) and internationally (Morgan, 2010). But perhaps more conventionally, tools like this can be used to keep tabs on what the surfing public is interested in.

Yahoo!'s Buzz index (http://buzzlog.yahoo.com/buzzlog) represents an interesting (albeit altogether less interactive) alternative to Google's efforts. Published Tuesday to Saturday, this edited list of stories is compiled from trends in Yahoo! Search log files is updated daily. At the time of writing there is no UK Buzzlog.

For tracking trends in Twitter there are a broad array of different tools, many freely available. While Twitter's web interface offers personalisable trends (by location), there are many other sources which offer a different approach, more granularity to your coverage, and more options. Once you've logged into your Twitter account via Twitscoop (http://www.twitscoop.com/) for example, you can watch trends come and go using their *Buzzing right now* tag cloud, or by using their interactive time line.

If you don't know why something is trending and don't have time to search around, What the Trend's UK index can be instructive:

http://whatthetrend.com/?woeid=23424975&place_name=United%20Kingdom

Trendsmap (http://trendsmap.com/) offers a visual means of tracking trends geographically. You can follow trends around your particular location, or a city, and additionally, for those not interested in big conurbations but in the spread of trends across a region, the interactive map offers a *My Region* option. This resource is fully interactive, allowing for zoom to varying degrees of granularity, with trends superimposed over the map.

Trendistic (now defunct) used to offer insights into historic Twitter trends. This tool was used to challenge the false assumption that criticism of journalist Jan Moir's controversial Daily Mail opinion piece on the death of Steven Gately was drummed up by celebrity tweeter Stephen Fry (Bradshaw, 2009).

As one of the most popular sites on the net, Wikipedia is a key source for tracking online 'buzz'. Trends in Wikipedia activity can be monitored using a number of different sources. Trending Topics (http://www.trendingtopics.org/) provides a list of the most commonly viewed Wikipedia content over the past 30 days, along with a Rising (24 hours) option. Content can be viewed according to a number of types, including People and Finance – and near-real-time monitoring is available via the current hour option (http://www.trendingtopics.org/hourly_trends).

Uptrends for topic pages (and associated popular pages) benchmarked over the past 24 hours can be viewed via Wikitrends (http://toolserver.org/~johang/wikitrends/english-uptrends-today.html). Wikirage (http://www.wikirage.com/) provides an overview of the most edited Wikipedia pages – over the past hour, six hours, day, three days, week and month. As for online video, the best way to keep on top of those most viewed, and most popular videos in Youtube, is via their Video channel (http://www.youtube.com/videos?s=mp); here you can browse the past day's, week's or month's content. Several categories are available, including News (http://www.youtube.com/news).

In terms of images, Google/Flickr Zeitgeist (http://fiveprime.org/zeitgeist.html) offers hourly updates of trending search terms, and the images they yield. For music, Next Big Sound (http://www.nextbigsound.com/) lets you see how popular musicians are proving, across those musical areas of the social web (such as Last.fm and Myspace) as well as Facebook, Twitter and Wikipedia. For consumer goods, Amazon's Movers and Shakers (http://www.amazon.com/gp/movers-and-shakers/) shows how popular different products are (including books, music and technology) – which can be a useful source for consumer journalism, and features.

For a very general overview of web-based trends, Trendsbuzz (http://trendsbuzz.com/) is a helpful resource, while SEOMoz popular searches (http://www.seomoz.org/popular-searches) provides a similar overview service across many different sources.

References

Bradshaw, Paul (2009) Notes on #janmoir – don’t ‘blame’ Fry, Online Journalism Blog, October 17:
http://onlinejournalismblog.com/2009/10/17/notes-on-janmoir-dont-blame-fry/

Bunz, Mercedes (2009) With its real-time search, Google is creating an archive of the present, The Guardian, December 08:
http://www.guardian.co.uk/technology/pda/2009/dec/08/real-time-search-google

Constine, Josh (2012) Where The Ladies At? Pinterest. 2 Million Daily Facebook Users, 97% Of Fans Are Women, TechCrunch, February 11:
http://techcrunch.com/2012/02/11/pinterest-stats/

Currah, James (2009) What's Happening to our News, Reuters Institute for the Study of Journalism: Oxford.

Forrestal, Valier (2011) Twitter RSS feed creation cheat sheet, Ridiculously Digitally Ubiquitous, June 02:

http://theinfobabe.blogspot.com/2011/06/twitter-rss-feed-creation-cheat-sheet.html

Galperin, Eva and York, Jillian (2011) Victory! Google Surrenders in the Nymwars, Electronic Frontiers Foundation, October 19: https://www.eff.org/deeplinks/2011/10/victory-google-surrenders-nymwars

Goodman, Eli (2009) What History Tells us About Facebook's Potential as a Search Engine, Part 1, April 26: http://searchenginewatch.com/article/2065524/What-History-Tells-us-About-Facebooks-Potential-as-a-Search-Engine-Part-1

Goad, Robin (2010, I) Social networks now more popular than search engines in the UK, June 08: http://weblogs.hitwise.com/robin-goad/2010/06/social_networks_overtake_search_engines.html

Goad, Robin (2010, II) Social networks now more popular than search engines in the UK, Hitwise, June 08: http://weblogs.hitwise.com/robin-goad/2010/06/social_networks_overtake_search_engines.html

Halavais, Alexander (2009) Search Engine Society (Digital Media and Society Series). Polity Press: London.

Halfacree, Gareth (2010) Google's Buzz causes privacy concerns, Bit-Tech, February 11: http://www.bit-tech.net/news/bits/2010/02/11/googles-buzz-causes-privacy-concerns/1

Hughes, Nicola (2011) Sorting the Social Media Chaos, Data Miner UK blog, May 30th: http://datamineruk.wordpress.com/2011/05/30/sorting-the-social-media-chaos/

Kollewe, Julia (2009) Twitter is not for teens, Morgan Stanley told by 15-year-old expert, The Guardian, July 13: http://www.guardian.co.uk/business/2009/jul/13/twitter-teenage-media-habits

McLaren, Nick (2011) Using Internet Search Data as economic indicators, Bank of England Quarterly Bulletin, Q2: http://www.bankofengland.co.uk/publications/quarterlybulletin/qb110206.pdf

Morgan, Kelli (2010) No. 1 Nation in Sexy Web Searches? Call it Pornistan, Fox News, July 13: http://www.foxnews.com/world/2010/07/12/data-shows-pakistan-googling-pornographic-material/#ixzz1PpQap5Aq

Myslewski, Rik (2010) Facebook tops half-billion users, wants your innermost thoughts, The Register, July 21: http://www.theregister.co.uk/2010/07/21/facebook_tops_billion_active_users/

Nielsen Wire (2009) The More Affluent and More Urban are More Likely to use Social Networks, September 25: http://blog.nielsen.com/nielsenwire/online_mobile/the-more-affluent-and-more-urban-are-more-likely-to-use-social-networks/

Perez, Sarah (2009) Ten ways to archive your tweets, Read Write Web, August 11:

http://www.readwriteweb.com/archives/10_ways_to_archive_your_tweets.php

Royal Pingdom (2009) Study: Males vs. females in social networks, Royal Pingdom, November 27:
http://royal.pingdom.com/2009/11/27/study-males-vs-females-in-social-networks/

Royal Pingdom (2010) Study: Ages of social network users, Royal Pingdom, February 16:
http://royal.pingdom.com/2010/02/16/study-ages-of-social-network-users/

Schofield, Jack (2008) Searching blogs by location, The Guardian, August 28:
http://www.guardian.co.uk/technology/askjack/2008/aug/28/searchingblogsbylocation

Schroeder, Stan (2009) Facebook Users Are Getting Older. Much Older, Mashable, July 07:
http://mashable.com/2009/07/07/facebook-users-older/

Smarty, Ann (2008) Blog Search Engines : The Complete Overview, Search Engine Journal, September 28:
http://www.searchenginejournal.com/blog-search-engines-the-complete-overview/7856/

Social Media Optimzation (2008) Social Network User Demographics, May 29:
http://social-media-optimization.com/2008/05/social-network-user-demographics/

Sparrow, Andrew (2009) Twitter users: young, metropolitan and angry about civil liberties, The Guardian, November 18:
http://www.guardian.co.uk/technology/2009/nov/18/twitter-users-survey-poll-britain

Su, Susan (2010 Who's Using Facebook Around the World? The Demographics of Facebook's Top 15 Country Markets, Inside Facebook, June 08:
http://www.insidefacebook.com/2010/06/08/whos-using-facebook-around-the-world-the-demographics-of-facebooks-top-15-country-markets/

Van Grove (2010) Bing Adds Facebook Updates and Links to Search Results, Mashable, June 09: http://mashable.com/2010/06/09/bing-adds-facebook/

Warman, Matthew (2010) General Election 2010: the Google trends during the leaders' debate, The Telegraph, April 16:
http://www.telegraph.co.uk/news/election-2010/7597345/General-Election-2010-the-Google-trends-during-the-leaders-debate.html

Wasserman, Todd (2012) Larry Page: Google+ Now Has 90 Million Users, Mashable, January 19:
http://mashable.com/2012/01/19/google-plus-90-million/

Chapter 6: Multimedia

Introduction

Few web users would struggle to name some common sources for finding video, images and audio online. But finding such material as a means of sourcing or substantiating news in a timely way, is an altogether more taxing matter.

Legal and ethical factors in the use of multimedia online

Time constraints are not the only pressure journalists experience in finding and selecting online multimedia - legal and ethical issues abound too. Those responsible for creating multimedia which appears online, retain rights over these items – so it behoves any journalist wishing to use such material to contact the author in order to request permission for use. If this is not possible, the journalist runs a risk of legal action – and while the law of fair use may be available to counter strictures in intellectual property rights in some situations (where applied to small portions of text for example), such use could nonetheless lead to legal action. In any event, fair use does not apply to all media – photography is exempt under UK definition of the concept (Banks and Hannah, 2009, p452). Where a third-party image is used to identify someone in the public eye, the public interest defence in using this image without permission may be stronger than where an image is merely decorative. But intellectual property rights are not the only legal barrier to using multimedia. Where a journalist does not factor in the running of active legal proceedings, the publication of multimedia evidence that may incriminate someone, or bias someone's legal position, can lead to action under Contempt laws.

The issue of ethics in handling social media explored in Chapter 4 applies equally to the use of multimedia found therein – as does the issue of taste and decency (whether in the form of industry regulation, or in-house media best practice, and guidance). In very simple terms, it is therefore necessary for journalists to avail themselves of best practice, regulatory regimes, and the criminal and civil laws of the domain in which they are working before setting out to use multimedia found online in their published journalism.

Here follows a range of sources to help find multi-media online.

Finding video

Today Youtube (http://www.youtube.com/) has grown in scale so much that it can be difficult to find content by means of conventional search. To help, the site offers filters, and ordering preferences when your search results are returned; it also supports some Boolean search, and some Google advanced operators too, including phrase searching, the wildcard keyword placeholder (*), and the *intitle:* command. Search aside, there are other methods which can help filter results. For example, it is possible to fast forward or rewind to a specific point in minutes and seconds within a YouTube video by appending some code to the end of the url: *#t=05m30s* (where #t=XXmZZs with XX representing minutes and ZZ seconds). This can be useful for sharing content hidden in long videos.

YouTube's chart option (http://www.youtube.com/charts) allows you to see what are the most popular, trending videos at any time. This can be fairly limited, not least due to the inconsistent (and sometime incorrect) way some people categorise their videos, but also due to the heavy US-bias of content sourced this way. For these reasons, it is perhaps best to refer to the 'most' options, rather than subject-specific content. Youtube's categories page (http://www.youtube.com/videos) offers (at the times of writing) the same content in an alternative layout.

While RSS feeds are available for categories in YouTube, it is a little-known fact that RSS feeds for keywords (or tags) are also available, if you are prepared to do some url-hacking:

http://gdata.youtube.com/feeds/videos/-/<keyword>

Simply copy the above url into a browser, and replace *<keyword>* with your tag of choice. If you would like to set up a feed for more than one keyword, put a forward slash after the first, and then type out the second.

Downloading YouTube videos is a precarious business in terms of copyright, and for the purposes of newsgathering (or sharing information) it is not necessary. But it is possible via a range of third party operators, and some browsers. In addition to YouTube, Google Videos (http://video.google.co.uk/) aggregates content from several online video providers in one place. It shares the same search syntax as YouTube, and RSS feeds can be obtained by scrolling to the bottom of results, selecting '*Create an email alert for...*', then selecting the 'feed' option in the *Deliver* to menu.

Blinkx (http://www.blinkx.com/) aggregates less content than Truveo (concentrating mostly on ITN and Press Association content), and as such suffers in terms of up-to-dateness. Nevertheless, it does offer a news category which provides the day's news in summary (http://www.blinkx.com/topics/news).

When it comes to searching the videos shared in social media, Viddler (http://www.viddler.com/) is a good starting place. As far as media shared across just Twitter is concerned, Twitvid (http://www.twitvid.com/) can be useful, albeit the search interface is minimal; you can't order results by time, let alone get an RSS feed from them.

Veoh (http://www.veoh.com/), a US-based aggregator, pulls together content from major film and TV production companies, independents and members of the public. The News section (http://www.veoh.com/list/videos/news) offers a means of tracking stories as they break, albeit only by language used (spoken, and in subtitles). There is a significant US bias, and few UK broadcasters have channels, which can equally be said for Daily Motion (http://www.dailymotion.com), despite its UK filter. Clipblast (http://clipblast.com) offer a more UK-friendly service, while Vimeo (http://www.vimeo.com) is more concerned with hosting amateur content (with, again, a significant US bias), Other social (and streaming) video services worthy of mention include:

Metacafe (http://www.metacafe.com/)
Videosurf (metasearch engine) (http://www.videosurf.com/)

In addition, there are many, many more video streaming sites (such as Hulu) which, for contractual reasons, require a US IP address to access.

A perfectly valid means of sourcing news, UK broadcasters have their own (limited) on-demand services online, which include:

BBC iPlayer http://www.bbc.co.uk/iplayer/
Sky Player (http://skyplayer.sky.com/vod/page/default/home.do)
4od (http://www.channel4.com/programmes/4od)
TVCatchup (http://www.tvcatchup.com/)

There are also several streaming services (aside from relative latecomers Youtube, Google+ and Facebook), including Livestream (http://www.livestream.com/) and Ustream (http://www.ustream.tv/).

Lastly, for sourcing older video archive, here are some useful professional (many of them subscription) sites:

Art Beats (http://www.artbeats.com/)
BBC Motion Gallery (http://www.bbcmotiongallery.com/)
Footage.net (http://www.footage.net/)
Free Stock Footage (http://www.freestockfootage.com/)
ITN Archive (http://www.itnsource.com/)
MovieTone (http://www.movietone.com/)

Finding images

Image search can inform several functions in journalism; from sparking ideas (creative thinking), to newsgathering, to (more conventionally) sourcing images to publish with written content. In terms of these first two categories of user need, there are many free alternatives from which to choose.

Google images (http://www.google.co.uk/imghp) remains the best mass-market image search engine available across the web. Its index far surpasses its nearest rivals, and it has become more utilitarian since it enabled reverse-image search, introduced in October 2011. This is a particularly useful tool for verifying images, and ensuring that images claimed to be original are not already in the public domain – as covered in a later chapter. Google's image results are more 'social' today than they were, by means of the *Google+* option. The assumptions driving Google's quest to bring ever increasing 'relevance' is tightening up their understanding of the intent behind search queries.

Exalead images (http://www.exalead.com/search/image/) works across a much smaller collection, of what are generally non-UK specific results (which are also less likely to come from social media accounts). All of Exalead's advanced operators can be used (at least in theory) in its image search. This platform offers an advanced operator which allows searchers looking for images of people to filter results by only bringing back 'face' shots (either select the option, or key in imagecontent:Face in addition to your search terms). However, unlike Google Images, Exalead will not auto-correct your search terms – so finding images of *Muhamed Quaddafi* (and the multiple ways in which his name was spelled) is not as straight forward as it could be.

Image and video hosting site Flickr has an emphasis much more on sharing content, than on professional photography (albeit there are many professionals and freelancers who publish their photography on the site). Flickr (along with del.icio.us) was an early pioneer of the 'folksonomy', encouraging the use of indexing based on bottom-up tagging by members, rather than top-down indexing by 'experts', or automated systems. As such it can be more intuitive to use (depending on what words and language you share with other users), but equally it can be rather hit-and-miss (with inconsistent, and incorrect indexing in some images, which won't necessarily be corrected by someone else in the Flickr community).

Flickr's advanced search (http://www.flickr.com/search/advanced/) is particularly useful in terms of sourcing fresh, original and easy-to-clear content. Besides offering search by update (which can be useful for news events), it also offers a check-box which can be used to limit results to images with a Creative Commons license, which are both easier to clear, and usually free to use (so long as credit is given) in news journalism – and so which are preferable to the time-intensive process of tracking a copyright-holder down, and negotiating permissions.

It should be noted that there are some limitations with Flickr's RSS options; despite claiming to offer RSS for photostreams, group & forum discussions, and specific tags, surfers may struggle to find these options for all content. Fortunately there are some third party alternatives, such as DeGreave's Flickr RSS Feed Generator (http://www.degraeve.com/), which can be used to set up feeds by date order (descending) for any Flickr tag. Fiveprime's Google/Flickr Zeitgeist (http://fiveprime.org/zeitgeist.html) is another hugely useful third party add-on for tapping into trends in pictures shared on Flickr.

Flickr has a number of competitors in the field of user-generated images; not least :

Instagram (http://instagram.com/)
Picasa (http://picasaweb.google.com)
Panoramio (http://www.panoramio.com/) and
Photobucket (http://photobucket.com/findstuff/)

In May 2011 Twitter launched its own photo-sharing services; but there remain several alternatives such as:

Twicsy (http://twicsy.com/) (offers search-based RSS)
Twitpic (http://twitpic.com/)
Yfrog (http://yfrog.com/)

There are also engines which allow viewing of images shared on Twitter in a real-time flow; such as Picfog (http://picfog.com/), Twitcaps (http://twitcaps.com/), and Pingwire for mobile (http://pingwire.com/) These search engines can be a good way to filter out non-visual tweets; and can be worked into real-time news-gathering system; but equally, they often contain material which is not safe for work.

If you plan to use images you've found online in your work, it is essential to make sure you have the permission of the rights-holder to use them. This is where Creative Commons content comes in useful. It is possible to filter advanced Google image results by usage rights, but there are also bespoke verticals which can help here, including:

Creative Commons image search (http://creativecommons.org/image/)
Everystockphoto (http://www.everystockphoto.com/index.php)

Clearing multimedia for use online can be an expensive process, especially if there is no Creative Commons license available. But if you have a working budget, there are a number of premium services which offer stock photography, which journalists may avail themselves of, including:

Corbis Images Royalty Free (http://www.corbisimages.com/stock-photo/royalty-free/)
Gettyimages royalty free (http://www.gettyimages.co.uk/CreativeImages/RoyaltyFree)
British Pathe Stills archive (http://www.britishpathe.com/)

For an exhaustive list of image sources see here:

http://randomknowledge.wordpress.com/2008/05/09/how-to-find-images-on-the-internet/

Finding music

Aside from iTunes (http://www.apple.com/itunes/) and Spotify (http://www.spotify.com/uk/), there are many other places to find and listen to music online, including:

Dogpile (http://www.dogpile.com/) (check the Audio filter)
Metacrawler (http://www.metacrawler.co.uk/) (check the Audio filter)
Mixcloud (http://www.mixcloud.com/) (especially good for playlists)
Soundcloud (http://soundcloud.com/)

In terms of recommendations, Amazon's listmania is particularly useful, albeit they can also be a little contrived. Blip.fm and Last.fm both offer more intuitive recommendations

There are also several sources online for finding audio and 'podcasts':

AudioBoo (http://audioboo.fm/)
Twaudio (http://twaud.io/)
Chirbit (www.chirbit.com/)

Using Google's *filetype:* extension (covered earlier in this book) can also be useful in this field, where you refine by audio filetypes. Using free-text search to find appropriate music for your package can throw up some interesting surprises; it can be helpful (and serendipitous) way of adding a touch of levity to your package. Google's domain filter (*site:*), when used to search across the domains of particular file sharing sites (like Rapidshare, Zshare, or Megaupload) can also yield useful results.

If you don’t have the money, time or patience to clear commercial music, there are no shortage of Production music (or 'mood music') companies online, where you can find suitable music to accompany your productions. Of course opinion is divided over the editorial value of this type of music (some believe it to be emotionally manipulative, and crass), however be aware that there are options out there:

Arcadia (http://www.arcadiamusic.com/) (account required)
Dewolfe (http://www.dewolfe.co.uk/)
Extreme (http://www.extrememusic.com/)
Sonoton (http://www.sonoton.com/)
Universal Publishing Production Music (http://www.unippm.com/)

Finding Sound effects is fairly straight-forward online, but clearance can be more complicated than music or TV archive. Many of these sites are amateur, and there is no guarantee that you won't stumble across material which has been lifted from others – so you may want to state your intentions with the Webmaster of the site you find them on:

Find Sounds (http://www.findsounds.com/)
Sounddogs (http://www.sounddogs.com/)

Alternatively, you can take a mini disk or MP3 recorder to the British Library Sound Archive, which is located at:

Sound Archive Information Service
The British Library
Sound Archive
96 Euston Road
London
NW1 2DB
United Kingdom
Tel: +44 (0)20 7412 7831
Fax: +44 (0)20 7412 7691

References

Banks, David, & Hanna, Mark (2009) McNae's Essential Law for Journalists. Oxford University Press.

Chapter 7: Bringing it all together: developing an online beat

The beat in a networked world

The concept of the 'beat' has a rich lexical history; an early etymological manifestation comes from the field of navigation, in the form of 'beating the bounds', meaning “to trace out boundaries in a perambulation, certain objects in the line of journey being formally struck, and sometimes also boys whipped to make them remember” (MacDonald, 1977, p113). Its application to journalism today has shed most of the masochism; it can best be summarised as “the routine path and set of locations that a reporter will visit each day, which brings them into contact with organisations that produce such news events” (Machin and Niblock, 2006, p72).

Fishman (1980), concerned with news-gathering in crime journalism, concluded that the news schedule (which demands a regular, and consistent flow of information, and stories) encourages journalistic dependence upon bureaucratic entities; councils, police forces, hospitals and other 'official' sources. The routinisation of news-gathering around these sources help journalists plan and prioritise the news (Harrison, 2006, p141), and bestow authority upon news output; taking the form of 'objective' sources of information (Tuchman, 1972), but at a cost (Foreman, 2010), for it leads to cynicism and disillusionment amongst audiences, as talking heads become ever more distanced from the communities they serve (Carey, 1999).

The ritualization of news-gathering and news sourcing is driven by a series of symbiotic and consensual processes wherein journalist and source both gain. Exploration of this sustenance of mutual dependence permeates much early literature in this field (Zelizer, 2004). McManus argues that our media thrive as an active arbiter within a market; bartering “access to the public to news sources in return for information needed to fill the paper or newscast.” (McManus, 1994 p5). But non-institutional organisations can take advantage of the diminution of editorial resources found in many modern newsrooms too (Davis, 2000); this is a market in which material wealth is not all that matters – social capital is a recognised currency too.

The beat is culturally relative; it does not exist in an hermetically sealed bubble. Commercial pressures shape it, and some argue, threatened its very existence. During the mid-1990s topic-teams emerged in US media, replacing beat reporters, and uncoupling the ties between local audiences and journalists – something many journalists of the era consider to be to journalism's detriment (Hansen et. Al, 1998). In the UK, the 1990s saw the rise of new technologies and marketing strategies, reinforcing the dominance of official voices in news. Industry consolidation (and resultant job cuts) impacted the beat. Fewer journalists began covering long-established geographical beats, work became consolidated around regional hubs and removed from local communities. The job cuts presaged a flight of journalists into public relations, who, in turn, began to feed those time-poor journalists who remain with easy-to-assemble news copy (Machin and Niblock, 2006, p70).

In the UK today, local beats are coming under pressure. Although local government data is available in unparalleled quantities, political decision-making (framed by the Local Government Act 2000), is more opaque than the committee system it replaced (Morrison, 2010). The rise of the council newspaper 'propaganda sheets', which are a competitive threat to local newspapers (Gilligan, 2009) may have abated, but the sweeping reforms promised before the 2010 election have not materialised (Department for Communities and Local Government, 2011). Another key 'beat' source, court reporting, is considered to be in decline; and certainly some newspaper groups struggle to justify it on cost-benefits analysis; leaving police forces to do a key job reporters used to do (BBC News, 2011).

In the early 1990s, 'public journalism' (or 'civic journalism') emerged as an alternative to the corporate-driven model which had swept away all before it in the American media market. Here was an attempt to de-institutionalise the profession. The rituals and practices concerning private (and protected) relationships in the old 'beat' were considered reactionary, standing counter to the public interest. The old relationships and routines which tie journalists to their beat were conceived of as a conspiracy against the people. The trust we place in our government has plummeted, as has the trust we place in their news media, leading to further experimentation with new journalistic forms, and means of engagement with the audience (Haas, 2003; Haas and Steiner, 2006).

Today we can look back on the burgeoning re-emergence of these experiments in online journalism, in the context of the rise of the network society. Nodes in the network are not of equal importance; and indeed, power relations are determined by inclusion or exclusion from the network (Castells, 2000, p15). This is problematic for public interest journalism in so far as social networks do not proportionally reflect wider society in terms of engagement or use – only one in 20 of those over 65 engage with social media frequently, and less than a third of 35-44 year-olds are active online (Ofcom, 2010). When it comes to news-gathering online, to developing sources, and reaching out to audiences, journalists must be careful how their networks develop, for pluralism is risked in an exclusively online network.

The rise of the network in online journalism, evidenced in everything from crowdsourcing via Twitter, to the processing of User Generated Content in the form multimedia, offers a means of extending public interest journalism, and of re-inventing the 'beat'. The network potential within the Internet can help bring about the re-constitution of journalism's traditional power-base, re-connecting journalists with their audience online within a wider social network.

Establishing an online beat

Aside from the search engines and social networks covered earlier, there are a range of bespoke verticals journalists (and news audiences) can use to access the news. Here follows a detailed look at some of the most useful, by function.

News search verticals

Google News (http://news.google.co.uk/) offers a relatively small number of generic news subject categories, but it is possible to create bespoke categories using Custom Sections:

> http://news.google.com/news/directory/createsection?cf=all&hl=en

This can be useful for journalists and other information professionals working in conceptual (or even geographical) beats, albeit it is not necessarily something the wider public shows any great demand for (Thurman, 2011).

Google News used to publicise the fact that it indexed 25,000 sources worldwide (with 4,500 news sources from the UK), but no official list is available, and indeed these figures have been disputed (Jarboe, 2007). The fact that many of these sources provide direct public relations copy has led some to question the blurring of lines between editorial and PR in online news aggregation (Orlowski, 2003).

Google News provides an advanced interface, and a range of filters including source, country, and byline. Headline search is available from the Occurrences drop-down. It is possible to search more than one location in Google News, but not from the advanced search panel. To do this, you must use the operator location: in an OR search, so for example:

"Celtic" location:Ireland

...will return results mentioning Celtic exclusively from this country (that is, search results will be sourced from publications within the country specified). In some searches this can help filter out irrelevant content (the above search likely won't return results about the *Boston Celtics*). Results can be ranked by date or relevance and results for major news stories are generally clustered (the source of much competition amongst online news publishers, vying for top position in the cluster). Search results are available in RSS or email alert formats. Google News contains only contains around 30 days worth of news, and most journalists will require access to much deeper archives available elsewhere.

Yahoo News (http://uk.news.yahoo.com/) offers a much more developed, hierarchical range of default categories than Google News does, but far, far fewer sources (it is unclear precisely how many). One fundamental difference between Yahoo News and Google News is that the former includes news published by Yahoo. It is important to use Yahoo News in addition to Google News not only because of differences in indexing and coverage, but also to offset any bias in results (Google News search results have proven to be more partisan, compared with sources found in Yahoo News (Ulken, 2005).

The advanced search options (http://search.news.yahoo.com/advanced) permits filtering by source and country. Bing offers a UK news category too (http://www.bing.com/news/) which has potential for news search.

Google News, Yahoo News and Bing News are all conventional search verticals, and functionality for each is optimised for searchers who know what they are looking for. But for some search needs, searching is neither optimal nor sometimes even possible. This is where other approaches are useful, such as browsing. This injects the potential for the sort of serendipity we experience when browsing newspapers, into search, and can be a much more intuitive way of sourcing news than search.

Newsnow (http://www.newsnow.co.uk/), with its easily-browsed (and detailed) hierarchy of topics, embodies this approach. Newsnow is more of a news directory, than a search engine, and its up-to-date (and densely populated) hierarchy of topics offers an intuitive touch, facilitating use in a very differeent way to the folksonomies found in some multimedia platforms. It is possible to search Newsnow, but only for one keyword at a time (for free), as using any more than one keyword will incur a charge. Some of the bigger UK news publishers have blocked Newsnow from indexing their content, but it remains a key resource for any local beat journalist.

Most online journalists will be familiar with seeing icons for...

Digg (http://digg.com/),
reddit (http://www.reddit.com/),
Stumbleupon (http://www.stumbleupon.com/)

...and other news sharing (or social news) sites peppered all around their copy. But there can be more to these sites than merely a means of distributing news; all of them aggregate news from around the web, making it easy to catch up on what online audiences are most interested in.

It is argued that the front-page of Digg, where the inclusion and arrangement of stories is determined by its users (via an algorithm), represents a means of sampling "what is interesting today?" (Halavais, 2009, p165). However, as with any statistical sample, skewing is possible. The popularity of content trending on these sites should always be taken with a pinch of salt as some politically motivated groups have been found to mobilise and effectively censor material they don't like on Digg (Halliday, 2010). For those who tire of visiting and re-visiting these social news sites, Popurls (http://popurls.com/) offers a trending list for each of them and more, presenting them in a relatively lo-fi, and unassuming way.

It has been argued that news verticals such as Google News assume journalistic functions (Machill et al., 2005), especially with regard to editorialisation on home pages, and ranking content by 'relevance'. It could also be argued that the format of conventional search engines does not lend itself well to different interpretations of the news. For these reasons, it is important for plurality in newsgathering, not to rely too much on one particular form of interface, or one way of interacting with search results.

As an alternative to the ten-results-in-a-list approach to new search engines, Silobreaker (http://www.silobreaker.com/) offers a unique news-gathering experience. Silobreaker is a semantic search engine; it understands key concepts in the news, in so far as it allows them to be isolated from the rest of news text. The service comprises news content which is filtered by a conventional search algorithm, but in conjunction with a community-constructed directory (or thesaurus) of names and concepts in the news. The Hotspots option (found along the top of the search bar) is one of the few remaining places online where geo-coded news can be found, in (visual) map form. The 'Network' option makes use of a thesaurus (or 'wiki') of proper nouns and descriptors, presenting the searcher with a social network diagram of all the companies, people, organisations, cities, key phrases and products which feature alongside your area of interest in the news. This is an excellent way to 'audit' news around key people/elements in the news, to ensure you haven't missed anything. This is just not possible in conventional search results.

Generic news aggregation – the news palate

Just as iGoogle offers a means of personalising web experience, so too are there various engines which offer an equivalent experience for news. Addictomatic (http://addictomatic.com/) encourages surfers to type in a keyword, and 'inhale' the web; it presents results from various verticals (spanning multimedia, blogs, social media and news) in one convenient place.

There are a number of other services which offer a similar news palate, albeit focussing on textual, conventional news, rather than multimedia and social media. Ensembli (account required) (http://ensembli.com/) offers a more linear, text-based news offering, and one for which there are far fewer timely results, but nevertheless, its uncomplicated interface has benefits.

EU Feeds (http://www.eufeeds.eu/) offers the latest 10 stories from key media for EU countries, coverage from all of which can be increased or decreased according to taste. This service is offered for all EU countries, state by state, organised by tab across the top of the page. For newspaper front pages, the imaginatively titled Front Pages Today (http://www.frontpagestoday.co.uk/) offers an overview of the UK nationals, while Newseum (http://www.newseum.org/todaysfrontpages/) sources international news titles (selection is more exploratory than comprehensive).

Daylife's topics (http://www.daylife.com/topics) offer a limited entry-point into current global (albeit dominated by the US) news themes. News is arranged around conventional topics, and the people/entities hitting the headlines within them. Pictorial stories tend to dominate, unsurprisingly given Daylife's business interests are closely aligned with those of Getty Images. Newser (http://www.newser.com/) offers an aesthetically appealing artists' palate view of the news (called 'the grid'). This user-curated aggregator shows what was trending on days in months gone by, though US news dominates (again). Newsvine (http://www.newsvine.com/) offers an easy-to-use social search news experience, albeit one which is severely limited if you are interested in a UK geographical beat – though several cities are covered, very few contributors specialise at this level of specificity. Newsvine (http://www.newsvine.com/) and Buzztracker (http://www.buzztracker.com/) offer alternatives, but both are US-centric in their coverage.

Automated news: email alerts

The search engines and aggregators introduced so far in this chapter can be classified as 'pull' technologies; that is, they pull the surfer toward them, in order to satisfy search needs. All of them serve a purpose as far as accessing a range of news sources is concerned; albeit it can be time-consuming and frustrating having to go back to each in turn as part of a regular routine, even moreso when following a breaking news story. To complement newsgathering, and add system to the process, it can be useful to set up 'push' technologies, like email alerts, which aren't so time-consuming.

Google Alerts UK (http://www.google.co.uk/alerts) (registration required) are the most comprehensive way of keeping up on keyword mentions across the Internet. These alerts are more customisable than the services' competitors, taking advantage of the speciality verticals Google excels in (including news, blogs, and video search). It is possible to set the frequency of these alerts to weekly, daily or as-they-occur – but of course this could quickly lead to a saturated email inbox. Boolean logic and some advanced operators (including the *site:* function) are supported in Google Alerts. There is an important caveat attached to alert results drawn exclusively from web, or news content, however. If material matching your search does not make it into the top 20 results (for the former) or top 10 results (for the later), then you won't get an alert – so less well-optimised content could slip under the radar.

The main competitor for Google Alerts in this market is Yahoo alerts (http://alerts.yahoo.com/) (registration required). For some searches it may be beneficial to run both at the same time (especially given differences in indexing across these two platforms). There are several other search and social monitoring services which offer email alerts as a means of keeping up with events which may be worth using, including:

Alerts.com (mobile market) (https://www.alerts.com/)
Giga Alert (subscription) (http://www.gigaalert.com/)
Topikality (semantic alerts engine) (http://www.topikality.com/)

Mailing lists and email newsgroups

Listservs, centrally organised email lists, have been a key part of journalistic news-gathering since email became the communicative medium of choice. Past literature in online journalism has focussed heavily on Listservs, however they are not without ethical considerations, nor controversy. The *Journolist* discussion group (whose technology rests on email-contributions), established for left-leaning American journalists in 2007, was closed down due to the leaking of critical remarks against conservative figureheads by contributor David Weigel, who subsequently resigned his post. Listservs are for some still a crucial component of research and news-gathering, but it is not always clear where to go to find them (certainly most of the publicly collected lists on the web have not been updated since the mid-2000s.

Listservs are particularly popular within specialist communities, such as investigative journalism, and the following two are highly regarded therein:

Global Investigative Journalism Network:
(http://www.globalinvestigativejournalism.org/)
JOURNET: (https://ls2.cmich.edu/cgi-bin/)

Callahan (1999) provides an extensive (if somewhat dated) list of listservs:

http://reporter.asu.edu/listserv.htm

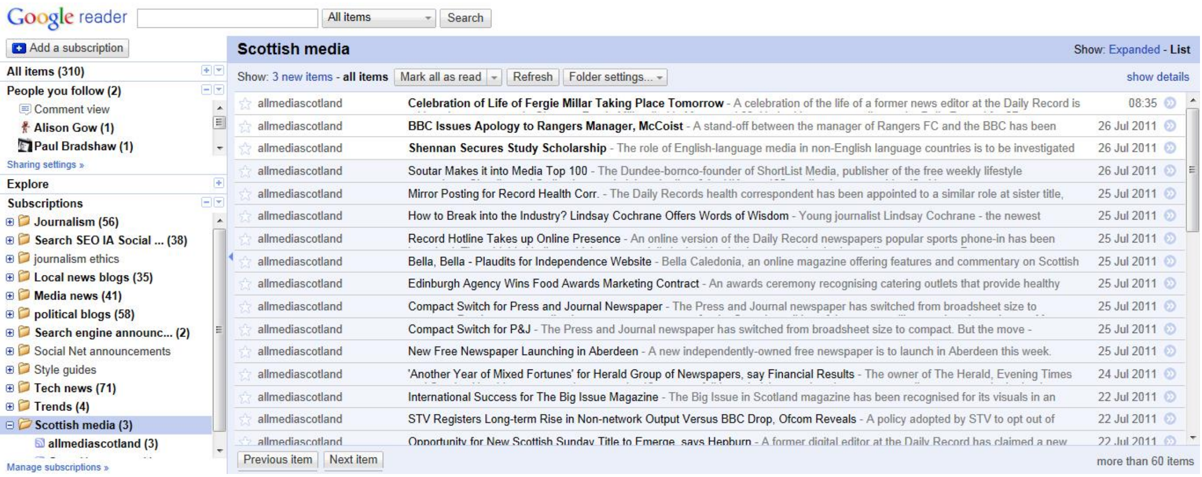
Really Simple Syndication - RSS

Taking the logic of 'push' technologies a stage further than email alerts, Really Simple Syndication (RSS) is perhaps the technology most closely aligned to the journalistic appetite for immediacy on the web. Put simply, RSS is a distribution technology which allows publishers of online content to instantly alert audiences when they have published new content.

In newsrooms up and down the country, organisations subscribe to Press Agency news wires which channel in breaking news from around the globe, provided to journalists via desktop applications. It is possible using RSS to develop a personalised newswire, helping journalists stay up-to-date with online news in their beat. This involves finding sources to follow, then setting up systems to follow them, usually via a feed-reader, or other RSS-compliant utility.

Most Internet browsers allow users to store RSS subscriptions locally, and will even allow other programs, including email clients, access to these feeds across the web (even if the latest versions of leading browsers no longer offer RSS alert icons in their default view). But for optimum accessibility (including, in some cases, via mobile devices) stand-alone platforms have their advantages:

iGoogle (http://www.google.com/ig)
My Yahoo (http://uk.my.yahoo.com/)
Net News Wire (Mac – download) (http://netnewswireapp.com/mac/)
Netvibes (http://www.netvibes.com/en)
Pageflakes (http://www.pageflakes.com/)
Reeder (Mobile app) (http://reederapp.com/)
Windows Live (http://explore.live.com/)



Google Reader's interface

Google Reader is perhaps the most comprehensive, and integrated of these. Once registered (and logged in) the site's navigation is comprised of three parts:

The *Add a subscription* box on the top-left of the screen.
The *Reading pane* in the middle of the screen.
The *Subscriptions* section at the bottom-left of the screen.

RSS feeds come in many forms, and can be found using a range of methods. Here are four:

Use Google Reader's search index (in the A*dd a subscription* box), for names or keywords you're interested in. When results are returned, click on the subscribe button (found under all search results). Each feed should indicate the number of subscribers it has. You may not know who they are, but their number may suggest how reliable a source they are, (and, importantly, whether or not the feed is still active.
Many websites contain RSS feeds, but these won't necessarily feature in Google Reader's index. Look for the orange RSS button in your browser's address bar (in newer releases of Internet Explorer, you may need to display the *Command Bar* to seek out RSS).
For some platforms, it is necessary to set up bespoke feeds (for example, YouTube used to require url-hacking, in order to set up tag-based RSS feeds.
It is possible to 'scrape' updates from web pages which don't provide RSS by using services like Page-2-Rss (http://page2rss.com/).

When adding more feeds to Google Reader, it will become increasingly necessary to organise them into folders. There are several keyboard shortcuts for navigating around Google Reader, and one of the most useful relates to reading. If you find the volume of results returned excessive, it may be sensible to adjust your reading view to the *List view* (as opposed to the *Expanded view*). This has the potential to save vast amounts of time, and reduce the wear and tear on many a journalists' wrists, caused by excessive scrolling .

The sharing options available in Google Reader set it apart from the rest; it is possible to share individual items with your Google contacts, or aggregate all of your feeds together in a public folder, or even export all of your feeds into an xml file to share manually with the world (see the *Import/export* tab in the *manage settings* section). Google Reader also provides recommendations, which is analagous to the 'similar' Google filter.

The future of RSS?

The demise of Bloglines in September 2010, and later, the decision to remove the RSS icon as a default view from version 4 of Firefox (Orchard, 2011) have led some to question whether RSS is still relevant, now that many people use Facebook and Twitter to keep up-to-date on news.

But it is very important to draw a distinction here; it is not the death of the technology itself which is implied by these recent events, RSS will continue to form part of the infrastructure which brings us news on any online platform. It is rather the death of consumer RSS readers which is suggested, something which has been addressed several times in the past few years (Diaz, 2009; Tartakoff, 2010; Kamen, 2011).

Whether or not there remains a viable commercial market for such readers, journalists, as professionals who deal primarily in information (one might add new information) will continue to be dependent upon such resources, along with technologists, researchers, publicity and marketing professionals, and government agencies.

But just because RSS readers are an important means of keeping up with the news today, doesn't mean that they will always be available, and even if they are, it is important to recognise the limitations inherent to this method of newsgathering. In the wider scheme of a journalistic beat (whether regional or subject-specific) there will inevitably be many online sources which do not carry RSS, because they provide alternatives services (such as email alerts), or just because they haven't enabled the technology.

Moreover, there are will be many beats for which there are few active and engaged online representatives, making crowdsourcing problematic. An alternative to available RSS feeds and crowdsourcing is therefore essential; the ability to scrape updates from websites becomes crucial.

At the most basic level, this would involve creating an RSS feeds for web pages which don't have them. There are many options which will help here, including:

Feed43 (http://feed43.com/)
FeedFire (http://www.feedfire.com/)
Feedity (full functionality requires subscription) (http://feedity.com/)
FeedYes (registration required) (http://www.feedyes.com/)
Page-2-RSS (http://page2rss.com/)

WebRSS (http://www.webrss.com/)

But it is important to recognise the limitations of such services; most notably, they only give updates on what has been added and what is new, not what has been deleted, changed, removed or hidden. For tracking the full gamut of changes to online sources, it is necessary to employ what Calishain refers to as 'information traps'. (Calishain, 2007).

These are effectively automated Page monitors which track changes over time. There are a number of freely available (and subscription) web-based tools which can help here, including:

Watch that page (http://watchthatpage.com/)
Trackengine (subscription) (http://www.trackengine.com)
Change Detect (http://www.changedetect.com/)
Trackle (subscription) (http://www.trackle.com/search/)

There are also a number of browser extensions which perform a similar function, including:

Update Scanner Addon for Firefox:
(https://addons.mozilla.org/en-US/firefox/addon/update-scanner/)
Update Scanner Addon for Chrome:
(http://updatescanner.mozdev.org/drupal/content/)
Outwit Hub Firefox Addon:
(https://addons.mozilla.org/en-us/firefox/addon/outwit-hub/)

In addition, for those who are fully engaged in this approach to news-gathering online, there are in addition a number of more powerful client-side screen scrapers (or 'web data extraction' software), most of which require subscription. These are generally more configurable and customisable than web-based offerings, they are better at avoiding false-positive changes and minor changes than web-based trackers, and some (like Outwit Hub) can perform a number of other analytical functions too, which can be highly valuable in online investigations (such as offering advanced filtering and notification, scheduling, and even support authorship of macros and 'regex', for bespoke scraping tasks, not to mention full technical support). These include:

Website watcher (subscription) (http://aignes.com)
Page Update Watcher (free) (http://download.cnet.com/Page-Update-Watcher/3000-2370_4-10468473.html)
Outwit Hub (subscription) (http://www.outwit.com/products/hub/)
ScreenScraper (http://www.screen-scraper.com/products/all.php)
Mozenda (subscription) (http://www.mozenda.com/web-data-scraping)

For further customisation and control over data scraping, it may be necessary to develop your own scrapers. This will involve learning programming code, whether iMacros (Shrenk, 2007) or Ruby (Spradlin, 2009), or even a high-level coding language. Today, in cities up and down the UK, a healthy culture of social media events is bringing together journalists (hacks) and programmers (hackers), whereby programming can be applied to journalistic problems, and so journalists might learn some of the skills required in the information age. For guidance see Hackers/Hackers (http://help.hackshackers.com/).

There are ethical issues concerning screen-scraping, aside from those issues which affect all informational resource. Extensive scraping of a website can place a burden on that site's servers and bandwidth, so professional screen scrapers will often contact the publishers of a site before starting work. Journalistic work in this area overlaps significantly with computer programming, for which reason an appeal to the ethical standards of this profession would be appropriate; especially the Association for Computing Machinery guidelines.

Newsgathering: a general framework

Once basic **push** and **pull** methods in online news-gathering are in place, it is necessary to consider what sources to follow. This can be a fairly subjective exercise. The following schematic is offered for the purposes of setting up a local news wire.

A model for hyperlocal online news-gathering: Rationale

The underlying philosophy behind this model is to look to the information outlets and sources in a local community 'beat' first, then follow up by finding these sources across the web. This approach is not presented as an alternative to conventional off-line news-gathering methods, nor as a replacement for them; there aren't enough people or information sources online (and there likely never will be) for this to be a realistic goal. This model is presented as a tool which may help journalists...

> To adapt from offline to online news-gathering.
> To systematise news-gathering online.
> To eke out stories which might otherwise be missed completely.
> To reinvigorate their relationship with readers.

In the process, this system will re-calibrate the relationship between journalist and reader; the 'beat', becomes a 'social network', which is open to new forms of measurement, audit, review and analysis.

Application

It is essential to choose a taxonomy-model which reflects the administrative bounds of the beat – in the case of a local reporter, the structure of the local authority is a good fit.

Once a model is established, keywords must be selected, which will be used to find sources and set up RSS feeds. For a geographical beat this will necessarily include a full run-down of all local place-names, but it will also (at different points in the news calendar) feature other terms too, including local people, organisations, companies, events, etc.. Many sources will not require keywords, merely automated RSS feeds.

This structure must be kept up-to-date, as it is subject to change over time; it needs to reflect the shifting nature of peoples' lives.

Some online sources have no RSS feeds, and scraping feeds for these sources can be problematic. Once relationships are established with sources, a social network begins to develop; such that journalists may find news more efficiently this way, than via official sources. For these reasons, this news-gathering model must comprise three parts;

> RSS feeds which are aggregated and filtered (in Yahoo Pipes, Feedrinse or similar),

A diary of sources which require systematic re-visit manually (ideally kept in a spreadsheet) and;
Membership of various social networks (including Twitter, Facebook, LinkedIn, Delicious etc.).

1. Meta-sources

Competitor news sources: Any beat will require several news search feeds – covering references to the local beat at local (competitor), regional, national and international news levels, and possibly references to the beat in the trade press too). Set up beat (and other) RSS feeds in:

Google News
Yahoo News
Bing News
Newsnow
Daylife

There will be some duplication, but this can be filtered out – ultimately these different services have different indexes, so it is important not to miss out

...if a subscription is available, then set up feeds in:

Factiva/Nexis etc.

Social news sites: of less importance to local journalism, but still worth monitoring:

Reddit
Stumbleupon
Digg

2. Administrative sources

Many 'official' news sources are available online; from local authorities, to local single-issue charity groups. These can be organised accordingly:

Governance sources (the authorities)

Council website RSS feeds
Council newspaper sheet (likely require monitoring)
Council committee minutes – via Openly Local (if available)
Online Forward Plan (likely require monitoring)
Hansard (covered by They Work For You searches for placenames and politician names within network)
European Parliament (no feeds on site, advanced search necessary)
Freedom of Information – What Do They Know feeds for all local public authorities (NHS trusts, schools, council etc.)

Political sources

Parish council (websites/SM groups, parish councillors across SM).
Local politics (committee SM groups, officers, councillors, mayor, opposition).
National politics (MPs, websites/SM presence party websites/SM presence).
International (MEPs websites/SM presence).

Policy areas (authorities, and alternative voices)

Community: council social services, animal welfare groups, asylum groups, carer groups, children groups, women's groups, minority groups, Surestart etc.

Leisure: local arts groups (theatres, arts centres, galleries, etc.), community centres, conservation groups, sports groups, teams and societies.

Health and social care: PCT management, carers groups, doctors GPs and hospitals, mental health groups, health awareness groups, patients groups.

Education and learning: Schools (primary, secondary, etc.), colleges, universities, council information, educational groups, teachers, parents associations.

Transport: transport committees and lobbies (motorists, cyclists, etc.).

Environment & planning: Planning committees, nimby groups, environment groups, big business.

Occupation/industry: employment agencies, business groups, businesses, local trade union reps etc.

Law and order: council advice, local courts, local police services, neighbourhood watch groups, local lawyers/solicitors, Citizens Advice Bureau, local arbitration.

3. Social sources

For all social sources, think about who (individuals, groups) in conjunction with either where (geography of beat) or what (topic of beat), whichever is more apt.

Social Networks

Facebook

Search for and keep up with people, pages, groups, and events.
For people concentrate on searching by Location/Education/Workplace
Speculative search using advanced search
Public updates and conversations available in IceRocket or advanced search Google. Account required.

Twitter (and other microblogs and Realtime services):

Set up place-name feeds
Set up Geo-location search feeds (watch for limitations)
Follow key/popular figures in beat, including famous sons and daughters (footballers, rugby players, other local celebrities – see knowhere and Wiki pages for suggestions)
For generic Twitter search feeds it may be necessary to filter for content that has been re- tweeted or which contain links, especially for breaking news.

Find followers in beat using geo-twitter and Tweepz searches (which provide feeds for new people).

Delicious

Sign up for tag-based RSS feeds. Quite a lot of companies, but the odd surprise. Account required.

LinkedIn

Keep up with companies within a beat
Search for and friend individuals working within beat. Account required.

Forums: Sometimes anonymity can be a strength, not just in terms of finding whistleblowers, but also in contentious/competitive fields – the main weakness with forums can be lack of RSS). Try geographical terms in:

Omgili
Google Groups
Yahoo groups
Boardtracker
Boardreader
Possibly some other bespoke site (like Mumsnet).

Blogs and websites

Blogs

Search for posts featuring local place-names via Google Blog search, Icerocket,or Social Mention.
Search for bloggers whose *About* includes mentions of a local place-name.

Websites

Conventional search in Google, Bing, Yahoo, Ask, Delicious etc. for local placenames.

Mutlimedia:

Video

Youtube
Blinkx
Google Videos,
Truveo

Images

Flickr
Panoramia
Twitpic (included in Twitter feeds)

Reference

Wikipedia: Set up searches to monitor all pages of interest in beat (places, people, subject matter etc.).
Google maps: need to do this manually now, RSS on Google maps is no longer.

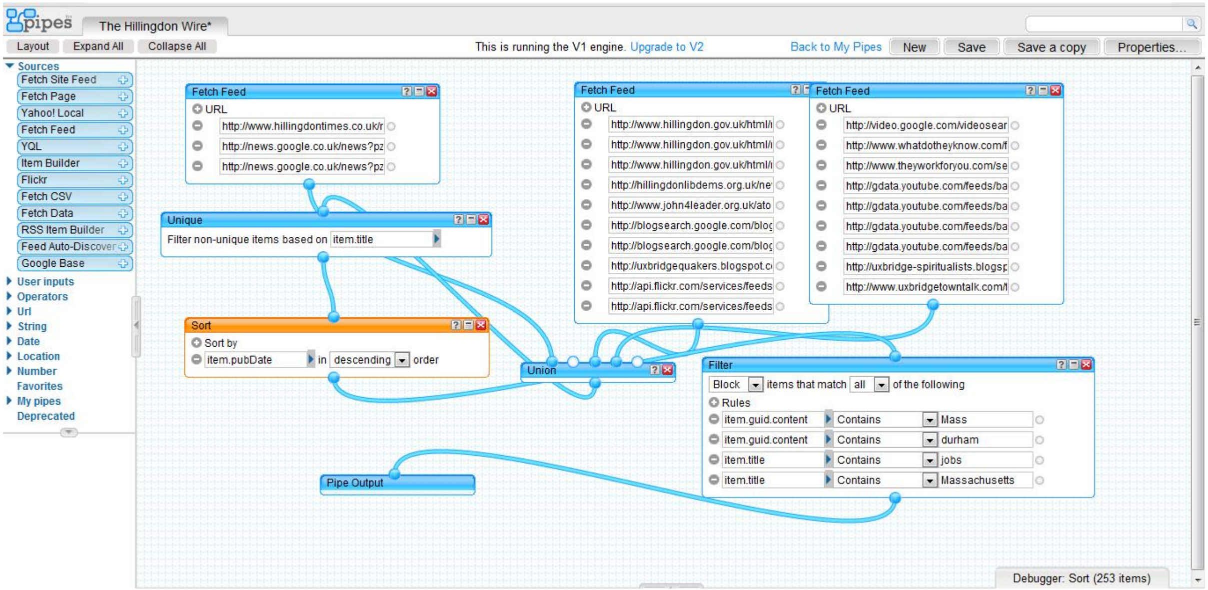
Follow other local themes in web 2.0, as they arise:

Crash map (http://www.crashmap.co.uk/)
Crime maps (http://www.police.uk/)
I want great care (https://www.iwantgreatcare.org/)

Getting organised

Once suitable sources for these themes (and types) are established, it is necessary to differentiate between those which can be effectively found online, and those which can't. Those which can't will have to be pursued via telephone, email and other offline methods. Those which can be found online, require differentiating between those which can be tracked using RSS and those which can't. Those which offer RSS can be followed in a feed reader. This is part of the newsgathering process which would ideally be shared within a team, and communicated via a shared diary (or calendar).

Meanwhile, it can be useful to aggregate all of your beat feeds into one place, where they can be filtered for non-relevant content, and parsed of duplicates. Yahoo Pipes is the most robust (but by no means the only) means of doing this.



Yahoo pipes interface: a simple pipe for gathering news

Functions in Yahoo Pipes

The key modules for a simple Yahoo Pipe are:

Fetch feed: paste your RSS feeds in this box – use the + sign to add more.
Filter: filters out terms associated with places, events and issues you don't want coming back. For example, there is an Uxbridge in Masaschussets – often referenced as *Mass* in blogs and news-copy. Use this module to block all references to this term.
Sort: this filter allows you to sort your content by publication date – obviously for a newswire we need all content to be distributed by publication date.

Union: use this module to channel several fetch feeds together into one output.
Unique: this filters out non-unique content, so if you find a story from a local newspaper which also crops up in an aggregator feed search, or in a blog search, it shouldn't publish all three.

Join these all together into a logical flow, and finally connect to the *Output* box, and with a little more tweaking, you should have your pipe up and running.

References

BBC News (2011) Court results put on Twitter by West Midlands Police, BBC, April 19: http://www.bbc.co.uk/news/uk-england-birmingham-13127533

Calishain, Tara (2007) Information Trapping: Real-Time Research on the Web. New Riders: New York

Callahan, Christopher (1999) A Journalists' Guide to the Internet: The Net as a Reporting Tool. Allyn and Bacon, New York.

Carey, James W (1999) 'In Defense of Public Journalism', in Glasser, Theodore L (Editor): The Idea of Public Journalism. The Guildford Press: London.

Castells, Manuel (2000) Material for an exploratory theory of the network society, British Journal of Sociology Vol No. 51, Issue No. 1 (January/March 2000) pp. 5-24.

Davis, Aeron (2000) Public relations, news production and changing patterns of source access in the British national press, Media, Culture and Society, Volume 22, pp39-59.

Diaz, Sam (2009) RSS: A good idea at the time but there are better ways now, ZDNet, August 25:
http://www.zdnet.com/blog/btl/rss-a-good-idea-at-the-time-but-there-are-better-ways-now/23276

Department for Communities and Local Government (2011) Code of Recommended Practice on Local Authority Publicity (The Publicity Code), DCLG, March 31:
http://www.communities.gov.uk/publications/localgovernment/publicitycode2011

Fishman, Mark (1980). Manufacturing the News. Austin, University of Texas Press

Foreman, Gene (2010) The Ethical Journalist: Making responsible decisions in the pursuit of news. Wiley-Blackwell: Chichester

Gilligan, Andrew (2009) The propaganda newspapers, The Evening Standard, July 27:
The propaganda newspapers

Haas, Tanni (2003) 'Importing Journalistic Ideals and Practices? The Case of Public Journalism in Denmark', The International Journal of Press/Politics 8: 90-97

Haas, Tanni and Steiner, Linda (2006) Public Journalism: A reply to critics, Journalism, Vol.

7, No. 2, pp238-254.

Halavais, Alexander (2009) Search Engine Society (Digital Media and Society Series). Polity Press: London.

Halliday, Josh (2010) Digg investigates claims of conservative 'censorship', The Guardian, August 06: http://www.guardian.co.uk/technology/2010/aug/06/digg-investigates-claims-conservative-censorship

Hansen, Kathleen; Neuzil, Mark and Ward, Jean 'Newsroom Topic Teams: Journalists' assessments of effects on news routines and newspaper quality. Journalism and Mass Communication Quarterly; Winter 1998; 75, 4; ABI/INFORM Research pg. 803-822.

Harrison, Jackie (2006) News. Routledge: London.

Jarboe, Greg (2007) Revealing The Sources Of Google News, Search Engine Land, May 31: http://searchengineland.com/revealing-the-sources-of-google-news-11353

Kamen, Croc (2011) RSS Is Dying Being Ignored, and You Should Be Very Worried, Camenblog, January 03: http://camendesign.com/blog/rss_is_dying

McManus, John (1994) Market-driven journalism: Let the citizen beware? Sage: California.

MacDonald, A.M. (1977) Chambers Twentieth Century Dictionary. New edition. W & R Chambers: London.

Machill, M., D. Lewandowski and S. Karzauninkat (2005) ‘Journalistische Aktualität im Internet. Ein Experiment mit den “News-Suchfunktionen” von Suchmaschinen’, pp. 105–64 in M. Machill and N. Schneider (eds) Suchmaschinen: Neue Herausforderungen für die Medienpolitik. Berlin: Vistas.

Machin, David and Niblock, Sarah (2006) News Production: Theory and Practice, Routledge: London.

Morrison, James (2010) Spin, smoke-filled rooms, and the decline of council reporting by local newspapers: the slow demise of town hall transparency. Presented at the 61st Political Studies Association Annual Conference:Transforming Politics: New Synergies: http://www.psa.ac.uk/journals/pdf/5/2010/612_322.pdf

Ofcom (2010) UK Adults' Media Literacy Report, May 17: http://stakeholders.ofcom.org.uk/market-data-research/media-literacy/medlitpub/medlitpubrss/adultmedialitreport/

Orchard, Leslie (2011) What happened to feed auto-discovery in Firefox 4?, Decafbad blog, January 15: http://decafbad.com/blog/2011/01/15/what-happened-to-feed-autodiscovery-in-firefox-4

Orlowski, Greg (2003) Google News: press releases are OK – Official, The Register, April

05: http://www.theregister.co.uk/2003/04/05/google_news_press_releases/

Schrenk, Michael (2007) Webbots, Spiders and Screen Scrapers. No Starch Press: New York.

Spradlin, Justin (2009) Ruby Screen Scraping with scRUBYt!, JustinSpradlin.com, April 27: http://www.justinspradlin.com/programming/ruby-screen-scraping-with-scrubyt/

Tartakoff, Joseph (2010) The Death Of The RSS Reader PaidContent, September 10: http://paidcontent.org/article/419-the-death-of-the-rss-reader/

Thurman, Neil. (2011) "Making 'The Daily Me': Technology, economics and habit in the mainstream assimilation of personalized news". Journalism: Theory, Practice & Criticism 12(4).

Tuchman, Gaye (1972) Objectivity as Strategic Ritual: An Examination of Newsmen's Notions of Objectivity, American Journal of Sociology, Vol. 77, No. 4 (Jan., 1972), pp. 660-679.

Ulken, Eric (2005) Non-traditional sources cloud Google News results, OJR, May 19: http://www.ojr.org/ojr/stories/050519ulken/

Zelizer, Barbara (2004) Taking Journalism Seriously: News and the Academy. Sage: Thousand Oaks

Chapter 8: Verifying online sources

The art of the online hoax

Journalists have been subject to hoaxes since the profession came to be. Sometimes hoaxers intended to change the course of history. In April 1887, the Times fell prey to one of the most infamous hoaxes of the nineteenth century, publishing fake letters purporting to show Charles Stewart Parnell's support for the Phoenix Park murders. On the other hand, some sophisticated hoaxes have been motivated by nothing more than the product of a bar-room bet. In 1953, The Atlanta Constitution carried a splash on a close encounter with extraterrestrial life. Two men had shot, shaved and cut the tail off a rhesus monkey, presenting what remained to the news desk as (late) life from Mars.

Debate abounds about the nature of such hoaxes, which in turn informs how journalists might deal with them. Some insist that the process of corroboration for online documents is materially different to the pre-online processes used, while others argue that there are offline precedents to all online hoaxes. This debate mirrors a much wider debate about the extent to which the internet is merely another tool for communicative purposes, or a new way of communicating. Friend (2007) suggests we should position ourselves somewhere in the middle-ground, which is a pragmatic position to take.

The advent of the internet, and the increasing reliance we all have upon electronic documents for unearthing information both factual and trivial, requires that the age-old checks and balances journalists have always used to verify the authenticity of their sources need updating. Here follows a series of suggestions which can help journalists tell fact from fiction, find out who is behind an electronic document, and ultimately, help verify that document's authenticity.

How to verify a website

Journalists often have to make snap judgements about the information contained in websites they use, and given how easy it is to replicate another website by lifting HTML code, this is a deeply problematic area.

Example: An unwelcome anniversary...

In December 2004, BBC World run a story on the aftermath of the Bhopal disaster 20 years on from the tragedy. A leak of chemicals from a factory managed by Union Carbide (an American petro-chemical company) late in 1984, led to the immediate death of 3,800 people, rising several thousands more in subsequent weeks and months.

In the intervening years, Union Carbide had been acquired by Dow Chemicals, whom a BBC journalist sought for interview. Unfortunately the journalist was fooled by a fake website run by anarchists The Yes Men, called dowethics.com. The journalist emailed the address on this site, and arranged an interview with the 'CEO of Dow Chemicals'. A Yes Men representative went on air live (masquerading as Dow's CEO) to publicly acknowledge Dow's 'responsibility' for the disaster, causing an immediate loss of $2 Billion from the share price of the company, and significant loss of face for the BBC (Wells and Ramjesh, 2004).

Websites are a key source of information for journalists, but it is important to approach unknown websites with a healthy dose of scepticism. A helpful way for journalists to weigh up the veracity of a website, and one that shouldn't prove onerous, is to employ the Journalists' Checklist...

Who... is responsible for the site? Is there an 'About' page, or an FAQ page which explains this? There may not be and of course in some situations it is perfectly reasonable and fitting that websites be published anonymously (especially when they are involved in whistleblowing). Nevertheless, if this is the case, this will necessarily impact its reliability as a primary source. If there isn't one, ask yourself why not. If there is one, are the details included legitimate (try searching for an address, names, telephone numbers, or anything else you can find using the various tools and methods elsewhere in this book).

Most reputable websites will have an about page, but if you are in any way suspicious, you can check ownership of a domain by searching in one of the many domain registration directories, such as domaintools' Whois Lookup (http://whois.domaintools.com/). Using the Whois domain directory it is possible to find out the name (or names) in which a website was registered, the emails used to register it, the date when it was last registered, where it was registered, and if you are prepared to pay a premium fee, you can even find out what other domains have been registered using the names you have found as registrant of the domain you are concerned with. This does not guarantee that the user has registered using legitimate information, and indeed it is possible and entirely legitimate to have some personal details hidden from public view from any Whois lookup service. But of course if this is the case, you may be drawn to certain conclusions about the reliability of the site in question.

What ...is the nature of the site? If it is a corporate site, its purpose is to sell you something, whereas if it is a governmental site, its purpose is to inform (some might say persuade). Is it obvious from content on the site what its purpose is?

Understanding basic conventions in web addresses can help here, especially with regard to upper level domains. All British universities have web addresses which end in *.ac.uk*, and government websites end in *.gov.uk*. Knowing that companies trade in *.com* and *.co.uk*, while advocacy groups, NGOs and hobbyists often use *.org* domains, can at least inform what you might expect from a website, at a very basic level. Stephen C. Miller developed a simple schematic (Miller, 2004) for determining the reliability for websites whereby Government, then University domains are ranked most trustworthy (albeit with relativist caveats; such as excluding non-democratic governments, and accounting for personal use of academic domain space). Wikipedia offers an exhaustive list of domains:

http://en.wikipedia.org/wiki/List_of_Internet_top-level_domains

Why would anyone trust this site – how popular is it? It is relatively easy to establish how large online (US) audiences are for some websites. This can in turn suggest how reliable information shared on said site may be. Of course popularity is certainly not a failsafe proxy for reliability and trustworthiness; after-all most major news outlets widely regarded as amongst the most reliable sources on the web, have been duped and hoaxed at some point. We should be wary of equating credibility with online popularity as the only means of verification, for this form of “pack mentality” is not full-proof (Friend, 2007, p69).

Nevertheless, if you have an idea of how popular a site is, this can add an extra dimension to the other means used to verify a claim made on a website.

Use the following tools to establish how big a sites' US following are (albeit be aware that these statistics are based on extrapolation of those who 'opt in' to using tracking software and services online, so they contain a bias):

Alexa (http://www.alexa.com/)
Quantcast (http://www.quantcast.com/)

When searching for the official websites of corporations and other large organisations, it is always worth looking for them within public directories (like those introduced in Chapter 3 of this book). The veracity and reliability of information about websites in these directories is managed by real human beings, who are 'experts' in the fields they manage within the directory. Search engines cannot offer such protection, so there is at least a modicum or reliability about the results you will find here:

DMOZ Open Directory: (http://dmoz.org/)
Yahoo Directory: (http://dir.yahoo.com/)

If you are in any way suspicious of a site, especially in relation to what might seem like doctored or purloined graphics or images, check the source code of the site in question (right-click on a page in any browser, and seek out 'view source' – you may then have to Control-F through the file, looking for file extensions, or duplicate code). If a site proclaiming to be a legitimate organisation has 'borrowed' from another site, or indeed if documents (say images for example) link through to other sites with an agenda, this might lead you to be sceptical about the authors.

Where can you find out what the website used to look like? The Way Back When machine (or Internet Archive) (http://www.archive.org/index.php) is a useful tool for prying into previous versions of web pages. This can be useful when pages disappear, move to new locations, or change content. While it is possible to design web pages in such a way that their pages are not indexed by the Way Back When machine, and while it is possible for web authors to request that content be removed from this site, it can never-the-less be helpful for uncovering lost or unavailable material.

A similar (albeit far more restricted) means of accessing past pages of websites is in the cached option you often see associated with Google results. Click on these links to see what the website looked like when it was last indexed (website indexing can be up weeks out of date), which goes some way to explaining why from time to time you click on a page from your search results which don't actually feature the terms you've searched for – that page is gone!

When... was the site last updated? Some sites will contain dates on pages, others will contain upload dates within the HTML of their pages, and others still contain dates in their web address (especially blogs). Google often includes a date of publication within search results for the web addresses of particular pages. In any case, reliability of online information is sometimes contingent upon timeliness, and there are other ways to check this.

By pasting the following javascript code into a browser, when looking at a particular page:

"javascript:alert(document.lastModified)"

...it is sometimes possible tell when a site was last updated (albeit this will not work on some browsers, nor where a web page contains dynamic content, which discounts a lot of modern web pages). If you wish to monitor a page over time to see if it changes, you can subscribe to a service such as Change Detection (http://www.changedetection.com/) (or any of the screen-scraping tools featured in the previous chapter).

Sometimes you can tell a lot about a site by which other sites are linking to it, which is where *Open Site Explorer* (http://www.opensiteexplorer.org/) can be a useful tool. Designed for webmasters who wish to check the link structure (and marketing) of their websites, this tool doubles up as a handy guide to who is linking to whom. Of course it is dangerous to presume online links are equivalent to an endorsement (many people link to sites they disagree with quite fundamentally). Nevertheless, all search engines use links to determine the PageRank of individual sites, and it is not uncommon for like-minded people to organise linking policies to improve their collective visibility in search (linking can be thought of as a form of social network). For this reason, Open Site Explorer (and to a lesser extent the *link:* function in Google search) are valuable tools in helping provide insight into those influences and lobbies which exist online, though neither can be considered to be comprehensive. As an alternative to Open Site Explorer, use Backlink Watch (http://www.backlinkwatch.com/).

Lastly, when a story appears too good to be true, try browsing for it through some of the established hoax-busting websites, to see if anyone else is discussing or alerting surfers to its unreliability:

Snopes (http://www.snopes.com/)
Urban Legends (http://urbanlegends.about.com/)
Museum of Hoaxes (http://www.museumofhoaxes.com/)

There are many ways to check the veracity of an internet source, but there is no substitute for the traditional, and entirely healthy scepticism most journalists will take to finding any source they might consider using.

How to corroborate news websites

Example: The vanishing newspaper

In early 2010, a new weekly newspaper launching in London: The London Weekly. At the time (and, some may say ever since) it was not easy to find work as a journalist, and many job-seeking journalists wondered why so few friends and contacts in the industry had heard of jobs being advertised for this new paper. Some forensic analysis later (undertaken by James Ball, journalists at Journalism.co.uk, and Media Guardian, not to mention contributors on Media Guardian), it was discovered that the newspaper was an elaborate PR exercise, rather than the fourth estate it purported to be. (Ball, 2010)

News websites (such as those covered in Chapter 7 of this book) are much like any other type of website when it comes to verification, and all of the approaches mentioned in relation to checking the veracity of websites more generally applies to this niche.

However, there are one or two other factors suggested in the literature, which may merit further consideration. Cooke (2004) incorporates them into an extensive check list for evaluating sources of news, reproduced in full here:

> What is the purpose of the site?
> What is the coverage of the site?
> What topics are covered, and are they covered comprehensively?
> Is the site an electronic version of a printed publication, or is it the site for a television or radio station? Does the site provide access to the whole content and if not how has the information been selected?
> What is the reputation and expertise of any individuals or organizations involved in the production of the site? Is this an authoritative source of information?
> What is the likely accuracy of the information?
> Has the information been through any quality-control processes, such as refereeing?
> Is the information likely to be biased by any individuals involved in its production?
> Is there an explicit date for the information?
> Is the information up-to-date?
> When was the information last updated?
> When will the information next be updated?
> How frequently is the information updated
> Is there a statement of policy regarding the frequency of updating and the updating process?
> Does the source need to be monitored or reassessed at a later date to ensure continued currency and maintenance?
> Is the site easily accessible?
> Is the information well presented and arranged?
> Is the site easy to use and are there any user support facilities?
> How does the site compare with other similar sites?
> What is your overall impression of the quality of the site?
>
> (Cooke, 2004, p135-6)

How to corroborate social media accounts and blogs

Example: Going off message...

In 2009 two students invented a fake Twitter account claiming to be then foreign secretary David Miliband. To make this profile as plausible as possible, they added comments patiently and accurately, monitoring the foreign secretaries real-world movements, and mirroring them in their Tweets. That was until the death of Michael Jackson, when they made some altogether more newsworthy comments, catching out much of the UK press in so doing (Anderson, 2009).

Example: The joy of texts...

Toothing is a digital means of engineering anonymous sexual encounters which relies on the short-range (often anonymous) messaging function available in Blue tooth technologies (such as Blackberries, and Palm top computers). However, it was first invented by TheTriforce.Com's Simon Byron, Ste Curran, and Dave Taurus, as a ruse to fool the media. The concept was brought to the masses in

March 2004 with the creation of a Toothing Blog, a social network where the experienced and uninitiated could come together. Involving a cast of imaginatively named characters (including Toothy Toothing, from Tooting) this 'community' was soon aided and abetted by curious and intrigued surfers from around the world, which in turn led to the establishment of further forums to serve the needs of Toothers internationally. Little wonder then, that the media would soon be drawn into this honey trap. In the month following the Toothing blog's debut, Reuters ran a lengthy article on the phenomenon, interviewing one of the hoaxers in the process. Wired News, The Independent (Arthur, 2004) and the BBC (Kelly, 2004) were also duped, and many of those who fell for it still hadn't acknowledged the fact a year later (Hanas, 2005), albeit perhaps with good cause.

Just as art imitates life, so the inverse is true of Toothing, which has subsequently moved beyond an acknowledged (and sometimes unacknowledged) urban myth into the realm of reality.

In some social media platforms, people can verify their social network accounts, but what if the news story in question is not contingent upon a celebrity or well-known figure? All social network accounts give off information trails (Bradshaw, 2009), and it is essential to approach these critically when trying to verify them. Here follow some suggestions for doing this.

All Tweets have timestamps, which can be used to trace the origins of a story, and help establish whether someone was part of this story as it broke, or if they are coming to the story 'downstream'. Contextualising every tweet you find against a users' history is also important in getting a sense of who they are; what are their interests, what language they use, whom they interact with. Sometimes people set up hoax accounts by starting to comment on unrelated topics – make sure you check fully how far back a users' comments go, and how prolific they are.

Profile information in any platform is crucial to reliability – occupation and other biographical information are always important when reading texts for parody, sarcasm, irony etc. However it is worth bearing in mind the nature of the medium. It is quick and easy to set up a Twitter account, and it is also much easier to have a one-to-many conversation with the world on this platform, than it is by using alternative social media, like Facebook. Twitter accounts with little history should therefore be taken with a pinch of salt. It is easy to tell long a Twitter account has been going,for as with most aspects of the service, someone has invented a third-party service which does this for you (such as How Long on Twitter: http://howlongontwitter.com/).

It is important to check profile information against what a user tweets, and against what else can be found online about a given individual. It can be difficult to remember different usernames and passwords for different social media sites, and for this reason many people use the same names, nicknames and 'handles' across different sites, which can in turn be searched.

In Twitter it is always helpful to source what other people are tweeting to the subject whose authenticity you are trying to check – use the advanced search command *to:(@username)* to find out (indeed many of the other advanced operators mentioned in Chapter 5 can be used to verify various aspects of a Twitter profile). This approach may give you an insight into friends in the real world too (though not necessarily), and may throw up evidence of veracity. But ultimately, directly interacting with the person behind the profile offers the most comprehensive means of proving their reliability.

Think about the age, and demographics of the profile owner (much of which is covered in Chapter 5 of this book). Does the user communicate in teen-speak, or are they middle-aged and middle-class, precious with punctuation, and formal in their grammar? Sometimes a professional will give away their trade in the technical language they use. Think about use of conventions. Twitter users know about hashtags, re-tweets, lists and other elements of the medium (such as Modified Tweets: MT), which we all have to learn. How likely is it that a novice Tweeter may use these elements?

It is always a good idea, when fishing for news in Twitter, to use an aggregator which instantly shows you the numbers of followers and followees a Twitter user has. The Twitter search option provided by IceRocket (http://www.icerocket.com/) lets you browse over these numbers (hover your mouse over the number to the left of usernames). Such tools are ideal for weeding out spam, but they are also a useful as a means of catching out online hoaxers.

Going further, to look at the names behind the numbers (of followers and followees), it is possible to see other hints about the trustworthiness of any given profile. If there are many obvious spam profiles amongst the users' follower, you can assume this user doesn't exercise quality control, and likewise, if they are following some random or unexpected accounts, you may draw your own conclusions.

There is no public, searchable directory of who owns blogs, nor blogging profiles; this would run entirely counter to the nature of the platform. However, it is sometimes possible to find out who a blogger is if they visit (and comment) on your own blog, if they have a blog of their own and use Google Analytics (the free web traffic metrics service) (Baio, 2011).

Wikipedia: curse of the obituarist

Example: Last of the summer wind-up

Following his death in October 2007, several national newspapers and media outlets ran with a surprising musical 'fact' from Ronnie Hazelhurst's impressively broad-ranging back-catalogue. Author of many instantly memorable sitcom theme tunes, including The Two Ronnies and Some Mothers Do 'Ave Em, he was, according to the obituaries columns of BBC News, Guardian online (Plunkett, 2007), the Independent, the Times, The Stage and Reuters, also author of S Club 7's 'Reach for the Stars'. Except that he wasn't – this gem was unearthed and regurgitated by many an obituary-writer, from a hoax entry in Wikipedia (Dick, 2008).

In 2009, lightening struck twice, as various official obituaries for French composer Maurice Jarre, including those for The Guardian, The Independent and The Sydney Morning Herald) carried quotes attributed to him, but which turned out to be the invention of a 22-year-old student at University College Dublin (Butterworth, 2009). Almost a year later, an unholy trinity was secured, when the authorship of '(There'll Be Bluebirds Over) The White Cliffs of Dover' was wrongly attributed to Norman

Wisdom in several of his official obituaries, including those found in The Guardian and The Mirror (Orlowski, 2010), a 'fact' eventually sourced to Wikipedia.

In regard to all three of these hoaxes (and those inevitably yet to be published), it is worthwhile considering how Wikipedia determines its frame of reference. The following, from the verifiability page of Wikipedia Policy, is instructive:

> The threshold for inclusion in Wikipedia is verifiability, not truth—whether readers can check that material in Wikipedia has already been published by a reliable source, not whether editors think it is true.
>
> To show that it is not original research, all material added to articles must be attributable to a reliable, published source appropriate for the content in question. In practice you do not need to attribute everything. This policy requires that all quotations and anything challenged or likely to be challenged be attributed in the form of an inline citation that directly supports the material. (Wikipedia, 2011)

Ultimately, if a 'fact' is found on Wikipedia which is un-sourced, it should not be taken at face value. The easiest method of verifying a reference found on Wikipedia, is to seek out the View history tab (found on the top right-hand corner in every page of Wikipedia). Look down through the most recent changes, and check to see when the change you are concerned with was made, and who made it. If it is not possible to tell from the summary provided whether the excerpt has been changed or not, it may be necessary to manually go through the most recent changes – which can be time consuming for some of the larger (or more controversial) topics. As a quick short-hand guide, it is fair to say that the more recently changes took place, the more likely they are to be erroneous.

Each change in Wikipedia is associated with an IP address; and many with usernames. Wikiscanner (http://wikiscanner.virgil.gr/) offers one way of shedding light on who has been amending Wikipedia entries is by the organisation they are associated with, though you have to search by subject or organisation (and associated IP range).

How to corroborate an email

Journalism is not just about getting out and ‘finding’ stories. Often stories find journalists, and email is one of the most convenient ways members of the public have of getting in touch. But how do you know when an email you have been sent is reliable or not? There are some useful guides available (BBC College of Journalism, 2011), and here follows some wider suggestions.

Language: Is the email written in a formal or informal way? What sort of words are used, and how are they used? How is punctuation used, and are there any obvious mistakes? In all cases answering these questions will give you an idea of how 'true' to the story your emailer is, and the extent to which they are who they claim to be.

Picture: Respond to your emailer asking them to send a picture of themselves – if they comply then they are more likely to be who they say they are. However, this means of verification can be problematic with breaking news.

People search: Where possible use a mix of proprietary and free online people-finders. In addition, think about your emailer's profile (and the context in which you are approaching them), and consider which social networks someone of that profile would most likely use, before undertaking a search. In short, it is a good idea to search for people by bearing in mind why they may be using the service. For a comprehensive list see Wikipedia:

http://en.wikipedia.org/wiki/List_of_social_networking_websites

Detail: always ask your emailer for more detail – any conflicting information may be helpful.

If you have undertaken all of these steps, and are still unsure, it can help to establish where an email has been sent from. Sometimes people will claim to be somewhere they are not, and such hoaxes can be easily spotted in a number of ways. For example, there are proprietary services which allow you to do this at a cost, such as eMailtrackerPro (http://www.emailtrackerpro.com/). But it is usually possible to check for this information free of charge. Use the following steps to verify an email sent to a Gmail account:

Log in to Gmail
Open the mail whose sender's location you wish to query
Click on the blue inverted pyramid on the right-hand side of the email navigation bar
Click on 'show original'
Look for line containing Received:from and look for IP address in square brackets (a series of four sets of numbers). If there is more than one, go through each, starting with the first; this is mapping out the route taken by each component within your email.
Ignore the domain name (which can easily be altered)
Search for IP addresses using a tracking service, such as IP-2-location (http://www.ip2location.com/).

It is important to note that this process is available in any email client which supports the viewing of email headers. Note also that emails sent from a Gmail user will not feature using this method within the Gmail client.

How to corroborate digital photographs

The sceptical amongst us know that cameras lie, or at least, photo-editing software can help to pull the wool over our eyes. Media organisations have developed sophisticated checks and balances to help verify the flood of user-generated images (BBC College of Journalism, 2009). Nonetheless, so successful have some hoaxes been, that we have a relatively new adjective for the process: to *photoshop*. And while some researchers have developed high-level software to help detect hoax multimedia (Zetter, 2007), it is nevertheless possible to use what is commonly available, when allied to sound editorial judgement, to do most of the fact-checking legwork you need.

Here follows some editorial and technical measures for use when it comes to verifying digital photography, to help journalists establish fact from fiction. If you receive an image via email, respond to the emailer asking to have a chat (via telephone). During the phone call ask the following:

What happened, what the image contains, and compare this with official sources for verification.
Who took the pictures, when and where. Check all of these details where possible

(i.e. check names on the web, check when and where with official accounts etc.). Where the accompanying email contains no text, or generic, non-specific text, then there's a good chance this is spam.

If the image contains any text or images, this should raise concerns. Beware of images contained in other files (Microsoft Word documents, PowerPoint etc.) – as they may have been doctored. Considering the image, think about the qualities of the photography – does it seem too 'professional'? If it is a series of pictures, is it logistically feasible that the same person could have taken them all?

Beyond these editorial checks, it is helpful to check the technical specifications of the images you are concerned with. You will need access either to Photoshop, or to alternative photo-editing software. Check the dimensions of the image – anything less than 2000 x 1200 psi will more than likely not be the original digital camera image, but a compressed version – and if so, you might want to ask yourself why it has been compressed and sent in this form. Of course there are all sorts of valid reasons why it may have been doctored, but never-the-less, this might give an insight.

Use your picture editor of choice to scrutinise the image. Using the zoom facility, inspect those areas of the picture where tones meet – you may see layers, or (suspiciously) straight lines, or pixelation (evidence of compression, which would imply a picture isn't in its original state).

Perhaps the quickest, and most through way of checking for photgraphic hoaxes is to use a reverse image search engine such as Tineye (http://www.tineye.com/) or Google Search By Image (http://www.google.com/insidesearch/searchbyimage.html). Tineye lets you upload an image, which is then used to scour its extensive database of images – if the image you use is found elsewhere, Tineye will show you the site – then you can check its veracity, its timeliness and other elements.

If the image concerns breaking news, use popular real-time image search engines such as Picfog (http://picfog.com/), Twitpic (http://twitpic.com/) and Yfrog (http://yfrog.com/) to see what else is being shared online.

The 'Properties' menu, accessed when right-clicking on any saved image, can give away useful information about when and where an image was taken. Creation and modification dates, GPS co-ordinates, and various other metadata are often available here, where digital cameras have been used. But it is also possible to find even more information about how, where and when an image was taken (not to mention the specification of the camera used), using bespoke software such as Exif Reader (for PCs) or EXIF Viewer (for Mac):

Exif Reader (program for PC) (www.takenet.or.jp/~ryuuji/minisoft/exifread/english)
GBO Imagehost (http://gbimg.org/)
Simple EXIF Viewer (program for Mac OS X) (homepage.mac.com/aozer/EV/)

The FXiF Firefox add-on (https://addons.mozilla.org/en-US/firefox/addon/fxif/) can also be used to check image metadata wherever you find them online.

However, a note of caution is necessary: it should be noted that this approach often yields inconclusive results. Research in 2009 showed that only 3.3% of images uploaded to various real-time image sharing services contain GPS coordinates (Flinn et al, 2010). If there is no Exif data associated with a digital image, then it is safe to say that the file was either taken with an analogue camera, or that it is a digital copy of an original, or that the camera owner has disabled some metadata capture. This doesn't mean it isn't genuine, but it does mean you can't verify its reliability without at least the help of other methods.

How to corroborate video files

As with images, all digital video is produced with operational metadata attached. As with images, some aggregators and platforms strip video of its metadata on upload, so it may not be possible to find in embedded material. But moving images can require a different level of analysis. It may be possible to find out if metadata information is available by trying any of the following pieces of (free) software:

Gspot (http://gspot.headbands.com/)
MediaInfo (http://mediainfo.sourceforge.net/en)
Video Inspector (http://www.kcsoftwares.com/?vtb#help)

At the BBC's User Generated Content Hub, staff double-check audience-submitted videos against places as they appear in digital maps, checking for shading, and other visual clues for inconsistencies (Murray, 2011).

How to corroborate electronic text documents

Example: All things to oil men...

Some of the biggest news stories start their life buried away in text documents – this is something most investigative journalists will attest to. And when these documents are in electronic form, the metadata hidden behind the words can give an insight into the origins and veracity of a document. In 2005 it was discovered that a document contributed by climate sceptic Lord (Nigel) Lawson's to a US Senate environment committee was not authored by him, but by an employee of London PR company Luthar Pendragon, who was formerly a spokesman for Exxon Mobile. This became clear after journalists checked the document's metdata (Vidal and Adam, 2005).

The 'Properties' option in either Microsoft Word or Adobe PDF Reader will betray some useful information (such as when a document was created, or modified, or where a username is associated with a file). But sometimes it is necessary to delve deeper.

There are a various pieces of software which can be used to find much more metadata; some are suitable for Mac, others for PC:

Docscrubber (program, PC) (http://www.javacoolsoftware.com/docscrubber.html)
Oracle Outside In Clean Content (program, PC) (http://www.oracle.com/technetwork/middleware/content-management/cleancontent-094288.html)

Something Clean Content divulges which Windows Explorer will not, is the file-path a file was saved in. Given that large organisations have large administrations, and that most employees have their own unique identities, it is possible to marry username and filepath, and find out who else (perhaps unnamed in the document) may have been involved in its drafting.

Once the software is downloaded, you can verify documents using Clean Content by following these steps:

Run the program.
Choose 'Select and process a single document'.
Locate document, and click 'Next'.
Click 'Generate a risk analysis of the document...', then click 'Report!'.
A browser window will open containing the program's findings.
It is possible to view changes and comments on drafts of word documents from the File menu, but only where they haven't been password protected.

References

Anderson, Kevin (2009) Fake David Miliband Twitter account dupes press, The Guardian, June 26:
http://www.guardian.co.uk/media/pda/2009/jun/26/twitter-michaeljackson-davidmiliband-hoax-journalism

Arthur, Charles (2004) Tuned in, turned on, The Independent, April 21:
http://news.independent.co.uk/sci_tech/article56861.ece

Ball, James (2010) The curious case of The London Weekly, JamesRB.co.uk February 02:
http://www.jamesrb.co.uk/?p=259

Baio, Andy (2011) Andy Baio: Think You Can Hide, Anonymous Blogger? Two Words: Google Analytics, Wired, November 15:
http://www.wired.com/epicenter/2011/11/goog-analytics-anony-bloggers/all/1

BBC College of Journalism (2011) Authenticating UGC Emails
http://www.bbc.co.uk/blogs/collegeofjournalism/skills/production/user-generated-content/authenticating-ugc-emails.shtml

BBC College of Journalism (2010) Hoax UGC Image Checks, BBC:
http://www.bbc.co.uk/blogs/collegeofjournalism/skills/production/user-generated-content/hoax-ugc-image-checks.shtml

Bradshaw, Paul (2009) How to spot a hoax Twitter account, Online Journalism Blog, October 16:
http://onlinejournalismblog.com/2009/10/16/how-to-spot-a-hoax-twitter-account-a-case-study/

Butterworth, Sibiohan (2009) Open door: The readers' editor on ... web hoaxes and the pitfalls of quick journalism, The Guardian, May 04:
http://www.guardian.co.uk/commentisfree/2009/may/04/journalism-obituaries-shane-fitzgerald

Cooke, Alison (2001) A guide to finding quality information on the internet: selection and evaluation strategies (second edition). Library Association Publishing: London.

Dick, Murray (2008) Is the future in bits? BBC News Magazine, December 03: http://news.bbc.co.uk/1/hi/technology/7761153.stm

Flinn, Michael, Teodorski, Christopher, Paullet, Karen (2009) 'Raising awareness: an examination of embedded GPS data in images posted to the social networking site Twitter', Issues in Information Systems, Volume XI, No. 1, 2010, pp432-438: http://www.iacis.org/iis/2010_iis/Table%20of%20Contents%20No1_files/432-438_LV2010_1370.pdf

Friend, Cecelia and Singer, Jane (2007) Online Journalism Ethics, Traditions and Transistions. ME Sharpe: New York.

Guardian PDA Blog (2009) Fake David Miliband Twitter account dupes press, The Guardian, June 26: http://www.guardian.co.uk/media/pda/2009/jun/26/twitter-michaeljackson-davidmiliband-hoax-journalism

Hanas, Jim (2005) UPDATE: Toothing Corrections? April 04: http://www.hanasiana.com/archives/000349.html

Kelly, Chris (2004) Biting into the new sex text craze, BBC News Online Bristol, May 07: http://news.bbc.co.uk/1/hi/england/3673093.stm

Miller, Stephen (2004) The Miller Internet Data Integrity Scale, New York Times: http://courses.ttu.edu/rreddick/ar/tools/MIDIS_handout.pdf

Murray, Alex (2011) #bbcsms: BBC processes for verifying social media content, BBC College of Journalism Blog, May 18: http://www.bbc.co.uk/journalism/blog/2011/05/bbcsms-bbc-procedures-for-veri.shtml

Orlowski, Andrew (2007) Braindead obituarists hoaxed by Wikipedia, The Register, October 03: http://www.theregister.co.uk/2007/10/03/wikipedia_obituary_cut_and_paste/

Orlowski, Andrew (2010) The Curse of Wikipedia strikes Norman Wisdom, The Register, October 06: http://www.theregister.co.uk/2010/10/06/wikipedia_norman_wisdom/

Plunkett, John (2007) Ronnie Hazlehurst - his greatest hits (obituary), The Guardian, October 02: http://blogs.guardian.co.uk/organgrinder/2007/10/ronnie_hazlehurst_his_greatest_1.html

Wells, Matt and Ramesh, Randeep (2004) BBC reputation hit by Bhopal interview hoax, The Guardian, December 04: http://www.guardian.co.uk/uk_news/story/0,3604,1366397,00.html

Timmer, John (2009) Wikipedia hoax points to limits of journalists' research, Ars Technica, May: http://arstechnica.com/media/news/2009/05/wikipedia-hoax-reveals-limits-of-journalists-research.ars

Vidal, John and Adam, David (2005) Eco Soundings, The Guardian, October 19: http://www.guardian.co.uk/society/2005/oct/19/environment

Wikipedia (2011) Policy page outlinging significance of verifiability: http://en.wikipedia.org/wiki/Wikipedia:Verifiability

Zetter, Kim (2007) Researcher's Analysis of al Qaeda Images Reveals Surprises — UPDATED, Wired.com, August 02: http://www.wired.com/threatlevel/2007/08/researchers-ana/

APPENDIX IV: INTERACTIVE INFOGRAPHICS AND NEWS VALUES

Murray Dick

This study is concerned with the news values and working practices that inform the creation of interactive infographics in UK online news. The author draws upon organisational theory in journalism studies, and considers how conventional journalistic news values compare with best practice as espoused in different literatures within this field. A series of open-ended, depth interviews with visual news journalists from the UK national media were undertaken, along with a short-term observation case study at a national online news publisher. Journalistic and organisational norms are found to shape the selection, production, and treatment of interactive graphics, and a degree of variation is found to exist amongst practioners as to definitions of quality in this field. Some news stories are considered to be better suited to rendering in interactive form than others. The availability of 'big data' does not drive decision making in itself, but some numbers are considered more newsworthy than others. Budgetary constraint drives practice and limits potential in this field. Risk aversion, embodied in various forms; from the use of templates, to a perceived need to avoid audience complaint, is found to dampen experimentation. Detailed audience research was found to inform the choice of methods used in data visualisation at one national news producer. This warrants further investigation as to how audiences engage with news interactives, and what the framing of news in certain (preferred) data visualisation formats means in terms of how news is understood.

KEYWORDS: interactivity, interactive graphics, interactives, infographics, news values, online journalism, digital journalism, visual journalism

Introduction

While interactive online journalism has been the subject of much analysis in contemporary communications studies (Boczkowski 2004; Bruns 2005; Deuze 2008), to date our understanding of the field of interactive visual journalism has been limited to output studies; including typological studies (Schroeder 2004; Quandt 2009), an audit of pre-existing materials in the context of newspaper workflow management routines (Giardina and Medina 2012) and a games-centric, end-use-focussed review (Bogost et al. 2010). But what of those individuals who create these innovative interactive story-telling devices? How they think and work is little understood in the literature.

The growing status of interactive graphics on our news and on the news profession more widely is evident within and beyond journalism practice. Today the Guardian has two interactive teams, one of which is concerned uniquely with its US audience. The BBC has a News Online Specials team, and The Times has a new Visual Journalism unit. Journalistic excellence in this field is officially recognised in the form of new professional awards such as The Data Journalism Awards (established in 2012 by The Global Editors' Network), and in new categories within existing industry awards, such as 'Digital innovation' in the Amnesty International Media Awards and 'The Digital Award' within The British Press Awards. There is a growing recognition of the potential for innovation, originality and excellence in the field.

The production of interactive visual journalism often involves bringing professionals from statistical, information design, and computing backgrounds into the newsroom. Do these professionals bring influences that in turn affect how interactive stories are selected, treated and published? Interactive news graphics inform the public sphere and the functioning of society just as traditional print and broadcast news formats do, so the professional norms at play are pertinent in any discussion of news values.

This exploratory study is concerned with establishing the routines, processes and relationships that shape the interactive news we consume. More specifically it aims to address a series of interconnected research questions:

What structure do interactives teams take, and what role and purpose do they serve in the modern newsroom?
How are news interactives ideas selected?
What creative processes are involved in the selection and treatment of news interactives?
What values inform the creation of news interactives?
Where does the audience factor within this creative process?
What distinguishes ‘good’ from ‘bad’ interactives?

This research is informed by three strands of scholarship; organisational studies of the newsroom, interactive online journalism studies, and a composite overview of best practice in interaction design and data visualisation (outlined in the *Methodology* section).

Definitions

The term 'interactives' is used throughout this study to describe news outputs in the sense set out by Nichani and Rajamanickam (2003): the purpose (or function) of these digital products is to facilitate explanation through interaction. The term is appended with 'news' simply to delineate the source of origination – ‘news interactives’ are created within the news industry, in a context of news production. It is acknowledged, however, that this is definition is coming under increasing pressure: for example, the ‘Hacks and Hackers’ social events, established in Massachusetts in 2009, and the Data Journalism Handbook (Gray et al. 2012) have problematized what it means to create interactive news, and what it means to be a journalist in this field.

Organisational studies of the newsroom

Journalists, it has long been argued, develop 'inferential frameworks' (Lang and Lang 1955), predicated on prevailing socio-cultural norms and developed through the course of experience in employment, which in turn informs the process of story selection, and the assembly and 'angle' of news. But as Schudson (1997) points out, this conception of the journalist tends to leave the information of which news is composed 'sociologically untouched'; as if it comes pre-prepared, rather than being subject to the process of negotiation.

Some scholars have observed that conflict has a normalising influence on the production of news by journalists (Bantz, in Berkowitz 1997). Conflict and the competitive effectiveness of a specialism can set the status of some professionals in the newsroom. Tunstall (1971) showed that TV news and current affairs producers have significant autonomy, and that autonomy is granted as a result of a valued specialism.

Many early sociological studies of news work converge around the importance of the deadline as an organisational norm, and a determining factor in how news is 'manufactured', a theme which dominated the field studies of the 1970s and 80s; (Epstein

1973; Schlesinger 1987; Tunstall 1971; Golding and Elliott 1979). But where space is boundless, and where deadlines are not constrained within a fixed publication schedule, such issues cannot continue to be considered to direct our understanding of the inclusion and/or prioritisation of certain stories in news production (Franklin 1997).

Influenced by scholarship in the sociology of work, several studies by Wilson Lowrey (1999; 2002a; 2002b; 2003) explore how the news we receive is shaped by the conflicting norms held by competing sub-groups within the fluid setting of the news room. Lowrey found that the demands of organizational need are not the only determining factors at play in news production; professional norms and values are important too, particularly where professionals from a range of backgrounds come together to produce the news. Lowrey's approach allows us to move beyond the management structures and ideological norms that shape behaviour, to see the autonomy exerted by individuals over the news we consume. But this approach is premised on a binary conflict between 'word' and 'picture' people from a pre-converged era. For the exemplar in the present study, we must consider the 'data person', and alternatively the statistician (or data journalist), the data visualisation expert (or graphic designer), and the computer programmer, all of whom may have their own pre-conceptions of what constitutes newsworthiness in data, and all of whom may exercise some degree of influence on the interactive graphics produced in today's news online. For this reason, in addition to establishing the role and functioning of interactive teams in the contemporary newsroom, the author also seeks to identify the extent to which 'interactive norms' exist within these professions, such as whether there is:

A common rationale to justify the creation of interactives;
Routinization of work processes (that may speak to organisational norms in this field)
Consistency in approaches to best practice in data visualisation (such as a common position on the 'chartjunk' debate in the literature), and;
Consistency in approaches to best practice in human-computer interaction (such as a common position on the significance of user-centredness, and the factoring of emotion into interaction design).

The rise of interactive online journalism studies

Interactive online journalism has its own field study literature, which in turn draws upon the organisational journalism studies literature of the past (Boczkowski 2004; Bruns 2005). However, interactive visual journalism has more often been researched from an 'outputs' perspective, which tells us little about the decision-making that goes into this field of news production. Schroder's (2004) survey of interactive infographics shows that the UK media lagged behind their European counterparts in this field, over the last decade. Similarly, Quandt (2008) pulled together the dominant themes in late 1990s literature on the formal and structural properties of news content online, to show that online media had an over-reliance on 'shovel ware' from print.

In the middle of the last decade, the rise of the network brought a new utopianism to journalism studies literature, bound up with the notion that "...the profession would have to articulate an equilibrium between its operationally closed working culture strictly relying on a ruling elite of 'experts', and a more collaborative, responsive, interactive or even dialogical journalistic culture" (Dueze 2003, 219). Supplanting the classic 'gatekeeping' role of the news journalist in the networked age, Bruns' theory of the 'gatewatcher' (which takes after Gans' 'multiperspectivality' [1979]) has at its heart a long-established US communications ideal: the marketplace of ideas. But this approach belies a significant reality: just because 'the former audience' now has access to (some of) the news source material journalists used to monopolise, does not mean that they have the time nor inclination to create or disseminate their own news.

A range of literature on the very general theme of interactivity in online news production has some bearing on the present study, but only at a tangent. These can broadly be classified as either functional studies (Kenney et al. 1999; Rosenberry 2005; Shultz 1999; Tankard and Ban 1998), concerned with how interactive elements in online news are presented; and perceived studies (Chung 2007; Larsson 2012) concerned with how interactivity is understood by the audience.

Methodology

This research employs two methodologies; a short-term observation case study, and a series of semi-structured, depth interviews. All subjects are visual journalists working in the UK news media, but they come from, alternatively; journalistic, graphical design, and programming and development professional backgrounds.

In total eight hours of observations were undertaken with the BBC News Online Specials Team over two separate days: Wednesday 15 August 2012 and Wednesday 22 August 2012. This team create interactive and non-interactive infographics that are published on the BBC News website. This team has editorial and budgetary autonomy within the news organisation, and take responsibility for management of their own workload. During observation, the author attended editorial and team meetings, and both observed and interacted informally with journalists, designers and programmers working across the team (involving ten staff in total).

Seven semi-structured, depth interviews were undertaken with six interactive journalists working across the UK national media. Two interviews were undertaken with Channel 4 News Online's Head of Infographics on Friday 27 July 2012, and Tuesday 16 October 2012, with further questions answered via email exchange between these dates. Two interviews were undertaken with the editorial lead of the BBC News Online Specials Team on Wednesday 15 August 2012 and on Wednesday 22 August 2012. A joint-interview with The Guardian's (now former) Head of Graphics, and a long-serving graphic and interactive designer was undertaken on Monday 23 July 2012. An interview was undertaken with former head of The Daily Telegraph's graphics team and current Director of the Society of Newspaper Design Region 15, on Thursday 30 August 2012. An interview with Head of Interactive at the Financial Times was undertaken on Friday 22 February 2013.

In total, these interviews comprised ten hours. Recordings were made, and notes taken during observations and interviews, and findings were summarised from all notes in around 12,000 words. The subjects chosen for this research were not selected at random – the field of interactive news graphics professionals in UK news is small, and the author was guided primarily by access.

Questions asked during observation and interviews covered many areas, including:

How do you decide what stories to cover?
What does the process of creating interactives involve?
How do you maintain the quality of your work?
Are some stories better suited to interactive graphics than others?
Do you work with some newsdesks or teams more than others?
Are some numbers more 'interesting' than others in terms of producing interactives?
Is there an ideal news subject/story type which particularly lends itself well to rendering in interactive graphics?
Do you have any assumptions about your audience which inform your interactives?

How to do you measure the success of your interactive graphics?

Questions were developed iteratively, and framed according to subjects' experience of terminology and process. Best practice in interaction design was derived from an international standard (ISO 2009), and from the literature, especially as regards the foregrounding of emotion in interactive design (Forlizzi and Battarbee 2004; Saffer, ed. 2010). Best practice in data visualisation was drawn from the works of Brinton (1914); Tufte (1983); Holmes (1984) and Tukey (1990) as well as a British Standard on the presentation of tables and graphs (BSI 1992). The debate over the merits of visual embellishment in data visualisation, or 'chartjunk' (Bateman et al. 2010) is revisited here, as is Tufte's 'Doctrine That Statistical Data Are Boring' in the newsroom. Journalistic norms are derived from work in print, broadcast and online journalism cultures; and are bound up with those conventional ideas of accuracy, fairness, detachment and objectivity long established in professional codes of conduct, and long critiqued in journalism and communications scholarship.

The weighting of responses is broadly proportionate to the range of participants, but reflects individual concerns. Thematic analysis was undertaken using a contextualist approach, with regards those themes found in the 'best practice' literature. This study takes a grounded, theory-building approach – it is acknowledged that the conclusions drawn here are not generalizable beyond the sample, though it is hoped that this research will inform future, more systematic studies.

Findings: observational study

The early morning editorial meeting for BBC News Online (held at 9:30AM) comprises members from each ‘index’ page, chaired by a central news editor. The meeting I observed started with a summary overview of the previous day's stories and audience statistics, then a general overview of the day's main stories. Each index page representative then explained their top two or three stories for the day. A home affairs story concerning Asil Nadir's court case was raised, a story for which the Specials team developed a (non-interactive) graphic:

http://www.bbc.co.uk/news/uk-19161940 (retrieved August 2013)

The Specials team representative talked through their two pieces of work for the day (the graphic above, and an interactive in preparation concerning youth unemployment). The meeting is declarative, descriptive and not exploratory or creative, and is expressly not about decision-making – this happens later via direct email or phonecall between interested parties, away from group dynamics. Everyone listens, but there is no real show of reaction. Later John Walton (deputy head of the Specials team, who represents at this meeting) explains that the selection of stories and methods of coverage used are driven by resources, and by potential for impact. Producing interactives can be time-consuming and resource-intensive, so ideas need to be feasible and credible. The team usually require contribution from subject specialist journalists on large projects. The Specials team are more likely to accept the offer of work on a story (a commission) where that story is likely to rise up the index pages to UK or World pages (the dominant pages in BBC News Online).

Project teams are tripartite, comprising data journalists, designers and developers, in numbers reflecting the volume of work anticipated – data journalists are ultimately tasked with editorial decision-making. At any one time one designer will be allocated to working on short-turnover projects while another will concentrate on features.

John spends much of the late morning reading a government document on youth unemployment: *Global Employment Trends for Youths*. He also spends considerable time on the phone (around forty minutes) speaking to an expert in the newsroom about alternative sources for data on this topic. One of the designers is re-working a satellite view map of Syria, where the Golan Heights hadn't been displayed properly. It is thought the map had been rushed because the previous day had been busy (and the team under-staffed), and subsequently questions were raised about the choice of colours used and the use of geographic space, which didn't give a clear impression of borders. The team spend a significant amount of time testing their output, and comparing it with the output of competitors. Bella (Hurrel, editorial lead) shows me a graphic produced two days previously by an American competitor, concerning charitable giving. Its strengths and weaknesses as a story-telling device are discussed informally with colleagues, in person and via email.

In keeping with all other news teams, the Specials team are inclined to develop user generated content in their output – one such 'high-end' example is given: Rover Panorama, a 360 degree visualisation:

http://www.bbc.co.uk/news/science-environment-19296006
(retrieved August 2013)

In this instance the Specials team worked with non-BBC developer Andrew Bodrov, who (using images sourced directly from NASA), created a sophisticated montage of space exploration. The Specials team added data points, in order to render the collection interactively. Rights issues have forced the team into acts of creative ingenuity. One of their developers has created a bespoke Google Chrome browser extension which allows for quick conversion of raw data into maps, making this tool available to BBC journalists via their staff intranet. This was done through necessity, due to prohibitive rights issues concering use of digital map imagery, and to bring technical support for the system in-house.

The Specials' recently appointed acting head of development said his news values come from BBC audience values, around accessibility and diversity. He mentioned that one of their current projects had raised ethical dilemmas around the sourcing of data. He received a large amount of proprietary data from an organisation who in return wanted publicity around the issue of music piracy. The team were planning to create an interactive area-map of music piracy across the UK. He said they had to make difficult decisions about this, and be careful not to use the data in a way that may work to the commercial advantage of the source who provided the data. He sees journalists as 'customers' – he works in web development first and foremost. He explains that there is some editorial over-lap between project managers and data journalists; suggesting the rise in tensions caused by the increasing fusion of form and content in online journalism.

Findings: interviews

Interviewee details

A Channel 4 interactive news specialism has existed since 2010. Ciaran Hughes' career background is in newspaper graphics.

The BBC Specials team has existed since 2003. Today it comprises eight journalists, four developers and four designers. Bella Hurrel's background is in arts journalism.

The Guardian's interactive team was created in 2001. Paddy Allen was a founder member of the team, and his background is in newspaper graphics. Michael Robinson's background is in newspaper design.

Michael Agar's background is in infographic design across various national newspapers. He was formerly head of graphics at the Telegraph between 2008 and 2012.

The FT Interactives team has existed since 2005. Today it comprises 9 members, equally weighted between data journalists, developers and designers. Emily Cadman's background is jointly in magazine and news journalism.

The status and function of interactives journalists within the newsroom

All participants are involved in editorial decision-making at their respective organisations. Drawing comparison with colleagues who work in TV graphics, Ciaran explains that his work is far more involved editorially, in generating ideas and in treatments: "I differ from the TV guys, in graphics; I won't react to a script – I'm much more involved from the very start in editorial and not just in deciding what we do and how we do it do, but in suggesting subjects, I'm involved in visual journalism" (Hughes, Ciaran. Interview by author. Digital recording. London, July 27, 2012). In the past, when working in newspaper graphics Ciaran's work would often provide support to the story of the day. Personal politics influence roles – some duty editors are more sympathetic to graphics (and interactives) than others, often depending on their own journalistic background. This is true of the BBC too – Bella explains that the placing of interactives within stories can in turn lead to disputes across departments, driven by scarcity of resources. Bella prefers the Specials team to create self-contained interactives rather than visual accompaniments to other team's news stories, because this involves retaining full editorial control. This approach creates impact; it yields page impressions, social media 'shares', and various other data (including interaction data), that can be directly attributed to the team's work. This audit trail can in turn re-enforce the team's editorial integrity and independence.

Emily explains that over the past three years it has become easier for her to express the importance of the interactives team at the FT as more and more journalists have become interested in interactive graphics as a journalistic medium, and as technological advances have made it easier to do their job. Similarly, Paddy explained that he finds it easier to convey the team's significance to the Guardian's editors today (having moved from print graphics to interactives), but that this has been a slow process over a long period. Seven years ago it had been very difficult, he explains:

> I used to have to rush around quoting all the time 'your story is 70%… more likely to be read if it has a graphic next to it', and it was a desperate attempt to talk people into why you should have these things… and quite often they would come up with briefs that were spectacularly inappropriate, but because they were higher up the food chain than you, they would want it like that (Robinson, Michael and Paddy Allen. Interview by author. Digital Recording. London, July 23, 2012).

Michael Robinson feels the main aspect which has changed such attitudes is the process of bringing designers and art directors into newspapers, a process which has loosened sub-editors' control of the visual space in news. He explains that his job was originally created (in 2002) in order to address tensions between the 'word' and 'picture' professionals at the Guardian. Poor time restraint, poor briefs, and poor scope were all considered to undermine graphical and design potential. He feels that the interactives team have benefited from these past arguments and resolutions. Paddy says that a positive consequence of working online is the potential to sit down and get on with work:

The demands for the web are very different in that you have limitless room, so you haven't got Adverts who might come in and say 'you haven't got room for this anymore'. It is far more open, in that if any of us has a half-decent idea, they will say 'yeah you can go ahead with that', and you can start there and then, but the paper will say 'well that might be a good idea, if we have room' (Ibid).

This in turn is important in terms of building self-confidence within the team, and professional cache outside the team - in print there is an ambient threat that a day's work may be discarded on account of the arbitrary intervention of a news event.

The BBC Specials' long-term strategic aim is to move beyond providing content for different editorial units, towards provide journalists with their own tools (an approach driven jointly by increasing interest in news interactives, and by tightening budgets). Similarly, Michael Agar's experience at the Telegraph involved building applications for journalists. Doing so allowed his former team to concentrate on longer-term projects, while empowering their non-specialist peers.

Who decides what stories become news interactives?

Commissioning in the BBC Specials team is approximately 40% internal (where a team member is commissioned to create infographics by the Specials editor) and 60% external (where other teams and departments engage the Specials editor with work), though this is prone to vary: “Sometimes it's more of a conversation” (Hurrel, Bella. Interview by author. Digital recording. London, August 15, 2012). Data journalists will select a newsworthy topic, or a large dataset, or at planning or commissioning meetings ideas will be raised. The team do their own newsgathering, but they don't have the budget to undertake investigative work: “If you have a big data set there's no point in going fishing in it... because you can waste a lot of time” (Ibid). On major stories the team are guided by experts, both in-house (in the form of correspondents and section editors), and externally by the Royal Statistical Society or the Office for National Statistics. At the FT there is a mixed approach to commissioning. Simple commissions that can be done in a day (such as timelines, or profile pictograms) are readily undertaken, but larger projects will only be started upon if they are likely to be carried on a main index page, or if the team stands to learn something from doing the project (which will, it is felt, in turn improve or streamline future coverage). Unlike the BBC Specials team, there is a propensity (and indeed a perceived necessity) to undertake 'fishing' exercises in large data sets, and so undertake original investigative work.

Paddy explains that the commissioning process at the Guardian is very open, and that some stories generate interesting commissioning situations. At the time of writing the political crisis in Syria was unfolding quickly, and commissioning has largely being driven by the live-blog of the story. Paddy had been working on a live interactive graphical map – updated day by day, and showing the changing state of conflict. This approach was in turn influenced by reporting decisions made during the Arab Spring the previous year, as Michael observed: ““if the coverage has worked in one part of the world on a certain aspect, then that is exactly what will kick off all the ideas for the next news event” (Robinson, interview, July 23, 2012).

News values: what to cover and how?

The values that inform Ciaran's approach to deciding what work to undertake, come from years of working in news rooms. The commissions he seeks (and accepts) are ultimately sanctioned and decided higher up the editorial chain, and his work (both interactive and non-interactive) is a product of this decision-making process. Similarly, Michael Agar's news values stem from his experiences of pitching ideas to news editors. His is a utilitarian approach: "I can't understand why people would spend time cleaning and processing data if there were no story in it" (Agar, Michael. Interview by author. Digital Recording. London, August 20, 2012). Bella also comes from a traditional journalism background that informs her news values, but she observes that the culture of developers and designers does have an effect on the team's collective approach. Editorial decision-making remains constant (structurally top-down) but non-journalists tend to question orthodoxies more: these professionals often have a different working culture and a different understanding of deadlines: "they question basic assumptions about the importance, and timing of coverage – they create a debate" (Hurrel, interview, August 15, 2012).

What is the basic justification for an interactive?

Ciaran feels that interactives work well on news stories and topics that contain comparisons, and they can be particularly effective on stories concerning many parties and their interconnectedness. He gives an example of an animation he put together on the financial crisis at Rangers Football Club in May 2012 that was repurposed into a video narrated by chief correspondent Alex Thomson, and then re-published on YouTube:

http://www.channel4.com/news/rangers-complex-connections-explained
(retrieved August 2013)

Interactives take one of three forms at the BBC Specials team:

'Big data' stories, concerning stories buried in large data sets…non-statistical process visualisations that help the audience understand an issue …and personalised information generators, such as calculators (Hurrel, interview, August 15, 2012).

Emily suggests that interactives that allow the user to explore something that affects them directly, or that encourage the exploration of complicated major topics are most important at FT.com. Alternatively, those stories which do not generally work so well as interactives are those "to go with a piece where there is a lot of reportage – a good, written piece based on interviews, anecdotes or interpretations" (Cadman, Emily. Interview by author. Digital Recording. London, February 22, 2013).

Are some numbers more newsworthy than others?

Ciaran thinks that some numbers are more interesting than others, and that some visualisations of information can work counter to conventional news values: "quite often when you visualise (data for a story), (the visualisation) goes against the story... a 2% increase or a 5% increase...you show it and it looks incremental" (Hughes, interview July 27, 2012). Sometimes the numbers don't look particularly newsworthy, but by plotting data over a longer timeframe, or by bring in another variable, a news story may emerge. Sometimes Ciaran works backwards (chronologically) from the numbers he is given, and finds stories to report on that are often political in nature (challenging the claims made by politicians). These are often published on the 'Fact Check' blog:

http://blogs.channel4.com/factcheck/
(retrieved August 2013).

Michael Agar said:

Most stories are designed to fit a headline anyway, so if the statistics are boring then you have no story, unless, you are hanging it on a story where a politician saying figures are amazing (when in fact nothing has changed), then that is the story" (Agar, interview August 20, 2012).

Bella said that 'outliers' often make for good stories, as do stories which involve processes or personalisable information (such as school league tables). At the Guardian there are competing views. Michael Robinson is wary of walking into a semantic trap in regards to the newsworthiness (or otherwise) of certain numbers: "The method you are using in difference or comparison may be wrong, it isn't the numbers, if it is your treatment of them. You could say 'there is no graphic in that' – but if it is relevant that it is flat, stability can be the issue. But it might be boring..." (Robinson, interview, July 23, 2012). Paddy argues that the outlier often makes things more newsworthy as "a crude rule of thumb" (Ibid), but Michael interjects: "that can be naughty – who spent the most, who spent the least… but if you are looking at statistics, you should chop off the extremes, when it comes to representative data" (Ibid). Taking a different approach to the other respondents, Emily explains that the FT's news values (and audience interest) is very much bound up in numbers which may appear 'boring' to some audiences; such as GDP figures that may change little, or that are rarely surprising, but that these are nonetheless the foundation for policy and political issues which affect everyone, and so are therefore crucial to the FT.com audience.

'Good' infographics

Bella says that the BBC Specials team are somewhere in the middle of the spectrum in terms of the 'chartjunk' debate: "we talk about it a lot – in terms of accessibility, feel, engagement we try to go for [David] McCandles, but it has to work - it has to be accurate… complaints would make it impossible [to experiment too freely]" (Hurrel, interview, August 15, 2012). Michael Robinson advocates Tufte's theories, but acknowledges he is "a bit boring… If you want a bible up there I would put his work above others, the others are more like wallpaper" (Robinson, interview, July 23, 2012).

Michael Agar is a fan of Tufte, but is more mindful of the work of Alberto Cairo, adding "I'd like to think you can back the academia up with a bit of fun" (Agar, interview, August 20, 2012). He believes emotion has an important function in news interactives, albeit in a pervasive way: "I think visually, we don't need to be as extreme with the emotion (as those writing the words) there still has to be an element of human interest there... but I can see where the conflict lies" (Agar, interview, August 20, 2012). Ciaran is influenced by Tufte, whom he describes as being "like an engineer – his work has a beauty in itself. Good design has an innate beauty, elegance" (Hughes, interview, July 27, 2012). Ciaran recognises the importance of emotion, but conceives of it as a first order component of the medium – he doesn't think there need be an impasse with objectivity (as is manifest, for example, in balance) in journalism:

If I'm telling you a story, the words I choose, the way I speak, the emphasis – it's all a matter of emphasis, it's always there – in any communication it's always there. The thing about engaging – it's actually like David McCandles (Information is Beautiful). It doesn't matter, it might be worthless – but if it is beautiful you will look at it, and realise it's worthless. If it doesn't engage you, if you don't get instant satisfaction (which is very important) then you will never get to the second layer: 'what is this showing?' (Hughes, interview, July 27, 2012).

'Bad' infographics

Ciaran explaines that an inability to grasp the difference between an information graphic and an illustration has been, in his experience, the source of bad interactive and infographic ideas in the past. According to Bella, some general issues may lead to bad commissions, for example over-prescriptiveness on the part of other teams and journalists in the corporation (commissioning an interactive that will enhance a story to make it more successful, but not to explain or explore the story or data better). Similarly, she explains that journalists (and readers) *like* maps, but the geographical location of a story is not always central to its significance, and the use of satellite imagery in maps for visual impact isn't always the most appropriate representation. Maps are nonetheless used to add verity to stories, to give the audience a sense of rootedness in their own space and geography, to connect them with news events, even when this is not appropriate to the story being told.

Michael Agar says that he had engaged in some bad commissions in the past in order to serve a purpose in the newsroom. He recalled a story comparing the cost of bread in the 1940s and today – involving a series of pictorial slides. This commission, from a senior news editor, served the purpose of encouraging the print newsroom to make more use of infographics as a means of online-only story publishing. Emily has been asked to undertake 'bad' commissions at the FT, but they've never been published. Sometimes she has questioned whether some commissions represent the best use of her time.

Newsroom relations

Ciaran does not work particularly closely with any other teams in the newsroom, though in the past, and as a non-interactive visual journalist, he spent much time with business journalists. Bella explains that a number of subject areas are more naturally aligned with the BBC Specials team than others, as much due to established relationships between staff, as by sympathetic subject matter or source types (she lists Science, Technology, Home Affairs and Business as regular commissioners). Paddy says that the environmental journalists at the Guardian have been particularly good partners in the development of interactives, because many of their stories are inherently visual. Michael Robinson is mindful of administrative issues that can drive workloads too – he mentions that his graphics team used to do very little for the Features desk on the Guardian's G2 magazine, but since they started attending the same meetings, they have started working more closely together. At the FT, the interactives team work most with the Investigations Unit, and with Public Policy and World News desks (the latter require a lot of maps to explain geo-political and globalisation stories, and key members of this team are enthusiastic about multimedia in general).

Interactive news style and audience

Channel 4 News Online have a style guide that dictates text (titles, capitalisation etc.), and a colour guide (created by Ciaran). Ciaran is mindful that audience assumptions are imagined assumptions, and he is cynical about the value of using focus groups (and user testing). At the FT most interactives are user-tested on spouses and colleagues around the interactives team – but not on the public. Nevertheless, the audience (and not functions, or systems) are central to development there.

The BBC Specials team use templates for regular story-types. Because they must appeal to all sections of society, Bella explains they try to make their work as simple as possible without being patronisingly reductive. Focus groups (with members of the public), guerilla testing and even long-term ethnographic research has been undertaken to find out more about their audiences. They have found that a portion of the general public are intimidated by charts of all kinds, because they remind them of unhappy memories of studying mathematics at school. Equally, some audiences find circles more visually appealing (they are less tainted by experience, or feelings of confusion or inadequacy). At the BBC it is considered vital to incorporate audience research into their decision-making and processes. Page impressions, volume of clicks, and levels of interaction on pages containing their interactives are all measured and analysed: the Specials team have been able to iterate their work based on this data. Most of the findings are 'common sense', but they can inform selections and routines in practice. A general rule they've found is that in taking a prescriptive narrative approach, more readers are likely to complete all stages of the interactive – so this approach is often preferred to experimental interactive graphics. The choice of degree of interaction employed is driven by perceptions of audience, not by an idealist notion of interactivity. BBC online style is set out in a Global Experience Language (http://www.bbc.co.uk/gel retrieved July 2013).

The Guardian have a graphical style, a colour palate and interactive templates that are driven by functionality. They have five different timeline types for different scenarios. The Guardian undertakes a good deal of user research, but little of it is presented directly to the Interactives team: “Sometimes you are so immersed in a culture that you take it on board via osmosis... this is probably a very bad thing...” (Allen, interview, July 23, 2012). Michael counters that “the assumptions about fitting content to an audience can be a big mistake, the audience are merely using a medium” (Ibid). Focus groups have been used, and the team receive feedback directly from readers. They undertake paper prototyping and user design on major projects. Responding to reader comments is considered to be more useful than market research. Michael is worried that the availability of interaction data may change the way they do their work, for the worse – he is wary that the availability of this data may bend their values in selecting and treating stories for interactives in a more a populist direction.

The FT interactives team have a basic style guide and colour palate (to which, from time to time they negotiate additions). They also use various templates, and will on occasion give access to these templates to particular journalists (they will also informally train interested journalists in various skills). They measure performance in terms of conventional web analytics, but they do not as yet collect interaction data (albeit Emily would like to). Feedback comes directly from readers, and both formally via an editors' letter every Friday night, reviewing the best story from the entire week's operation, and informally from colleagues around the organisation. Given the highly-educated (and numerate) nature of the FT audience, Emily does not feel constrained by complexity when it comes to designing interactives.

Analysis

Those driving the creation of interactive graphics in our news “talk the talk” (Lowrey 2002a, 419) of professional journalism because many of them come from journalistic, rather than data, graphical or programming backgrounds. As such those conventional journalistic values concerning accuracy, fairness, detachment and objectivity dominate the way in which work is undertaken in this field, in the UK media. Professional values from other domains find voice, concerning attitudes to deadlines, and the questioning of conventions and reasoning around coverage or treatment of news stories. The notion of journalists as ‘customers’ is held by some development professionals within this field – which in turn poses further questions about the nature of interactives journalism, and its place in the newsroom. But for the most part, non-journalistic professional values are dampened by the culture of the newsroom, and the mix of positivist journalistic and organisational values.

Interactives journalists do not prioritise pleasing other journalists over the efficacy of their own work, and evidence of a “service-department” (Lowrey 2003, 138) mentality, as found in non-interactive graphic work, was not found here. Today’s interactives professionals are becoming more self-confident, more creative, and less 'auxiliary' than their non-interactive predecessors in UK newsrooms. These journalists seldom compromise their professional integrity on account of what they perceive to be 'bad ideas', or the perceived innumeracy of others in the newsroom.

Where conventional journalistic norms do not dominate, organisational norms prevail. This is found most clearly where professional norms are not fixed. For example, not all interactive teams comply with best practice in terms of user-centred design. There is some variation in the perception of what constitutes ‘good’ interactive graphics, within the wider debate concerning ‘chartjunk’. Most participants recognise the merit in Tufte’s work and values, but it is also clear that the importance of visual impact in the sense set out by Tukey, represents a potential clash with Tuftian minimalism. Some practioners see themselves as working somewhere between the minimalism of Tufte and the experimentalism of David McCandles. Some numbers are considered more newsworthy than others – an approach that accords broadly with Tufte’s positivist stance: “If the statistics are boring, then you've got the wrong numbers” (Tufte 1983, 80). Ultimately budgetary constraint casts a shadow over practice and potential in this field. The development and use of templates serves to de-specialise the specialist. Risk aversion, embodied in the fear of audience complaint, and in the use of detailed, prescriptive audience research, dampens experimentation.

But on the other hand, budgetary constraint also gifts interactives teams with a powerful bargaining mechanism within the editorial structures of the media organisations in which they practice, allowing them to challenge orthodoxies, and contribute to the news agenda. Today's interactives teams not only provide education and guidance, but produce bespoke software and provide ad hoc training for non-specialists to do their own work – their authority is based not just upon how they do their own work, but on how they can facilitate others to more effectively and efficiently do their work too.

There is no agreed framework amongst interactives practioners on what constitutes a justification for the production of a news story in interactive form, nor on what forms are appropriate in particular news story scenarios, though it is clear that the issue of quality (however conceived) is a driving factor in practice. The competing views and theories presented in the Findings here offer a very different perspective from those theories and typologies that are focussed on end-use (Bogost et al. 2010).

The rise of the network can be said to account in part for the raised status of these professionals, and this new journalistic form. Some reach out to the audience as co-producers and involve them in design. On the other hand, audience behaviour is measured, and use is made of audience metrics towards informing future ideas for coverage. Clearly some interactives journalists are become gatewatchers, but others maintain a more traditional approach to their audience; basing creative judgements on ‘gut’ instinct (Schultz 2007), while responding to the audience primarily via the route of editorial feedback. For some the only influence from outside the newsroom comes from statistical and data visualisation experts, precluding all but a tiny number of audience members from participation in this field of news production. Equilibrium between the operationally closed newsroom and the ‘former audience’ in this field seems unrealistic in the near term.

Interactives are more suited to some news stories than to others, on account of narrative type and the dynamics involved. Interactives are not necessarily more suited to particular subjects or niches than others, but in practice they are more likely to emerge in some fields than others. This is due in part to the availability of 'newsworthy' data, and in part to newsroom dynamics. The resource-intensive nature of producing interactive graphics limits what stories may be selected for coverage, and ‘important’ stories are more likely to be selected for coverage. Only one of the organisations covered in this study (with a niche, rather than general audience) provide their interactives teams with the budget to investigative or delve deeply into large data sets. It would be wrong to presume, therefore, that the recent explosion in availability of data from government may lead to greater scrutiny within the conventional fourth estate. The rise of the network has brought ‘big data' into the newsroom, but the availability of ‘big data’ is not a reason for investigation or coverage in itself. On the other hand, some numbers are considered more newsworthy than others amongst professionals in this field –so it may be that studies of news values (or story selection criteria) based upon normative lists (Galtung and Ruge 1965; Golding and Elliott 1979; Allern 2002) may require re-appraisal in light of this.

Some of the most popular applications of interactivity in this field, such as personalisable calculators, represent an up-to-date form of 'news you can use', and some conventional graphical formats (such as charts) are avoided because they 'turn off', or alienate audiences. Together, these two findings warrant further investigation in the context of how audiences interact with news. The pursuit of certain popular data visualisation formats, and the active avoidance of others, for fear of alienating news audiences, may yet have consequences for how we engage with and understand online news.

Future studies in this field may profitably interrogate the thoughts of non-interactive journalists on their interactives-producing peers (with particular emphasis on definitions of quality). A content analysis of output in this field may yield further insights into the values at play in news interactives, especially as they relate to the theories shared by professionals here. The rise of this non-journalistic form outside of the newsroom may yield valuable information too, especially in terms of how non-journalists influence and inform the work of professional interactive journalists.

References

Allern, Sigurd. 2002. "Journalistic and commercial news values." *Nordicom review* 23: 137-152.

Bantz, Charles. 1997. “News Organizations: Conflict as a Crafted Cultural Norm”. In *Social Meanings of News: A Text Reader* edited by Berkowitz, Dan. London: Sage.

Bateman, Scott; Regan L. Mandryk; Carl Gutwin; Aaron Genest; David McDine and Christopher Brooks. 2010. “Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts”, *CHI 2010*, April 10–15, 2010, Atlanta, Georgia, USA.

Boczkowski, Pablo. 2004. “The processes of adopting multimedia and interactivity in three online newsrooms”. *Journal of Communication*, 54: 197-213.

Bogost, Ian; Simon Ferrari and Bobby Schweizer. 2010. *Newsgames : journalism at play*. Massachusetts: The MIT Press.

Brinton, Willard C. 1914. *Graphic Methods for Presenting Facts*. New York, The Engineering magazine company.

Bruns, Axel. 2005. *Gatewatching: Collaborative Online News Production*. New York: Peter Lang.

BSI (British Standards Institution). 1992. *British Standard 7581:1992 A Guide to the Presentation of Tables & Graphs*. London: British Standards Institute.

Chung, Deborah S. 2007. “Profits and perils: Online news producers' perceptions of interactivity and uses of interactive features”. *Convergence*, 13(1): 43-61.

Deuze, Mark. 2008. “The Changing Context of News Work: liquid journalism and monitorial citizenship”, *International Journal of Communication*, 2: 848-865.
Epstein, Edward J. 1973. *News From Nowhere*. New York: Random House.
Forlizzi, Jodi and Katja Battarbee. 2004. "Understanding Experience in Interactive Systems”. *Human-Computer Interaction Institute*. Paper 46.
Franklin, Bob. 1997. *Newszak and News Media*. London: Arnold.
Galtung, Johan and Mari H. Ruge. 1965. “The Structure of Foreign News. The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers”, *Journal of Peace Research*, 2: 64-91.
Gans, Herbert J. 1979. *Deciding What's News A Study of CBS Evening News, NBC Nightly News, Newsweek and Time*. New York: Pantheon Books.
Giardina, Marco and Pablo Medina. 2012. “Information graphics design challenges and workflow management”. Paper presented at the International Conference on Communication, Media, Technology and Design, (ICCMTD), 09-11 May, 2012 in Istanbul, Turkey.
Golding, Peter and Phillip Elliott. 1979. *Making the News*. London: Longman.
Holmes, Nigel. 1984. *Designer's Guide to Creating Charts and Diagrams*. New York: Watson-Guptill Publications.
Gray, Jonathan, Lucy Chambers and Liliana Bounegru. 2012. *The Data Journalism Handbook*. London: O'Reilly.
ISO (International Organization for Standardization). 2009. *International Standard 9241-210 Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems*. Geneva: International Organization for Standardization.
Kenney, Keith; Alexander Gorelik and Sam Mwangi. 1999. “Interactive Features of Online Newspapers”, *First Monday*, 5(1).
Lang, Kurt and Gladys Lang. 1955. “The Inferential Structure of Political Communications: A Study in Unwitting Bias”, *Public Opinion Quarterly*, 19: 168-183.
Larsson, Anders Olof. 2012. “Interactivity on Swedish newspaper websites: what kind, how much and why?”, *Convergence*, 18(2): 195-212.
Lowrey, Wilson. 1999. “Routine news: The power of the organization in visual journalism”, *Visual Communication Quarterly*, 6(2): 10-15.
Lowrey, Wilson. 2002a. “Word People vs. Picture People: Normative Differences and Strategies for Control Over Work Among Newsroom Subgroups”, *Mass Communication and Society*, 5(4): 411-432.
Lowrey, Wilson. 2002b. “The Influential Designer: Explaining Variability in Control Over News Presentation Work”, *Visual Communication Quarterly*, 9(1): 4-13.
Lowrey, Wilson. 2003. “Normative Conflict in the Newsroom: The Case of Digital Photo Manipulation”, *Journal of Mass Media Ethics*, 18(2), 123-142.
McMillan, Sally J. 2002. “A Four-part model of cyber-interactivity: Some Cyber-places are more interactive than others”, *New Media and Society,* 4(2): 271-291.
Nichani, Maish and Venkat Rajamanickam. 2003. “Interactive Visual Explainers — A Simple Classification,” eLearningPost, September 1: http://www.elearningpost.com/articles/archives/interactive_visual_explainers_a_simple_classification/ (retrieved August 2013).
Prabu David. 1992. "Accuracy of Visual Perception of Quantitative Graphics: An Exploratory Study," *Journalism Quarterly*, Summer: 273-292.
Quandt, Thorsten. 2008. “(No) News on the world wide web?”, *Journalism Studies*, 9(5): 717-738.
Rosenberry, Jack. 2005. “Few Papers use online techniques to improve public communication”, *Newspaper research journal*, 26(4): 61-73.
Saffer, Dan, ed. 2010. *Designing for Interaction Second Edition: Creating Innovative Applications and Devices*. Berkley: New Riders.

Schlesinger, Philip. 1987. *Putting 'reality' together: BBC news*. London: Methuen.
Schroeder, Roland. 2004. “Interactive Info Graphics in Europe--added value to online mass media: a preliminary survey”, *Journalism Studies*, 5(4): 563-570.
Schudson, Michael. 1997. “The Sociology of News Production”. In *Social Meanings of News: A Text Reader*, edited by Berkowitz, Dan. New York: Sage.
Schultz, Ida. 2007. “The Journalistic gut feeling”, *Journalism Practice*, 1(2): 190-207.
Shultz, Tanjev. 1999. “Interactive options in online journalism: a content analysis of 100 US newspapers”, *Journal of Computer Mediated Communications*, 5(1): http://jcmc.indiana.edu/vol5/issue1/schultz.html (retrieved August 2013).
Tankard, James and Howard Ban. 1998. “Online Newspapers: Living up their Potential?”. Paper presented at AEJMC conference Baltimore, MD., August.
Tukey, John W. 1990. "Data-based graphics: visual display in the decades to come." *Statistical Science*, 5(3): 327-339.
Tufte, Edward. 1983. *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press.
Tunstall, Jeremy. 1971. *Journalists at work*. London: Constable.

APPENDIX VI: 'JUST FANCY THAT'
An analysis of infographic propaganda in The Daily Express, 1956-1959

Murray Dick

This research finds that the emergence of infographics as a regular phenomenon in UK news can be traced back to The Daily Express of the mid-1950s. These 'Expressographs' were often used not as a means of conveying data accurately and objectively, but in order to propagate the paper's editorial line, and to further Lord Beaverbook's political interests. A series of editorial narratives are established via the literature on The Daily Express, and its proprietor Lord Beaverbrook. These narratives are used as a framework to analyse statistical infographics published in The Daily Express between January 1956 and October 1959, by means of a combined content analysis and structural semiotic analysis. Best practice as espoused in the information design literature is used to identify misleading graphical methods in the sample, which were then analysed in the context of the editorial narratives identified. This study finds that infographics in UK news were the product of a lavishly-financed organisation whose key decision-makers were deeply concerned with the impact of the visual in news. The purpose of these infographics was to perpetuate their employer's idiosyncratic view of how the world should be. Occupational norms and practices may account for some of the biases identified, but cannot account for the breadth, range and consistency of bias found across the sample, which constitutes an example of mid-twentieth century propaganda.

KEYWORDS Daily Express; infographics; information graphics; newspaper graphics; newspaper history; propaganda

Introduction

The centrality of design and graphical presentation to our interpretation and understanding of the news is long-established in communications studies literature (Pasternack & Utt 1986; Griffin & Stevenson 1996; Middlestadt & Barnhurst 1999). And yet the history and function of the infographic in UK news has remained the preserve of practioner texts to date. Infographics are here defined as visual representations of concepts, whether in the form of information or data (Chandler and Munday 2011; Stevenson 2010). Within scientific and social scientific literatures, infographics are championed as a means by which statistical truths about the world may be accurately conveyed (Tufte 1983). And yet some of the most celebrated examples, including work by William Playfair and Florence Nightingale were more concerned with persuading the audience of a particular cause or belief, than they were with objectively explaining reality (Small 1998). So what might we say of the antecedents, history, and function of the form in UK newspaper journalism – when did infographics become part of our news, what purpose did they serve, and to what end?

The history of newspaper infographics

In the US the rise of newspaper infographics during the 1930s (Meyer 1997) has been attributed to those graphic artists, inspired by Swiss modern design, who moved from Europe to America, bringing the form from the sciences and social sciences into popular culture (Bogost et al 2010) and enriching the field of American communications along the way (Meggs and Purvis 2011). As for UK newspapers, the only two detailed works written on this topic (both of which were authored or co-authored by celebrated former Sunday Times journalists) agree that before the 1960s “little of note” (Evans and Taylor 1997: 287) had been done, and certainly “nothing apart from a half-hearted use of maps and graphs” (Sullivan 1987: 17). And yet, on the contrary, anyone with access to the many digitised UK newspaper archives available today will find that several sophisticated infographics were published in British newspapers from the mid-18th century onwards. These include, for example, an area map of the UK, detailing the journey of mail coaches from the capital to the provinces (St. James's Chronicle or the British Evening Post, September 30, 1786 - October 3, 1786; Issue 3990), and a reproduction of Playfairs' seminal 'ballance of payments in England' line graph (The Scots Magazine, January 01 1787, 55). Moreover, the present research finds that infographics actually started appearing routinely in The Daily Express during the mid-1950s.

But why in the modern era should it take two decades for an increasingly popular innovation in one modern anglophone western country to find voice in another? It is argued that technological factors may be discounted, as infographics were routinely published in other UK media at the time, particularly in advertising (Sullivan 1987: 17). Sullivan goes on to suggest that the political and educational presumptions that dominated thinking in UK newsrooms stifled the rise of the form (17). That professional emnities may constrain innovative journalistic forms has been found in organisational studies of graphic artists in the newsroom (Lowrey 1999; 2002), and from a study of a contingent field (photojournalism) during the 1930s (Zelizer 1995). But just as our understanding of the form must accommodate those factors that held it (and its practioners) back, so too must we understand the mind-set of the newspaper, its editorial staff and owner, who first saw the communicative potential in infographics for a UK audience.

Beaverbrook, The Daily Express and propaganda

The commercial rise of The Daily Express under Lord Beaverbrook which peaked in the 1950s, has been attributed to the newspaper's pioneering visual allure (Greenslade 2004: 65). As early as the 1940s, due care was seen to be applied consistently across all areas of the paper, from front page to back (Chisholm and Davie 1992: 212). Beaverbrook was popular amongst his journalists, who recognised his preparedness to invest heavily in news production (Greenslade 2004: 165). This largess was also recognised at the time by the Economist Intelligence Unit, which criticised Beaverbrook's titles as being the most over-manned, wasteful, and inefficient in Fleet Street – lending much to the notion that Beaverbrook invested in newspapers as a means of achieving personal ends, rather than as a financial means to an end (197). That he used his newspapers as vehicles for propaganda is well established in the literature (Greenslade 2004; Chapman and Nuttal 2011), and is indeed a matter of public record. His contribution to the Royal Commission on the Press in 1947 contained the infamous admission that he ran the paper “purely for the purpose of making propaganda” (Greenslade 2004: 63), albeit he qualified this by saying that it was carried in his newspapers' opinion columns only (wherein several infographics identified in the present study were also found). During his tenure as Minister for Information in 1918, Beaverbrook employed many propagandistic methods in media manipulation, some of which were considered pioneering (Taylor 1972: 137). That he recognised the significance of the visual in propaganda is evidenced by the fact that sponsorship of art increased greatly during his time in this role (Messinger 1992: 128).

Methodology

This research involves a combined content analysis and structural semiotic analysis, a method which was developed iteratively. In order to obtain a reliable sample for this study, the author searched *The UK Press Online* commercial newspaper database, which contains a near-complete collection of archives for The Daily Express from 1900 until the present day. Several generic keywords associated with news infographics were used in the first instance, to obtain an initial sample. These included; *an analysis of, analysis, axis, calculated, calculation, diagram, explains, guide, map, number of, our analysis, overview, plan, show, shows, shows that,* and *working*. The infographics and associated stories found using these terms were then analysed for more suitable terminology. The term 'expressograph' was found as were the bylines of several infographic artists (Michael Rand, John Bodle and Derek Dale), and these terms were used to increase the sample. In order to ensure the sample was truly representative of the period chosen, key time-frames established in the initial analysis were re-visited by means of browsing. For example, all published output between August 1956 and April 1958 was manually browsed in order to test an initial hypothesis that no infographics supporting the narrative 'Britain is still a global power' were published during that period.

To establish a representative coding framework for the sample, random examples were analysed for basic, non-optional categories of information. These included: year, date, page number, topic (and sub-topic), type (of infographic), caption (or headline) and summary. As the topicality of news stories can be a slippery concept (Saab 1990), it was decided that an additional, optional narrative 'summary' category was desirable. The classification system used for the infographics found in the sample is based upon BSI 7581:1992 (for statistical infographics). It comprises: tables, bar graphs, line graphs, area graphs, pie graphs, isotype graphs, scatter graphs, histograms, three-dimensional graphs, superimposed graphs, thematic maps, illustrated graphs, and pictorial graphs.

A number of further optional categories were developed, in order to facilitate analysis of the sample. The recurring narratives found in this study were established by means of a critical reading of histories of The Daily Express (Smith et al 1975; Greenslade 2004), biographies of Lord Beaverbrook (Taylor 1972; Chisholm and Davie 1992), and social, economic and political histories of the 1950s (Montgomery 1965; Peden 1985; Messinger 1992; Pearce and Stuart 1992; Shaw 1996). Categories of numerical bias and semiotic bias were informed by best practice in infographic design and data visualisation. A framework of common bad practice elements (which may in turn be used as a means of delivering misleading propaganda) was sourced from Tufte (1983); Tukey (1990); Holmes (1991); and The BSI (1992).

Limitations

It is acknowledged that the process of digitising these documents using Optical Character Recognition technology can fail to identify words correctly, and so may impede the full reach of this research, . It is also acknowledged that because some infographics stand alone, and some are neither bylined nor adequately key-worded, that this may also impede the range of the sample identified. Non-statistical graphics (including non-statistical maps) have not been included in this study, as it may be impossible to construct a detailed, normative framework with which to analyse these forms. It is acknowledged that some of the 'best practice' texts identified (such as Tufte 1983) lack empirical basis (David 1992), and indeed that contemporary empirical research shows that, contrary to Tufte's criticisms, people prefer (Inbar et al 2007), are better able to interpret and recall data associated with (Bateman et al 2010) and in any case are not hampered from interpreting (Spence 1990; Blasio and Bisantz 2002; Kulla-Mader 2007) data presented in 'chartjunk' format, a style commonly associated with newspaper infographics. It is further acknowledged that some of the models of excellence used are predicated on non-complimentary theoretical approaches. For example, Tukey's identification of impact as central to best practice presents a counterpoint to the Bauhausian minimalism of Tufte's data:pixel ratio theory (Wainer 1990: 341). In short, excellence in data visualisation is circumscribed by the limitations of positivism.

Findings: Propaganda narratives

The narratives established in this study are all realms of public affairs in which The Daily Express sought influence, namely:

The economy:

<u>For empire and protectionism</u>
<u>Anti-Common Market, anti-free trade</u>

Beaverbrook's first forays into British politics were driven by tariff-reform and Imperial Preference (Chisholm and Davie 1992: 70). In founding the Empire Free Trade party in 1929, he sought to establish a trade agreement between countries within the Empire in order to counter rival trading blocks. During this period The Daily Express ran an 'Empire Crusade' campaign, encouraging its readers to buy 'imperially' at Christmas (290), and both management and editorial staff at the newspaper were drafted into this 'crusade' directly (285). These themes continued to find expression in The Daily Express into the 1950s – for example, the paper celebrated Empire day in 1957 by distributing 1.5 million buttonhole flags (Daily Express, May 24 1957, 1). The Express remained, as one leader column slogan declared: "The paper that keeps faith with the Empire" (Daily Express, October 30 1959, 8).

Beaverbrook was implacably opposed to the European Common Market (Chisholm and Davie 1992: 517). Throughout this period The Daily Express devoted much space to content that championed trade with empire countries (Daily Express, October 20 1958, 11; Daily Express, 04 June 1959 Page 8; Daily Express, January 12 1958, 2), while undermining the Common Market (Daily Express, January 21 1957, 4-5); denouncing the perceived damage done to UK farming by free trade in bacon (Daily Express, February 04 1958, 6); and decrying the alleged 'dumping' of non-Empire butter on the UK market (Daily Express, March March 24 1958, 2).

Society:

<u>For the (consumerist) individual</u>
<u>Anti-collective (anti-trade union)</u>

The Express's outlook during this period was, in the words of the editor of its main rival: "fundamentally materialist... appealing to the young man on the way up, encouraging him to accumulate possessions on route... preaching that hard work and application can enable every ambitious reader to make his pile like Lord Beaverbrook" (Greenslade 2004: 168). The political consensus of the 1950s that saw increasing trade union involvement in government decision-making, caused consternation amongst those on the right of British politics (Pearce and Stewart 1992: 460). Throughout this period The Daily Express ran many stories critical of trade unions (Daily Express, April 30 1957, 2); critical of trade unions' demands (Daily Express, July 03 1957, 2; Daily Express, July 10 1957, 7); critical of industrial action (Daily Express, July 24 1956, 1), and even critical of the perceived inequality of collective bargaining (Daily Express, March 27 1958, 2). This opposition was not pragmatic, but was rather borne of an ideology predicated on the individual over the collective. This has been observed of the paper's coverage of other news events elsewhere; for example, improvements to public health under the NHS were attributed to individuals rather than the collective; to lawmakers, politicians, economists (Smith et al 1975: 148).

Standard of living

'You've never had it so good'

The increasing cost of living

The period saw a major re-structuring of the country's domestic economy; by 1954 hire purchase was being paid on goods worth £750,000,000 by eight out of ten Britons (Montgomery 1965: 269). Consumption was perceived as key to economic progress, and the lifting of government restraint on hire purchase was embraced in at least one Express leader column (Daily Express, November 03, 1958, 8). And yet the collectivist pressures of wages-driven inflation, and increased taxes to pay for the new welfare state were routinely used as vehicles to decry a perceived increase in the cost of living throughout the period.

Britain's place in the world

Britain is losing ground to competitors

Britain is still a global power

Britain's place in the world was a significant issue during the 1950s not just because of fears for the future in a post-Empire, globalised world, but also because of economic under-performance by comparison with rival economies (Montgomery 1956: 283). This was compounded by the humiliating climb-down over Suez, that saw Prime Minister Eden depart in ill-health, and with him any pretensions towards a global policing role for the UK (Pearce and Stewart 1992: 473-4). On the other hand, imperial preference required that people still recognise the virtues and reach of the empire.

Politics

For Conservatives

Anti-Labour

Given Beaverbrook’s personal politics, it should come as no surprise that The Daily Express was broadly supportive of The Conservative Party and also broadly critical of the Labour Party; this will become apparent during the polling case study presented later in these findings.

These narratives are not mutually exclusive, and some infographics embody more than one narrative. Where this is apparent the dominant value was chosen as preferred category. In the majority of cases, data and signs in these infographics are affected in various ways to support editorial position on each of these narratives. Other minor and infrequent narratives were found (such as support for royalty), but given their limited reach, they offered little towards the analysis presented here.

Findings: Content analysis

In total, 215 infographics were found to have been published between January 17 1956 and October 22 1959. Of these 45 were found to be non-statistical in nature, leaving a sample of 170 statistical infographics in the sample. From July 1956 infographics start appearing at a rate of more than two per month, and from this point until October 1959 in only one month (November 1957) were less than 2 infographics published. Some 41 of these statistical infographics appeared on the front page of the newspaper; 24% of the total sample, the first being a bar chart that conveyed survey findings on Sir Anthony Eden's handling of the Suez crisis (Daily Express, September 10 1956, 1). Some 19 of the 45 non-statistical graphics found in this survey (42%) also featured on the front page. The tally of statistical infographics for each year was as follows:

29 were published in 1956
51 were published in 1957
50 were published in 1958
40 were published to October 1959

The most popular topic was The Economy (54), followed by; Consumer Affairs (45), Politics (32) and International Affairs (18). All other topics were represented in low single-digits. The most popular infographic form in the sample was the line graph (60) followed by table (36), pictorial graph (34) and bar graph (21). Mixed formats accounted for 14 examples, and no other graphical format was represented by more than two examples. Of these infographic forms, it was found that bias in the representation of numbers occurred in 36 (or 60%) of line graphs; in 19 (or 54%) of pictorial graphs, and in 11 (or 52%) of bar graphs. Over time it was found that bias in the representation of numbers peaked in 1957, and decreased in subsequent years.

The most popular narratives amongst these infographics were found to be:

Narrative	Number	Percentage of total
For Conservatives	28	75%
The increasing cost of living	25	44%
Consumer culture	23	26%
You've never had it so good	21	57%
Britain is still a global power	18	28%

Seven of the sample (or 0.04%) were found not to conform to any of the narratives identified, but two of these were found to contain a bias towards recent change. Fourteen infographics (8%) across the sample range were found to accompany opinion columns. None concerning the narrative 'Britain is still a global power' were published between August 1956 and April 1958. None concerning the narrative 'Britain is losing ground to its competitors' were published between March 1958 and October 1959. Of the 21 infographics concerned with the theme ‘You've never had it so good’, eight were published during 1959.

Findings: methods of bias in infographics

Tables

Bias in the implication of order

A table (Figure 1) is used to detail the average pay for three non-professional employment sectors in four countries (France, West Germany, Italy and Britain). Britain is listed bottom in this table, but no clear method of organisation is chosen to justify this position (in terms of alphabetical, or scale ordering). Britain has the highest average earnings for all three of these sectors compared with the others, and yet (in keeping with the anti-collective narrative) it remains at the bottom of the table. Throughout the sample several infographics challenge non-professional wage-bargaining; questioning the motives of trades unions (Daily Express, July 10 1957, 7), and airing scepticism about the ultimate value of increased wages, given the rising cost of living relative to other nations (Daily Express, August 27 1957, 2). Elsewhere in the sample scale-order is routinely used in tables to support other narratives; to convey a selection of non-professional job-types in order of wages, high to low (Daily Express, March 26 1957, 13); and to detail overpaid tax as expressed in magnitude of weekly earnings, low to high (Daily Express, 06 June 1957, 6).

Pictorial graphs

Improper scaling

In 19 of the 35 pictorial graphs found in the sample (54%) the use of two-dimensional images or figures to represent single variables was found to distort or bias data. A series of dual-pictograms is used to represent the proportions of those in and out of work during the Austin and Morris car strike at twelve different plants (Figure 2). Those who chose not to strike narrowly outnumbered those who opted to strike, according to the employer's data cited here, and data published elsewhere (The Daily Mirror, August 11 1956, 1), but the method used to display this information shows a clear visual advantage to those who broke the strike. On the first day a large majority in some of the largest plants were on strike (i.e. Fisher and Ludlow, at 90%), but this visual symbol is significantly smaller than the largest single figure on the page (a working man) used to represent one of the smallest plants (MG Cars, Abingdon). It is also smaller than a strike-breaking icon used to represent Morris Radiators (Oxford) 90% of whose workforce broke the strike. This strike-breaking figure is both taller and broader than the Fisher and Ludlow striker figure, though they represent the same percentage value. The image used to represent those who broke the strike is of a worker holding a work-tool aloft with legs spread apart, which has greater visual impact than the slouching, downcast striker with legs closed together, and with hands in pockets.

Elsewhere in the sample a correlation between increases in pay and decreases in days lost to strikes is (falsely) implied via a table containing pairs of three-dimensional symbols (a pile of pound notes to represent settlements, and stop watches to represent time lost to strike action) (Figure 3). The size of the clock used to represent 1954 (427,000) is only slightly smaller than that representing 1953 (481,000), but the clock representing 1955 (395,000 hours) is disproportionately smaller than either. The visual confusion in this representation is compounded because it seems that both the diameter and depth of these images is being used to convey scale.

General methods of bias in statistical graphs

Truncated axis

Fluctuations in the price of petrol between 1950 and 1957 are exaggerated by means of a truncated vertical axis on a line graph (Figure 4). In the absence of any scale along the vertical axis the value jumps from around a fifth of the way up the vertical axis (at 3s 0d, in 1950) until it breaks through the top of the chart (at 6s 11d, in November 1956, during the Suez crisis). Were this line plotted on a chart whose vertical axis started at zero, this rise would begin around half-way up, and so would look significantly less dramatic. Elsewhere data on car ownership are biased towards the UK (and against

Common Market competitors) in the form a bar chart with a series of ratios representing countries, where the vertical axis shows rates of car ownership as a ratio decreasing upwards (Figure 5). The upper limit of the chart is 60 rather than 100, which exaggerates the lead the UK holds. This example of bias is compounded by the fact that the data presented here is used elsewhere in the sample (Figure 6), but in the former Luxembourg, which outstripped Britain in terms of car ownership, is omitted.

The second of these two infographics was published across a gatefold opinion column concerning the perils of joining the Common Market. Five bar graphs convey the comparative strength of the UK against its European rivals in terms of; average earnings, car ownership, unemployment, prosperity and standard of diet (Daily Express, January 21 1957, 4-5). All but one of these (unemployment) exaggerates the UK's advantage over its competitors, by truncating the value axis.

Irregularity in scale

The issue of infant mortality is expressed by means of a line graph containing five different variables labelled as 'social groups' (Figure 7). The time series starts at 1911 on the horizontal axis, and runs to 1950 (left-to-right), while the vertical axis is marked at irregular intervals of deaths per 1,000 live births (4, 6, 8, 10, 20, 40, 60, 80, 120, from bottom-to-top) with inconsistently scaled gaps between these values (the gap between 120 and 80, for example, is approximately a third of the size of the gap between 20 and 10). This irregularity is most pronounced amongst the highest numbers, and so the visual effect is to harmonise the five trend-lines, and so make the decline in mortality look broadly consistent across all groups over time. This approach dampens the positive visual effect of public health expenditure (and the advent of the NHS) on mortality rates for the poorest in society. It also informs a misleading reading of the data, as expressed in the accompanying opinion piece: "…in spite of the levelling influences of 'free' medical service, subsidised milk, cheap vitamins, the soaking of the rich by income tax, babies born to the better-off still stand a much better chance of survival than those born to poorer parents" (Daily Express, June 04 1959, 8).

Bias in icons, labels, language

A graphic illustrating how trade unions were increasingly demanding a shorter working week rather than higher earnings was found in the sample (Daily Express, July 10 1957, 7). Here workers are represented as either examining their watches before a clocking-in machine, or as holding out their hand before a payment kiosk– both of these images speak to the perception of demands rather than desserts, re-capitulated in the text, which speaks of "the new challenge" (Daily Express, July 10 1957, 7) for government.

The most consistently used bias in the use of descriptors concerns the labelling of the Labour Party as 'Socialists' (covered in the following polling infographics case study). But elsewhere, a theme emerged across a range of infographics dealing with the Suez crisis - the day Nasser took control of the canal is described in labels (Daily Express, August 08 1956, 1; Daily Express September 09 1956, 1) as a 'grab day' and in the standfirsts and text accompanying infographics in a further article, reference is made to 'Nasser's grab' (Daily Express, July 28 1956, 4).

Boundary bias

Where one variable is seen to break through the boundaries of an infographic, this may bias a comparison with other variables, and may imply something editorially that the data does not support. Variables break the boundaries of several infographics in the sample; for example, a three-dimensional pictorial graph is used to explore the costs associated with Egyptian ownership of the Suez canal – where the costs associated with an alternative route is seen to burst through the edge of the chart, thus metaphorically supporting the notion that Prime Minister MacMillan's decision to deal with Nasser was a victory for pragmatism, rather than a necessary move following diplomatic humiliation (Daily Express, May 15 1957, 2). This approach is also used to exaggerate the increase in the cost of living for those earning £2,000 in 1938 (Daily Express, July 29 1957, 2).

Inconsistent placement of labels or symbols

A horizontal bar chart is used to compare the ratio of air miles to employees for several international airlines, in the context of industrial action at BOAC airlines (Figure 8). The symbol of a plane is used at the end of each bar, with the plane representing American Airlines being marginally longer than the others, thus exaggerating their advantage.

Lack of labelling on axes

In the example used above, the bars (and pictures) used to represent the number of employees at various airlines are difficult to compare for lack of values (and grid lines) on the horizontal axis – allowing for the alteration of the lengths of each bar. If the scale for these bars were taken from the length of the value attributed to American Airlines (which, at a magnification level of 125% of the original was measured at 10.8cm, representing 155,000 air miles), then while the UK figure is more or less visually representative (4.7cm representing 68,000 air miles), those other competitors; KLM Royal Dutch (5.3cm representing 71,100); Air France (4.2cm representing 58,000); and SAS Scandinavian (6.6cm representing 85,000) should all be smaller than they actually are (they should be 4.9 cm; 4.0 cm and 5.9 cm respectively), thus giving a wholly false sense of comparison biased against the UK airline (and its striking workers).

Elsewhere in the sample a lack of labelling on axes was found to impede accurate comparison of rising wages against the cost of living (Daily Express, 28 July 1959, 1) and satisfaction with Eden’s approach to the Suez Crisis (Daily Express, 12 December, 1956, 1).

Methods of bias in line graphs

Tear or line-breaks between disparate trend lines

Two values, one each for Britain's weekly consumption (in tonnes) of coal and oil, are plotted within a line graph in the sample, but their values are so far apart that a vertical line break is introduced between the two (Figure 9). This renders the grid lines used in this graph meaningless, and gives an inflated sense of significance to the rise in consumption of oil, at the expense of coal, an industry whose workers (and leading trade union) were often criticised in the Daily Express during the period.

Making trend lines too thick

By making the trend line of a line graphic very thick, it is possible to skew interpretation of the data relative to the values on the axes (or grid-lines). Unemployment is presented as having fallen from 2.8% to 2.5% between December 1958 and March 1959 in one infographic (Figure 10), but the trend line is relatively thick and includes an arrow on the end. In conjunction with the truncated vertical axis, its downward trajectory is over-emphasised. A similar approach is taken elsewhere (Figure 11), where a fall in the unemployment rate from 620,786 in January 1959 to 413,000 in June 1956, is (due to the truncated axis) presented in an even more exaggerated way.   These examples fall between the 'For Conservative' and 'You've never had it so good' narratives.

Extrapolation

The extrapolation of data without explanation is considered bad practice in the literature, but equally other manifestations of exaggeration when combined with extrapolation (especially where based upon a single unverifiable source) may also be considered to embody bias. Data based upon a plan set forth by the head of a Canadian Trade Council are projected using a line graph, whose horizontal axis starts at £100 million, rather than zero, thus artificially inflating the rise in recent imports, and exaggerating the predicted rise in trade (Daily Express, November 22 1957, 2).

Bias towards recent change

Bias towards recent change is found commonly throughout the sample, and indeed is found (by degrees of application) in most of the time series infographics already mentioned in this analysis. Further examples range from increases in the price of commodities (Daily Express, September 19, 1); rises in the price of oil after the Suez crisis (Daily Express, August 08 1956, 1); rises in the price of petrol during the Iraq crisis (Daily Express, July 16 1958, 2); and increases in the price of turkey in the run up to Christmas (Daily Express, December 21, 1956, 1).

Non-graphical forms of bias

Other forms of bias may be seen at be at work within this sample that do not have a graphical basis. For example, the sourcing of data in several examples may be construed as biased. In industrial relations stories, the data used comes from employers (Daily Express, July 24 1956, 1; Daily Express, July 25 1956, 2), but not trade unions, while on tax issues, CBI data are used (Daily Express, January 02 1956, 2); but no counter-veiling view is offered. The selective interpretation of data may also be observed in several examples. A pictorial graph is used to convey comparative international unemployment rates (Daily Express, March 18 1958, 2), but regional variations, which accounted for significant differences (that allowed pockets of long-term unemployment to develop during the period, despite the aggregate figures) were not explored in the sample. The timing of inclusion of certain stories may also be seen to embody a form of bias, as may the process of omitting certain variables (or stories) altogether.

Findings: Political bias in polling infographics (case study)

Selection of time series starting point and intervals

The starting point used in polling infographics found in the sample varies, but in most cases, the choice (and regularity) of time intervals favours The Conservative Party. Several line graphs use the 1955 election as their starting point, despite all other featured intervals being much later (for example: Daily Express, October 08 1956, 2; Daily Express, October 29 1956, 9; Daily Express September 30 1957, 1). In each of these cases the Labour Party was at the time leading the poll in question, and by including the result from the last election the reader is visually reminded that the Tories once led, and Labour’s lead is visually dampened. On the other hand, several polls do not include the previous election result in their schedule, one only displays Eden's popularity, with no counter-trend representing Labour (Daily Express, October 03 1956, 2); one shows overwhelming public support for Eden's decision over Suez (Daily Express, November 24 1956, 1); one shows the Tories overtaking Labour (Daily Express, November 27 1956, 1), and one includes the November 1956 result, where the Conservatives led Labour (Daily Express, January 11 1957, 1). In each of these cases, there was a contemporary electoral advantage for the Conservatives.

The careful selection of time intervals can also be used to effectively erase from history those periods when support for The Conservatives dropped. In one infographic five time intervals are presented to indicate which party the public would vote for at the next election, over time; 'General Election 1955', 'August 1958', 'September 1958', 'October 1958' and 'now' (mid-October) (Figure12). This schedule elides two years of polling data, during much of which time Labour held a consistent, and at times substantial lead. Other examples of this phenomenon can be found elsewhere in the sample (Figure13).

Bias in selection of polling answers represented

The nature of polling questions was found to vary in this sample, but the dominant three questions, at least one of which appear in all of the political polls identified, may be classified in the following way:

> The 'general election question': *If a General Election were held immediately for which party would you vote?*
> *The 'government question': In general, are you satisfied or dissatisfied with the Government's conduct of affairs?*
> The 'leader question': *In general, are you satisfied or dissatisfied with <insert leader name> as Prime Minister?*

The choice of which issues to present visually often present some form of positive finding for The Conservative Party, and often fail to report polling success for the Labour Party, excepting where they lead in all questions asked, or they will provide a 'least bad' result for the Conservatives. For example, in one infographic all three questions were asked, but the one presented in line graph form (in response to the 'leader question') contains a marginal gap (3 points) between those 'dissatisfied' and those 'satisfied', compared with a 19 point lead for dis-satisfaction over satisfaction in the 'government question', and a 9 point lead for Labour in the 'general election question' (Daily Express, March 07 1958, 8). Elsewhere the 'government question' is plotted, showing a narrow lead for satisfaction over dissatisfaction, by 1.5 points – albeit the 'dissatisfied' trend is not plotted (Daily Express, October 03 1956, 2). A 16.5% lead for dissatisfaction in the 'leader question' in this poll was not presented. In another infographic in the sample, two polls are presented in line graph form; the 'leader question' and the 'government question'; both showing Conservative leads, but the 'general election question' (which favoured Labour) is not presented (Daily Express, September 12 1956, 1).

Bias in the representation of polling answers

In one infographic the answers to different polling questions are presented on the same line graph. Satisfaction polled in relation to the 'leader question' (which favoured the Conservatives) was plotted against voting intention in the 'election question' (against Labour) (Figure14). Had these answers been plotted in graphs against alternative responses to the same question, then a less favourable results for the Conservatives would have been shown in both cases. In another infographic responses to the 'leader question' were not plotted against each other within the same graphic space, but rather they were presented separately one on top of the other; with 'satisfied' (top) followed by 'dis-satisfied' and then 'don't know' at the bottom (Figure15). This renders comparison of these trends over time impossible (the values show that dis-satisfaction was higher than satisfaction for three of the six intervals presented in this graph; 'Late March'; Early April' and 'June'; but because of the placement of these trend lines, all instances of dis-satisfaction are presented as lower than satisfaction).

Omission

Between December 1956 and July 1957, no political polling graphics were published in The Daily Express. During this time Labour extended its lead in other opinion polls, such as Gallup, by up to ten points during some months (Pack, 2011).

Stigmatising the opposition

All polling infographs in the sample were found to contain use of the word ‘Socialists’ to describe the Labour Party (with both ‘Tory’ and ‘Conservatives’ used alternately to describe The Conservatives) until early September 1959, and the dissolution of parliament prior to the 1959 General Election. From this point until the election the term ‘Labour’ was used exclusively in all polling infographics.

.

Discussion

That over 99% of the statistical infographics found in this study can be attributed to a handful of narratives suggests that the news values involved in selecting data and stories to represent were very limited indeed. It may be argued that selection may have been limited by the availability of data and sources, but this would not account for the examples of missing (but available) variables in some representations, nor in the use of speculative, and single-source datasets identified in the sample, all of which support one (or more) identified narrative. That almost 8% of the infographics identified in this study were accompaniments to opinion pieces shows that those responsible for the production of the newspaper recognised the power of information design to persuade as well as to inform. Infographics in UK news were borne of a lavishly-financed news organisation whose key decision-makers were deeply concerned with the impact of the visual in news.

That the 'Britain is still a global power' narrative disappeared from the Daily Express news agenda between August 1956 and April 1958 may be attributed to the paper’s “vigorous and loyal” (Shaw 1996: 31) support for Eden's actions in the early stages of The Suez Crisis, which would in hindsight have been a source of embarrassment for its proprietor. But on the other hand, not all incidences of these narratives may be attributed to bias. For example, the ‘You've never had it so good’ narrative featured eight times during 1959 (in the run up to the general election) and 'The increasing cost of living' narrative was published only once that year. During this period unemployment and inflation were both (even by comparison with the wider period) particularly low.

Organisational needs (and norms) and production values may account for some elements of ‘bad practice’ found in the sample; in the irregular or inconsistent use of values and intervals in line graphics; in the use of thick trend lines, in order to facilitate better quality reproduction, or indeed in “exaggerating the rate of recent change” (Tufte 1983: 76) in order to pursue a particular news angle. It may be argued that newspapers have an innate bias towards recent change, and so we shouldn’t be surprised to find many examples of this phenomenon (however defined) in a study such as this. Creative expression in newspapers serves as a function of white space, and this constraint in turn informs the news values of the form (Sullivan 1987). Other pressures in the form of budgetary and staffing constraints, and in organisational politics (Lowrey, 1999; 2002) may have consequences too. But these issues alone cannot account for the consistency of bias found in this study nor can they account for the use of semantic bias in symbols, labelling and in the ordering of variables.

One aspect of propaganda in this sample (the representation of numbers) declined over time. It cannot be known with any certainty why this may be; the graphic artists may have improved their craft over time; response to reader complaints (or actions by a fledgling press regulator) may have encouraged change, as may a creeping acknowledgement of improving numeracy amongst the paper's readership. That use of the label 'Socialists' in reference to Labour in polling infographics stopped during the period between the dissolution of parliament, and the election in 1959, is, however, telling. It suggests that an editorial decision was made to present polling infographics in a balanced, ‘objective’ way during this period. This example hints at how decisions were made as to the form and content of Daily Express infographics, and gives an insight into the balance between professional and organisational norms at play.

Further study

Future research may profitably explore the practice and function of the infographic artist in historical (and current) newsrooms, by drawing upon archive sources that give an insight into the pressures of editorial decision making in practice. Further research is required into why it took so long for infographics to take hold in UK media. More generally, the findings of this study indicate that infographics in British news media is a topic ripe for critical review.

References

Blasio, Anthony J. and Ann M. Bisantz. 2002 . “A comparison of the effects of data-ink ratio on performance with dynamic displays in a monitoring task”, Journal of Industrial Ergonomics 30 (2): 89-101.

Bogost, Ian; Simon Ferrari and Bobby Schweizer. 2010. Newsgames : journalism at play. Massachusetts: The MIT Press.

BSI (British Standards Institution). 1992. British Standard 7581:1992 A Guide to the Presentation of Tables & Graphs. London: British Standards Institute.

Chandler, Daniel and Rod Munday. 2011. A Dictionary of Media and Communication. Oxford: Open University Press

Chapman, Jane, and Nick Nuttall. 2011. Journalism today: a themed history. London: John Wiley & Sons.

Chisholm, Anne, and Michael Davie. 1992. Beaverbrook: a life. London: Random House.

David, Prabu. 1992. "Accuracy of Visual Perception of Quantitative Graphics: An Exploratory Study." Journalism Quarterly. 69: 273-292.

Evans, Harold, and Edwin Taylor. 1997. Pictures on a page: photo-journalism, graphics and picture editing. London: Pimlico.

Greenslade, Roy. 2004. Press Gang: how newspapers make profits from propaganda. London: Pan Macmillan.

Griffin, Jeffrey L. and Robert L. Stevenson. 1996. “The influence of statistical graphics on newspaper reader recall”, Visual Communication Quarterly, 3 (3): 9–11.

Holmes, Nigel. 1984. Designer's Guide to Creating Charts and Diagrams. New York: Watson-Guptill Publications Inc.

Kulla-Mader, J. 2007 “Graphs via Ink: Understanding How the Amount of Non-data-ink in a Graph Affects Perception and Learning”. Master‘s Thesis, Dept. of Information and Library Science, University of North Carolina.
Inbar, Ohad, Noam Tractinsky and Joachim Meyer. 2007. “Minimalism in Information Visualization - Attitudes towards Maximizing the Data-Ink Ratio”. Proc. ECCE, ACM Press: 185-188.
Lowrey, Wilson. 1999. “Routine news: The power of the organization in visual journalism”, Visual Communication Quarterly, 6 (2): 10-15
Lowrey, Wilson. 2002. “Word People vs. Picture People: Normative Differences and Strategies for Control Over Work Among Newsroom Subgroups”, Mass Communication and Society, 5 (4): 411-432
Meggs, Philip B. and Alison W. Purvis. 2011. Meggs' history of graphic design. Hoboken, New Jersey: John Wiley & Sons.
Messinger, Gary S. 1992. British propaganda and the state in the First World War. Manchester: Manchester University Press.
Meyer, Eric K. 1997. Designing infographics. Indianapolis: Hayden Books.
Middlestadt, Susan E. and Kevin G. Barnhurst. 1999. “The influence of layout on the perceived tone of news articles”, Journalism and Mass Communication Quarterly, 76 (2): 264–276.
Montgomery, John. 1965. The Fifties. London: Allen and Unwin.
Pack, Mark. 2011. “A psephology present for the new year: opinion poll data 1943-2011”, December 31:
http://www.markpack.org.uk/28444/a-psephology-present-for-the-new-year-opinion-poll-data-1943-2011/ (checked November 2013)
Pasternack, Steve, and Sandra H. Utt. 1986. “Subject perception of newspaper characteristics based on front page design”, Newspaper Research Journal, 8 (1) (Fall): 29–35.
Pearce, Malcolm, and Geoffrey Stewart. 1992. British political history, 1867-1990: democracy and decline. London: Routledge.
Peden, George C. 1985. British economic and social policy: Lloyd George to Margaret Thatcher. London: P. Allan.
Shaw, Tony. 1996. Eden, Suez and the mass media: propaganda and persuasion during the Suez crisis. London: IB Tauris Publishers.
Small, Hugh. 1998. “Florence Nightingale's Statistical Diagrams”, paper from Stats & Lamps Research Conference, organised by the Florence Nightingale Museum at St. Thomas' Hospital, 18th March.
Smith, Anthony Charles H.; Elizabeth Immirzi and Trevor Blackwell. 1975. Paper voices: The popular press and social change, 1935-1965. Totawa, NJ: Rowman and Littlefield.
Spence, Ian. 1990. “Visual Psychophysics of Simple Graphical Elements”. JEP: Human Perception and Performance, 16 (4): 683-692.
Stevenson, Angus (editor). 2010. Oxford Dictionary of English (Third edition). Oxford: Oxford University Press
Staab, Joachim F. 1990. "The role of news factors in news selection: A theoretical reconsideration." European Journal of Communication 5 (4): 423-443.
Sullivan, Peter. 1987. Newspaper Graphics. Darmstadt, IFRA.
Taylor, Alan J. P. 1972. Beaverbrook. London: Simon and Schuster.
Tufte, Edward, R. 1983. The Visual Display of Quantitative Information. Cheshire, Conn: Graphics Press.
Tukey, John W. 1990. "Data-based graphics: visual display in the decades to come." Statistical Science, 5 (3): 327-339.
Wainer, Howard. 1990. “Graphical Visions from William Playfair to John Tukey”, Statistical

Science, 5 (3): 340-346.
Zelizer, Barbara. 1995. “Journalism’s last stand: Wirephoto and the discourse of resistance”, Journal of Communication, 45 (2) (Spring) 78- 92.

Figure captions

Figure 1: Britain is placed bottom of this league table, despite having the highest rates of pay (Daily Express, May 12 1958, 2). Express Newspapers/Express Syndication.

Figure 2: The scaling of icons used in this pictogram is skewed in favour of strike-breakers (Daily Express, July 24 1956, 1). Express Newspapers/Express Syndication.

Figure 3: A (false) causal 'link' is presented between increases in wages and days lost through strikes (Daily Express, April 30 1957, 2). Express Newspapers/Express Syndication.

Figure 4: Truncated vertical axis (and a lack of scaling) gives a false impression of petrol prices between 1950 and 1958. (Daily Express, July 16 1958, 2). Express Newspapers/Express Syndication.

Figure 5: Ratios of car ownership in various European countries with the UK's lead inflated by means of a truncated axis (Daily Express, March 16 1957, 2). Express Newspapers/Express Syndication.

Figure 6: Based on the data as Figure 5, this bar graph includes Luxembourg, a close competitor of the UK (Daily Express, January 21 1957, 4). Express Newspapers/Express Syndication.

Figure 7: Scaling such that all 'social groups' appear to decline relative to each other. If scaled regularly, this graph would show a comparatively dramatic decline in infant mortality amongst the poorest social group (Daily Express, June 04 1959, 8). Express Newspapers/Express Syndication.

Figure 8: Symbols and scaling are used here that give a misleading (negative) sense of BOAC's comparative performance (Daily Express, October 15 1958, 8). Express Newspapers/Express Syndication.

Figure 9: A tear across the middle of the axis here allows encourages a misleading comparison of two variables (Daily Express, December 01 1958, 2). Express Newspapers/Express Syndication.

Figure 10: A line graph whose line-width skews interpretation (Daily Express, March 19 1959, 2). Express Newspapers/Express Syndication.

Figure 11: A combination of excessive line width and truncated axis implies unemployment is at zero (Daily Express, June 30, 2). Express Newspapers/Express Syndication.

Figures 12: The selection of time intervals in these chart erases many months of Labour Party leads (Daily Express, October 21 1958, 2). Express Newspapers/Express Syndication.

Figure 13: The horizontal axis in this line graph starts at 1955, then jumps to 'Early June 1959' – thus eliding several years of data containing Labour polling leads (Daily Express, September 10 1959, 1). Express Newspapers/Express Syndication.

Figure 14: Responses from two different polling questions are compared in the same graph, favouring The Conservatives Express Newspapers/Express Syndication

Figure 15: Variables presented one on top of the other, regardless of their relative values, favours the Conservatives (Daily Express, July 09 1957, 1). Express Newspapers/Express Syndication.

Figures

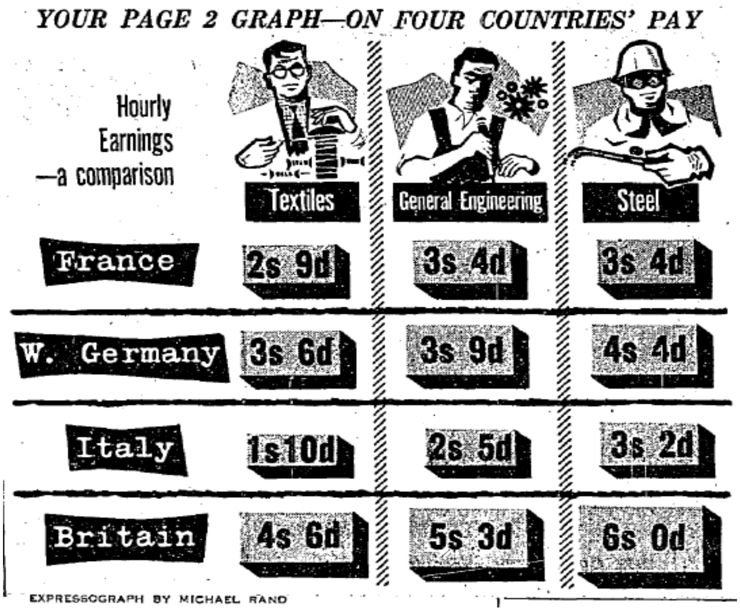


Figure 1

Figure 2

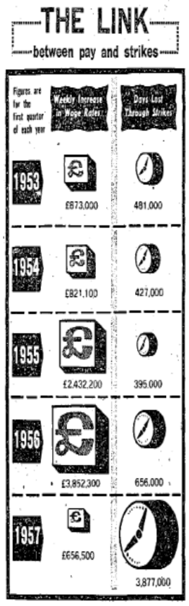


Figure 3

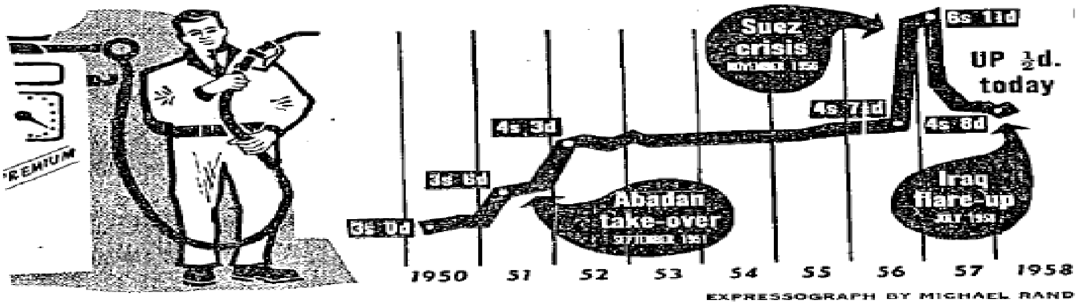
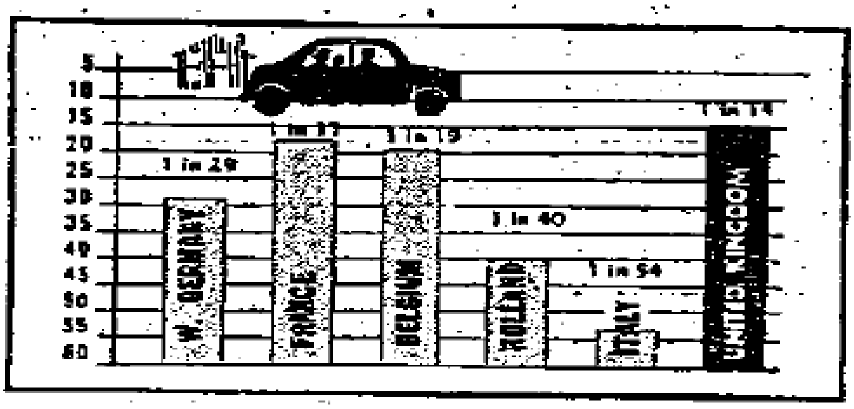


Figure 4



THIS GRAPH in the Daily Express pamphlet "Your Future?" is Mr. R. A. Butler's source for his figures on car ownership.

Figure 5

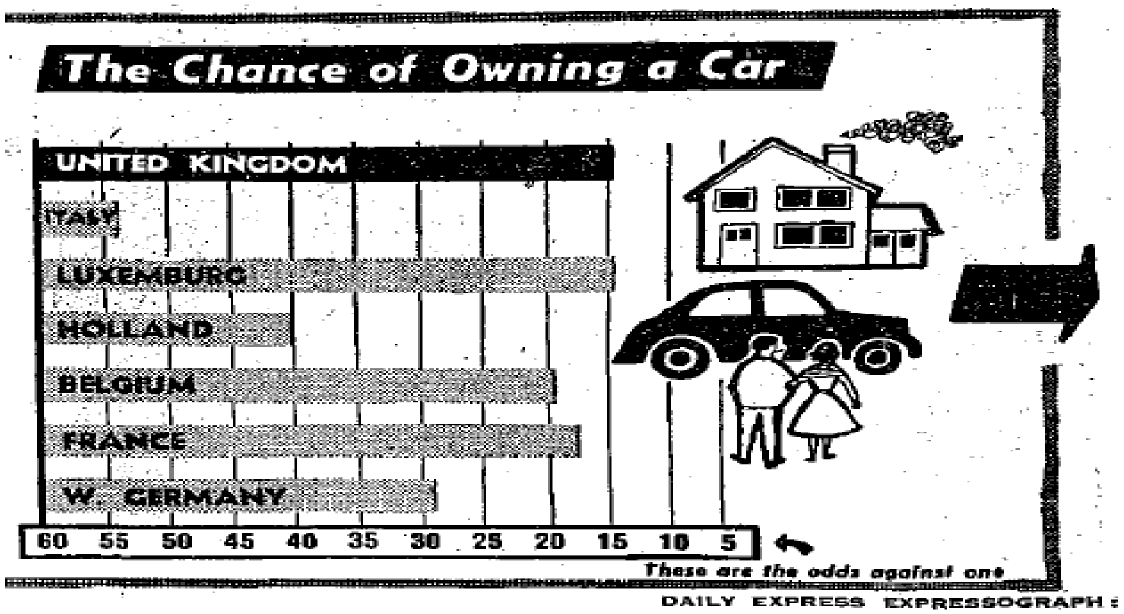
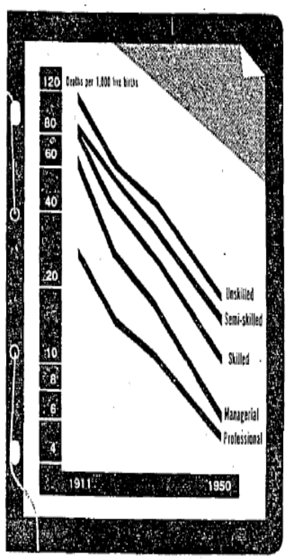


Figure 6



THIS MICHAEL RAND CHART SHOWS THE DECLINE IN THE NUMBER OF NEWLY-BORN BABIES WHO DIED, DIVIDED INTO SOCIAL GROUPS

Figure 7

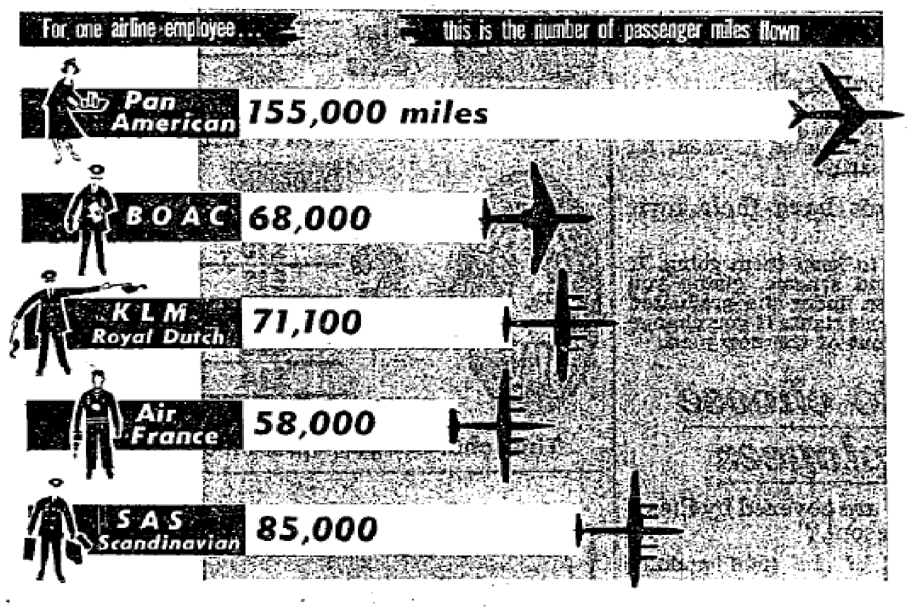


Figure 8

Figure 9

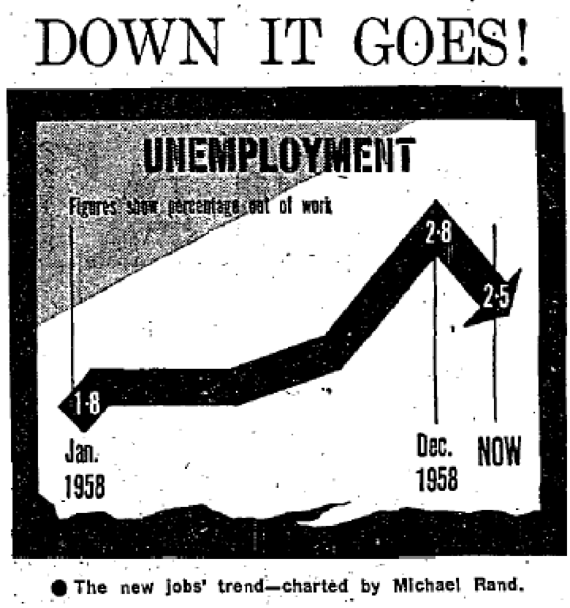


Figure 10

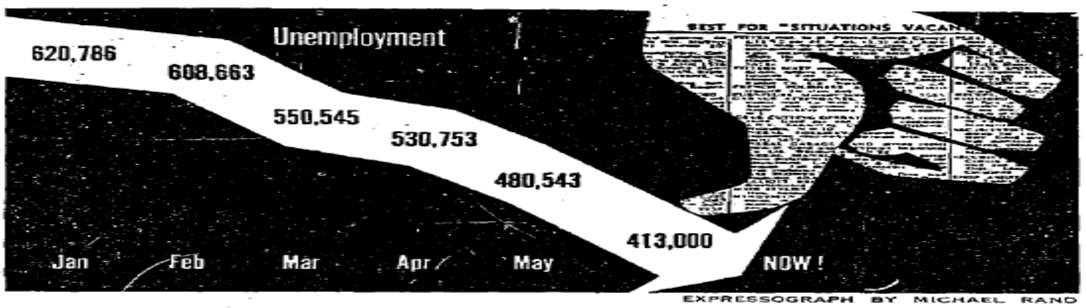


Figure 11

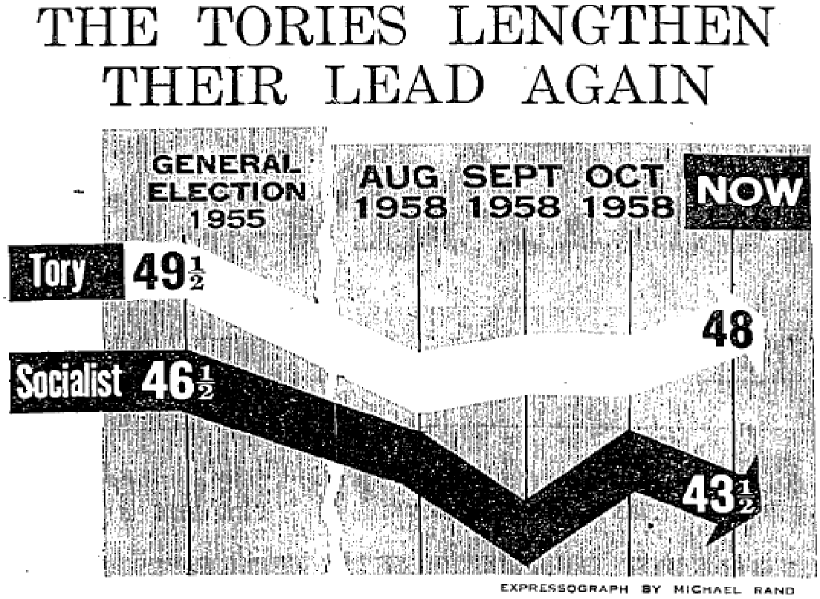


Figure 12

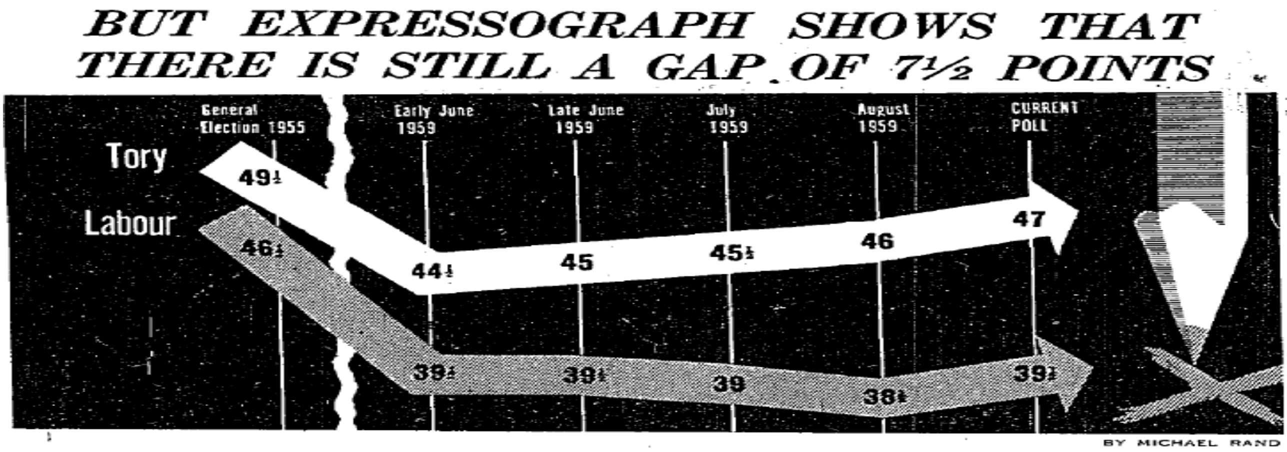


Figure 13

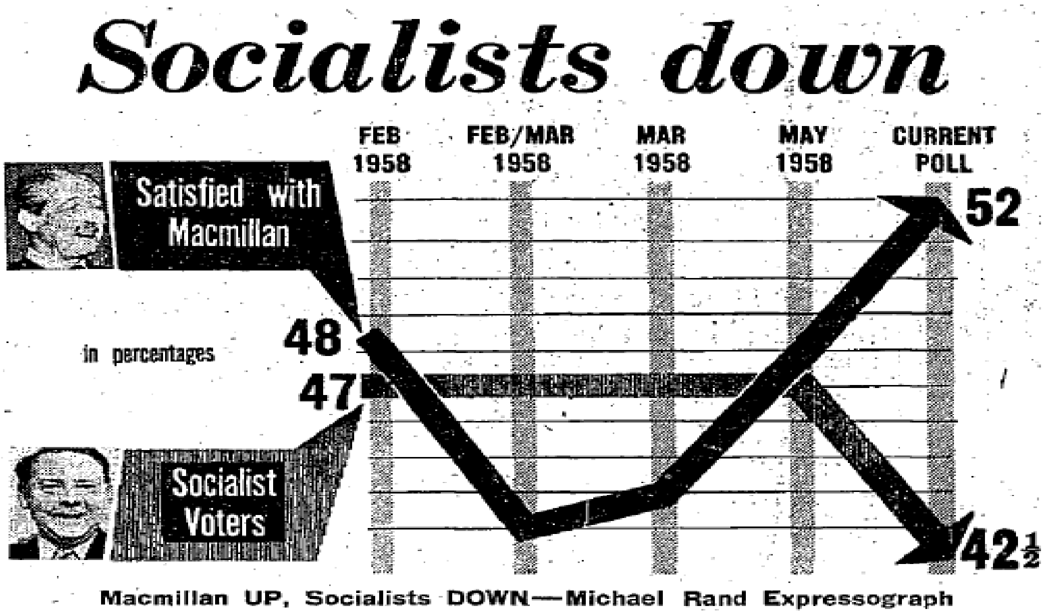


Figure 14

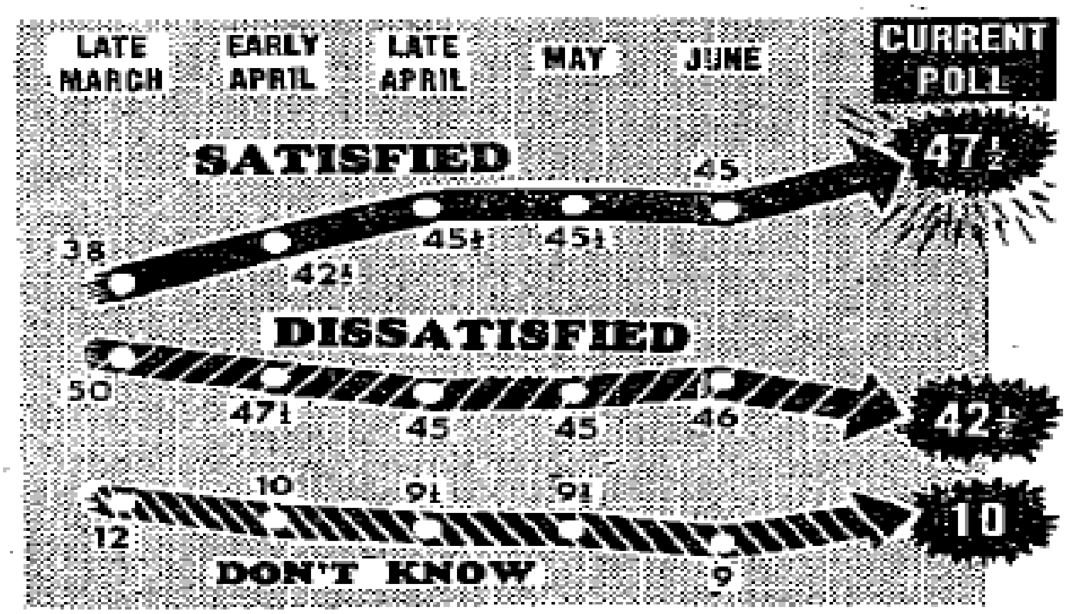


Figure 15

Zelizer, B. and Allan, S. 2010. Keywords in news and journalism studies. McGraw-Hill Education.