

# Exploring Future Challenges for Big Data in the Humanitarian Domain

David Bell <sup>a,\*</sup>, Mark Lycett<sup>b</sup>, Alaa Marshan<sup>a</sup>, Asmat Monaghan<sup>b</sup>

<sup>a</sup> *Department of Computer Science, Brunel University London, Uxbridge, UB8 3PH, United Kingdom; emails:*

[david.bell@brunel.ac.uk](mailto:david.bell@brunel.ac.uk) (\*corresponding author) (Tel: +44 (0)1895 265845)

[alaa.marshan@brunel.ac.uk](mailto:alaa.marshan@brunel.ac.uk)

<sup>b</sup> *School of Management, Royal Holloway, University of London, Egham Hill, Egham, Surrey, TW20 0EX, United Kingdom; emails:*

[mark.lycett@rhul.ac.uk](mailto:mark.lycett@rhul.ac.uk)

[asmat.monaghan@rhul.ac.uk](mailto:asmat.monaghan@rhul.ac.uk)

## Abstract

This paper examines the challenges of leveraging big data in the humanitarian sector in support of UN Sustainable Development Goal 17 “Partnerships for the Goals”. The full promise of Big Data is underpinned by a tacit assumption that the heterogeneous ‘exhaust trail’ of data is contextually relevant and sufficiently granular to be mined for value. This promise, however, relies on relationality – that patterns can be derived from combining different pieces of data that are of corresponding detail or that there are effective mechanisms to resolve differences in detail. Here, we present empirical work integrating eight heterogeneous datasets from the humanitarian domain to provide evidence of the inherent challenge of complexity resulting from differing levels of data granularity. In clarifying this challenge, we explore the reasons why it is manifest, discuss strategies for addressing it and, as our principal contribution, identify five propositions to guide future research.

**Keywords:** Big Data; Veracity; Granularity, Heterogeneous Datasets, Humanitarian, Value

# Exploring Future Challenges for Big Data in the Humanitarian Domain

---

# Exploring Future Challenges for Big Data in the Humanitarian Domain

## Introduction

Aid is typically perceived to be of two basic types, development aid and humanitarian aid. Development aid concentrates on systematic change over the long-term in support of developing countries. In contrast, humanitarian aid concentrates on providing highly time-sensitive and context-driven assistance to address the urgent needs of those affected by disaster. This distinction is increasingly challenged. From a practice perspective, for example, humanitarian actors often extend the boundaries of post emergency aid – potentially merging the original intent and the ultimate application of the aid provided (Buchanan-Smith & Fabbri, 2005; Riddell, 2008). From a policy perspective, the United Nations General Assembly has established 17 interconnected United Nations Sustainable Development Goals (UNSDGs) to extend global efforts beyond the flow of aid (United Nations, 2015c). Inherent in these goals is a need for joined-up thinking, made explicit by UNSDG 17 “*Partnerships for the Goals*” (United Nations, 2015b), which further blurs the line between development aid and humanitarian aid. As noted by the UN Secretary General – “Humanitarian actors need to move beyond repeatedly carrying out short-term interventions year after year towards contributing to the achievement of longer-term development results” (Ki-moon, 2016 p. 32). This is a direction of travel that is promoted by some (Grogan & Strohmeyer, 2016; UN News, 2018) and viewed with trepidation by others (DuBois, 2016).

Implicit in the drive to meet policy goals is an imperative for “quality, accessible, timely and reliable dis-aggregated data” (United Nations, 2015c Para. 48). The UN Secretary General’s Report to the 2016 World Humanitarian Summit repeatedly stresses the importance of systematically tracking and collecting reliable data with deference to privacy as well as the need to securely collate and analyse said data (Ki-moon, 2016 Paragraphs 58, 60, 80, 93, 120, 122–124, 128, 137, 150, 156 and Annexe pages 52, 56, 57–59). In practice, however, such imperatives present significant challenges that hinge around how to extract *value* from big data. In recognising such challenges, UNSDG targets 17.18 and 17.19 note the need for capacity building re (big) data, monitoring and accountability (United Nations, 2015b). Furthermore, and importantly, the UNSDG push on data represents a tacit

acknowledgement that joined-up *thinking* needs joined-up *data* (Lisowska, 2017; Orrell & Lisowska, 2017). Here there is a recognition that the spectrum of domain sources will be of significant variety, volume and/or velocity – i.e., ‘big data’ (Laney, 2001; United Nations, 2015d).

In learning from industrial work in this area, studies note that a primary barrier to achieving value stems from a naive assumption that instrumentation alone is sufficient to deliver value (Lavalle et al., 2011). The value of data and analytics as strategic assets that can help organizations achieve their goals “depends on the questions that leaders ask and the capabilities they are looking to cultivate” (Lavalle et al., 2011). The causal link between analytics and value concepts such as organizational performance is neither clear nor straightforward and this is an issue that has attracted increasing attention (e.g., Huberty, 2015; Popovič et al., 2018; Sharma et al., 2014). Addressing this link, Sharma, Mithas, & Kankanhalli (2014) make several important observations. First, they question an (implicit) assumption in the extant analytics literature that organizations can function as they did before – proposing that decision making and resource allocation processes will need to transform if organizational performance is to benefit from analytics. This is a point that is strongly echoed in Goal 17 (United Nations, 2015b). Second, they propose that insight does not emerge mechanically from the analytical process; rather it comes from an active process of engagement. Third, they propose that the pathway between insight and action is not obvious, involving creative, contextual and psychological factors in many instances. Last, they question a strong belief in the literature that, if the quality of decisions can be improved by analytics, then deriving value from those decisions is trivial – here uncertainties of successfully implementing decisions and of the success of strategic actions are raised.

Underpinning these issues, however, is another that concerns the heterogeneous nature of big data – its generation and sharing mechanisms challenge the technical and expertise-based models that have evolved to guide data practices and organisational intelligence to-date (Constantiou & Kallinikos, 2015). Mismatches in granularity (level of detail) and the inability to predict a priori what data assets/resources may have to be integrated to answer certain questions significantly muddy the analytical picture (Almeida & Calistru, 2012; Sivarajah et al., 2017). In concentrating on value as an important outcome of (big) data analytics, we critically explore the intersection between the points raised above – in essence

to answer the question of *how does the vision of (big) data analytics translate to practice in the humanitarian domain*. The UDSGD and policy on data makes strong demands in this regard and it is important to examine how and where the state-of-the-art falls short. In doing this, the paper is structured as follows. First, we review the big data literature with a particular focus on the causal link between instrumentation, analytics and humanitarian 'value'. Second, we explore that causal link empirically via the synthesis of eight heterogeneous data sets from the humanitarian aid domain. We do this to illustrate the issues that arise from a lack of attention to the need for heterogeneous data integration and the resulting effect that has on the questions that can be asked of such data and subsequent limitation on value. Last, we discuss how the issues we identify can be addressed, presenting a call to arms through five propositions that form the basis of an agenda for future research. It is this call to arms, firmly based in issues arising from our empirical work that provides the principal contribution of the work.

### **The Concept of Value in 'Big' Data**

In 2010, The Economist noted that data are becoming the 'new raw material of business: an economic input almost on a par with capital and labour' (The Economist, 2010). This was an observation on both the inexorable rise in the volume of data being generated and the potential of this data to do things that could not previously be done – unlocking new sources of economic and social value, providing fresh scientific insights and holding governments to account in doing so for example. The potential for 'big data' to generate value across a multitude of domains is echoed in management reports that argue it to be a key basis for competition – providing the foundation for ongoing productivity, innovation and competitiveness (Manyika et al., 2011). Industry interest and investment in big data technologies and skills has been significant as a consequence and, concomitantly, there has been an exponential rise in publications related to big data (see Ekbia et al., 2015).

Rather than having a concrete definition – the understanding of big data revolves around its characteristics (the so called 'V's'), which have expanded to include volume, velocity, variety, veracity, value, visibility, viability and variability (Bedi et al., 2014; Grimes, 2013; IBM, 2013; Laney, 2001, 2012, 2013; White, 2012). While all characteristics are arguably important, value has been highlighted as a key concern. From a business perspective, this concern is targeted at the conversion process between big data (and its associated technologies) and

improved firm performance – proceeding on the assumption that big data adds value indirectly (Ylijoki, 2018). The insight from early studies examining this phenomena clearly connected (effective) analytics with performance – with top performing respondents proposing that it was their use of data and analytics that provided a key differentiator with industry competitors (Lavalle et al., 2011). Findings here (ibid) were that top performers (as defined): (a) Made decisions based on rigorous analysis; (b) were twice as likely to use analytics to guide future strategies than poor performers; and (c) were twice as likely to use analytics to guide day-to-day-operations than poor performers. In the extant literature, the way in which value manifests itself includes improved product development and faster time-to-market cycles in product/service innovation cycles (Manyika et al., 2011); inventory accuracy (Bärenfänger et al., 2014); delivery accuracy (Dutta & Bose, 2015); improved customer relationship management (McAfee & Brynjolfsson, 2012).

In the world of commerce, the perception of value from analytics is ultimately linked to the continued viability and success of a commercial entity (H. Chen et al., 2012; Sharma et al., 2014). In the humanitarian aid sector, however, the value proposition of big data is inextricably linked to human survival and well-being (Kamper, 2019; Meier, 2015). In this sector, amongst other value-laden possibilities, big data analytics is seen as offering the means to overcome inherent weaknesses in humanitarian aid supply chains by improving visibility and coordination and thereby yielding the intrinsic value of saving lives and reducing the suffering of survivors (Dubey et al., 2018). Additionally, before, during and after a crisis event, big data analytics holds the promise of descriptive (e.g. crises mapping), prescriptive (e.g. routing) and predictive (e.g. shelter planning) solutions across humanitarian operations (Lycett & Monaghan, 2013; Swaminathan, 2018). Nonetheless, while the value dimension of business analytics may be anchored to human needs in the humanitarian domain, the underlying concepts, constructs, paradigms and limitations of business analytics remain applicable (Gandomi & Haider, 2015; Sharma et al., 2014).

Researchers have suggested that business analytics can help organisations to better compete in the marketplace by understanding its needs and leveraging opportunities embedded in large amount of data to make fast and informed decisions (H. Chen et al., 2012; Lavalle et al., 2011). Studies have also noted, however, that the causal link between analytics and value concepts such as organizational performance is neither clear nor

straightforward. As noted in the introduction, Sharma et al. (2014) argue that business performance gains are not a direct result of business analytics: Instead, value is an outcome of the decisions that are based on meaningful insights that can be generated from data (Gandomi & Haider, 2015; Sharma et al., 2014): Perhaps, more properly, it is the actions resulting from such decisions. The impediments to direct performance gain are many, including: Developing organisational strategy appropriate to big data (Constantiou & Kallinikos, 2015; Elgendy & Elragal, 2014; Gandomi & Haider, 2015; Gantz & Reinsel, 2011); organisational social capital and dynamic capabilities (Blyler & Coff, 2003); the quality of decision making (Sharma et al., 2014); data quality (Hazen et al., 2017; Merino et al., 2016; Vidgen et al., 2017) and data maturity (Anand et al., 2016; Comuzzi & Patel, 2016); resource availability (T. Davenport, 2014; Janssen et al., 2017); organizational culture (Anand et al., 2016; Sandberg & Aarikka-Stenroos, 2014); privacy and security concerns (Clarke, 2016; Newell & Marabelli, 2015); and regulations (Keen et al., 2013; Truyens & Eecke, 2014).

These nuances of transforming data into value are developed as a process model by Ylijoki and Porrás (2018), to better explain how big data can be transformed into economic value (see Figure 1). The model is founded in the observation that, with the exception of value measurement, data meets the International Financial Reporting Standard (IFRS) criteria for an intangible asset – in that data are ‘identifiable, non-monetary, non-physical, potentially valuable resources that a firm produces or harvests from different sources’ (Ylijoki, 2018 p.73). The four key sub-processes in this model are as follows:

- Asset creation process. This sub-process refers to what Lycett (2013) discusses as the dematerialisation aspect of datafication – essentially, the ability to separate the informational aspect of an asset/resource and its use in context from the physical world. Data are not objectively ‘out there’ and harvesting them is not quite the same as picking fruit from trees – thus the conceptualisation and codification of data is important (e.g., via schemas), including metadata aspects such time and location that relate to contextualisation (Piccoli & Pigni, 2013).
- Capability creation process. This sub-process refers to the need to develop resources and capabilities that (dynamically) support the data asset creation. First, via investment in supporting technology assets (e.g., infrastructure) that help automate dematerialisation and the storage of resulting data assets spanning data cleansing,

storage and distribution for example. Second, developing human-centric analytic capabilities (which may be embedded in algorithmic implementation) that are important in turning information into knowledge (Mithas et al., 2011). Third, developing innovation capabilities (Dyer et al., 2009) to provide the conditions necessary for effective digital transformation, including aspects such as information management, governance, dynamic resource allocation etc.

- **Transformation process.** In this sub-process, firms combine their assets, capabilities and knowledge to produce outcomes of value (of the forms discussed above). This refers to what Lycett (2013) describes as ‘densities’ in the datafication process – the (re)combination of assets, resources and capabilities, mobilised for a particular context of application/use. Potential impacts here are strongly influenced by the innovation capabilities, which may serve as a seedbed for new business models as we discuss in more detail below.
- **Completion process.** This sub-process addresses the earlier point that (big) data value is indirect and impacts need to be linked to concrete performance metrics. Firm performance depends on competition, which connects a firm to its wider ecosystem – e.g., the competitors, suppliers and customers relevant within its prescribed industry segments. While (big) data impacts might be a necessary condition for firm performance, they are not sufficient as the competition process is subject to a number of intervening factors.

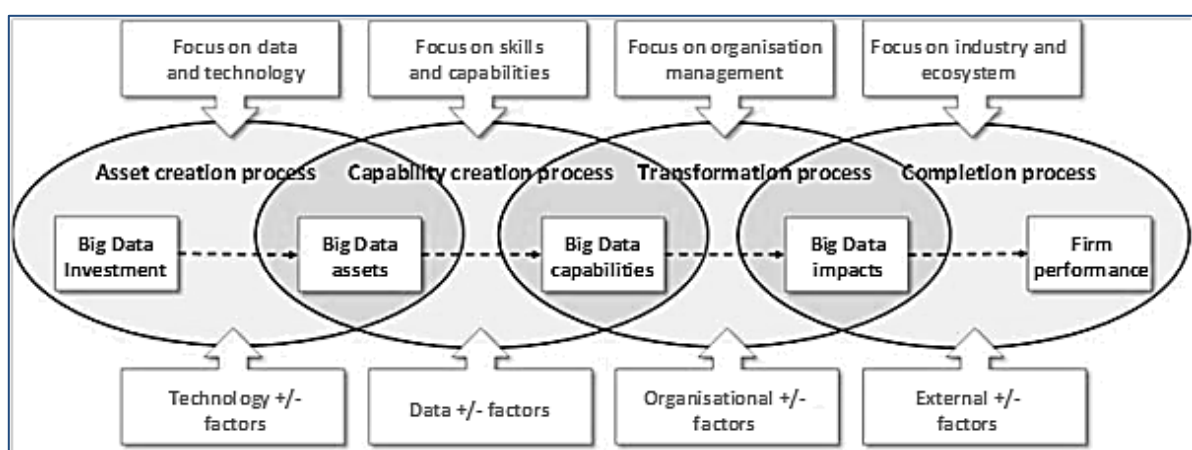


Figure 1: Value creation with big data (adapted from Ylijoki & Porras (2018))

The transformation sub-process highlights the importance of big data in the creation of new business models (e.g., Davenport, Barth, & Bean, 2012; Oestreicher-Singer & Zalmanson,



2013). In its idealised form, big data sees organisations feeding from ecosystems of digital data (e.g., Constantiou & Kallinikos, 2015; Davenport, 2014). Business models are fundamental as, crudely, they determine how an organisation creates and captures value (Chesbrough, 2010) and are seen as a boundary-spanning unit-of-analysis (Zott & Amit, 2010) – describing the system of interdependent activities that are performed by a focal firm and its partners alongside the mechanisms that link such activities together.

In this context, Huberty (2015) makes an important observation re the process of refining data into business value. The core of his argument is that the successful business models that have been built on big data (e.g., Google, Amazon, Facebook etc.) are successful because: (a) They are self-referential – they use data about how people search, shop etc. to create or improve the focal firm’s services related to searching and shopping etc.; and (b) the consequences of failure are fairly low because, importantly, they can be internally managed by the focal firm concerned. In reality, therefore, big data appears more as an evolution than revolution – improving on the existing business models of a focal firm via description, prediction and inference rather than creating radically new business models. The point made (ibid) is that this happens because “having promised a first-order world, big data has delivered a third-order reality” (p.42). Value is not created from the direct exchange between producer and consumer but, rather, via a web of transactions that are removed from the creation of the data itself by several orders. His conclusion is that the raw materials required for big data engineering solutions need to originate from a first-order approach to studying and understanding the world.

This is an important point, as it illustrates that some distance remains between the vision of big data and the current reality – it does not transcend borders quite in the manner that capital and labour currently do and is not as reusable as ‘new oil’ metaphors might suggest. Boyd & Crawford (2012) argue that big data is not of note because of its size but, rather, because of its relationality – its value comes from the patterns that can be derived from combining different pieces of data. Indeed, the value of big data for innovation increases when datasets are combined across organizations and sectors (van den Broek & van Veenstra, 2018). Huberty’s points arguably indicate foundational difficulties in this relationality however. In its idealised form, big data is somewhat haphazard in nature, very heterogeneous and, often, trivial, messy and agnostic (Anderson, 2008). In the value

creation process noted above (and in the asset sub-process in particular), big data can be made business relevant, but this relevance may not derive in a straightforward manner from the original data records – in the majority of circumstances, big data is produced via a very heterogeneous user base and its generation and sharing mechanisms challenge the technical and expertise-based models that have evolved to guide data practices and organisational intelligence to-date (Constantiou & Kallinikos, 2015).

The literature is waking up to these challenges noting, for example, that the heterogeneity of data and mismatches in granularity among and between internal and external (e.g., online) resources present significant data integration issues for big data projects (Almeida & Calistru, 2012; Elgendy & Elragal, 2014; Sagiroglu & Sinanc, 2013; Sivarajah et al., 2017). As data could be of any format, structured or unstructured, extracting appropriate metadata from the underlying resource is a necessary task (Sivarajah et al., 2017) – but heterogeneity poses a considerable challenge related to the automated generation of such metadata (Jagadish et al., 2014). In turn, heterogeneity necessitates the careful integration and aggregation of noisy, dynamic and inconsistent datasets (Khan et al., 2014) for meaningful representation (Sivarajah et al., 2017). Such process, however, often yields derived data that can pose additional challenges in tracking and representation (Jagadish et al., 2014). Thus demands are placed on capabilities related to the representation and analysis of big data and its interpretation in given contexts (Günther et al., 2017; Jagadish et al., 2014). Similarly, consideration needs to be given as to what to record and store (alongside what to forget) and deriving the measures that define the value of data itself presents a challenge (Jagadish et al., 2014). While such challenges are emerging, however, there is little to no work that examines their effects empirically and draws together strands of work that may present solutions.

### **Big Data in the Humanitarian Domain**

It is a valuable exercise to examine such challenges in practice and understand their implications. In doing so, we draw on data from the humanitarian domain where the consequences of failure are significant and the responsibility for success is spread across a multitude of actors. We do this by attempting to integrate multiple datasets in that domain to examine whether global data on the occurrence and impact of humanitarian crises, and the monies that flow in response to these crises, can reveal the impact of given aid. This is

an important question in the domain because tens of billions of dollars are spent every year in the humanitarian response to disasters (Lattimer et al., 2016), but shortfalls prevail – resulting in decisions to stop or cut humanitarian relief programmes (van der Zee, 2015). Resource limitations heighten pressure to achieve more with less humanitarian funding which, in turn, increases focus on measuring the impact of the humanitarian aid given – with a view to using this knowledge to improve the likelihood of survival and address the welfare of victims in future disasters (Moorhead & Sandler Clarke, 2015; Purvis, 2015; Völz, 2005). An assessment of the impact of humanitarian disasters also underpins initiatives in disaster risk reduction and their relationship to UN Sustainable Development Goals (IAEG-SDGs, 2017; UNSD, 2018; Wahlström, 2015).

Methodologically, we followed a standard data science lifecycle of obtaining data, scrubbing (cleaning the data), exploring the data to understand its properties and limitations, model construction and interpretation (e.g., see Lau, 2019). At the outset, the following six sources were selected because they hold differing disaster loss data, based on disaster type or scale:

1. Emergency Events Database (EM-DAT) for humanitarian crises most of which are categorised as *natural* or *technological* disasters (Guha-Sapir et al., 2017);
2. Disaster Inventory System (DesInventar), which holds disaster types that are believed to be of a similar ilk to those in EM-DAT, but unlike EM-DAT, DesInventar does not restrict the data collected based on inclusion criteria. Additionally, DesInventar is an open source solution that is deployed in multiple countries or districts, therefore it is a collection of databases and not centrally curated (DesInventar, 2017);
3. Uppsala Conflict Data Program (UCDP), which provides a perspective of conflict-type global humanitarian crises for all countries except Syria, which is excluded because of unverifiable sources (UCDP, 2017);
4. Violations Documentation Center in Syria, which plugs the gap in UCDP by providing data of the human effect of the conflict in Syria (VDC-SY) (VDC-SY, 2016);
5. Global Terrorism Database (GTD), which provides a perspective of global terrorism-type humanitarian crises and the losses caused by these crises (GTD, 2017);
6. United Nations High Commissioner for Refugees (UNHCR), which provides a view of people uprooted from their homes and/or their countries (UNHCR, 2017).

The remaining two sources selected offer humanitarian aid flow data:

7. Financial Tracking Services (FTS), which tracks the flow of international humanitarian aid and is managed by the United Nations Office for the Coordination of Humanitarian Affairs (FTS, 2017; UNOCHA, 2017).
8. International Development Statistics (IDS) online databases, which also tracks humanitarian aid and is managed by the Organisation for Economic Co-operation and Development (OECD, 2017b).

Scrubbing and data exploration involved rationalising, reclassifying, consolidating and aggregating data from these sources into a cohesive whole. This inevitably meant sacrificing detail and nuance as potentially valuable data points, discarded for consistency and coherence and reclassifications and summations, concealed the possibility of useful micro-patterns in the data. Figure 2 illustrates the steps taken to build a coherent view of disasters and marry these integrated data to available humanitarian aid data. At each step, visibility was lost, either because the source withholds potentially valuable information *or* the process of integration necessitates decisions in favour of compatibility and coherence.

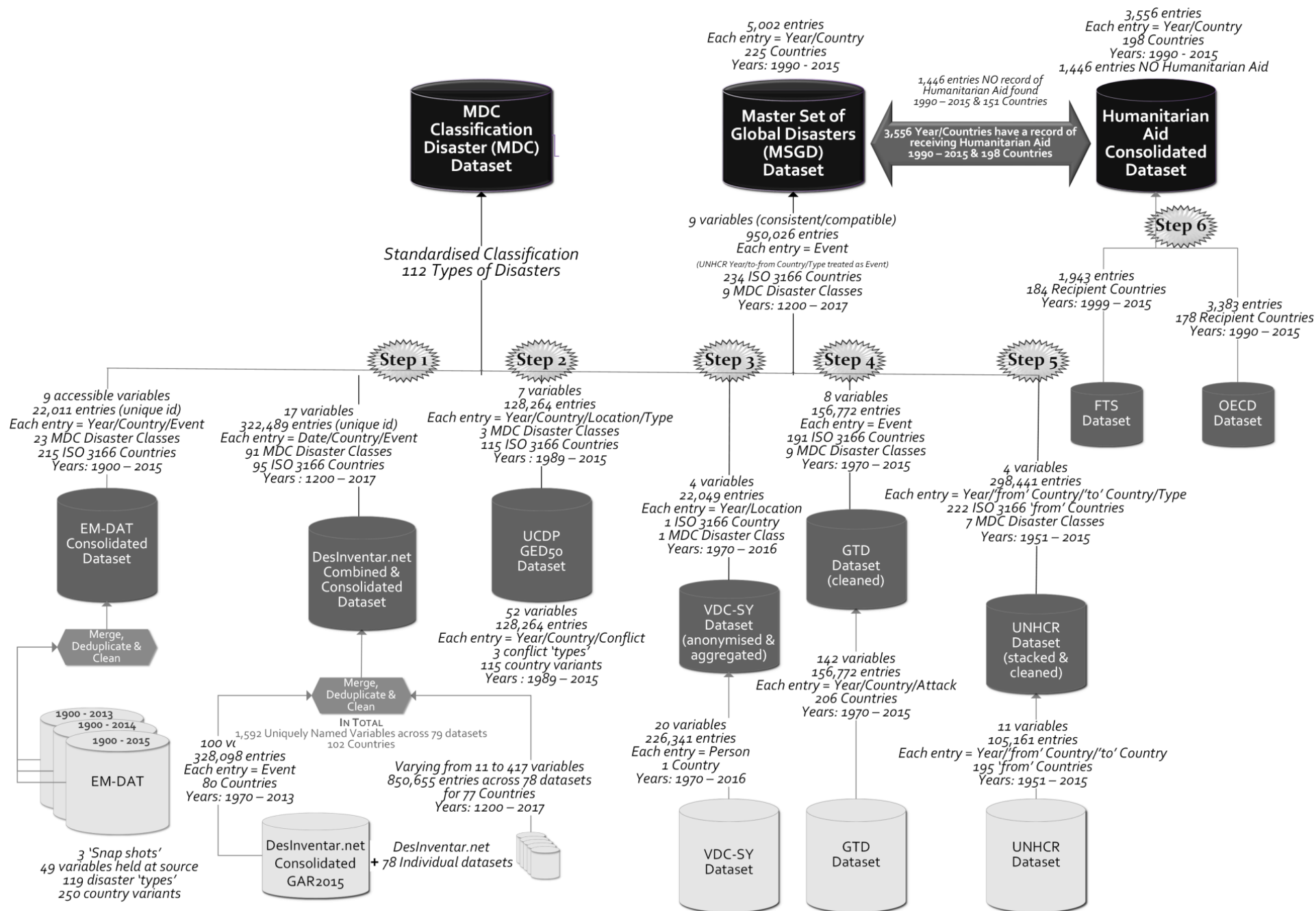
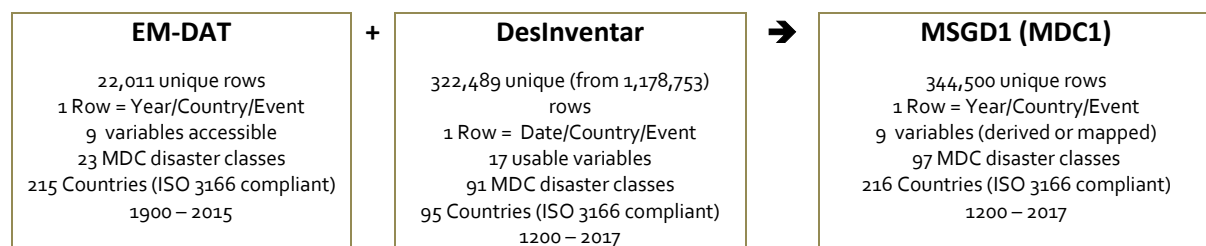


Figure 2: Bringing Together Disaster Data and Humanitarian Aid Data

To demonstrate the scale of data abandoned or modified in pursuit of conformity and cohesion we elucidate the steps taken in synthesising the datasets.

### Step 1: Developing a Global View of Disasters caused by Nature or Accidents

This step entailed bringing together EM-DAT and the DesInventar datasets to create an initial Master Set of Global Disasters (MSGD) and a standardised Master Disaster Classification (MDC) dataset (DesInventar, 2017; Guha-Sapir et al., 2017). The latter of these, the MDC, being based on the Integrated Research on Disaster Risk Peril Classification and Hazard Glossary (IRDR, 2014).



**Figure 3:** Humanitarian Crises caused by Nature or Accidents → Baseline MSGD and MDC

Notably these two datasets were combined to provide a more comprehensive view of disasters cause by nature as both datasets have a limited overlap in the scale of disasters recorded. Ironically, the data loss in combining these datasets was significant and in showing this we classify the type of loss. Losses from EM-DAT:

- Only 9 accessible/usable variables from 49 documented variables (*loss of visibility*).
- Only *Year* is ‘reliably’ usable temporal data (*loss of specificity*).
- 119 disaster ‘types’ standardised to 23 MDC classifications (*loss of nuance*).
- 250 country variants standardised to 215 *ISO 3166* country codes (ISO, 2017) (*loss of historical changes to sovereignty/convention*).

Losses from DesInventar:

- Only 17 usable variables from the 1,592 uniquely named variables across the various datasets (*loss of visibility and nuance*).
- Date is not consistently populated and only *Year* is ‘reliably’ usable temporal data (*loss of specificity*).
- *Deaths* and *Missing* combined to ‘Total Deaths’ for compatibility (*loss of specificity and detail*).

- Injured, Victims, Affected, Relocated and Evacuated combined to ‘Total Affected’ for compatibility (loss of specificity and detail).
- 75,631 disaster classes/descriptions standardised to 91 MDC classifications (*loss of nuance*).
- 102 country variants standardised to 95 ISO 3166 country codes (ISO, 2017) (*loss of historical changes to sovereignty/convention*).

### Step 2: Expanding the Global View of Disasters to include Conflict-Related Disaster

This step entailed integrating the UCDP dataset with the MSGD dataset and expanding the MCD dataset created in Step 1 as the previously combined disaster impact datasets do not routinely record humanitarian crises caused by conflict (UCDP, 2017).

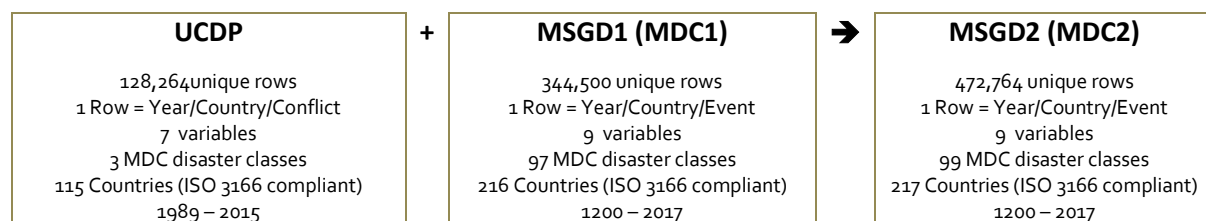


Figure 4: Humanitarian Crises caused by Conflict ➔ Expanded MSGD and MDC

Again, the data loss in adding the UCDP data was significant:

- Only 7 the 52 variables available were usable (*loss of visibility and nuance*).
- Two usable variables hold deaths-related information, *best\_est* and *high\_est*, if *best\_est* = 0 then *high\_est* is used (*loss of visibility*).

### Step 3: Expanding the Global View of Disasters to Include the Conflict in Syria

- This step entailed integrating the VDC-SY dataset with the MSGD dataset and expand the MCD dataset created in Step 2 (VDC-SY, 2016). This dataset was needed because the main conflict dataset, UCDP, explicitly excludes the human impact of the current on-going conflict in Syria (Croicu & Sundberg, 2015).

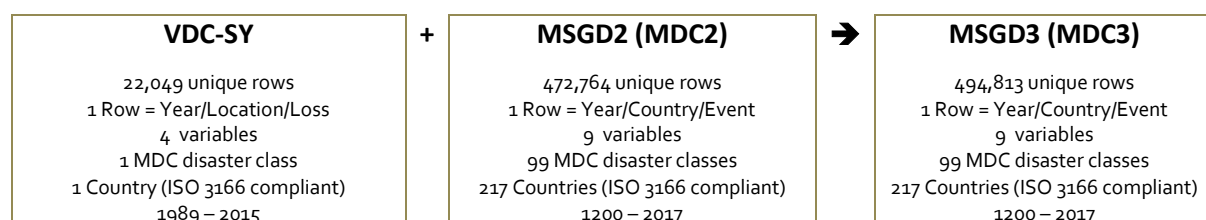


Figure 5: Humanitarian Crises caused by the Conflict in Syria ➔ Expanded MSGD

The data loss in adding the VDC-SY data was as follows:

- The 226,341 entries in VDC-SY are at a ‘*per person*’ level, this needed to be aggregated and anonymised to create 22,049 entries (*loss of specificity and detail*).
- Only 4 variables of 20 available variables used, of these 4 two are aggregations – Total Deaths and Total Affected to allow compatibility (*loss of visibility, specificity, detail and nuance*).
- The type of victim, i.e. *killed, killed, detained* mapped to 1 MDC disaster class (*loss of specificity, detail and nuance*).

#### Step 4: Expanding the Global View of Disasters to Include Acts of Terror

This step entailed integrating the GTD dataset with the MSGD dataset and expand the MCD dataset created in Step 3 (GTD, 2017). This dataset was needed because UCDP, in which an event represents an action between two actors, does not include terrorist attacks (Croicu & Sundberg, 2015).

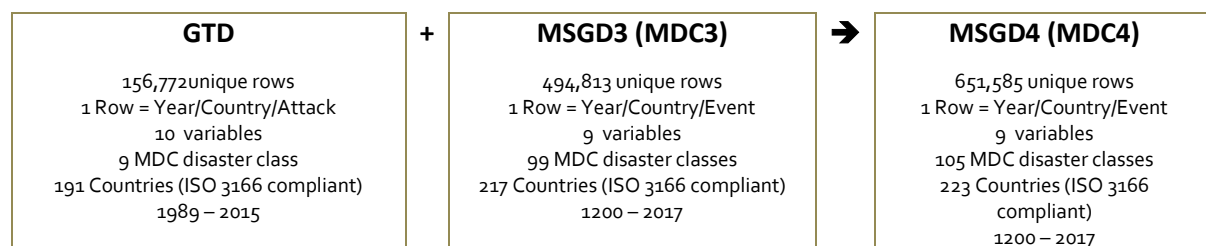


Figure 6: Humanitarian Crises caused by Acts of Terror → Expanded MSGD and MDC

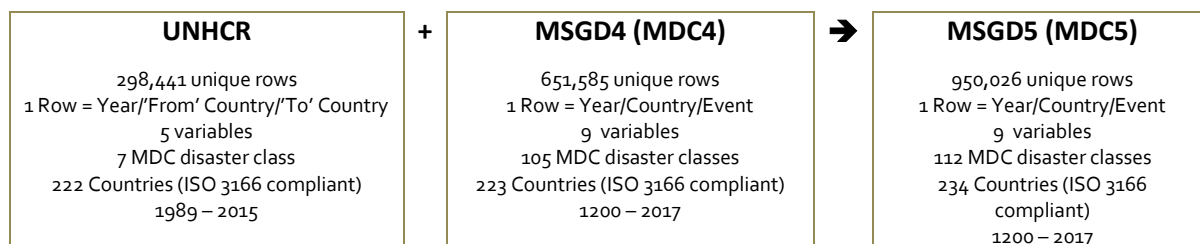
The loss in adding the GTD data was as follows:

- Only 10 variables of 142 available variables are usable due to compatibility (*loss of visibility*).
- 119 types of terrorist attack mapped to 9 MDC disasters classes (*loss of specificity, detail and nuance*).
- 206 country variants standardised to 191 *ISO 3166* country codes (ISO, 2017) (*loss of historical changes to sovereignty/convention*).

#### Step 5: Expanding the Global View of Disasters to Include Refugee Crises

This step entailed integrating the UNHCR dataset with the MSGD dataset and expand the MCD dataset created in Step 4 (UNHCR, 2017). This dataset was needed as all the disaster impact datasets integrated up to this step do not include the humanitarian crises of deracination.





**Figure 7:** Humanitarian Crises of Refugees → Expanded MSGD and MDC

The loss in adding the UNHCR data was as follows:

- Only 5 variables of 11 available variables are usable due to compatibility (*loss of visibility*).
- 195 country variants standardised to 222 *ISO 3166* country codes (ISO, 2017) (*loss of historical changes to sovereignty/convention*).

#### **Step 6: Integrating Humanitarian Aid Flow Data to Evaluate against Disaster Loss Data**

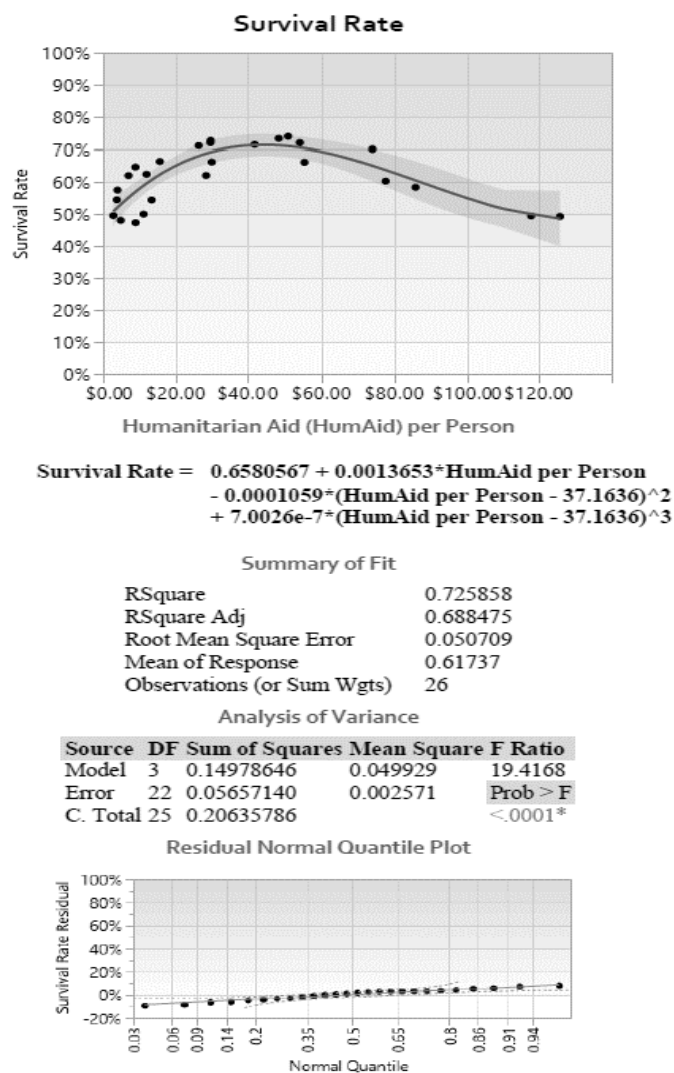
The final step required integrating the humanitarian aid data from FTS and OECD (FTS, 2017; OECD, 2017b). The data from these sources are not provided at a level or with the references that would reconcile easily with humanitarian crises data. The version of OECD data that identified recipient countries can only be obtained as annualised values: As a result, for compatibility, the FTS data was also taken at the year/recipient country level for consistency. Even consolidated, these two datasets only provided a view of the annualised flow of humanitarian aid monies to 198 countries from 1990 to 2015, a total of 3,556 records.

Ultimately, to examine the impact of humanitarian aid involved rationalising, reclassifying, consolidating and aggregating data from multiple sources into a cohesive whole of 950,026 records that were in and of themselves distilled views of disaster data. To marry the available humanitarian aid data to this distilled view of disaster data required further loss of detail through aggregation of losses to the year/country level and a restricted view of years 1990 – 2015, resulting in 5,002 records. Of these another 1,446 records had to be discarded as they do not have any corresponding flows of humanitarian aid.

Returning to the research question of empirically assessing the impact of aid by finding patterns in integrated disaster effect and humanitarian aid data, this question was constrained to 3,556 aggregated records limited to year, country and additive variables that can be aggregated. Even with these records, anomalies were found in humanitarian aid data

with some country/year combinations exhibiting disproportionate aid flows. For example, 30 year/country combinations appear to have received aid of over US\$250,000 per person affected by disasters and 146 year/country combinations appear to have received aid of less than US\$1 per person affected by disasters. This suggests a possible misallocation of aid flows that makes analysis even at the year/country level of aggregation non-viable. In the end this dataset was used to calculate mean survival rate (*people affected expressed as a percentage of total human effect, i.e. killed and affected*) against humanitarian aid per person (*calculated as total humanitarian aid divided by total human effect*).

Figure 8 displays a scatter plot and polynomial line of fit of mean survival rate and humanitarian aid per person. R<sup>2</sup>, F-Ratio, and Prob>F indicate that the polynomial line of fit is statistically significant.



**Figure 8:** Mean Survival Rate x Humanitarian Aid per Person  
Polynomial Line of Fit (Degree=3)

From the plot it appears that the *mean survival rate* peaks at 70% with ~US\$50 per person and then steadily falls to 50% as humanitarian aid moves towards ~US\$120 per person.

From this model it appears that increasing humanitarian aid above US\$50 per person does not necessarily result in a higher likelihood of surviving a disaster. This is worthy of further investigation, which in turn accentuates the data losses that had to be accepted when integrating datasets and the role reconcilable and granular data could have played in explaining the model. As a result, it can reasonably be argued that only a fragment of the much of the hoped for value proposition of analysing this data is realised.

### **Five Core Research Propositions (RPs)**

In discussing datafication, Lycett (2013) concludes that it is a technology-driven *sense-making* process, two interlinked aspects of which have particular salience – plausibility and enactment. Plausibility is embodied in accounts (Maitlis, 2005), which construct order among sets of entities (e.g., data assets), making tangible a perceived reality that can be enacted, e.g., via Ylijoki & Porras's (2018) process model. The nature of big data hampers plausibility, however, and the empirical example provides ample illustration of the noisy, dynamic and inconsistent nature of the datasets that needed to be integrated in attempting to answer the research question posed (i.e., whether global data on the occurrence and impact of humanitarian crises, and the monies that flow in response to these crises, can reveal the impact of donated aid). This, in point of fact, is the limitations of relationality at work (Boyd & Crawford, 2012a) as the account of value here is severely limited by the foundational differences of the data available. In short, the innovative possibilities and/or insights generated by big data in the humanitarian domain are left wanting. As a consequence, the ability for the humanitarian domain to learn and predict from data is an obstacle to the UN call to contribute to the achievement of longer-term sustainable development.

In coming face-to-face with the limitations of relationality, we have illustrated that a number of different types of 'loss' were experienced. These types were as follows and their scope shown in Table 1:

- a) *Detail, e.g. when records were aggregated to achieve compatibility;*
- b) *Visibility, e.g. when variables were abandoned for the sake of consistency;*
- c) *Specificity, e.g. when dates were discarded for the sake of cohesion;*

- d) Nuance, e.g. when variables were merged and events reclassified for the sake of comparability;
- e) History, e.g. when historically accurate country names were updated for current-day relevance.

Data Source	Loss of...				
	Detail	Visibility	Specificity	Nuance	History
	e.g. because of....				
	aggregation	discarded variables	discarded dates	grouping & reclassification	discarded geopolitical history
EM-DAT		✗	✗	✗	✗
DesInventar	✗	✗	✗	✗	✗
UCDP		✗		✗	
VDC-SY	✗	✗	✗	✗	
GTD	✗	✗	✗	✗	✗
UNHCR		✗			✗

Table 1: Scope of loss in pursuit of data relationality

These losses stem from the fact that the individual datasets concerned were not designed (and/or developed) with external integration in mind: Thus, our key point of learning from the empirical work is that, in the context of the UNSDGs, the asset creation and capability creation processes (at least) need to be considered and addressed from an ecosystem perspective, rather than from the perspective of a focal organisation (where transformation and completion processes are more likely to be enacted). Unsurprisingly perhaps, the big data vision of ‘ecosystems of digital data’ requires that assets and capabilities (to some degree) are addressed from that same perspective if (potential) value is to be maximised. An added challenge in this respect, however, is that the ways that big data assets might be used are not necessarily known at the time the assets is created – that challenge places additional burden on capabilities.

Thus, to guide future research in this area and as our principal contribution, we outline five research propositions (RPs) in Table 2 that we see as core research propositions. In relation to the value process model (Figure 1), we restrict these propositions to the first two of the value creation sub-processes, namely:

- Capability Creation: as related to the *governance of data assets* such that they can be trusted and managed outside their initial context of use; and
- Asset Creation: as related to the *construction of data assets* such that they can be shared effectively outside their initial context of use.

We then discuss the (limited) existing work that is relevant to each proposition to provide a platform upon which future research can build.

RP	Research Proposition	Sub-Process		Literature
		Capability	Asset	
1	<p><b><u>Ecosystem Data Governance</u></b></p> <p>Models of governance need to be developed that are appropriate to the sharing of data across humanitarian ecosystems</p>	✓		(Chatterjee & Ravichandran, 2013; Hansen & Porter, 2017; Lycett & Monaghan, 2013; Marelli et al., 2017; Provan & Kenis, 2008; Riddell, 2008; Stephenson, 2005; The Lancet, 2010; United Nations, 2015d; van den Broek & van Veenstra, 2018)
2	<p><b><u>Evaluating Data Veracity</u></b></p> <p>The veracity of data assets requires transparent evaluation if trust is to be established and maintained.</p>	✓	✓	(Berti-Equille & Borge-Holthoefer, 2015; Berti-Equille & Lamine Ba, 2016; Boyd & Crawford, 2012b; Clarke, 2016; Claverie-Berge, 2012; D’Mello, 2016; Grimes, 2013; Laney, 2013; Lee et al., 2002; Lukoianova & Rubin, 2014; Normandeau, 2013; Powers Dorette, 2016; Rubin & Vashchilko, 2012; Schroeck, 2012; Strong et al., 1997; Wand & Wang, 1996; R. Y. Wang & Strong, 1996)
3	<p><b><u>Value and Data Privacy</u></b></p> <p>Mechanisms are required for addressing the conflict between value in the humanitarian domain and the privacy, autonomy and liberty of individuals</p>	✓		(Boyd & Crawford, 2012b; Cooper, 2015; Cukier, 2019; Dubey et al., 2018; Ki-moon, 2016 Paragraphs 58, 60, 80, 93, 120, 122–124, 128, 137, 150, 156 and Annexe pages 52, 56, 57–59; Lycett & Monaghan, 2013; Marelli et al., 2017; Nissenbaum, 2004, 2010, 2018; Privacy International, 2018; Raymond et al., 2016; Swaminathan, 2018; UNHCR, 2017; United Nations, 2015a; VDC-SY, 2016; Winter & Davidson, 2019)
4	<p><b><u>Data Transparency</u></b></p> <p>Mechanisms are required to manage and maintain data transparency across ecosystems.</p>	✓	✓	(EDRIS, 2017; FTS, 2017; Global Campaign for Aid Transparency, 2017; HDX, 2017; HXL, 2019; IATI, 2017; IRDR, 2014; Lattimer et al., 2016; Maiers et al., 2005; Mayer et al., 2016; O’Neill, 2018; OECD, 2017a; Proctor, 2006, 2008; Riddell, 2008; Tatham & Hughes, 2011; The Lancet, 2010; William & Mary University, 2018)
5	<p><b><u>Data Granularity and Boundaries</u></b></p> <p>Data assets require well-defined boundaries and clear interfaces to the data they contain to deliver data at clearly defined levels of granularity</p>		✓	(Baldwin & Clark, 1997; Pal et al., 2015; Reyes-Galaviz & Pedrycz, 2015; G. Wang et al., 2017; Yao, 2000).

**Table 2:** Five Core Research Propositions

## RP 1: Ecosystem Data Governance

Our empirical example illustrates that big data in the humanitarian and developmental domains involves relationships that span organisational and national borders that, aside from the material data assets discussed, also tread upon many norms and ideologies (Hansen & Porter, 2017). Moving toward the UN 2030 vision requires bringing together big data from diverse jurisdictions and deploying the outcomes of data analysis transnationally (United Nations, 2015d). In broad terms, big data in the humanitarian domain makes strong demands on our current conceptions of data governance. Existing work that addresses inter-organisational governance defines it as – *the arrangement of institutions and structures ensuring that individual parties behave in line with collective goals with means of conflict resolution in place and collective resources effectively and fairly used* (Provan & Kenis, 2008; van den Broek & van Veenstra, 2018). In achieving the promise of big data, and in the empirical case here however, those (implicitly) participating in data integration have no formal authority over each other (Chatterjee & Ravichandran, 2013) and the ways that their (open) data assets were combined in the attempt to produce insight/innovation were not definable a priori. Arguably, therefore, work is required on models of data sharing/governance that are appropriate from an ecosystem perspective.

The humanitarian ecosystem as a whole can be argued to contain three core groups of actors defined by their relationship to the flow of aid: (1) Donors, the governmental and multilateral agencies, as well as the individuals, who are the sources of aid; (2) executors, the non-governmental organisations and not-for-profit entities who compete for aid and on receiving it control its distribution and allocation; and (3) beneficiaries, the victims of disaster who are in need of aid. Unsurprisingly, the data assets in this ecosystem are predominantly generated, collated, held and selectively shared by actors from Groups (1) and (2). Now consider some of the relevant characteristics of this ecosystem: Vying for aid (a finite resource) from Group 1 (donors) by Group 2 (executors) to address the needs of a seemingly limitless Group 3 (beneficiaries) creates a dynamic that engenders competition when sourcing *supply* (i.e. aid and resources) and not when securing *demand* (i.e., the needs of victims) (Lycett & Monaghan, 2013). Group 2 (executors) must therefore strive to distinguish themselves – e.g., through reputation, impact, innovation, trust, transparency, flexibility etc. – to be seen as worthy of being entrusted with other people’s money to

satisfy the needs of those struggling to survive (Riddell, 2008; Stephenson, 2005; The Lancet, 2010). Consequently, this is not an ecosystem that is conducive to data sharing when there is a possibility of competitive disadvantage in the acquisition of aid and resources.

Other factors likely to affect the fit of *a humanitarian ecosystem data governance model* are the need for privacy and the effect of power. While there is a genuine need for data privacy (and security), especially where the data pertains to the extreme vulnerabilities of the victims of disasters that it may help locate and identify (Marelli et al., 2017), the effect of power can be less obvious. When it comes to money, common sense dictates that the flow of power correlates with the flow of aid (funds), with the source of aid – i.e., Group 1 (donors), having the most power and Group 3 (beneficiaries) having the least. Data assets however can obscure the power of those who: (a) Create them; (b) intervene to recalibrate them; (c) use them to reinforce and/or leverage their power; or (d) use the data that is generated to extend their power into new spaces (Hansen & Porter, 2017). Arguably, and likely unwittingly, the limitations of the data assets integrated in our empirical case provides good example of the exercise of such power that limits the greater good in relation to the analysis and action in the humanitarian domain.

Governance models from an ecosystem perspective have received little consideration to date. In inter-organisational situations the extant literature points to four potential models of governance that differ in their balance between the need for control (e.g. privacy and security; development and enforcement of standards) and fostering innovation (e.g. flexibility; collaboration; independence). Table 3 maps facets of an ecosystem to the types of data governance models that may suit.

Facets	Type of Model			
	Market	Bazaar	Hierarchy	Network
<b>Normative Basis</b>	Intellectual property	Open license	Bureaucracy/ centralised control	Social contracts
<b>Incentives</b>	Competition	Community reputation	Power/Position	Trust
<b>Control</b>	Moderate/Mixed: dyadic contracts	Low: reputation in the community	High: administrative power	Moderate: reciprocity and social contracts
<b>Reasons for Adoption</b>	Low coordination costs; high flexibility in participants	Innovation and low coordination costs	Negotiation position; strategic differentiation	Low cost access to resources; joint solutions

<b>Flexibility of the Collaboration</b>	High	High	Low	Moderate
<b>Duration</b>	Short-term	Unlimited	Unlimited	Long-term
<b>Social Contract</b>	Formal, distrust	Informal, focus on public value	Formal, bureaucratic	Informal, focus on common objectives
<b>Relation between Network Members</b>	Independent	Independent	Dependent	Interdependent

**Table 3:** Inter-organisational Governance Arrangements (from van den Broek & van Veenstra, 2018)

The humanitarian ecosystem is not a neat fit to any of the four models. For example:

- The need for control in support of data privacy and security would suggest a *hierarchy* model, but there is no overarching transnational bureaucracy in place to enforce this control.
- The competition for aid and resources would suggest a *market* model, but there is an increasing demand from donors for transparency and accountability, which makes questionable typical methods of facilitating competitive advantage, such as the proprietary withholding of data as intellectual property.
- The need to innovate in order to build and maintain reputation requires levels of flexibility and independence that would suggest a *bazaar* model, but this model challenges the levels of control needed for data privacy and security.
- The need to meet a common objective (i.e. the survival of those affected by disaster) leans towards a *network* model, but competition for aid and resources challenges levels of trust needed to make the exclusive application of this model viable.

This leads us to the following future research proposition:

1. *Models of governance need to be developed that are appropriate to the sharing of data across humanitarian ecosystems*

## **RP 2: Evaluating Data Veracity**

Data veracity is a key determinant of the level of credence and trust that can be placed on the models and insights that data enables, which in turn suggests the value the data can yield is reliant on establishing the veracity of the data employed (Lukoianova & Rubin, 2014). Our empirical example showed that, in bring together data from numerous disparate sources, there was a significant sacrifice in levels of detail and nuance. The folding-in of each unrelated set of disaster data to form a cohesive whole of 950,026 disaster entries required rationalising, reclassifying, consolidating and aggregating data from six sources. With each



*keep, change and/or drop* decision the consolidated disaster dataset's veracity was in jeopardy of errors of omission, misrepresentation and obfuscation. Going on to align disasters with aid required further aggregation and the discarding of more source information, the final usable dataset being no more than 3,556 records of indeterminate veracity. In our case, as the overall distillation, consolidation and aggregation of the data could yield only macro-level informational, as opposed to actionable insights, the entries assessed as being of questionable veracity were flagged to allow appropriate levels of credence to be placed on outputs from the data.

It is worthy of note that there is risk associated with any tacit assumption that data from highly reliable sources will be equally as reliable. Questions of veracity can extend beyond those related to the mechanics of consolidating and aggregating structured data from disparate sources. Even with governed data from credible sources, pre-existing conditions of poor data veracity related to errors introduced during data creation, collection and/or curation are not unknown. If social media data is included in the search for insights it is important to cognisant of challenges to veracity caused by inherent bias and falsehoods that can be human and/or machine generated (Rubin & Vashchilko, 2012). As noted by Boyd & Crawford (2012 p. 668), "Large data sets from Internet sources are often unreliable, prone to outages and losses, and these errors and gaps are magnified when multiple data sets are used together". Therefore, establishing veracity should be a prerequisite to the pursuit for value from data, big or small (Lukoianova & Rubin, 2014).

Reliance on the existence and accessibility of data alone falls short of offering credible and actionable insights in support of the UNSDGs if its veracity cannot be evaluated - eroding trust in outputs created from data that are of questionable or unknown veracity. Logic dictates that data that are not veracious are likely to lead to insights that are misleading or incorrect (Clarke, 2016). Creating algorithms, models and visualisations without understanding the veracity of the data on which these artefacts are based is to risk analyses that form false perceptions, or potentially worse, trigger erroneous actions (D'Mello, 2016). Hence, regardless of the scale or magnitude of the data being analysed there is a need to (openly and transparently) gauge the extent to which the data can be trusted and identify any weaknesses that must be addressed or accepted (or that may in fact render the data unfit for purpose).

Having stressed the importance of knowing the veracity of the data being used, the extant literature offers no consensus as to the definition of data veracity. The concept is often described in the negative – e.g., ‘misinformation’, ‘data in doubt’, ‘uncertain’, ‘abnormal’, ‘untruthful’, ‘noisy’ and ‘biased’ (Berti-Equille & Borge-Holthoefer, 2015; Berti-Equille & Lamine Ba, 2016; Claverie-Berge, 2012; Lukoianova & Rubin, 2014; Normandeau, 2013; Schroeck, 2012). As the state-of-the-art stands there is a paucity of work oriented at: (a) Providing foundational definitions; and (b), more importantly, the development of models, methods and/or technologies to determine the veracity of data. Arguably, the foundations for such work exists in seminal work in data quality (Lee et al., 2002; Strong et al., 1997; Wand & Wang, 1996; R. Y. Wang & Strong, 1996) and concepts from the big data literature (Berti-Equille & Lamine Ba, 2016; Grimes, 2013; Laney, 2013; Lukoianova & Rubin, 2014; Normandeau, 2013; Powers Durette, 2016). Neither offer complete answers, however, so the humanitarian ecosystem currently lacks a means of consistently and comprehensively evaluating the veracity of its data assets, calling into question the credibility of the data used to support the UNSDGs.

This leads us to the following future research proposition:

2. The veracity of data assets requires transparent evaluation if trust is to be established and maintained.

### **RP 3: Value and Data Privacy**

There are significant opportunities to create value from data within the humanitarian ecosystem. These include descriptive, prescriptive and predictive possibilities across humanitarian operations (Dubey et al., 2018; Lycett & Monaghan, 2013; Swaminathan, 2018) and the capture and sharing of critical data to support the UNSDGs (Ki-moon, 2016 Paragraphs 58, 60, 80, 93, 120, 122–124, 128, 137, 150, 156 and Annexe pages 52, 56, 57–59). The juxtaposition of privacy with value however highlights that these can be conflicting needs that raise questions of balance and compromise in the humanitarian ecosystem. Simply put, there is a need to carefully consider the circumstances in which it is appropriate to infringe privacy in pursuit of value or sacrifice value to protect privacy, when value in this context is likely to be measurable in terms of human survival and well-being. Consider the differing approaches of two of the six disaster data sources used in our empirical example:

(I) **UNHCR**: persons of concern, annualised totals of displaced populations (UNHCR, 2017).

This dataset enables aggregate analysis of scale and trends as it provides totals for displaced populations by origin and destination, subdivide into values for leaving or returning refugees, leaving or returning internally displaced persons (IDPs), asylum-seekers, stateless people and *other*. Where any of the numbers are fewer than 4, an asterisk is published instead of a number to reduce the risk of exposing the identities of those concerned.

(II) **VDC-SY**: individuals killed, detained or are missing in the Syrian conflict (VDC-SY, 2016).

Personal and identifying information is collected and made openly available via a searchable database, e.g. the victim's name; mother's name; province; cause of death; location of death or where last seen, marital status, civilian/non-civilian, who were the actors involved etc.

The former dataset makes macro-level analysis feasible – e.g. annual patterns of forced displacement across African countries. This limits the value to be gained by not being able to examine aspects such as: (a) Patterns of displaced people associated with specific humanitarian crises; (b) people who were displaced because of insufficient aid in response to a crisis; (c) whether those that flee crises are also counted as those affected; (d) the per day/week/month persons of concern displacement per country; and/or (e) how many women and children are in each of the displaced groups. Answers to such questions are dependent on more granular data than contained in UNHCR's 'persons of interest' dataset and the ability to coherently combine this data to other datasets – the relationality point noted by Boyd & Crawford (2012 p.663).

Conversely, the latter dataset above is designed to be specific to each victim, thus humanising those who have died or suffered in Syria because of human rights violations (VDC-SY, 2016). Analysis here can be at a micro-level, where patterns of killing, detentions and disappearances can be examined using dimensions such as the victim's age, gender, marital status, regional affiliation, civilian or non-civilian status, where they were killed, detained, or last seen, who were the perpetrators etc. The ability to gain value from this level of granularity is considerable, but comes at a cost to privacy, which is generally held as sacrosanct. For example UNSDG 16 "*Peace, Justice and Strong Institutions*", Target 16.9 is based on the premise that controlling our identity is crucial to our ability to exercise our

other fundamental rights (United Nations, 2015a). Nevertheless, as solutions such as identity cards, biometrics and blockchain are tested, which take the UN notion of identity beyond recording births, understandable fears of surveillance, tracking and profiling emerge (Privacy International, 2018). In addition, such fears do not apply exclusively to the personal data of/associated with individuals, as even access to seemingly peripheral information may risk the safety of collectives such as refugee caravans, villages and religious factions (Marelli et al., 2017).

Arguably, what lies beneath the surface here is the significance of context to both value and privacy. Boyd & Crawford (2012), argue that (big) data taken out of context loses *meaning* as data is not generic and retaining context is critical. Real-world scenarios, particularly in the humanitarian domain, are often highly nuanced raising complex questions about the right to and appropriate use of data (Cukier, 2019; Raymond et al., 2016). Thus, common conceptions of privacy are not always easily aligned to the intricate, and sometimes incompatible, ethical and moral challenges within the humanitarian domain where the value of data is measured in terms of human lives. Outside of the humanitarian domain, some have started to recognise this issue. Nissenbaum's (2004, 2010) theory of contextual integrity with respect to privacy, for example, challenges the more commonly perceived proprietary relationship between the individual and privacy. In essence, the theory posits that the infringement of privacy is caused by a failure in contextual integrity, where context is defined as a social domain (Cooper, 2015; Nissenbaum, 2018). Nissenbaum stresses the importance of social values and examines contextual integrity as a product of three independent (mandatory) components: (1) Actors – the sender, recipient and subject of the data; (2) the information – the data; and (3) the transmission principle – the terms or constraints of the flow (Nissenbaum, 2018). The sense that privacy has been compromised is argued by Nissenbaum (2018) to occur when contextual integrity has been violated. Put simply, privacy is compromised when the informational norms of *appropriateness* (what may be shared), *flow* (how it may be shared) or *distribution* (by and with whom it may be shared) are violated (Cooper, 2015). The challenge in the humanitarian ecosystem, therefore, is to maintain contextual integrity without compromising value through obfuscation and aggregation (Winter & Davidson, 2019).

This leads us to the following research proposition:

3. Mechanisms are required for addressing the conflict between value in the humanitarian domain and the privacy, autonomy and liberty of individuals.

#### **RP 4: Data Transparency**

Our empirical example underscores the difficulties of gaining a comprehensive, coherent and contextually-informed view of the global patterns of disasters and human survival over time by integrating published data from (only) six different sources. In working with the datasets it became clear that, in some instances, more detailed data existed but was not publicly available. Consequently, a more general issue of data transparency is apparent – which, in essence, relates to how parties within the domain can openly share data without damaging themselves, or allowing data breaches to happen, while meeting legislative requirements. There are several aspects of transparency to note here, which straddle policy and technical conceptions of governance. First, that there needs to be an ecosystem-based consensus in relation to domain understanding – the things that matter within the domain, the relationships between those things, general and party specific policy constraints on the representation and use of knowledge etc. These are elements which, though instantiated as data or capability assets, are crucial to transformation and completion processes. The immediate difficulty here is that even the most obvious efforts that may facilitate ecosystem-wide transparency – such as domain repositories for data sharing, a domain-specific language, a domain-wide ontology and disaster classification standards – where they exist, have been less than consistently or comprehensively employed in the humanitarian domain to-date (EDRIS, 2017; FTS, 2017; Global Campaign for Aid Transparency, 2017; HDX, 2017; HXL, 2019; IATI, 2017; IRDR, 2014; Lattimer et al., 2016; OECD, 2017a; William & Mary University, 2018).

Second, greater consideration needs to be given to metadata, particularly in the context of integration. Though not well addressed in the academic literature, there is an increasing realisation in the commercial big data world that significant insight can be gained from data about (big) data – i.e., data that describes the content/raw data such as origin, time, date, format etc. (Mayer et al., 2016), for example, illustrate the insight that can be gleaned from mobile call metadata and, indeed, the unintentional infringement of privacy that can occur via that insight. The aspects that we point researchers to in this respect are the need to capture metadata related to the data assets themselves such that judgements can be made

about them in relation to the: (a) Veracity of the data; and (b) the process associated with contextual integrity if a balance between privacy and appropriateness is to be achieved. In addition, metadata needs to be captured re the process of integration itself, such that other parties can understand and reproduce the (big) data integration process (such as the integration in our empirical example). We see these as important aspects of digital asset management, which is a subject that does not appear to have been addressed in the humanitarian domain.

Third, there is the consideration of data voids – that no one party gets to hold data ‘hostage’ without undue reason. While the logistical difficulties in post-disaster data collection are relatively easy to understand for example (Maiers et al., 2005; Riddell, 2008; Tatham & Hughes, 2011), the prospect of instinctive or intentional data voids are no less real. In this respect, the study of ignorance (agnotology), may be of use here to explore the existence and effect of data voids in the humanitarian domain (Proctor, 2006, 2008). Proctor (2008 p.3) delineates ignorance into three forms: (1) Ignorance as a native state; (2) ignorance as choice; (3) ignorance as an active construct. All three forms of ignorance may exist in the humanitarian domain. In terms of native state, it is entirely possible that humanitarian aid executors are ignorant of what data should be collected. In terms of choice, there is the possibility that data will not be collected or shared as there is no benefit to the collector (*choice*). For example, data that might suggest poor use of funds or poor outcomes is likely to have a significant detrimental effect on future funding (The Lancet, 2010). Last, in terms of active construct, it may be the case that data are intentionally concealed, such as the ‘sex for food’ shame of well-known charities (O’Neill, 2018). Any one of these forms of ignorance may distort the humanitarian landscape; therefore there may be merit in studying the absence of data and its causes, distribution and scale. Paradoxically, a study of ignorance may contribute knowledge about the humanitarian domain and yield opportunities to actively address data voids through methodological change (e.g. education and incentivisation) and/or technological innovation (e.g. algorithmic extrapolation to bridge data gaps).

This leads us to the following research proposition:

4. Mechanisms are required to manage and maintain data transparency across ecosystems.

## RP 5: Data Granularity and Boundaries

From our empirical work, it is apparent that what a data asset is remains loosely defined – in this case an asset equates to the dataset that a particular organisation makes available in essence. This position is the case more broadly as the state-of-the-art stands – an asset ranging, at its most general, from any entity that comprises data through to the result of taking the ‘raw material’ of operational data and transforming it in some way in the data warehousing world. Within organisational confines this is less of an issue as data integration is a well-trodden path (a situation that also applies, to a lesser extent, between collaborating organisations). The clear advantage within such contexts is that a focal organisation has control over both the asset and capability creation processes – the purpose for asset creation (whatever that might be) is normally well-defined and the capabilities to (dynamically) deliver it in place. Where organisations collaborate and share data, the grounds for doing that are either agreed and/or a focal organisation dictates the systems and/or standards for sharing.

From an ecosystem perspective, where control cannot be leveraged in the same manner, the situation is much more problematic. At the outset, as illustrated by the empirical case here, there are issues of data availability and the granularity of the data that is available. From a computational perspective, modularity provides a general set of principles for managing complexity – viewing a (eco)system as a complex system of discrete entities that communicate/interoperate through standardised interfaces and architectures (Baldwin & Clark, 1997). Such approaches provide a plausible means of integration within ecosystems but, arguably, additional attention is required re the granularity of content (where content is not controlled by any one party). The lesson from data warehousing is that integrated data needs to be of the same grain (level of detail) and analysis at different levels of abstraction is dealt with by storing data at the lowest level of detail, pre-aggregating data at higher levels and using the principles of modularity for access. From our empirical work, we have found little evidence that even these common approaches are being applied. This is of importance in the context of big data analysis due to the multiple data granularities that have to be dealt with while extracting value from big data (Reyes-Galaviz & Pedrycz, 2015). In addition to common practice, the concept of granular computing has recently emerged – broadly predicated on the observation that different and interesting patterns or regularities

appear in the data at different levels of granularity and that real-world problems require processing at different levels of granularity/abstraction (Reyes-Galaviz & Pedrycz, 2015). Research in this area relates to the granularisation on data values, variables (features), systems and concepts. Granular computing (GrC) is an umbrella term that cover any theories, methodologies, techniques, and tools that make use of information granules in problem solving (Pal et al., 2015; Yao, 2000). The idea behind GrC is to mimic how human brain works to use a structured representation of the real-world problem to understand and process different levels of information granularity and abstraction related to that problem using AI techniques (such as fuzzy and rough sets technologies). The granular approach is in its infancy, however, and significant research is required to create an effective dynamic model for the intelligent processing multi-source heterogeneous big data (G. Wang et al., 2017).

This leads us to the following future research proposition:

5. Data assets require well-defined boundaries and clear interfaces to the data they contain to deliver data at clearly defined levels of granularity.

### **Limitations**

It is important to note that these research propositions are neither exhaustive nor mutually exclusive. There exist areas of overlap and elements of interconnection that viably link these research propositions to each other and as such may also be worthy of study. Additionally, the empirical work carried out here is limited to static historic data readily available via the web. We note that, were the study extended to the consolidation and alignment of proprietary data held by multilateral, governmental and non-governmental organisations, potential opportunities for additional research propositions may result. Last, the increasing use of technologies/approaches such as crowdsourcing of data, drones, sensors and distributed ledgers in the humanitarian domain will undoubtedly expose research opportunities in the data analytics of multi-sourced live-steamed data: Though these technologies are beyond the scope of the empirical work done here, we acknowledge the point and refer the reader to articles that are of starting to emerge in these areas (Chen et al., 2015; Estrada & Ndoma, 2019; Quinn et al., 2018; Sandvik & Lohne, 2014; Zwitter & Boisse-Despiaux, 2018).



## Conclusion

In the humanitarian and development domains, more widely available (big) data has the potential to drive a better integrative and predictive understanding yielding the intrinsic value of saving lives and reducing human suffering. Further, such an understanding would provide a platform for developing a greater degree of resilience and preventing complex emergencies that would move toward the UN goal of the humanitarian sector contributing to the achievement of longer-term sustainable development. Guided by the UN imperative for 'quality, accessible, timely and reliable dis-aggregated data' our work here has explored the difficulties in extracting value from data in the humanitarian domain. Guided by a model that explains how big data can be transformed to derive value, we propose that significant distance remains between the 'vision' of big data (analytics) and the current reality. Big data is not of note because of its size but, rather, because of its relationality – value comes from the ability to integrate heterogeneous data effectively and that remains challenging.

Our empirical example evidenced these challenges, drawing eight distinct sources of data together in an attempt to gain a perspective of global disasters and aid sent in response. What transpired from that analysis were significant losses in relation to the detail, visibility, specificity, nuance and history of that data. These losses severely impacted the analytical value of the integrated data and are rooted in the fact that the individual datasets concerned were neither designed nor developed with integration in mind – thus, our key observation is that an ecosystem perspective need to be adopted and that data asset creation and capability processes need to be considered from that perspective. In such a context, the vehicles of integration are not definable a priori and, thus, an added burden is placed upon designing for relationality.

In addressing the issues that arose in our attempt to integrate heterogeneous humanitarian datasets, we distilled five propositions as a means of directing future research to areas that we argue will bear fruit in relation to the UNSDGs aspirations re 'dis-aggregated' (big) data. These propositions covered the need for research related to: (a) Models of governance appropriate to ecosystems; (b) means for evaluating the veracity of data; (c) addressing the tension between value and the need for privacy; (d) developing mechanisms to manage and maintain the transparency of data; and (e) the need to better define data assets and the ways in which they can interact. We do not suggest that our propositions are the only ones

of importance re the use of (big) data in the humanitarian domain but, rather, we have sought to confine our observations to those that have emerged from our empirical work – which we approached with the ideals of big data in mind.

## References

- Almeida, F. L. F., & Calistru, C. (2012). The main challenges and issues of big data management. *International Journal of Research Studies in Computing*, 2(1), 11–20. <https://doi.org/10.5861/ijrsc.2012.209>
- Anand, A., Coltman, T., & Sharma, R. (2016). Four steps to realizing business value from digital data streams. *MIS Quarterly Executive*, 15(4), 259–277.
- Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete (C. Anderson, Ed.). Retrieved from Wired Magazine website: <https://www.wired.com/2008/06/pb-theory/>
- Baldwin, C. Y., & Clark, K. B. (1997). Managing In An Age Of Modularity. *Harvard Business Review*, 75(5), 84–93. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=9709112720&site=ehost-live>
- Bärenfänger, R., Otto, B., & Österle, H. (2014). Business value of in-memory technology – multiple-case study insights. *Industrial Management & Data Systems*, 114(9), 1396–1414. <https://doi.org/10.1108/imds-07-2014-0212>
- Bedi, P., Jindal, V., & Gautam, A. (2014). Beginning with big data simplified. *2014 International Conference on Data Mining and Intelligent Computing (ICDMIC)*, 1–7. <https://doi.org/10.1109/icdmic.2014.6954229>
- Berti-Equille, L., & Borge-Holthoefer, J. (2015). Veracity of data: From truth discovery computation algorithms to models of misinformation dynamics. In *Synthesis Lectures on Data Management* (pp. 1–9). Morgan & Claypool.
- Berti-Equille, L., & Lamine Ba, M. (2016). Veracity of Big Data: Challenges of Cross-Modal Truth Discovery. *Journal of Data and Information Quality (JDIQ)*, 7(3), 1–3. <https://doi.org/10.1145/2935753>
- Blyler, M., & Coff, R. W. (2003). Dynamic capabilities, social capital, and rent appropriation: Ties that split pies. *Strategic Management Journal*, 24(7), 677–686. <https://doi.org/10.1002/smj.327>
- Boyd, D., & Crawford, K. (2012a). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Boyd, D., & Crawford, K. (2012b). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Buchanan-Smith, M., & Fabbri, P. (2005). Linking relief, rehabilitation and development. A review of the debate. In T. E. Coalition (Ed.), *Tsunami Evaluation Coalition*. Retrieved from <http://www.alnap.org/pool/files/IRRD-review-debate.pdf>
- Chatterjee, D., & Ravichandran, T. (2013). Governance of interorganizational information systems: A resource dependence perspective. *Information Systems Research*, 24(2), 261–278. <https://doi.org/10.1287/isre.1120.0432>
- Chen, H., Chiang, R. H. L., Storey, V. C., & Robinson, J. M. (2012). Business Intelligence and Analytics: From Big Data To Big Impact. *MIS Quarterly*, 36(4), 1165–1188. Retrieved from [www.freakonomics.com/2008/02/25/hal-varian-answers-your-questions/](http://www.freakonomics.com/2008/02/25/hal-varian-answers-your-questions/)
- Chen, J., Wang, C., & Bai, Y. (2015). CrowdMR: Integrating crowdsourcing with MapReduce for AI-hard problems. *Proceedings of the National Conference on Artificial Intelligence*, 6(3), 4272–4273.

- Chesbrough, H. (2010). Business model innovation: Opportunities and barriers. *Long Range Planning*, 43(2–3), 354–363. <https://doi.org/10.1016/j.lrp.2009.07.010>
- Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, 26(1), 77–90. <https://doi.org/10.1111/isj.12088>
- Claverie-Berge, I. (2012). Solutions Big Data (IBM, Ed.). Retrieved from Netezza website: [http://www-05.ibm.com/fr/events/netezzaDM\\_2012/Solutions\\_Big\\_Data.pdf](http://www-05.ibm.com/fr/events/netezzaDM_2012/Solutions_Big_Data.pdf)
- Comuzzi, M., & Patel, A. (2016). How organisations leverage Big Data: a maturity model. *IMDS* 116,8 1468 Received 9 December 2015 Revised 25 February 2016 29 March 2016 Accepted 30 March 2016 How Organisations Leverage Big Data: A Maturity Model Marco Comuzzi School of Management Engineering, Ulsan National Institute of Science and Technology, 116(8), 1468–1492. <https://doi.org/10.1108/IMDS-07-2014-0202>
- Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, 30(1), 44–57. <https://doi.org/10.1057/jit.2014.17>
- Cooper, G. (2015). Hurricanes and hashtags: How the media and NGOs treat citizens' voices online in humanitarian emergencies. *Interactions: Studies in Communication & Culture*, 6(2), 233–244. [https://doi.org/10.1386/iscc.6.2.233\\_1](https://doi.org/10.1386/iscc.6.2.233_1)
- Croicu, M., & Sundberg, R. (2015). *UCDP GED Codebook version 5.0*. Retrieved from Department of Peace and Conflict Research, Uppsala University website: <http://ucdp.uu.se/downloads/ged/ucdp-ged-50-codebook.pdf>
- Cukier, K. (2019). Babbage: Data to the rescue (K. Cukier, Ed.). Retrieved from Babbage, The Economist website: <https://play.acast.com/s/theeconomistbabbage/1481f4c7-5690-4f8f-8044-36c7c8818dd8>
- D'Mello, N. (2016). 5 cases where big data was a big flop. *CIO*. Retrieved from <http://www.cio.in/feature/5-cases-where-big-data-was-big-flop>
- Davenport, T. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Boston, MA: Harvard Business Review Press.
- Davenport, T. H., Barth, P. F. P., & Bean, R. (2012). How 'Big Data' is Different. *MIT Sloan Management Review*, 54(1), 43–46. Retrieved from <https://www.semanticscholar.org/paper/How-'Big-Data'-is-Different-FALL-2012-Davenport-Barth/b82282089d3ff97f133a172ee38804d2357c71b1>
- DesInventar. (2017). Disaster Information Management System - OnLine Database (LA RED, Ed.). Retrieved January 7, 2017, from Network of Social Studies in the Prevention of Disasters in Latin America (Red de Estudios Sociales en Prevención de Desastres en América Latina - LA RED) website: <http://www.desinventar.net>
- Dubey, R., Luo, Z., Gunasekaran, A., Akter, S., Hazen, B. T., & Douglas, M. A. (2018). Big data and predictive analytics in humanitarian supply chains: Enabling visibility and coordination in the presence of swift trust. *The International Journal of Logistics Management*, 29(2), 485–512. <https://doi.org/10.1108/ijlm-02-2017-0039>
- DuBois, M. (2016). Don't blur the lines between development and humanitarian work. *The Guardian*. Retrieved from <https://www.theguardian.com/global-development-professionals-network/2016/may/12/dont-blur-the-lines-between-development-and-humanitarian-work>
- Dutta, D., & Bose, I. (2015). Managing a Big Data project: The case of Ramco Cements Limited. *International Journal of Production Economics*, 165, 293–306. <https://doi.org/https://doi.org/10.1016/j.ijpe.2014.12.032>

- Dyer, J. H., Gregersen, H. B., & Christensen, C. M. (2009). The Innovator's DNA. *Harvard Business Review*, (December), 1–9.
- EDRIS. (2017). European Disaster Response Information System (EDRIS) (Directorate-General for European Civil Protection and Humanitarian Aid Operations (ECHO);, Ed.). Retrieved July 12, 2016, from <http://ec.europa.eu/echo/>
- Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., ... Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, 66(8), 1523–1545. <https://doi.org/10.1002/asi.23294>
- Elgendy, N., & Elragal, A. (2014). Advances in Data Mining. Applications and Theoretical Aspects. *Proceedings of The14th Industrial Conference, ICDM, 7987*. <https://doi.org/10.1007/978-3-642-39736-3>
- Estrada, M. A. R., & Ndoma, A. (2019). The uses of unmanned aerial vehicles –UAV's- (or drones) in social logistic: Natural disasters response and humanitarian relief aid. *Procedia Computer Science*, 149, 375–383. <https://doi.org/https://doi.org/10.1016/j.procs.2019.01.151>
- FTS. (2017). Financial Tracking Services (Humanitarian Programme Cycle Information Services Unit (HPC-IS), Ed.). Retrieved May 30, 2017, from Tracking Humanitarian Aid Flows website: <http://fts.unocha.org/>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gantz, J., & Reinsel, D. (2011). Extracting Value from Chaos State of the Universe. *IDC IView*, (June), 1–12. <https://doi.org/10.1007/s10916-016-0565-7>
- Global Campaign for Aid Transparency (Ed.). (2017). Aid Transparency Index (ATI). Retrieved from Publish What you Fund website: <http://ati.publishwhatyoufund.org/>
- Grimes, S. (2013). Big Data: Avoid “Wanna V” Confusion (I. Week, Ed.). Retrieved March 3, 2017, from InformationWeek website: <http://www.informationweek.com/big-data/big-data-analytics/big-data-avoid-wanna-v-confusion/d/d-id/1111077>
- Grogan, B., & Strohmeyer, H. (2016). *Leaving No One Behind* (P. D. and S. B. OCHA Policy Analysis and Innovation Section, Ed.). Retrieved from [https://www.unocha.org/sites/dms/Documents/HEStudy\\_20151109\\_Small.pdf](https://www.unocha.org/sites/dms/Documents/HEStudy_20151109_Small.pdf)
- GTD. (2017). Global Terrorism Database (GTD), National Consortium for the Study of Terrorism and Responses to Terrorism (START) (University of Maryland, Ed.). Retrieved January 7, 2017, from START: A Center of Excellence of the U.S. Department of Homeland Security website: <http://www.start.umd.edu/gtd/>
- Guha-Sapir, D., Below, R., & Hoyois, P. (2017). Emergency Events Database (EM-DAT): The CRED/OFDA International Disaster Database (U. C. de Louvain, Ed.). Retrieved August 16, 2016, from EM-DAT website: <http://emdat.be>
- Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems*, 26(3), 191–209. <https://doi.org/10.1016/j.jsis.2017.07.003>
- Hansen, H. K., & Porter, T. (2017). What do big data do in global governance? *Global Governance*, 23(1), 31–42. Retrieved from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85012226949&partnerID=40&md5=d1c86788e793119f93cbc8663a13a11d>

- Hazen, B. T., Bradley, R. V., Bell, J. E., In, J., & Byrd, T. A. (2017). Enterprise architecture: A competence-based approach to achieving agility and firm performance. *International Journal of Production Economics*, 193(July), 566–577. <https://doi.org/10.1016/j.ijpe.2017.08.022>
- HDX. (2017). Humanitarian Data Exchange (HDX). Retrieved January 6, 2017, from <http://data.humdata.org/>
- Huberty, M. (2015). Awaiting the Second Big Data Revolution: From Digital Noise to Value Creation. *Journal of Industry, Competition and Trade*, 15(1), 35–47. <https://doi.org/10.1007/s10842-014-0190-4>
- HXL. (2019). The Humanitarian Exchange Language. Retrieved August 1, 2019, from <http://hxlstandard.org/>
- IAEG-SDGs. (2017). Sendai Framework for Disaster Risk Reduction 2015-2030 & 2030 Agenda for Sustainable Development-Multi-Purpose Data. In *Inter-agency and Expert Group on Sustainable Development Goal Indicators (IAEG-SDGs)*. Retrieved from United Nations Statistics Division (UNSD) website: <https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-06/5>. UNISDR Disaster Risk Reduction Indicators.pdf
- IATI (Ed.). (2017). International Aid Transparency Initiative (IATI). Retrieved February 1, 2017, from International Aid Transparency Initiative website: <http://www.aidtransparency.net/>
- IBM. (2013). The Four V's of Big Data (IBM Big Data & Analytics Hub, Ed.). Retrieved May 12, 2014, from IBM Big Data & Analytics Hub website: <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>
- IRDR. (2014). Peril Classification and Hazard Glossary. In Integrated Research on Disaster Risk (IRDR) (Ed.), *Data Project Report n°1*. Retrieved from International Social Science Council (ISSC); United Nations International Strategy for Disaster Reduction (UNISDR) website: [http://www.irdrinternational.org/wp-content/uploads/2014/04/IRDR\\_DATA-Project-Report-No.-1.pdf](http://www.irdrinternational.org/wp-content/uploads/2014/04/IRDR_DATA-Project-Report-No.-1.pdf)
- ISO. (2017). International Organization for Standardization (International Organization for Standardization, Ed.). Retrieved January 9, 2017, from <http://www.iso.org/iso/home.html>
- Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86–94. <https://doi.org/10.1145/2611567>
- Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, 70, 338–345. <https://doi.org/10.1016/j.jbusres.2016.08.007>
- Kamper, H. (2019). Digital Humanitarianism: Using Big Data. Retrieved from <https://borgenproject.org/digital-humanitarianism/>
- Keen, J., Calinescu, R., Paige, R., & Rooksby, J. (2013). Big data + politics = open data: The case of health care data in England. *Policy and Internet*, 5(2), 228–243. <https://doi.org/10.1002/1944-2866.POI330>
- Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., ... Gani, A. (2014). Big data: Survey, technologies, opportunities, and challenges. *Scientific World Journal*, 2014. <https://doi.org/10.1155/2014/712826>
- Ki-moon, B. (2016). One humanity: Shared responsibility (Report of the Secretary-General for the World Humanitarian Summit). In *World Humanitarian Summit*. Retrieved from <https://sgreport.worldhumanitariansummit.org/>

- Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. In M. Group (Ed.), *Application Delivery Strategies* (Vol. 949). Retrieved from Meta Group website: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Laney, D. (2012). Deja VVVu: Others Claiming Gartner's Construct for Big Data (Gartner, Ed.). Retrieved March 1, 2017, from Gartner Blog Network website: <http://blogs.gartner.com/doug-laney/deja-yyyue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/>
- Laney, D. (2013). Batman on Big Data (Gartner, Ed.). Retrieved March 1, 2017, from Gartner Blog Network website: <http://blogs.gartner.com/doug-laney/batman-on-big-data/>
- Lattimer, C., Swithern, S., Tuchel, L., Beecher, J., Cecilia Caio, Collins, R., ... Watts, R. (2016). Global Humanitarian Assistance Report 2016. In G. H. A. (GHA) (Ed.), *Global Humanitarian Assistance Reports* (Vol. 2016). Retrieved from Global Humanitarian Assistance (GHA) website: <http://www.globalhumanitarianassistance.org/wp.../2016/07/GHA-report-2016-full-report.pdf>
- Lau, C. H. (2019). 5 Steps of a Data Science Project Lifecycle. Retrieved February 7, 2020, from Towards Data Science website: <https://towardsdatascience.com/5-steps-of-a-data-science-project-lifecycle-26c50372b492>
- Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, (52205). <https://doi.org/10.0000/PMID57750728>
- Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information Management*, 40(2), 133–146. [https://doi.org/10.1016/s0378-7206\(02\)00043-5](https://doi.org/10.1016/s0378-7206(02)00043-5)
- Lisowska, B. (2017). Joining up data in the humanitarian–development nexus: Why does it matter? (U N Sustainable Development Goals, Ed.). Retrieved May 19, 2019, from Silo Fighters website: [https://undg.org/silofighters\\_blog/joining-up-data-in-the-humanitarian-development-nexus-why-does-it-matter/](https://undg.org/silofighters_blog/joining-up-data-in-the-humanitarian-development-nexus-why-does-it-matter/)
- Lukoianova, T., & Rubin, V. L. (2014). Veracity Roadmap: Is Big Data Objective, Truthful and Credible? *Advances in Classification Research Online; 24th ASIS SIG/CR Classification Research Workshop*, 24(1), 4–15. Retrieved from <http://journals.lib.washington.edu/index.php/acro/article/view/14671/12311>
- Lycett, M. (2013). 'Datafication': making sense of (big) data in a complex world. *European Journal of Information Systems*, 22(4), 381–386. <https://doi.org/10.1057/ejis.2013.10>
- Lycett, M., & Monaghan, A. (2013). *Big data and humanitarian supply networks: Can Big Data give voice to the voiceless?* 432–437. IEEE.
- Maiers, C., Reynolds, M., & Haselkorn, M. (2005). Challenges to effective information and communication systems in humanitarian relief organizations. *Professional Communication Conference, 2005. IPCC 2005. Proceedings. International*, pp.82-91. <https://doi.org/10.1109/ipcc.2005.1494163>
- Maitlis, S. (2005). The social processes of sensemaking. *The Academy of Management Journal*, 48(1), 21–49. <https://doi.org/10.5465/amj.2005.15993111>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big Data: the next frontier for innovation, competition, and productivity* (M. G. Institute, Ed.). Retrieved from McKinsey Global Institute website: [http://www.mckinsey.com/~media/McKinsey/Business Functions/McKinsey Digital/Our Insights/Big data The next frontier for innovation/MGI\\_big\\_data\\_exec\\_summary.ashx](http://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx)

- Marelli, M., Apraxine, P., Bircher, R., Kuner, C., Papakonstantinou, V., Jasmontaite, L., ... (VUB), B. P. H. (BPH) an entity of V. U. B. (2017). Handbook on Data Protection in Humanitarian Action. In C. Kuner & M. Marelli (Eds.), *ICRC*. Retrieved from International Committee of the Red Cross (ICRC); Brussels Privacy Hub (BPH) an entity of Vrije Universiteit Brussel (VUB) website: <https://www.icrc.org/en/publication/handbook-data-protection-humanitarian-action>
- Mayer, J., Mutchler, P., & Mitchell, J. C. (2016). Evaluating the privacy properties of telephone metadata. *Proceedings of the National Academy of Sciences*, *113*(20), 5536. <https://doi.org/10.1073/pnas.1508081113>
- McAfee, A., & Brynjolfsson, E. (2012). Big Data : The Management Revolution. *Harvard Business Review*, (October), 1–12. Retrieved from <http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf>
- Meier, P. (2015). Digital humanitarians, big data and disaster response (T. B. Institution, Ed.). Retrieved from The Brookings Institution website: <https://www.brookings.edu/blog/techtank/2015/02/19/digital-humanitarians-big-data-and-disaster-response/>
- Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2016). A Data Quality in Use model for Big Data. *Future Generation Computer Systems*, *63*, 123–130. <https://doi.org/10.1016/j.future.2015.11.024>
- Mithas, Ramasubbu, & Sambamurthy. (2011). How Information Management Capability Influences Firm Performance. *MIS Quarterly*, *35*(1), 237–256. <https://doi.org/10.2307/23043496>
- Moorhead, J., & Sandler Clarke, J. (2015). The global humanitarian crisis “can’t depend on a continuous increase in funding.” *The Guardian*. Retrieved from <http://www.theguardian.com/global-development-professionals-network/2015/dec/10/global-humanitarian-crisis-continuous-increase-funding>
- Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of “datification.” *Journal of Strategic Information Systems*, *24*(1), 3–14. <https://doi.org/10.1016/j.jsis.2015.02.001>
- Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review, Symposium - Technology, Values, and the Justice System*, *79*, 119–158. Retrieved from <https://heinonline.org/HOL/P?h=hein.journals/washlr79&i=129>
- Nissenbaum, H. (2010). *Privacy in context : technology, policy, and the integrity of social life*. Stanford, Calif.: Stanford, Calif. : Stanford Law.
- Nissenbaum, H. (2018). Respecting context to protect privacy: Why meaning matters. *Science and Engineering Ethics*, *24*(3), 831–852.
- Normandeau, K. (2013). Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity (InsideBigData, Ed.). Retrieved from InsideBigData website: <http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>
- O’Neill, S. (2018). UN knew of sex-for-food scandal at top charities. *The Sunday Times*. Retrieved from <http://www.thetimes.co.uk/edition/news/un-knew-of-sex-for-food-scandal-at-top-charities-xlkgkt0kb>
- OECD (Ed.). (2017a). Humanitarian Assistance. Retrieved from Development finance statistics website: <http://www.oecd.org/dac/stats/humanitarian-assistance.htm>
- OECD. (2017b). Organisation for Economic Co-operation and Development (OECD) (Organisation for Economic Co-operation and Development (OECD), Ed.). Retrieved January 8, 2017, from <http://www.oecd.org/>



- Oestreicher-Singer, G., & Zalmanson, L. (2013). Content or Community? A Digital Business Strategy for Content. *Management Information Systems Quarterly*, 37(2), 591–616. Retrieved from <http://aisel.aisnet.org/misq/vol37/iss2/15>
- Orrell, T., & Lisowska, B. (2017). What are the principles of joined-up data? (Open Data Watch, Ed.). Retrieved May 19, 2019, from Open Data Watch website: <https://opendatawatch.com/blog/what-are-the-principles-of-joined-up-data/>
- Pal, S. K., Meher, S. K., & Skowron, A. (2015). Data science, big data and granular mining. *Pattern Recognition Letters*, 67, 109–112. <https://doi.org/10.1016/j.patrec.2015.08.001>
- Piccoli, G., & Pigni, F. (2013). Harvesting external data: The potential of digital data streams. *MIS Quarterly Executive*, 12(1).
- Popovič, A., Hackney, R., Tassabehji, R., & Castelli, M. (2018). The impact of big data analytics on firms' high value business performance. *Information Systems Frontiers*, 20(2), 209–222.
- Powers Drette, D. (2016). Why the Veracity of Data Matters in Health Care Research. *The Open Journal of Occupational Therapy*, 4(4), 1–4. <https://doi.org/10.15453/2168-6408.1324>
- Privacy International. (2018). The Sustainable Development Goals, Identity, and Privacy: Does their implementation risk human rights? Identity and its Dimensions. Retrieved July 1, 2019, from Privacy International website: <https://privacyinternational.org/feature/2237/sustainable-development-goals-identity-and-privacy-does-their-implementation-risk>
- Proctor, R. N. (2006). "Everyone knew but no one had proof": tobacco industry use of medical history expertise in US courts, 1990–2002. *Tobacco Control*, 15(suppl 4), iv117–iv125.
- Proctor, R. N. (2008). Agnotology, A Missing Term to Describe the Cultural Production of Ignorance (and Its Study). In R. N. Proctor & L. Schiebinger (Eds.), *Agnotology: The making and unmaking of ignorance* (1st editio, p. 312). Stanford, California USA: Stanford University Press.
- Provan, K. G., & Kenis, P. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of Public Administration Research and Theory*, 18(2), 229–252. <https://doi.org/10.1093/jopart/mum015>
- Purvis, K. (2015). Delivering aid: agencies struggle with increasing needs, costs and insecurity. *The Guardian*. Retrieved from <http://www.theguardian.com/global-development-professionals-network/2015/dec/16/delivering-humanitarian-aid-increasing-needs-costs-insecurity-conflict->
- Quinn, J. A., Nyhan, M. M., Navarro, C., Coluccia, D., Bromley, L., & Luengo-Oroz, M. (2018). Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), 20170363. <https://doi.org/10.1098/rsta.2017.0363>
- Raymond, N., Al Achkar, Z., Verhulst, S., Berens, J., Barajas, L., & Easton, M. (2016). Building data responsibility into humanitarian action. Retrieved July 1, 2019, from OCHA Policy and Studies Series website: <https://www.unocha.org/publication/policy-briefs-studies/building-data-responsibility-humanitarian-action>
- Reyes-Galaviz, O. F., & Pedrycz, W. (2015). Granular fuzzy models: Analysis, design, and evaluation. *International Journal of Approximate Reasoning*, 64, 1–19. <https://doi.org/10.1016/j.ijar.2015.06.005>
- Riddell, R. C. (2008). The growth of emergencies and the humanitarian response. In *Does Foreign Aid Really Work?* (pp. 311–324). Oxford, United Kingdom: Oxford University Press.

- Rubin, V. L., & Vashchilko, T. (2012). Extending information quality assessment methodology: A new veracity/deception dimension and its measures. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1–6. <https://doi.org/10.1002/meet.14504901350>
- Sagiroglu, S., & Sinanc, D. (2013). Big data - A review. *Proceedings of the International Conference on Collaboration Tech-nologies and Systems*, 42–47. <https://doi.org/10.1109/CTS.2013.6567202>
- Sandberg, B., & Aarikka-Stenroos, L. (2014). What makes it so difficult? A systematic review on barriers to radical innovation. *Industrial Marketing Management*, 43(8), 1293–1305. <https://doi.org/10.1016/j.indmarman.2014.08.003>
- Sandvik, K. B., & Lohne, K. (2014). The Rise of the Humanitarian Drone: Giving Content to an Emerging Concept. *Millennium*, 43(1), 145–164. <https://doi.org/10.1177/0305829814529470>
- Schroeck, M. (2012). Analytics: the real-world use of Big Data. How innovative enterprises extract value from uncertain data. IBM Global Business Services. In *IBM Institute for Business Value*. Retrieved from <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=GBE03519USEN>
- Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, 23(4), 433–441. <https://doi.org/10.1057/ejis.2014.17>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Stephenson, M. (2005). Making humanitarian relief networks more effective: operational coordination, trust and sense making. *Disasters*, 29(4), 337–350. <https://doi.org/10.1111/j.0361-3666.2005.00296.x>
- Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data Quality in Context. *Commun. ACM*, 40(5), 103–110. <https://doi.org/10.1145/253769.253804>
- Swaminathan, J. M. (2018). Big Data Analytics for Rapid, Impactful, Sustained, and Efficient (RISE) Humanitarian Operations. *Production and Operations Management*, 27(9), 1696–1700. <https://doi.org/10.1111/poms.12840>
- Tatham, P., & Hughes, K. (2011). Humanitarian logistics metrics : where we are and how we might improve. In M. Christopher & P. Tatham (Eds.), *Humanitarian logistics : meeting the challenge of preparing for and responding to disasters* (1st ed., p. Pages 65-84). London, United Kingdom: Kogan Page Limited.
- The Economist. (2010). *Data , data everywhere*.
- The Lancet. (2010). Growth of aid and the decline of humanitarianism. *The Lancet*, 375, 253 (Editorial). Retrieved from [http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(10\)60110-9/abstract](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(10)60110-9/abstract)
- Truyens, M., & Eecke, P. Van. (2014). Legal aspects of text mining. *Computer Law & Security Review*, 30(2), 153–170. <https://doi.org/10.1016/j.clsr.2014.01.009>
- UCDP. (2017). Department of Peace and Conflict Research, Uppsala University (Uppsala University, Ed.). Retrieved January 9, 2017, from Uppsala Conflict Data Program (UCDP) website: <http://www.pcr.uu.se/>
- UN News. (2018). ‘Fire-fighting approach’ to humanitarian aid ‘not sustainable.’ Retrieved May 19, 2019, from UN News website: <https://news.un.org/en/story/2018/12/1028421>

- UNHCR. (2017). United Nations High Commissioner for Refugees (UNHCR), Population Statistics Database Overview (UNHCR, Ed.). Retrieved January 7, 2017, from Population Statistics website: <http://popstats.unhcr.org/en/overview>
- United Nations. (2015a). Sustainable Development Goal 16: Peace, Justice and Strong Institutions (U. Nations, Ed.). Retrieved from United Nations website: <https://sustainabledevelopment.un.org/sdg16>
- United Nations. (2015b). Sustainable Development Goal 17: Partnerships for the Goals (United Nations, Ed.). Retrieved May 19, 2019, from United Nations website: <https://sustainabledevelopment.un.org/sdg17>
- United Nations. (2015c). Sustainable Development Goals (United Nations, Ed.). Retrieved May 19, 2019, from UNSDG website: <https://www.un.org/sustainabledevelopment/>
- United Nations. (2015d). Transforming our world: The 2030 agenda for sustainable development. In *Resolution adopted by the General Assembly*. Retrieved from United Nations General Assembly website: [https://www.un.org/ga/search/view\\_doc.asp?symbol=A/RES/70/1&Lang=E](https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E); [https://sustainabledevelopment.un.org/content/documents/21252030 Agenda for Sustainable Development web.pdf](https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf)
- UNOCHA. (2017). United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) (Humanitarian Programme Cycle Information Services Unit (HPC-IS), Ed.). Retrieved May 30, 2017, from <http://www.unocha.org/>
- UNSD. (2018). *United Nations Sustainable Development Goal indicators*. Retrieved from United Nations Statistics Division website: <https://unstats.un.org/sdgs/report/2018/overview/>, <https://unstats.un.org/sdgs/files/report/2018/TheSustainableDevelopmentGoalsReport2018-EN.pdf>
- van den Broek, T., & van Veenstra, A. F. (2018). Governance of big data collaborations: How to balance regulatory compliance and disruptive innovation. *Technological Forecasting and Social Change*, 129, 330–338. <https://doi.org/10.1016/j.techfore.2017.09.040>
- van der Zee, B. (2015). UN launches its biggest ever annual appeal for humanitarian funding. *The Guardian*. Retrieved from <http://www.theguardian.com/global-development-professionals-network/2015/dec/07/un-launches-biggest-ever-annual-appeal-for-humanitarian-funding>
- VDC-SY (Ed.). (2016). Violations Documentation Center in Syria. Retrieved July 10, 2016, from VDC-SY website: <http://vdc-sy.net/en/>
- Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, 261(2), 626–639. <https://doi.org/10.1016/j.ejor.2017.02.023>
- Völz, C. (2005). Humanitarian coordination in Indonesia: an NGO viewpoint. In *Forced Migration Review* (Vol. 18). Retrieved from <http://www.fmreview.org/sites/fmr/files/FMRdownloads/en/FMRpdfs/Tsunami/11.pdf>
- Wahlström, M. (2015). Sendai Framework for Disaster Risk Reduction 2015-2030. *Third United Nations World Conference on Disaster Risk Reduction*. Sendai, Miyagi, Japan: United Nations (UN).
- Wand, Y., & Wang, R. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11), 86–95. <https://doi.org/10.1145/240455.240479>
- Wang, G., Yang, J., & Xu, J. (2017). Granular computing: from granularity optimization to multi-granularity joint problem solving. *Granular Computing*, 2(3), 105–120. <https://doi.org/10.1007/s41066-016-0032-3>

- Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5–33. <https://doi.org/10.1080/07421222.1996.11518099>
- White, M. (2012). Digital workplaces: Vision and reality. *Business Information Review*, 29(4), 205–214. <https://doi.org/10.1177/0266382112470412>
- William & Mary University (Ed.). (2018). AidData. Retrieved September 5, 2018, from <https://www.aiddata.org/>
- Winter, J. S., & Davidson, E. (2019). Big data governance of personal health information and challenges to contextual integrity. *The Information Society*, 35(1), 36–51. <https://doi.org/10.1080/01972243.2018.1542648>
- Yao, Y. Y. (2000). Granular Computing: Basic Issues and Possible Solutions. *Proceedings of the 5th Joint Conference on Information Sciences*, 1, 186–189.
- Ylijoki, O. (2018). A recipe for big data value creation. *Business Process Management Journal*, 4(1), 69–91. <https://doi.org/http://dx.doi.org/10.1108/MRR-09-2015-0216>
- Zott, C., & Amit, R. (2010). Business model design: An activity system perspective. *Long Range Planning*, 43(2–3), 216–226. <https://doi.org/10.1016/j.lrp.2009.07.004>
- Zwitter, A., & Boisse-Despiaux, M. (2018). Blockchain for humanitarian action and development aid. *Journal of International Humanitarian Action*, 3(1), 16. <https://doi.org/10.1186/s41018-018-0044-5>