

Published in *Information, Communication and Society* online first on 18 September 2018,

DOI:10.1080/1369118X.2018.1521457

Pre-printed version of the authors, accepted 20 August 2018

## **Deconstructing the Data Life-cycle in Digital Humanitarianism**

**Silke Roth, University of Southampton  
Associate Professor of Sociology  
Department of Sociology, Social Policy and Criminology  
Southampton, SO17 1BJ, UK**

**Contact: [silke.roth@soton.ac.uk](mailto:silke.roth@soton.ac.uk)**

**and**

**Markus Luczak-Roesch, Victoria University of Wellington  
Senior Lecturer in Information Systems  
School of Information Management  
Victoria University of Wellington  
Wellington 6140, New Zealand**

### **Abstract**

The role that technologies have historically played in producing and reproducing global inequalities is well documented. Although technological innovation is associated with progress that does not mean that it necessarily narrows the gap between rich and poor, instead technological inequalities tend to exacerbate other inequalities. This applies also to information and communication technologies (ICT) and Big Data, which play an increasingly important role in humanitarianism. In this article, we address the socio-technical work that is necessary to acquire, process, store and use data and study the power relations that are embedded in these processes. We focus in particular on the use of Big Data in digital humanitarianism and argue that at each stage of the digital data life-cycle (data acquisition, data processing, data storage, and data usage and decision making) different resources are required. These include not only access to hardware, software and connectivity but also the ability to make use of the affordances of digital technologies. We posit that in the context of humanitarianism, ICT and Big Data are a particularly intriguing to study due to their ambivalent position of seeking to address inequalities while at the same time perpetuating them.

Keywords: Big Data, data life-cycle, digital humanitarianism, inequality, power

## Deconstructing the Data Life-cycle in Digital Humanitarianism

### 1) Introduction

Digital humanitarianism -- connecting populations in regions affected by disasters with spatially and socially distanced humanitarians -- is on the rise (Meier, 2015). Very broadly conceived, digital humanitarianism contributes to the quick dissemination of information and aid. More specifically, the use of ICT in humanitarianism ranges from fundraising efforts through social media, the communication between head office and personnel based in the field, to a multitude of interventions making use of consciously and unconsciously created data that can be used to identify and track disasters and their victims and dispatch aid. UN agencies, bi-lateral and non-governmental organisations have adopted digital technologies for fundraising, advocacy and the delivery of programmes. These traditional aid organisations are joined by newly emerging organizations of digital humanitarians such as the Standby Task Force, the Humanitarian Open Street Map Team and the Digital Humanitarian Network that are engaged in data gathering on platforms such as Ushahidi, OpenStreetMap, Tomnod or Shana (Burns, 2014). The Haiti Earthquake of 2010 was one of the first disasters in which digital humanitarianism<sup>1</sup> played an important role (Meier, 2015).

Humanitarian initiatives which employ digital technologies make use of small scale and 'small data' as well as Big Data. Given our interest in the construction and use of data, in this paper we focus particularly on Big Data, though we will also explicitly point out when the aspects we discuss apply to small data contexts. The emergence and availability of Big Data are associated with the massive growth of transaction data from consumers and the explosion of new data sources including social media and mobile devices (Gandoni & Haider, 2015). Big Data are usually defined by volume, variety,

---

<sup>1</sup> The relationship between development and humanitarianism is long and complex (Roth 2015). This also applies to the relationship between ICT for Development (ICT4D) and digital humanitarianism, the latter can be seen as a subcategory of the former. For a critical assessment (Unwin 2017, ch 6) and post-colonial review of ICT4D see Chipidza and Leidner (2017).

velocity, veracity, variability and value. Volume refers to the sheer mass of data that is produced either consciously through posts on social media or as by-product of digital communication and transactions (for example the transmission of geo-data during the use of mobile devices). Variety stands for the structural heterogeneity of data of which 95% are unstructured. Velocity refers to the rate at which such data are generated. Veracity highlights the uncertainty and unreliability of such data. Variability emphasizes the variation in data flow rates. Value is relative to volume (for a detailed discussion of these six aspects see Gandoni & Haider, 2015). However, Kitchin and McArdle (2016) criticize the prevailing definitions of Big Data and point out that they lack ontological clarity. They argue that different 'species' of Big Data can be distinguished, and note that what distinguishes varieties of Big Data from small data are velocity and exhaustivity (Kitchin and McArdle, 2016: 8). The increasing use of ICT in humanitarianism through social media and digital transactions contributes to the growth and availability of varieties of Big Data which can be of further use for digital humanitarianism, but also raises new challenges related to data sovereignty and use in potentially disparate socio-cultural contexts, which we will discuss in more detail below.

In the context of digital humanitarianism, Big Data raise important questions with respect to data management and data analytics which affect decision making of those involved in the provision of aid. Especially in contexts and relationships that are characterized by inequality and power such as aid relationships, it is important to consider who is included and who is excluded from decision making processes and with what consequences. Inclusion and exclusion reflect power differences and other aspects of inequality. We argue that it is therefore important to pay attention to the data life cycle and analyse the socio-technical work that is necessary in each of its four stages (data acquisition, data processing, data storage, and data usage and decision making) in order to understand how inequalities and power relations are reproduced, transformed or overcome. Such an analysis needs to pay attention to different types of knowledge (technical knowledge, cultural

knowledge) that are needed to create, process, store, analyse and interpret data. Our paper is conceptual and based on a literature review.

The remaining paper is structured as follows: First, we outline our theoretical framework. Second, we discuss the relationship between technology, progress and humanitarianism and different types of knowledge. Third, we survey critical perspectives on digital humanitarianism and Big Data in aid contexts. Fourth, we turn to the data life cycle and identify the inequalities as they emerge during each stage. We conclude that the involvement of different types of actors in the various stages of the data life cycle has significant consequences for the decision making of humanitarian actors and the aid provided to beneficiaries.

## **2) A Constructivist Perspective on Data, Information and Knowledge**

First, we need to clarify what we mean by data, information and knowledge, and how they are constructed. Data are signals or symbols that allow for representing some difference, whereas information is data that is well-formed and meaningfully structured according to some syntax and semantics (Floridi, 2010). According to these definitions, information is genuinely socially constructed, whereas unstructured data are regarded as pseudo-information (Floridi 2010). We understand knowledge as socially constructed (Berger & Luckmann, 1966). Knowledge reflects and underpins power relations (Foucault, 1977, 1980; Rouse, 1994) and is always situated (Harding, 1986; de Sousa Santos 2014; Go, 2016). Science and Technology Studies (STS) highlight that knowledge is relational and performative, it is an outcome of and maintained through practices (Law, 2008). Given that knowledge is contextual and socially embedded, it is therefore important to distinguish different forms of knowledge, how they are ordered and how they interact. In this paper, we are particularly interested in two types of knowledge. On the one hand, we are interested in cultural and indigenous knowledge of the life-worlds that are addressed by humanitarian assistance. Such knowledge includes language skills and the familiarity with social practices, for example

migration patterns associated with work or holidays or the sharing of communication devices. On the other hand, we are interested in the technical knowledge that is needed to carry out various processes of data acquisition, preparation, storage and usage. This knowledge might be more or less complex. We note the cultural nature of technical knowledge and the overlap between the two types of knowledge. How technical and cultural knowledge are distributed among humanitarian actors and beneficiaries, and how these types of knowledge are valued reflects power relations and inequality (Roth 2012).

It is important to consider how the “information poor” and “information poverty” are discursively constructed as a result from growing technology and knowledge gaps, thus equating “knowledge” with academic thinking and technological advances in high-income countries while at the same time marginalizing indigenous knowledge (Wilson, 2003, de Sousa Santos, 2014). Critical disaster studies identify a number of risks putting too much trust in data produced through new information technologies (Crawford & Finn, 2015). It is important to understand how Big Data are used to construct information and knowledge, for example who is using social media and how the use of social media reflects cultural contexts and changes during events. In order to provide meaningful information that informs the provision of aid, the analysis of social media data in disaster contexts needs to acknowledge the source of information and be culturally responsive and socio-culturally specific. A lack of detailed information about the local context might result in a misallocation of aid in crises situations (for example Ebola outbreak, earthquakes in Haiti or Nepal).

Concerning the use of digital technologies much attention has been paid to the ‘digital divide’ (DiMaggio & Hargittai, 2001) or how unequal access to digital communication reflects and perpetuates other inequalities. In addition, to the first-level digital divide concerning access, additional levels of digital inequality have been identified (Scheerder et. al. 2017). The second level-divide concerns Internet skills and use. Halford and Savage (2010) analyse practices of using

technology, how affordances<sup>2</sup> of digital technologies are employed and how this use affects and intersects with various forms of inequality. Less attention has been paid to the third-level digital divide, which tries to assess the outcomes of Internet use (Scheerder et. al., 2017). Our deconstruction of the digital data lifecycle in digital humanitarianism concerns all three levels – access, use and outcomes. Moreover, we argue that digital inequalities, which shape digital humanitarianism, need to be understood in a broader historical perspective to which we turn in the next section.

### **3) Technology, Modernization and Humanitarianism**

Technology plays a central role in modernization processes, and social change and technological development frequently appear synonymous with progress.<sup>3</sup> Historically, Portuguese colonialism was made possible through the development of new ships and instruments – and the training to use them (Law 1986). Technologies (including ICT) enable the production and reproduction of global inequalities and power relations, and modernisation processes are inextricably intertwined with the construction of differences between “developed” and “un(der)developed” societies. Technology enabled and justified the dominance of European societies which persisted well beyond independence including the transfer of knowledge or “technical assistance” from highly developed countries to less developed countries in order to aid transformation of the later. Aid relationships imply unequal relations benevolent donors and “beneficiaries” in need of assistance. Like earlier technological interventions, ICT and Big Data are employed to provide aid to lower income and presumably less developed countries.

---

<sup>2</sup> Following Hutchby (2001) we define affordances as possibilities that enable and constrain action.

<sup>3</sup> Of course, post-modern theory provides critical perspectives on technology with respect to power relations and the impact on the environment.

Technology has always played a central role for humanitarianism and is characterized by frequent innovations and shifting agendas, which in recent years have included an emphasis on participation, empowerment, gender mainstreaming, sustainability, peace-building, human rights, internationalization, diversity (Cornwall & Brock, 2005; Cornwall, 2007) and resilience (Manyena 2006). However, these semantic shifts do not necessarily change the practices of aid organizations. On the contrary, continuities are easily detected. For example, Cooke (2003) identifies similarities between contemporary participatory approaches and indirect rule during colonialism. Of course, there are also discontinuities. Whereas the training of colonial officers included language skills and ethnology as well as valued regional specialism and in-depth knowledge of other places (Kirk-Greene, 1999), these forms of knowledge were replaced by (presumably) universal technical expertise. Professionalisation processes and the introduction of standards such as the Sphere Handbook have raised concerns that they might marginalise and weaken local actors (Roth, 2012). Furthermore, supplementary feeding programmes represent hierarchical and paternalistic control rather than self-government (Scott-Smith, 2015), while the modular standardized kits used by Medecins sans Frontières (MSF) represent the opposite of “local knowledge” (Redfield, 2008). Professionalisation processes and the adoption of new practices and technologies overall tend to strengthen international rather than local actors. The emphasis on technical knowledge and the marginalisation of local knowledge are replicated in digital humanitarianism. We argue that the digital divide reinforces existing inequalities within the aid system (Read, Thaite and Mac Ginty, 2016). In fact, digital humanitarianism represents just the latest example of “humanitarian neophilia” (Scott-Smith, 2016), a perpetual quest for innovation that leaves the existing aid and power relations intact.

As we discuss in the remainder of this paper, digital humanitarians, local, national and international aid organisations, beneficiaries, and resource-poor and crises-affected areas differ widely in their access to digital communication and their practices. Moreover, they also vary significantly in the

understanding of local power relations and culture. The positionality of those engaged of producing, processing, storing and analysing Big Data (and small data) thus matters for the information it yields and how this information affects the decisions of donors, aid organisations and other actors. Non-local actors tend to have better access to the Internet and the World Wide Web, but are less likely to have the necessary indigenous knowledge needed to interpret data and information to make meaningful decisions (Swidler & Watkins, 2017). They may lack an understanding of the appropriate practices for collecting and using data and how they are related to traditional knowledge and cultural heritage (Kukutai and Taylor, 2016). We argue that the in- and exclusion of local and non-local actors has significant impact on digital humanitarianism not only from a short-term perspective of technical effectiveness, but also in terms of long-term socio-cultural consequences. In the next section, we provide a brief review of the impact of ICT in aid contexts.

#### **4) The promises and perils of ICT in Aid Contexts**

The innovative potential of ICT includes their contribution to the dissemination of information, their potential for collaboration and cooperation concerning software and content creation, the democratisation of digital technology and the potential for citizen journalism (Kleine & Unwin, 2009).<sup>4</sup> Furthermore, new media provide opportunities for new global voices and serve as a democratizing agent (Murthy, 2013). This includes the empowerment of women in developing countries supporting community building, political organisation and the participation in feminist movements (Hilbert, 2011). Furthermore, ICT have found attention in peace-building (Welch et al., 2015). However, the optimism expressed by some of the actors in the aid industry concerning these new developments is accompanied by a critical assessment of the affordances of ICT (Unwin, 2017). This critique concerns inequality in access, the transformation of the aid systems through the influx of new actors, commercialisation, socio-cultural complexity and ethical considerations. These points concern first-, second- and third-level digital divides and we will discuss them in turn.

---

<sup>4</sup> However, social media have also been used by authoritarian regimes (cf. Tufekci, 2017).



## Access

Firstly, how ICT can support humanitarianism depends on network access, which varies across countries and regions. The first-level digital divide needs to be understood in the context of the electricity divide as many low-income countries lack access to reliable electricity, which is a precondition for the operation of ICT-related equipment (Arney & Hosman, 2016). Due to this electricity divide, access to the Internet across the African continent is extremely unequal, whereas North Africa and South Africa are most connected, conflict affected regions such as the Democratic Republic of Congo and Somalia had the lowest connectivity (Polikanov & Abramova, 2003).<sup>5</sup> In addition to electricity and connectivity, the cost of devices and data is an obstacle in low-income countries. Internet access varies significantly depending on devices, software and networks (Donner, 2015). This means that unequal access within countries also needs to be considered, for example rural/urban differences, socio-economic, age, and gender differences. Although women are as interested in ICT as men, they tend to lack the resources (time, money, education) to use ICT (Hilbert 2011). Burns (2018) shows that digital humanitarians neither solicited input from local residents concerning data collection and visualisation, nor made the collected data available which was placed behind a paywall instead of empowering a disaster affected community. Moreover, difficulties solving software problems contribute to shifting control to outside actors and empower experts and educated lay persons, thus undermining the self-reliance of communities (Brandusescu et al., 2016). Equally, the second-level digital divide is an expression of existing inequalities and “those experiencing the most severe deprivation tend to communicate less” (Resnyansky, 2015, p. 205). In fact, “digital inequalities amplify social inequalities and compound the effects of delayed recovery”

---

<sup>5</sup> The proportion of individuals in Africa using the Internet has increased from 2.1% in 2005 to 21.8% in 2017 Internet access and is lowest compared to all other world regions. Moreover, it is vastly unequal within Africa, in 2016, Internet access of individuals varied between 1.18% in Eritrea and 1.88% in Somalia, 11.77% in Sierra Leone, 20% in Rwanda, 26% in Kenya, 34.67% in Ghana, 48.05% in Gabon, and 54% in South Africa. <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> (accessed 4 April 2018).

(Madianou, 2015, p. 9) thus contributing to the third-level digital divide. As we show below, these first-, second- and third-level digital divides matter at the different stages of the digital life cycle.

#### *New actors and commercialization of aid.*

ICT and Big Data in humanitarianism are associated with the involvement of new, including commercial, actors. Private-public partnerships and interventions of corporations are closely linked with the exploration of new markets. Examples include India's biometric identification system *Unique Identification Authority of India (UIDAI)*, IBM's project *Lucy* in Kenya and *Facebook Zero* which provides free Internet access to the most populous emerging economies (Taylor and Broeders, 2015). The resulting data are owned privately, which raises questions concerning access. Furthermore, initiatives such as *Facebook Zero* shape the way information on the World Wide Web can be accessed and used and have clear profit motives, which are known to spark novel (algorithmic) biases and have significant impact on information diffusion (Lerman, 2016). And even if the use of the Internet and social media are offered for free, they require hardware to access it. Moreover, the introduction of ICT and social media platforms in middle and low-income regions might contribute to "extending a markedly North American world view into the developmental sphere" (Thompson, 2004, p. 1). Furthermore, the dissemination of large commercial platforms such as Facebook tends to displace indigenous platforms, and shapes and transforms online social interaction.

#### *Ethical questions*

The involvement of private actors is associated with the "*datafication*" in low and middle income countries, which results in the shift from official statistics to corporate-led data collection and analytics transforming individuals "from citizen to data subject" (Taylor & Broeders, 2015, p. 231). Whereas citizens provide data knowingly (for example by consciously providing census data), ICT users inadvertently (and perhaps unknowingly) create Big Data that can be mapped and

manipulated. This is particularly problematic as the “overwhelming majority of states without privacy or data-protection laws are low and lower-middle-income countries” (Taylor & Broeders, 2015, p. 236). The use of Big Data in emergency contexts require rethinking privacy and ethical questions (Burns & Thatcher, 2015). Storing tweets containing personal information provided in a situation of need in databases raises questions concerning ethical issues and meaningful informed consent (Crawford & Finn, 2015; Kinder-Kurlanda et. al. 2017). In particular, in political contexts such as the Libya Crisis, information provided on public accessible maps can put those volunteering this information at risk (Burns, 2014). These ethical questions raise important issues concerning the outcomes of Internet use or the third-level digital divide.

It is important to note that the rise of digital humanitarianism intersects with increasing numbers of attacks on humanitarian aid workers. The introduction of digital and remote interventions thus coincides with responses to security threats such as the removal of international staff from the field and seeking security in fortified compounds (Duffield, 2012, 2016). This means that data are produced by local actors but analysed in head offices excluding “small-scale actors with local knowledge and understanding” (Taylor & Broeders, 2015, p. 232) from analysis and decision-making. There is a risk that local communities do not trust interventions that fail to build relationships and ignore local conditions. The withdrawal of international actors results in a “resilience of the ruins” (Duffield, 2016), meaning that disaster-affected communities are forced to survive accessing support through electronic interfaces such as cash-transfer programmes but are otherwise abandoned by aid agencies. Although disaster risk reduction based on the analysis of Big Data may allow (and require) adaptation of communities at risk and thus enhance “community resilience”, this does not address root causes of disasters (Chandler, 2015).

Thus, we identify a number of issues that matter with respect to the contributions that ICT and Big Data can make to humanitarianism (i) Who is involved in generating the data? (ii) How does this

shape the content of available information? (iii) What are the implications for intervention? We suggest that these questions need to be raised for the four stages of the data life-cycle, which comprises key tasks such as 1) data acquisition, 2) data processing and analysis, 3) data storage and curating and 4) data visualisation and decision making. Analysing the socio-technical work that is going on at the different stages of the data life-cycle allows us to demonstrate that access to the various stages matters for information and knowledge that are produced and how they are used in decision-making.

**Figure 1 about here**

### **5) The Digital Data Life-Cycle – Participation, Inclusion and Exclusion**

Data, whether consciously created in the context of research projects, official statistics or as user data (for example resulting in Big Data), goes through different stages in the data life-cycle.<sup>6</sup> The processes and activities for the creation and maintenance of data are usually organised as two types of data life-cycles. First we distinguish data life-cycles that represent, link and share data from a particular knowledge domain, for example in the biomedical field, where domain experts assisted by technical experts and sophisticated tools create so called Web ontologies that allow for representing data with rich semantics and to make logical inferences (Simperl & Luczak-Rösch, 2014). This type of life-cycle often involves dedicated processes to reuse or develop an adequate schema or controlled vocabulary and is known to ease data sharing and integration. Second, we recognize digital data life-cycles which involve the analysis of large amounts of instance data from heterogeneous sources - often referred to as Big Data life-cycles (Jagadish et al., 2014). Digital humanitarianism combines these two types of data life-cycles and includes individuals (e.g. data journalists) processing and

---

<sup>6</sup> For an illustration of the data life-cycle see the advice for researchers how to create and manage data which can be deposited and reused provided by the UK Data Archive. <http://www.data-archive.ac.uk/create-manage/life-cycle>

visualising data which are then used by other actors as well as groups collaborating in some formal or informal organisational form (e.g. the Humanitarian OpenStreetMap Team).

Each stage in the digital data life-cycle requires a different set of resources and capacities. By *resources* we mean the devices that enable technical access, for example mobile phones, smartphones and computers of varying power as well as access to the Internet through broadband or satellites. By *capacities*, we mean the intellectual capacity and knowledge required to use these resources. Having access to devices obviously allows for the development of capabilities that can be acquired through use, however, depending on the extent of digital literacy owning a device does not necessarily mean making use of its affordances (Halford & Savage, 2010). But being able to analyse data through powerful statistical programmes does not suffice for a meaningful interpretation of data which requires familiarity with the life-worlds of user groups. Thus two types of knowledge – on the one hand the technological know-how (for example the ability to use sophisticated statistical programmes), on the other hand the context specific knowledge to interpret the findings – are required to understand and use data for informed decisions. We argue that it is important to pay attention to i) these different kinds of knowledge, ii) who has access to these different types of knowledge and iii) to what extent different groups are involved in the different stages of the data life-cycle in digital humanitarianism. How different types of knowledge are valued, determines the role they play for the decision making of actors in the aid world and beyond (Smith, 2013). In the next section of this paper, we present our analysis of the resources and capacities that are needed to make use of the affordances of Big Data and social media. We assess how inclusive these technologies and practices are and how they shape the participation at each stage of the data life-cycle.

### *Data Acquisition*

We distinguish data that are consciously produced by users and those that are created as a by-product of public or private transactions. Consciously produced data include micro-posts or text messages and contributions to crowd-sourced maps via platforms like Ushahidi. Other data are created as by-product for example geo-data that results from making phone-call and accessing the Internet. This also includes customer- and citizen-data that are created when using services that leave digital traces, for example consumer data, access to government websites or the use of search engines. Both types of data production – conscious and unconscious – require access to digital forms of communication, though not necessarily owning devices. Those who have more access to devices have the ability to leave more traces whereas those who have less access to devices might remain invisible. In addition, those who share devices muddle the information provided through digital traces – while the device can be traced, the user might not be identifiable. Thus, the more autonomy and access individuals and groups have in making calls, sending (and receiving) text messages or using social media the more they contribute to crisis information by producing more digital data. Sharing devices does not only constrict access temporally, but might also constrain users to share information as it might be monitored by (co-)owners of such devices which serve as gate-keepers to digital communication. With respect to data production we thus can – drawing on Madianou (2015) – distinguish between four different groups: “media poor”, “minimal owners”, “moderate owners” and “media rich”. With respect to data that is inadvertently and involuntarily created as by-product of transactions, we can also distinguish two user groups – those who leave digital traces and those who don’t. This can result in “black holes” – while some communities and areas are visible, others aren’t. Moreover, when devices such as mobile and smart phones are co-owned, this results in methodological challenges for the analysis and interpretation of mobile phone data (Tatem et al., 2014; Wesolowski et al., 2014), since it cannot be assumed that a digital trace stems from just one single device user.

Thus access to technology which allows to engage in digital communication is a necessary but not sufficient condition for involvement in data acquisition. What matters as much or even more is the ability to skilfully employ digital technologies, not only by making use of the affordances of various platforms and devices, but also by knowing whom to contact and what information to provide and how to leave - or avoid to leave - a digital footprint. Moreover, participation does not result in equal digital representation: not every tweet is re-tweeted, rather certain users have larger groups of followers and thus potentially more influence than others whose micro-blogs are not picked up (Tinati et al., 2014; Lerman, 2016). However, those who are digital media savvy are not necessarily those who are most familiar with the local conditions and culture of those affected by poverty and crisis.

The first stage of the data life-cycle provides the broadest opportunity for participation. Data can be submitted by a wide range of devices which might be shared rather than owned. However, variations in access to ICT and the savvy to use devices results in unequal contributions to data production. Exclusion from this stage of the data life-cycle affects the data that is produced and has significant consequences for the next stages.

### *Data Processing*

The second stage of the data life-cycle concerns the processing, filtering and cleaning of data before they are put in a data storage solution on which data analysis can be carried out for sense and decision making. Tweets or Ushahidi posts for example need to be filtered, translated, mapped and verified. False and biased data need to be identified. These are very often tasks for crowdsourcing in which large numbers of paid or unpaid crowd workers engage in tagging tweets and translating Ushahidi post from local and vernacular languages into – mostly – English, the language of the aid industry and the web (Olteanu et al., 2015). Once translated such crowds-sourced data might no longer be accessible to those who provided initially information in local languages (Crawford & Finn,

2015). In addition to human coding, algorithmic methods are also employed at this stage (Purohit et al., 2014). Data classification also includes the classification of images that were provided by satellites or drones.

The capacity to engage in data preparation and the use of algorithms require specialised skills. However, these technical skills need to be combined with social and cultural knowledge. A lack of familiarity with culture and local conditions might result in mistakes in data processing, for example exclusion of vital information because it was falsely considered as irrelevant. Sampling, filtering and coding of raw data has significant impact on the data set. In the case of the Haiti earthquake, SMS message gateways received requests for help and information about environmental disruptions out of a crisis region. These messages were analysed and translated from Haitian Kreyole and French into English (the language spoken by the primary emergency response teams) by volunteers in a ten minutes turnaround time on average (Munro, 2010). However, there was no capacity to translate any information from English into the local languages to make sure local people stay informed as well. The fact that the translators' resources were employed to serve the response teams' purposes highlights that limited resources may harm the ability to provide a two-way flow of information equally in and out of the crisis region.

While individuals in crisis affected regions might be providers of information, they are less likely to engage in the data processing or have access to processed data. This inequality has consequences for the following stages of the data life-cycle. Local people and groups are successively excluded from the information flow (Mulder et al., 2016). Moreover, even highly skilled local tech workers responsible for the implementation of digital humanitarian projects experience marginalization "used and discarded as quickly as the technology they are hired to test" (Ong and Combinido, 2018, p. 100).



### *Data Storage*

The third stage of the data life-cycle concerns how data are stored and who has access to them. Data access constitutes a complexity continuum comprising different dimensions including data formats, legal and ethical issues. Proprietary or uncommon data formats prevent the use of data or make it very costly. In practice we witness the complexity that arises from heterogeneous data formats when looking at the Humanitarian Data Exchange project<sup>7</sup>. As long as crisis data are shared in CSSV, Excel, or JSON format, links and mappings between data from different sources are missing and it requires effort to integrate them with other data. The humanitarian community could greatly benefit from adopting the Linked Data principles (Heath & Bizer, 2011) to upgrade the quality of data by following a common standard that increases interoperability (Steiner & Verborgh, 2015).

With respect to data storage, the ownership of data represents another problem for digital humanitarians. Data that are a by-product of commercial transactions (e.g. mobile phone usage) are private and access is likely to be restricted requiring negotiation. Similarly, data collected by public administration (for example health services, social services, criminal justice system) as well as by non-governmental organisations cannot be shared as this would violate privacy.

Data storage is a multi-faceted problem that embodies the danger of exclusion a) technically, since data formats or the volume may exceed users' capabilities and b) legally, since the diverse terms of use may cause legal uncertainty and threats to the users holding the data. Even data that were consciously and deliberately shared by individuals on platforms such as Twitter and Facebook cannot simply be publicly provided elsewhere as this violates the terms and conditions of those commercial platform providers and raise ethical issues. Similar to the previous stage (data processing), data storage requires more resources and technical capacities than the first stage in the data life-cycle,

---

<sup>7</sup> <https://data.humdata.org/> accessed and investigated on 11 March 2018.

data acquisition. It is therefore exclusionary in that it prevents those who contributed to the production of data from accessing it.

### *Data Usage and Decision Making*

The last stage of the data life-cycle comprises analysis and visualisation which inform decision making. Especially for large data sets, this means access to powerful computers or compute clusters. In addition, it requires the skills to employ data mining, statistical programming, machine learning and data visualisation strategies. Even though we see an increasing number of open source and free software tools becoming available for data analysis and visualization tasks, these do not immediately empower all people equally (Donner, 2015). Very often the free versions of these tools are only available with limited cloud resources and running them at significant scales requires to subscribe to paid plans or the actual value is not in the raw availability of a tool or algorithm but the actual model derived from extensive training with real data. When these resources are not made available alongside each other, the most promising algorithm can be useless for stakeholders that lack the ability to understand or reproduce a useful model for the problem at hand.

Meaningful data analysis requires time and humanitarian aid is a time critical matter of significant scale. Crowdsourcing thus plays a crucial role in the creation of data sets using Big Data. However, information flows in disaster settings can differ significantly. During the Haiti Earthquake of 2010, information flowed from affected people to translators to digital humanitarians. In contrast, during the earthquake in Nepal in 2015 information flows reached back to the affected communities (Mulder et al., 2016). However, even though in Nepal significant efforts were made to involve grassroots groups, these groups lacked the adequate resources. Furthermore, international organisations did not use the crowdsourced data (Mulder et al, 2016).

## 6) Discussion

Digital inequality goes beyond differences in access and includes making use of the affordances of digital technologies (Halford & Savage, 2010). Thus, how individuals and groups engage with ICT needs to be understood within social, cultural, political and economic contexts. Moreover, in particular in the context of digital humanitarianism, it is important to acknowledge the importance of local and culturally-specific knowledge, which is crucial for creating meaningful web-based services and adequately interpreting Big Data in emergency contexts. While international actors and “experts” might have more technical knowledge, which local actors tend to lack, the latter are more likely to possess crucial context knowledge.

Our analysis of the four stages of the data life-cycle describes what resources and capacities are necessary at each stage. The four stages vary with respect to inclusion and participation. We conclude that the stage of data acquisition is most inclusive and allows for the involvement of a wide range of participants. These participants include those in resource-poor regions who are affected by humanitarian crises and those in resource-rich areas who are providing support through donations and digital humanitarianism. At the same time, this is also the stage with the largest risk of exploiting the most vulnerable, who may not have had their voices heard during data collection, who may no longer have access to the data they have provided in the subsequent stages, and who might not even know that they contributed to data production through the use of ICT. The following stages – processing, storage, and usage and decision making – become successively more exclusionary as they require having access to data and powerful analytical tools to analyse them. We argue that the exclusion of local actors from the analysis of data is problematic as long as the results of the analysis shape the decisions of aid organisations without taking into consideration culturally and context specific knowledge. Furthermore, differences between local actors, which include “media rich” and “media poor” groups (Madianou 2015) need to be considered.

The widespread availability of new information technologies even in the remotest areas is intertwined with the increased emphasis on security which in some regions resulted in “bunkerization” (Duffield, 2010) and shapes the relationship between international staff and their colleagues and the local population. International aid personnel are increasingly separated from the local population through the use of armed vehicles, guards and barbed wires. Access to information technologies and security measures thus detach international aid personnel from local realities and local countries. Digital humanitarians – whether hyper-bunkered in disaster regions or in head offices far removed from such sites – engage in “crisis informatics” and analyse data that has been produced by using ICT and social media (Duffield, 2016). This means that rather than self-governance, ICT and Big Data offer opportunities for “immense marketing, behavioural management and security surveillance possibilities for organisations, corporations and states” (Duffield, 2016: 152). This includes biometric registration, enabled through a lack data of regulation and privacy laws, which is implemented by local actors but utilised by remote actors. We do not see any data-centric solutions that address crises in a more holistic way and with the goal to provide local actors with capacities to help themselves before, during and after the actual event. In fact, Chandler (2016) warns that

“Big Data as a technique of knowledge production and of governance, profoundly constrain the possibilities for politics: reducing governance to an ongoing and technical process of adaptation, accepting the world as it is” (Chandler, 2016: 835)

## **7) Conclusions: Back to the Future – Technology, Progress and Global Inequalities**

In this paper, we have argued that more attention needs to be paid to the socio-technical work in digital humanitarianism and how it reinforces existing inequalities. Our analysis suggests that the introduction of innovative disaster response platforms and the use of Big Data in digital humanitarianism are undersupplied with bottom up pathways into the data life-cycle. We therefore propose to develop bottom up approaches that complement the current top down digital

humanitarianism by providing local actors in low-income countries with technologies that build on indigenous knowledge, respect indigenous data sovereignty, and enhance their abilities to articulate and voice their needs and build the capacity of local organisations which are so far marginalised in the aid system. Top-down methods of collecting data, for example in HIV/AIDs research risk dismissing and ignoring the experiences and perceptions of local communities (Chilisa, 2012). In order to include local sources, there should be support for the development and distribution of technologies that allow local actors to communicate their needs and become co-producers of knowledge. This includes issues of access to devices, electricity and connectivity, software that adapts to a range of languages beyond English as well as training to make use of ICT and Big Data at all stages of the data life-cycle. In particular, the stage of data usage and decision making needs to become much more inclusive in order to address global inequality and power differences.

Moreover, addressing the socio-technical work that is going on in the data life-cycle in digital humanitarianism and acknowledging how it is shaped by and shaping these inequalities means putting the use of ICT and Big Data in humanitarianism into a broader political and historical context. Technology driven interventions – including ICT and Big Data – cannot solve deep-rooted problems associated with global inequality and injustice, but perpetuate the “international denial of responsibilities and interconnections” (Chandler, 2016, p. 395). Moving beyond the “resilience of the ruins” (Duffield, 2016) requires not only transparency and access to all stages of the data life-cycle, but also resources and power to address the underlying causes of poverty, disaster and conflict. As long as technical innovation is driven by organizations from the Global North who have the means to employ new approaches and implement them in low(er) income countries (Krause 2013), global inequalities will be perpetuated.

## References

Armey, L. E. & Hosman, L. (2016). The centrality of electricity to ICT use in low-income countries.

*Telecommunications Policy*, 40(7), 617–627.

Atzori, L., Iera, A. & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*

54(15), 2787–2805.

Berger, P. L. & Luckmann, T. (1966). *The Social Construction of Reality. A Treatise in the Sociology of*

*Knowledge*. New York: Doubleday.

Brandusescu, A., Sieber, R.E. & Jochems, S. (2016). Confronting the hype: The use of crisis mapping

for community development. *Convergence: The International Journal of Research into New Media*

*Technologies*, 20(6), 616–632.

Burns, R. (2014). Moments of closure in the knowledge politics of digital humanitarianism.

*Geoforum*, 53(0), 51–62.

Burns, R. (2015). Rethinking big data in digital humanitarianism: practices, epistemologies, and social

relations. *GeoJournal* 80(4), 477–490.

Burns, R. (2018). Datafying Disaster: Institutional Framings of Data Production Following Superstorm

Sandy. *Annals of the American Association of Geographers* 108(2): 569-578.

Burns, R. & Thatcher, J. (2015). Guest Editorial: What's so big about Big Data? Finding the spaces and

perils of Big Data. *GeoJournal*, 80(4), 445–448.

Chandler, D. (2015). A World without Causation: Big Data and the Coming of Age of Posthumanism. *Millennium - Journal of International Studies* 43(3), 833–851.

Chilisa, B. (2012). *Indigenous Research Methodologies*. Los Angeles: Sage.

Chipidza, W. & D. Leidner (2017). ICT4D Research – Literature Review and Conflict Perspective. *Twenty-third Americas Conference on Information Systems, Boston 2017*.

Cooke, B. (2003). A new continuity with colonial administration: participation in development management. *Third World Quarterly*, 24(1), 47–61.

Cornwall, A. (2007). Buzzwords and fuzzwords: deconstructing development discourse. *Development in Practice*, 17(4), 471–484.

Cornwall, A. & Brock, K. (2005). What Do Buzzwords Do for Development Policy: A Critical Look at 'Participation', 'Empowerment' and 'Poverty Reduction'. *Third World Quarterly*, 26(7), 1043–1060.

Crawford, K. & Finn, M. (2015). The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal*, 80(4), 391–402.

de Sousa Santos, B. (2014). *Epistemologies of the South: Justice against Epistemicide*. Boulder, CO, Paradigm Publishers.

Donner, J. (2015). *After Access. Inclusion, Development, and a More Mobile Internet*. Cambridge, MA: MIT Press.

Duffield, M. (2010). Risk-Management and the Fortified Aid Compound: Everyday Life in Post-Interventionary Society. *Journal of Intervention and Statebuilding*, 4(4), 453–474.

Duffield, M. (2012). Challenging environments: Danger, resilience and the aid industry. *Security Dialogue*, 43(5), 475–492.

Duffield, M. (2016). The resilience of the ruins: towards a critique of digital humanitarianism. *Resilience*, 4(3), 147–165.

Floridi, L. (2010). *Information: A very short introduction*. Oxford: Oxford University Press.

Foucault, M. (1977) *Discipline and Punish*. New York: Pantheon.

Go, J. (2016). *Postcolonial Thought and Social Theory*. New York, Oxford University Press.

Halford, S. & Savage, M. (2010). Reconceptualizing Digital Social Inequality. *Information, Communication & Society*, 13(7), 937–955.

Harding, S. (1986). *The Science Question in Feminism*. Ithaca, NY: Cornell.

Heath, T. & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1), 1–136.

Hilbert, M. (2011). Digital gender divide or technologically empowered women in developing countries? A typical case of lies, damned lies, and statistics. *Women's Studies International Forum*, 34(6), 479–489.



Hutchby, I. (2001). Technologies, Texts and Affordances. *Sociology*, 35(2), 441–456.

Jagadish, H., 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.

Kirk-Greene, A. (1999). Public administration and the colonial administrator. *Public Administration and Development*, 19, 507–519.

Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J. & Morstatter, F., 2017. Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*, 4(2).

Kitchin, R. & G. McArdle (2016) What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3 (1): 2053951716631130.,

Kleine, D. & Unwin, T. (2009). Technological Revolution, Evolution and New Dependencies: what's new about ict4d? *Third World Quarterly*, 30(5), 1045–1067.

Kshetri, N. (2017). The economics of the Internet of Things in the Global South. *Third World Quarterly*, 38(2), 311-339.

Kukutai, T. & J. Taylor (eds.) (2016) *Indigenous data sovereignty: Toward an agenda*. Canberra: Anu Press.

Law, J. (1986). On the methods of long-distance control: vessels, navigation and the Portuguese route to India. *The Sociological Review*, 32:234 –263.

Law, J (2008). On sociology and STS. *The Sociological Review* 56(4): 623–649.

Lerman, K. (2016). Information is not a virus, and other consequences of human cognitive limits. *Future Internet*, 8(2), 21.

Madianou, M. (2015). Digital Inequality and Second-Order Disasters: Social Media in the Typhoon Haiyan Recovery. *Social Media + Society*, 1(2). First published date: September-11-2015. 10.1177/2056305115603386.

Manyena, S. B. (2006). The concept of resilience revisited. *Disasters*, 30(4), 434–450.

Meier, P. (2015). *Digital Humanitarians. How BIG DATA is changing the face of humanitarian response*. London: CRC Press.

Mulder, F., Ferguson, J., Groenewegen, P. Boersma, K. & Wolbers, J. (2016). Questioning Big Data: Crowdsourcing crisis data towards an inclusive humanitarian response. *Big Data & Society*, 3(2). First Published August -1- 2016 10.1177/2053951716662054

Munro, R. (2010). Crowdsourced translation for emergency response in Haiti: the global collaboration of local knowledge. *AMTA Workshop on Collaborative Crowdsourcing for Translation*, 1-4.

Murthy, D. (2013). New Media and Natural Disasters. *Information, Communication & Society*, 16(7), 1176–1192.

Olteanu, A., Vieweg, S. & Castillo, C. (2015). What to expect when the unexpected happens: Social media communications across crises. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, 994–1009.

Ong, J. C., & Combinido, P. (2018). Local aid workers in the digital humanitarian project: between 'second class citizens' and 'entrepreneurial survivors'. *Critical Asian Studies*, 50(10), 86–102.

Polikanov, D. & Abramova, I. (2003). Africa and ICT: A Chance for Breakthrough? *Information, Communication & Society*, 6(1), 42–56.

Purohit, H., Hampton, A., Bhatt, S., Shalin, V. L., Sheth, A. P. & Flach, J. M. (2014). Identifying seekers and suppliers in social media communities to support crisis coordination. *Computer Supported Cooperative Work (CSCW)*, 23(4-6), 513–545.

Read, R., Taihe, B. & Mac Ginty, R. (2016). Data hubris? Humanitarian information systems and the mirage of technology. *Third World Quarterly*, 37(8), 1314–1331.

Redfield, P. (2008). Vital Mobility and the Humanitarian Kit. In A. Lakoff & S. Collier (Eds.), *Biosecurity Interventions: Global Health and Security in Question*. (pp. 147–171) New York, NY: Columbia University Press.

Resnyansky, L. (2015). Social media data in the disaster context. *Prometheus*, 33(2), 187–212.

Roth, S. (2012). Professionalisation Trends and Inequality: experiences and practices in aid relationships. *Third World Quarterly*, 33(8), 1459–1474.

Roth, S. (2015). *Paradoxes of Aid Work. Passionate Professionals*. London/New York: Routledge.

Rouse, J. (1994). Power/Knowledge In G. Gutting (Ed.), *The Cambridge Companion to Foucault* (pp. 92-114). Cambridge, UK: Cambridge University Press.

Scott-Smith, T. (2015). Control and Biopower in Contemporary Humanitarian Aid: The Case of Supplementary Feeding. *Journal of Refugee Studies*, 28(1), 21–37.

Scott-Smith, T. (2016). Humanitarian neophilia: the ‘innovation turn’ and its implications. *Third World Quarterly*, 37(12), 2229–2251.

Scheerder, A., A. van Deursen and J. van Dijk (2017). "Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide." *Telematics and Informatics* 34(8): 1607-1624.

Simperl, E. & Luczak-Rösch, M. (2014). Collaborative ontology engineering: a survey. *The Knowledge Engineering Review*, 29(01), 101–131.

Smith, L. T. (2013). *Decolonizing methodologies: Research and indigenous peoples*. London: Zed Books.

Steiner, T. & Verborgh, R. (2015). Disaster monitoring with Wikipedia and online social networking sites: structured data and linked data fragments to the rescue? *AAAI Spring Symposium 2015 arXiv: 1501.06329*.

Swidler, A. & Watkins, S.C. (2017). *A Fraught Embrace. The Romance and Reality of AIDS Altruism in Africa*. Princeton: Princeton University Press.

Tadele, F. & Manyena, S. B. (2009). Building disaster resilience through capacity building in Ethiopia. *Disaster Prevention and Management*, 18(3), 317–326.

Tatem, A. J., Huang, Z., Narib, C., Kumar, U., Kandula, D., Pindolia, D. K, ....Lourenço, C. (2014). Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria Journal*, 13(1), 1–16.

Taylor, L. & Broeders, D. (2015). In the name of Development: Power, profit and the datafication of the global South. *Geoforum*, 64, 229–237.

Thompson, M. (2004). ICT, power, and developmental discourse: A critical analysis. *The Electronic Journal on Information Systems in Developing Countries*, 20(4), 1–26.

Tinati, R., Halford, S., Carr, L. & Pope, C. (2014). Big Data: Methodological Challenges and Approaches for Sociological Analysis. *Sociology*, 48(4), 663–681.

Tufekci, Z. (2017) *Twitter and Tear Gas. The Power and Fragility of Networked Protest*. New Haven: Yale University Press.

Unwin, T. (2017) *Reclaiming Information and Communication Technologies for Development*. Oxford: Oxford University Press.

Welch, J. R., Halford, S. & Weal, M. (2015). Conceptualising the web for post-conflict governance building. *Peacebuilding*, 3(1), 58–74.

Wesolowski, A., Buckee, C.O., Bengtsson, K. Wetter, E., Lu, X. & Tatem, A. J. (2014). Commentary: Containing the Ebola Outbreak - the Potential and Challenge of Mobile Network Data. *PLoS Currents* 6: ecurrents.outbreaks.0177e0177fcf52217b52218b634376e634372f634373efc634375e.

Wilson, M. (2003). Understanding the international ICT and development discourse: Assumptions and implications. *The Southern African Journal of Information and Communication*, 3(1), 80–93.

Figure 1 The Digital Data Life-cycle

