

Questioning Big Data: Crowdsourcing crisis data towards an inclusive humanitarian response

Femke Mulder, Julie Ferguson, Peter Groenewegen,
Kees Boersma and Jeroen Wolbers

Big Data & Society

July–December 2016: 1–13

© The Author(s) 2016

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/2053951716662054

bds.sagepub.com



Abstract

The aim of this paper is to critically explore whether crowdsourced Big Data enables an inclusive humanitarian response at times of crisis. We argue that all data, including Big Data, are socially constructed artefacts that reflect the contexts and processes of their creation. To support our argument, we qualitatively analysed the process of 'Big Data making' that occurred by way of crowdsourcing through open data platforms, in the context of two specific humanitarian crises, namely the 2010 earthquake in Haiti and the 2015 earthquake in Nepal. We show that the process of creating Big Data from local and global sources of knowledge entails the transformation of information as it moves from one distinct group of contributors to the next. The implication of this transformation is that locally based, affected people and often the original 'crowd' are excluded from the information flow, and from the interpretation process of crowdsourced crisis knowledge, as used by formal responding organizations, and are marginalized in their ability to benefit from Big Data in support of their own means. Our paper contributes a critical perspective to the debate on participatory Big Data, by explaining the process of in and exclusion during data making, towards more responsive humanitarian relief.

Keywords

Disaster, crowdsourcing, Nepal, Haiti, humanitarian, Big Data

Introduction

In this paper we analyse the Big Data making processes deployed by a new type of (semi-formal) humanitarian organization in disaster settings: the geographically dispersed networks of *digital humanitarians* (Meier, 2015). These humanitarian networks of technology volunteers rely on crowdsourcing and are organized around open civic technology platforms. Global attention to this way of making big crisis data was catalysed by the 2010 earthquake in Haiti. During this disaster, the open source platform Ushahidi was used to produce a crisis map based on volunteer-driven information sharing, generating around 50,000 incident reports (Gao et al., 2011; Morrow et al., 2011). At the time of the crisis, this map was proclaimed by the US Federal Emergency Management Authority (FEMA) as the 'the most comprehensive and up-to-date source of information on Haiti for the humanitarian community'

(as reported in Heinzelman and Waters (2010: 9)). Thus, this new way of organizing at times of disaster is often celebrated as a Big Data success, with the potential to more effectively address on-the-ground needs, while also enabling people's agency (e.g., World Bank, 2013).

Crowdsourcing and open data platforms in humanitarian disaster settings are widely advocated in both scholarly and practitioner domains. Whilst such interventions can play an important role in acquiring vital situational knowledge to aid response efforts, the underlying assumption that such methods by default

Vrije Universiteit, The Netherlands

Corresponding author:

Femke Mulder, Department of Organization Sciences, De Boelelaan 1081,
1081 HV Amsterdam, The Netherlands.

Email: f.mulder@vu.nl



Creative Commons NonCommercial-NoDerivs CC-BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License (<http://www.creativecommons.org/licenses/by-nc-nd/3.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

strengthen citizen's participation in a humanitarian response is worth critically examining. We show how data collected through crowdsourcing by digital humanitarians is often transferred and mutated through editing by different actors in the social process that links crisis knowledge derived from 'the crowd', via volunteer efforts, to formal aid organizations. In other words, people directly affected by the crisis are often excluded from information processing and interpretation, and marginalized in subsequent response decision-making that affect their very lives.

By critically examining the processes through which digital humanitarians create Big Data, we aim to provide more clarity on how and to what degree affected citizens can contribute to and access humanitarian crisis information. Such understanding is important to help this new type of humanitarian organization improve the preconditions for enabling citizen participation through Big Data in a humanitarian response. We flip the conventional view of data as the raw building blocks of knowledge, arguing instead that data is generated from different sources of knowledge. Extending this perspective, we explain the social process that underlies the creation, editing and translation of data and show how exclusion can inadvertently result from crowdsourcing initiatives that aspire to realizing precisely the opposite.

The paper is based on our familiarity with the 2010 Haiti crisis and on fieldwork in Nepal, carried out two months after the 2015 earthquake and a second visit a year later. By exploring crowdsourcing and open data initiatives that were used in the immediate aftermath of these crises, we aim to establish how community participation in creating – and using – crisis data was enabled and hindered in these crowdsourced aid efforts, what efforts were undertaken by crowdsourcing platforms to address barriers to inclusion and what challenges still remain. Thus, the paper is guided by the overall research question: *How can the promise of participatory Big Data be understood from a critical view on inclusive humanitarian data making?*

First we introduce and take position within the current debate on Big Data within the context of humanitarian response. Next, we outline how crowdsourcing and open access platforms were used in the immediate aftermath of the earthquake in Haiti, drawing on the examples of Open Street Map (OSM) and Ushahidi. We then compare the response to Haiti with the Nepal earthquake disaster and in particular the humanitarian aid efforts undertaken by Kathmandu Living Labs (KLL), Code for Nepal and the Mobile Citizens Helpdesks. Finally, we discuss the implications of our study for research and practice, and conclude with an overall assessment of the extent to which crowdsourcing humanitarian Big Data enables different

socially constituted groups to participate in the creation and management of crisis data. That is, we analyse to what extent crowdsourcing Big Data contributes to an inclusive humanitarian response.

Community participation in Big Data

Digital humanitarians believe in the empowering potential of open Big Data (Meier, 2015). Especially open, crowdsourced Big Data is seen to have the potential to foster democracy and innovation due to its transparency and broad stakeholder base (Baack, 2015). In this paper, we interpret big crisis data as referring to all datafied – and datafiable – information about the disaster at hand that people knowingly (and unknowingly) share electronically – either directly or through intermediaries. It covers a wide array of channels and formats, ranging from Facebook and Twitter messages, to (online) government records, to content created by formal media outlets, to responses and comments recorded in needs assessment surveys NGOs (summarize and) post online (Kitchin, 2014).

Open data is data that is made freely available, accessible for all to use as needed. In order to facilitate empowerment through open data, activists have developed civic technologies, specialized applications run on open data platforms that aim to connect people in order to 'improve public services or help citizens to coordinate with each other to solve problems together' (Baack, 2015: 7). In other words, by opening up data, citizens are expected to gain access to information that informs policies, strategies and actions taken by governments, firms and other organizations and become more actively engaged in decision-making processes that affect them (e.g., Baack, 2015). This perspective builds on the long-standing participation debate, aimed at fostering aid efforts that are more inclusive of intended beneficiaries (Hickey and Mohan, 2004), and which has also been addressed in the context of ICT-enabled information exchange (Ferguson and Soekijad, 2015).

In addition to aiming to provide citizens with access to data that informs the decisions made by governments and other formal organizations, civic technology platforms sometimes set out to enable citizens to add to and edit this data. This is often done through crowdsourcing, a method of data creation originally defined as 'the act of taking a job traditionally performed by a designated agent [...] and outsourcing it to an undefined, generally large group of people in the form of an open call' (Howe, 2008: 99). In this paper we distinguish between limited, functional data sets and Big Data: whereas the information contained in the former could in theory be processed by a small team with a desktop computer, Big Data is so vast and

abundant that ‘we have to turn to other means of analysis: people working together, or [with multiple] computers, or both’ (Horowitz, 2008).

Crowdsourcing in our humanitarian cases refers to a method of creating a limited, functional data set from Big Data, both by filtering and processing existing shared, verbalized information and by encouraging the creation of new data from hitherto unshared knowledge. Its distinguishing feature is that its method is social and that the resulting data set is user generated, as opposed to the product of a formal institution (Kaplan and Haenlein, 2010). Crowdsourcing, then, refers to both the social creation and filtering of Big Data. The open data humanitarian civic technologies we discuss in this paper aim to use crowdsourcing to enable communities to share their knowledge about the (local impact of) the disaster and their needs with the wider world.

The ambition to empower communities by including their local knowledge into data sets used by government and other organizations is well established, yet contested. Indeed, the ideal of development is often expressed as people’s freedom to pursue their interests independently and be able to fulfil their fundamental human needs (Sen, 1999). This refers to both material interests and needs (such as a dependable livelihood and a clean environment) and non-material interests and needs (such as social inclusion and the ability to participate in politics) (Madon, 2000). Given that the fulfilment of the interests and needs of intended beneficiaries is increasingly recognized as a key indicator of development, the inclusion of local communities in development planning and agenda setting is at the heart of the academic and practitioners’ debate on participatory aid and development (Cooke and Kothari, 2001; Hickey and Mohan, 2004).

Central to such an approach is the idea that people’s local knowledge be incorporated into project planning (Mosse, 2001). However, the idea that local knowledge or individual needs simply exist – to be recorded and taken into account in humanitarian or development programme planning – is problematic, as for instance Mosse (2001) contends: ‘local knowledge (such as community needs, interests, priorities and plans) is a construct of the [programme] planning context, behind which is concealed a complex [...] politics of knowledge production and use’ (2001: 387). Indeed, local knowledge that is created in a development context for the purpose of a specific project will often end up reflecting both local power dynamics as well as outside agendas, such as those of the international agencies that fund the project (Ferguson et al., 2010). The reason for this is the fact that communities often strategically choose to present a list of ‘local needs’ that meet certain criteria (i.e. that are perceived as ‘legitimate’ needs by

the agency and its global donor) and that fit the community’s perception of what the project is able to offer (Mosse, 2001). As such, local knowledge tends to be the product of both local and global factors.

We adopt a similar approach, recognizing knowledge as a contested social construct (Brown and Duguid, 2001), to explore how the context and process of data creation shape the content of local knowledge in big crisis data. Our starting point is that all data, including Big Data, are human artefacts. We draw on this perspective to explore the co-creation of Big Data by groups of actors in the aftermath of two recent humanitarian crises, the 2010 earthquake in Haiti and the 2015 earthquakes in Nepal. Approaching Big Data as a social construct allows us to shed light on how the process of Big Data making shapes the data set as well as citizens’ ability to use this data set for their own ends. Our aim in examining Big Data making is to identify and explain the social processes that enable citizen participation – as well as those that limit such inclusion.

Crowdsourcing in disaster settings

When a major crisis unfolds, it is a tremendous challenge to satisfy the information needs of humanitarian responders. In particular, access to up-to-date data on the physical layout of the affected area, the location of vital infrastructure and services is critical. Moreover, to develop the situational awareness that is needed to act, responders need information on what assistance is required where, and what has already been done (Stanton et al., 2001). For this reason, maps are of great importance during crisis response (Meier, 2015). However, in rapidly urbanizing developing countries, existing maps and data sets held by governments or private firms tend to be of limited use, if they are made available at all. One reason for this is that particularly in developing countries, built-up areas generally grow and morph organically, often without formal registration or in ways that do not correspond with official planning. As a result, existing data sets become quickly outdated and do not reflect people’s experience of their current physical environment. The context of Haiti provides a frequently cited example of such a situation, similarly to the present-day situation of Nepal, both of which comprise the setting of our study.

Crowdsourcing in post-earthquake Haiti: From the margins to the centre of relief work

Haiti was struck by a 7.0 magnitude earthquake on 12 January 2010. At this time, crowdsourcing for social ends was already well established, but fairly

marginal to humanitarian relief work in disaster settings. However, the deployment of crowdsourcing platforms, especially OSMs and Ushahidi, during the immediate aftermath of the Haiti earthquake gave this approach enormous momentum.

OSMs, known as ‘the Wikipedia of maps’, is a volunteer-driven platform that aims to make crowd-sourced geospatial data freely accessible. The Ushahidi platform was created in 2008 to enable the mapping of crowdsourced information about the violence that followed the 2007–8 elections in Kenya. The platform enables the datafication of information pulled from online community platforms such as Facebook, Twitter and blogs, as well as information received via text message. On the basis of this data, reports can be created and categorized according to their content.

In the direct aftermath of the Haiti crisis, crowdsourcing through these platforms addressed a significant crisis information gap that up to that point had remained unfilled. For the first time humanitarian organizations had systematic access to live, local situational knowledge. The volunteers were credited with creating the most up to date, reliable and detailed map of Port au Prince’s downtown area, providing more detail than Google Maps or the maps used by FEMA at the time (Meier and Munro, 2010). Thus, the Haiti crisis catalysed the idea and momentum behind crisis data crowdsourcing, and many responding organizations are now seeking to systematically incorporate this innovation in their approach (e.g., OCHA, UNICEF (Batty, 2010), the UN Logistics Base and the IOM (Soden and Palen, 2014)).

Crowdsourcing in post-earthquake Haiti: Information flows

When the earthquake struck, 600 remotely located OSM volunteers came online and quickly built a base layer map of the affected areas, almost from scratch (Soden and Palen, 2014). On the basis of satellite images, digital humanitarians (Meier, 2015) identified roads, damaged buildings and camps for internally displaced persons. The Ushahidi platform was also quickly deployed, initially run by a small group of volunteer crisis mappers who scraped social and traditional media sources for actionable pieces of information. After four days, the technology volunteers set up a telephone number – 4636 – where Haitians could send information via text (SMS) for free. They further developed an internet-based system called Mission4636 to process all the incoming text messages (Heinzelman and Waters, 2010; Meier, 2015). The free phone number linked to Mission4636 received 1000–2000 text messages per day from affected citizens (Heinzelman and Waters, 2010: 7). As these text

messages were mostly written in Creole, a language most global volunteers could not read, Ushahidi crowd-sourced over 1000 volunteers from the Haitian diaspora, living in the USA or Canada, to translate these messages into English.

Mission4636 was published widely through social networks. People were encouraged to share crisis relevant information through the website via text message or email. Members of the Haitian diaspora posted information they had received through relatives. Volunteers processed all the different types of information into crisis reports, which they categorized by topic, such as medical emergencies, trapped individuals and specific needs. They subsequently located the GPS coordinates of the physical location of the situation described in the report and, where possible, mapped it using the OSM platform. Reports containing commentary but no actionable data were filed away under the category ‘insufficient data’ (Sutherland, 2013: 402).

The crowdsourcing process in Haiti was marked by three distinct crowds: the digital humanitarians, the people based in Haiti affected by the earthquake and global volunteer translators (Sutherland, 2013). Figure 1 presents a screenshot of Mission4636, showing the categories and geo-locations of the crisis reports. Access to these reports and maps was in theory open to all who had an internet connection.

The information flows marking the crowdsourcing process in Haiti are represented in Figure 2. Although information sourcing was relatively efficient, the figure shows that there were very few feedback loops, and information sharing enabled by the platform tended to flow in one direction only. First, the information provided by affected people based in Haiti (online crowd 1) flowed to translators engaged via the platform as volunteers (online crowd 2), who converted it into English. This is indicated by a large red arrow, indicating one-way flow of information as translations were only carried out in one direction, from Creole to English, but not vice versa. Second, translated information flowed from the translators to the digital humanitarians (online crowd 3), and again not vice versa. Finally, digital humanitarians converted the translated messages into reports and put these on a map, but did not send reports back to the translators to be converted into Creole for the benefit of local Haitians.

Naturally, there was some overlap between the three crowds, e.g. some volunteers from the diaspora both carried out translations and helped with the mapping of data reports. Furthermore, there were pre-existing, personal connections and linkages between members of the different groups. These personal networks enabled two-way flows of information within the crowdsourcing project, as marked by the green arrows. However, these feedback loops had not been purposefully designed into

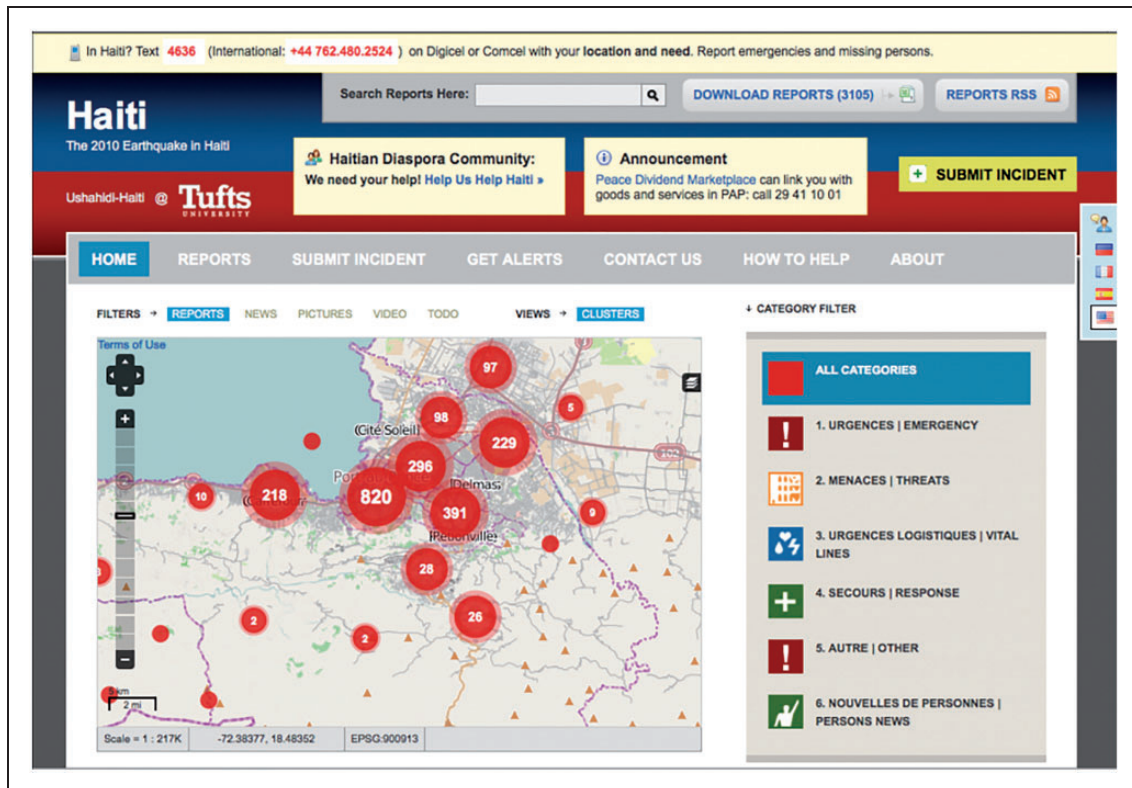


Figure 1. Screenshot of Mission4636 (run on the Ushahidi platform using OSM) for the 2010 Haiti Earthquake.

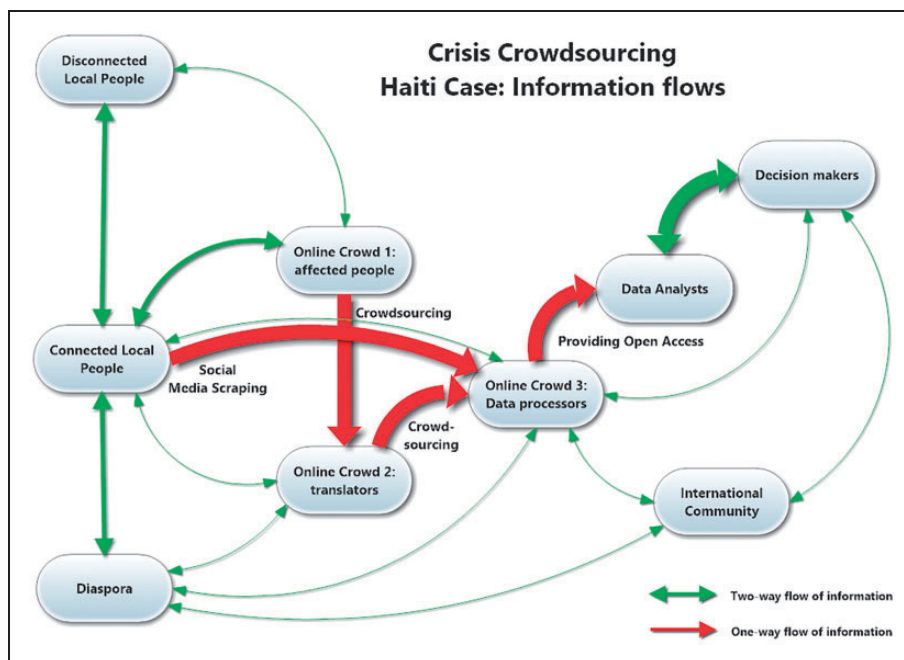


Figure 2. Information flows in post-earthquake crisis data crowdsourcing, Haiti 2010.

the crowdsourcing process and were not specifically supported by the platform. As such, they were sparse and serendipitous.

Crowdsourcing in post-earthquake Nepal: Small-scale Nepali initiatives

Five years after the earthquake in Haiti, Nepal was struck by two major earthquakes in quick succession. The first, on 25 April 2015, reached a magnitude of 7.8 on the Richter scale, and the second, on 12 May, reached 7.3 (Nepal Risk Reduction Portal). At this point, information crowdsourcing was fairly established in disaster settings, but largely driven by open data activists rather than formal humanitarian organizations. Since Haiti, the World Bank had played an active role in supporting open data crowdsourcing projects towards humanitarian aid. Indeed, two of the three initiatives discussed in this section, KLL and Code for Nepal, are (indirectly) linked to the World Bank in that founding members and current leaders at these organizations also work(ed) for the World Bank. The key lesson for the World Bank based on the deployment of open data crowdsourcing platforms in Haiti was that the effectiveness and sustainability of such an approach depended to a significant extent on local ownership, that is to say, on the leadership of locally based stakeholders.

The OSM community shared this view. They had come to regard community participation as the best way of ensuring that maps had actual relevance and were up to date with local knowledge. Furthermore, they saw the active involvement of affected stakeholders in the platform as ethically appropriate, given their significance as intended beneficiaries (Soden and Palen, 2014). The main difference, then, between the two settings, is that the information flows that mark the crowdsourcing process in Haiti were mainly linear whereas in the Nepali crowdsourcing initiatives discussed below, an effort was made to create information loops that linked back to the affected communities. Furthermore, the efforts in Haiti were primarily led by remotely based volunteers with no personal links to Haiti, whereas the projects discussed below were initiated and led by Nepalis.

KLL

When the first earthquake struck Nepal in April 2015, thousands of remotely located volunteers from the Humanitarian OSM Team used satellite imagery in order to rapidly complete the maps of Nepal the Nepali OSM community had been developing. A leading organization in the Nepali OSM community is KLL, which was founded in 2013 by a group of

people (predominantly Nepali open data enthusiasts) who had previously worked together on an Open Cities Project in Kathmandu initiated by the World Bank.

In the immediate aftermath of the first earthquake, KLL quickly developed QuakeMap, a civic technology that runs on the Ushahidi platform. Through QuakeMap, KLL crowdsourced information on local needs: affected people could report their requirements via a hotline, SMS or through an online form. KLL then checked this data (often by telephone) and created a crisis data report, which it categorized and placed on a map. The aim of QuakeMap was to connect people affected by earthquakes with responding organizations. Both data reports and map were freely accessible online. Some of the first responders who used the data were in-country volunteers who organized themselves around the Yellow House guesthouse. These volunteers were recruited via word of mouth and through a Facebook page called Himalayan Disaster Relief Volunteer Group. The Yellow House volunteers were among the first to send supplies to western Sindhupalchowk and Gorkha, some of the worst hit regions in the country (Streep, 2015).

Quakemap itself was available in English, Nepalese and Hindi, but the reports were in English and remained untranslated. Nevertheless, unlike Haiti where non-English speakers were practically cut-off from 'their' crisis information once it had been translated, the people at QuakeMap actively sought to link back to the people who had originally provided the data. By telephone they checked the accuracy of the data at the opening of a report, while the report remained open, and at its closing. As the people who reported needs to QuakeMap – or the humanitarian responders who used this data – often did not provide updates on aid received or delivered, KLL took a very active role in creating feedback loops by chasing this data.

The screenshot of QuakeMap (Figure 3) shows that the vast majority of data reports were geo-tagged as in the Kathmandu valley. However, this predominance is not due to this region suffering the greatest impact: in fact, the rural areas adjacent to the Kathmandu valley were most heavily affected: Gorkha, Dhading, Nuwakot, Rasuwa, Kavrepalanchok, Dolakha and Sindhupalchok (OSOCC, 2015). Instead, the map shows the infrastructure of earthquake-related big crisis data, specifically the locational density of online participation, as will be discussed below.

Code for Nepal

Humanitarian response organizations are under tremendous pressure to be seen to be doing something to address a crisis, especially during the first few

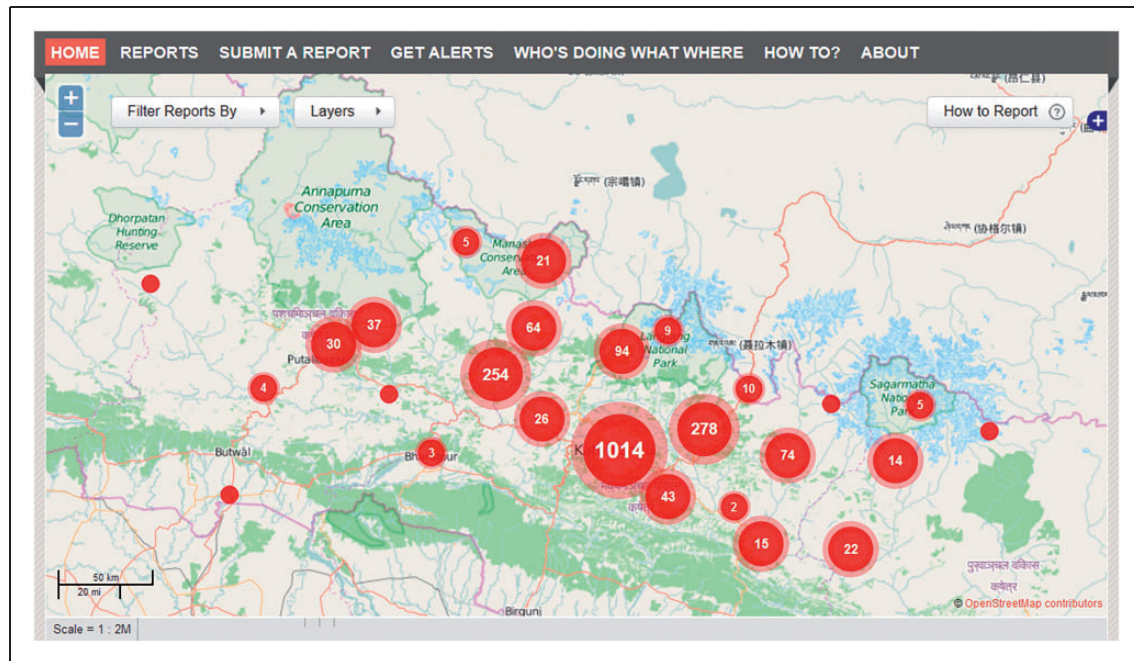


Figure 3. Screenshot of QuakeMap (run on the Ushahidi Platform using OSM) for the 2015 Nepal Earthquakes.

weeks following a disaster. In Nepal, a disproportionate amount of information about the earthquake came from and was focused on the Kathmandu valley area during this initial period, which resulted in many agencies focusing on this region. This imbalance was of course recognized by a number of formal and grassroots organizations who were familiar with the Nepali context, such as Code for Nepal ('Code') and Accountability Lab, discussed below. Founded in 2014, 'Code' is a small US non-profit organization, staffed predominantly by members of the Nepali diaspora. Code lobbies for open data and aims to address the digital inequalities that mark that country. Code projects are aimed at making public data (e.g., about the humanitarian response) available and accessible to Nepali citizens.

When the first earthquake struck Nepal, the team at Code was keen to shift the focus of the humanitarian response towards other regions beyond Kathmandu that were badly affected (CodeForNepal, 2015; Kumar, 2015). In order to achieve this, the team turned to crowdsourcing, deliberately using a low-tech digital approach in order to lower barriers to participation. They relied on commonly used digital tools and mainstream social media, launching a Google Document on Facebook that listed needs and resources by region in order to connect volunteers and people that had been affected. Instructions on how to suggest edits and/or verify the data contained within the document were included in the document. Anyone with moderate digital literacy and familiarity with Google

Docs could use it: no additional training or specialist ICT knowledge was required. Indeed, the document was shared more than 7500 times.

Mobile citizen helpdesks (MCH)

MCH was initiated by the NGOs Accountability Lab and Local Interventions Group, in partnership with the Nepali government. Like Code, MCH aimed to close 'the loop on information related to the earthquake response to ensure relief efforts reach those most in need' (Mobile Citizen Helpdesks, 2015). Unlike Code, MCH combined both online and offline methods for collecting and sharing crisis relevant information. The project provided a platform for affected communities and responders to report gaps at the last mile of humanitarian relief distribution. The MCH targeted the 15 most affected districts outside the capital city. They were run by district coordinators and volunteers in Kathmandu. The project was supported by a toll free number for SMS text messages and a 1234 hotline, manned by volunteers based at the Nepalese Home Ministry. The aim of the initiative was to facilitate a two-way flow of information: MCH monitored the overall response and gathered information at the local level. They then used these insights to help local people obtain information they needed and to explain the decisions donors and the Nepalese government had taken.

In sum, unlike the Haiti mapping efforts, the three initiatives described above were marked by two-way

feedback loops between the crowd and crowdsourced information. However, it is worth noting that, despite their value, these grassroots initiatives lacked the scale required to record or address more than a fraction of all humanitarian needs. In addition, the value of their data sets appears to have been greatest to humanitarian responders during the earliest phase of the relief efforts and diminished once formal organizations had their own information systems up and running. It also seems that the UN agency that coordinated the initial phase of the response in Nepal (the UN Office for the Coordination of Humanitarian Affairs) had not been able to make effective use of the data collected through crowdsourcing. Indeed, one year after the earthquake many affected citizens (particularly in remote areas) still awaited formal assistance from humanitarian, government or grassroots initiatives, relying largely on their own networks towards their recovery.

Crowdsourcing in post-earthquake Nepal: Information flows

Figure 4 represents a synthesis of the communication patterns found in the three Nepali cases discussed. Compared to the Haiti case, volunteers powering the online platforms in Nepal took on a far broader role: in addition to collecting and processing local crisis data, they also sought to generate feedback loops to locally based people, communicating directly with them about their needs and broadcasting information to them about the response. Moreover, they explicitly

attempted to reach affected communities that were not online, for instance by sending volunteers into the communities or using low-tech collection methods. Furthermore, in contrast to Haiti, the Nepali initiatives did not draw on a separate crowd of translators: information was either left untranslated, or was translated into English by digital humanitarians or affected people who were bilingual. Hence, the main flow of information throughout the big crisis data making process was similar to Haiti (i.e. from the affected online people, via the data processing volunteers, to the analysts and decision makers at formal organizations, as indicated by the large red arrows), but there were also smaller, two-way flows of information between the different groups of actors. This constitutes an important difference in terms of the reliability of the data set and the potential of crowdsourcing initiatives to include local people, as we explain in the next section.

Interpretation

Explaining the transformation process: From local knowledge to Big Data

Data – including Big Data – is not simply found, but constructed by people in specific contexts (Tuomi, 1999). Indeed, the creation of data to be processed by computers starts with a person's knowledge of the phenomenon or situation at hand, drawing on both tacit and explicit sources (Tuomi, 1999: 5). This local knowledge then undergoes a number of mutations as it is

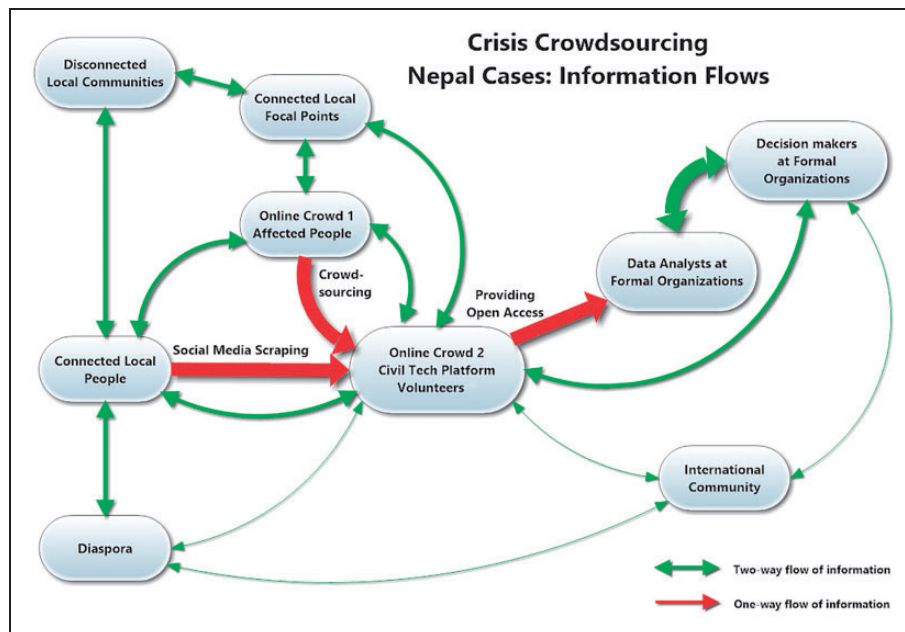


Figure 4. Information flows in post-earthquake crisis data crowdsourcing, Nepal 2015.

processed and transferred between the different groups that contribute to Big Data making. In the open data crowdsourcing initiatives thus far described, crisis-affected citizens were able to share their knowledge through crowdsourcing platforms by verbally transforming their explicit and tacit knowledge of the crisis into written or spoken words. This involved making judgements about what contextual or background knowledge had to be included in the text versus what could be left out as superfluous.

The transformation of explicit and tacit knowledge into verbalized information – whether into spoken words or writing – was the first mutation of knowledge in the datafication process. The second mutation occurred when the verbalized information was translated into English, making it accessible to data processors and analysts without knowledge of the local language. This one-way translation can lead to the immediate exclusion of affected people without English language skills. Once translated, the information was processed by technology volunteers, which involved another substantial mutation because the information was now morphed to fit a pre-existing data structure. This required data processing volunteers to make numerous judgements about categorization, labelling and filing. Finally, analysts at responding organizations interpreted the data, combining it with other data, in order to provide decision makers with insights and advice. The migration and transformation of knowledge in a triple-crowd crowdsourcing setting (as in Haiti where the translators constituted a separate crowd) is depicted in Figure 5.

Data creation thus involves many choices and judgements, and draws heavily on actors' pre-existing knowledge and way of seeing the world. As data migrates from data originators, to data processors, to data users it gets (re)interpreted in a series of different contexts by actors who may be unfamiliar with the affected people's individual and shared needs. As a consequence, crowd-sourced Big Data gets mutated during every step of the

process, and does not accurately convey what affected people intended to communicate, but is nonetheless used as a basis for decisions, sometimes even reinforcing rather than mitigating inequalities, as we now explain.

Digital inequalities deriving from crowdsourced Big Data

The in- and exclusions that characterize Big Data making in the aftermath of a disaster do not affect all groups of potential contributors evenly. Whereas some groups of affected citizens experience no difficulty in accessing and using crowdsourced open crisis data, others lose access at some point during the knowledge transformation process (e.g., through translation), and some are unable to contribute their knowledge at all. Both Haiti and Nepal are marked by numerous digital inequalities. In both countries online participation appears to be primarily determined by knowledge and skills, rather than by access to physical equipment. Digital inequality is strongly correlated with economic inequality, physical location (e.g., rural versus urban) and socially constituted identity markers, such as caste and gender.¹ As such, those Haitians and Nepalis who shared their crisis knowledge in the aftermath of the disaster through crowdsourcing were not necessarily representative of their communities as a whole. Indeed, socially produced crisis maps sometimes ended up reflecting, in part, the density of people who were able to participate online by region, rather than the severity of needs. In our study, we saw that KLL' QuakeMap had a disproportionately large number of data reports geo-tagged to the Kathmandu valley area. Generally, urban Kathmandu residents are significantly more educated than the rest of the country. Their online presence is further strengthened by the fact that they are more used to – and confident in – voicing their views than rural Nepalis. Moreover, they are more often connected to influential people – most of whom are based in Kathmandu – to champion their needs.

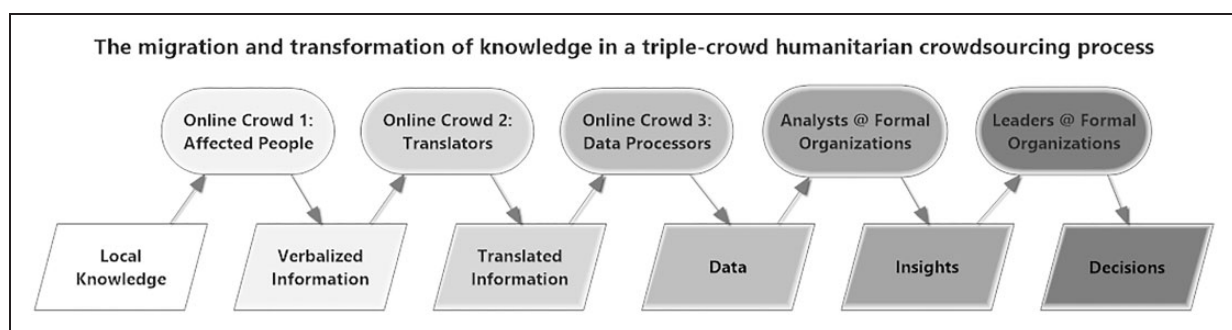


Figure 5. The migration and transformation of knowledge in a triple-crowd humanitarian crowdsourcing process.

Discussion: Transformations in the social construction of Big Data

In this paper, we explain how digital inequalities are created during the Big Data making process, showing that crowdsourced local crisis information becomes gradually less accessible to certain affected communities as ‘local knowledge’ is transformed through various stages into ‘data’. We demonstrate that crowdsourced crisis data can end up reflecting societal inequalities due to the fact that, as a result of digital and virtual barriers, certain population groups have less of an online presence (Johnson, 2014). This raises the concern that reliance on crowdsourced crisis data can result in these inequalities being replicated, especially if marginalized communities are underrepresented by or excluded from data (Crutcher and Zook, 2009; Elwood, 2007). This is especially an issue when responders overlook local patterns of exclusion, either because they are unaware of them (e.g., because they are foreign temporary staff) or because they share the underlying biases that give rise to them.

Our study shows how the process of Big Data making in Haiti and Nepal was partly shaped by those nations’ pre-existing patterns of socio-economic inequalities. These patterns are the product of the co-evolution of those societies with their environments (Oliver-Smith and Hoffman, 2002) as well as the (geo)political relationships that mark those countries’ histories. It is beyond the scope of this paper to describe the unique socio-political context of humanitarian aid in these two countries in detail, but it is worth highlighting that – from macro to micro level – humanitarian aid is deeply political and highly contested (e.g., Duffield, 2016; Escobar, 2011; Ferguson et al., 2010). This paper’s main contribution is an analysis of how a new type of (semi-formal) humanitarian organization operates in this arena: one that comprises a geographically dispersed network of digital humanitarians.

The central aim of digital humanitarians is to facilitate the creation of timely, useful and actionable information so as to allow responders to be more adaptive to the situation on the ground. Crowdsourcing is central to the ways in which this new type of organization makes data. However, in this paper we show that crowdsourcing can result in those people who would benefit most from external intervention being rendered less visible – an hence – overlooked by responders. For instance, the QuakeMap example shows that crowdsourcing of crisis information can result in data sets that reflect existing inequalities, especially digital divides (Crutcher and Zook, 2009; Elwood, 2007; Goodchild, 2007). We argue throughout this paper that big crisis data is not an objective entity, but is created in a social process involving a wide variety of heterogeneous actors. We

explain how the social process of transformation that occurs in conjunction with information transfer and translation during Big Data making leads to the in- and exclusion of specific stakeholders.

Our argument extends beyond a simple binary ‘included’ versus ‘excluded’ from data making and data management as a result of access to ICT equipment and (digital) literacy. As Graham (2011) points out, resolving issues of access to technology and of digital literacy does not enable people to connect with anyone they want, as other (virtual or non-virtual) divides may remain (Graham, 2011). For instance, people are generally restricted to those sections of cyberspace that are available in a language they command. They also need knowledge and skills to navigate digital information. Vulnerable communities may, for example, lack access to sources and connections informing them of the existence of online platforms through which they can share and access crisis information.

The fact that humanitarian crowdsourcing initiatives may inadvertently exclude certain communities from access to their Big Data sets is problematic for another reason: if people with relevant knowledge about affected regions cannot check the data that has been posted about these areas, the validity of the data set as a whole is negatively affected. This point relates to the usefulness and reliability of Big Data in the context of crises. Some responders are reluctant to incorporate data from community platforms into their work practices because such data is seen as containing too much misinformation (Dailey and Starbird, 2014; Hughes and Palen, 2012). Ensuring that affected communities can access and contribute to big crisis data sets would make it possible to engage local people in the ongoing and live triangulation of humanitarian crisis data. Indeed, it has been argued that this is one of the strongest potential contributions of open data platforms to a humanitarian response, enhancing the reliability of the crowdsourced data sets (Vieweg et al., 2008).

Implications of our study and future research

Our analysis explains the transformation process that marks Big Data making in a humanitarian context. We build on prior debates on the social effects of datafication (Baack, 2015; Sutherin, 2013), which emphasize that seemingly neutral processes, catalysed by the engagement of technologies, are in fact highly contested and can have far-reaching social implications. This idea, that data is socially constructed to reflect particular forms of knowledge has also been debated in other scholarly domains, such as new media studies (e.g., Parks, 2009), critical cartography (e.g., Crampton et al., 2013), geoweb studies (e.g., Burns, 2015; Shelton et al., 2014), spatial technologies (e.g., Elwood, 2007),

development studies (Avgerou, 2008) and organization sciences (Leonardi, 2011). However, beyond asserting that data is socially constructed, we explain the social process through which crowdsourced crisis data is constructed. Moreover, we show the important implications for a new type of semi-formal organization that has entered the humanitarian domain: the digital humanitarians.

This paper thereby contributes to humanitarian literature on open Big Data (Meier, 2015), in that we explain on the one hand how humanitarian crowdsourcing efforts can sometimes yield counter-effective outcomes towards an inclusive, participatory humanitarian response and on the other hand, how these effects can be counteracted, leading to forms of aid that are more responsive to the affected people in need who comprise the original data source.

Extending our study, future research on the topic of crowdsourced Big Data in disaster settings would be valuable. This might comprise for instance analysis of the interplay between the geopolitics of humanitarian aid and the actions of geographically dispersed humanitarian crowdsourcing networks. Alternatively studies might further elaborate the hypothesized correlation between inclusion in big crisis data processes and coping and adaptive capacity of affected communities. Another important domain of future research extending from this study relates to the ambiguous effects of technology (cf. Morozov, 2011) and crowdsourcing in disaster settings. Governments may, for example, base decisions on crowdsourced data for the sake of legitimacy (Lehdonvirta and Bright, 2015). Hence, if crowdsourced data sets do not reflect what the original crowd of affected people intended to communicate, resulting policies may prove suboptimal. Analysis of these ambiguities and the politics involved in the processes of data creation, representation and response, can help further fine-tune the critical perspective on Big Data making introduced in this paper.

Conclusion

In this paper, we analysed the process of Big Data making through crowdsourcing and open data platforms in order to explore what barriers stand in the way of this approach enabling citizens' agency in the aftermath of a crisis. Following Tuomi (1999) we flipped the conventional view of data as the raw building blocks of knowledge, arguing instead that different sources of knowledge constitute the building blocks of data. Through a detailed analysis of two humanitarian cases we showed how Big Data making comprises transformation of information, attending to the hidden normative and power issues of Big Data making. Namely, our study disclosed that Big Data making in a humanitarian

context sometimes results in counter-effective outcomes, marginalizing rather than enabling the original crowd at the heart of humanitarian aid efforts. As such, our paper provides an important critical perspective on Big Data in a humanitarian setting, as well as other domains oriented towards facilitating citizens' agency through participatory Big Data.

Acknowledgements

The authors wish to thank the people at Kathmandu Living Labs, Code for Nepal, Accountability Lab, Local Interventions Group, Oxfam Novib, Cordaid and all other respondents who generously shared their time with us towards realizing this study. We also wish to extend our gratitude to the editors and the anonymous reviewers for their invaluable insights and comments.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This project was sponsored by the Netherlands Organization for Scientific Research (NWO), Division of Social Sciences, Smart Disaster Governance (Project 409-14-003), under the supervision of principal investigator dr. ir. Kees Boersma.

Notes

1. Digital inequalities reflect other socio-economic inequalities: at the time of the 2010 earthquake, 61.7% of Haitians lived on less than US\$1.25 per day (UNICEF, 2013) and only 48.7% were literate (UNESCO Institute for Statistics, 2012). Even though roughly 85% of households had access to a mobile phone (Pereira, 2010), only 8.4% of Haitians were internet users, that is, people who had used the internet in the previous 12 months from any device (e.g., computers, mobile phones, etc.) (World Bank, 2015). Likewise, in 2014, 82.5% of Nepalis had a mobile phone subscription but only 15.4% were internet users (World Bank, 2015). The discrepancy between the male and female literacy rates in Nepal is similar to the one found in Haiti (75.1% versus 57.4%, respectively). This suggests that in both countries digital inequalities are highly gendered (Central Bureau of Statistics, 2014). Literacy rates also vary widely by district: the overall literacy rate in Kathmandu is for example 86.3% whereas in Rautahat (a rural district bordering India) it is only 42%. The latter is also marked by exceptionally high levels of both gender and caste-based discrimination (UNFCO, 2013).

References

- Avgerou C (2008) Information systems in developing countries: A critical research review. *Journal of Information Technology* 23(3): 133–146.
- Baack S (2015) Datafication and empowerment: How the open data movement re-articulates notions of democracy,

- participation, and journalism. *Big Data and Society* 2(2): 1–11.
- Batty P (2010) OpenStreetMap in Haiti Part 1. On: Youtube. Available at: <http://www.youtube.com/watch?v=PyMTKABxaw4> (accessed 4 October 2015).
- Brown JS and Duguid P (2001) Knowledge and organization: A social-practice perspective. *Organization Science* 12(2): 198–213.
- Burns R (2015) Rethinking big data in digital humanitarianism: Practices, epistemologies, and social relations. *GeoJournal* 80: 477–490.
- Central Bureau of Statistics (2014) Population Atlas of Nepal. Available at: <http://cbs.gov.np/atlas/tables.html?chapter=4&table=4.2> (accessed 31 December 2015).
- CodeForNepal (2015) Map of reported #Nepal earthquake deaths, injured and houses damaged. Available at: <http://www.codefornepal.org/2015/04/map-of-reported-nepal-earthquake-deaths> (accessed 24 December 2015).
- Cooke B and Kothari U (2001) *Participation: The New Tyranny*. London and New York: Zed Books.
- Crampton JW, Graham M, Poorthuis A, et al. (2013) Beyond the geotag: Situating ‘Big Data’ and leveraging the potential of the geoweb. *Cartography and Geographic Information Science* 40(2): 130–139.
- Crutcher M and Zook M (2009) Placemarks and waterlines: Racialized cyberscapes in Post-Katrina Google Earth. *Geoforum* 40: 523–534.
- Dailey D and Starbird K (2014) Visible skepticism: Community vetting after Hurricane Irene. In: Hiltz SR, Pfaff MS, Plotnick L and Robinson AC (Eds.), *Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management ISCRAM 2014*. University Park, Pennsylvania, May 2014, 772–776.
- Duffield M (2016) The resilience of the ruins: Towards a critique of digital humanitarianism. *Resilience*. Epub ahead of print 15 March 2016. DOI: 10.1080/21693293.2016.1153772.
- Elwood S (2007) Volunteered geographic information: Future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal* 72: 173–183.
- Escobar A (2011) *Encountering Development: The Making and Unmaking of the Third World*. Princeton, NJ: Princeton University Press.
- Ferguson J, Huysman M and Soekijad M (2010) Knowledge management in practice: Pitfalls and potentials for development. *World Development* 38(12): 1797–1810.
- Ferguson J and Soekijad M (2015) Multiple interests or unified voice and quest; online communities as intermediary spaces for development. *Journal of Information Technology* 1–24.
- Gao H, Wang X, Barbier G, et al. (2011) Promoting coordination for disaster relief—From crowdsourcing to coordination. In: Salerno J, Yang SJ, Nau D, et al. (eds) *Social Computing, Behavioral-Cultural Modeling and Prediction*. Berlin Heidelberg: Springer, pp. 197–204.
- Goodchild M (2007) Citizens as sensors: The world of volunteered geography. *GeoJournal* 69(4): 211–221.
- Graham M (2011) Time machines and virtual portals the spatialities of the digital divide. *Progress in Development Studies* 11(3): 211–227.
- Heinzelman J and Waters C (2010) *Crowdsourcing Crisis Information in Disaster-Affected Haiti*. Washington, DC: US Institute of Peace.
- Hickey S and Mohan G (2004) *Participation—From Tyranny to Transformation?: Exploring New Approaches to Participation in Development*. London and New York: Zed Books.
- Horowitz M (2008) Visualizing Big Data: Bar charts for words. *Wired*. Available at: www.wired.com/science/discoveries/magazine/16-07/pb_visualizing (accessed 4 October 2015).
- Howe J (2008) *Crowdsourcing: How the power of the crowd is driving the future of business*. New York: Random House.
- Hughes AL and Palen L (2012) The evolving role of the public information officer: An examination of social media in emergency management. *Journal of Homeland Security and Emergency Management* 9(1): 1–20.
- Johnson J (2014) From open data to information justice. *Ethics and Information Technology* 16(4): 263–274.
- Kaplan AM and Haenlein M (2010) Users of the world, unite! The challenges and opportunities of social media. *Business Horizons* 53(1): 59–68.
- Kitchin R (2014) *The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences*. Thousand Oaks, CA: Sage.
- Kumar R (2015) Use data, not nepotism, to deliver aid in Nepal. *Time Magazine*, 5 May, 15.
- Laney D (2001) 3D data management: Controlling data volume, velocity, and variety. *Gartner Blog Network*. Available at: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-ManagementControlling-Data-Volume-Velocity-and-Variety.pdf> (accessed 4 October 2015).
- Lehdonvirta V and Bright J (2015) Crowdsourcing for public policy and government. *Policy and Internet* 7(3): 263–267.
- Leonardi PM (2011) When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly* 35(1): 147–167.
- Madon S (2000) The internet and socio-economic development: Exploring the interaction. *Information Technology and People* 13(2): 85–101.
- Meier P (2015) *Digital Humanitarians. How Big Data Is Changing the Face of Humanitarian*. Boca Raton, London and New York: CRC Press.
- Meier P and Munro R (2010) The unprecedented role of SMS in disaster response: Learning from Haiti. *SAIS Review of International Affairs* 30(2): 91–103.
- Mobile Citizen Helpdesks (2015) *Mission Statement*. Available at: <http://www.quakehelpdesk.org/what.php> (accessed 23 July 2016).
- Morozov E (2011) *The Net Delusion: How Not to Liberate The World*. London, UK: Penguin.
- Morrow N, Mock N, Papendieck A, et al. (2011) Independent evaluation of the Ushahidi Haiti Project. Ushahidi Haiti independent evaluation. Available at: http://sites.google.com/site/haitiushahidieval/documents/Ushahidi_Haiti_Eval_final.pdf?attredirects=0 (accessed 4 October 2015).
- Mosse D (2001) People’s knowledge’, participation and patronage: Operations and representations in rural

- development. In: Cooke B and Kothari U (eds) *Participation: The New Tyranny*. London and New York: Zed Books, pp. 16–35.
- Nepal disaster risk reduction portal. Incident report of earthquake 2015. Available at: <http://drrportal.gov.np> (accessed 23 May 2016).
- Oliver-Smith A and Hoffman SM (eds) (2002) *Catastrophe and Culture: The Anthropology of Disaster*. Santa Fe: School for Advanced Research Press.
- OSOCC (2015) Situation analysis Nepal earthquake 15.05.2015. Available at: https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/assessments/150515_nepal_situation_analysis_osocc_assessment_cell_-_final_final.pdf (accessed 4 October 2015).
- Parks L (2009) Digging into Google Earth: An analysis of “Crisis in Darfur”. *Geoforum* 40(4): 535–545.
- Pereira E (2010) The race (for marketshare) is on: Commercial launch of mobile banking in Haiti. *Forbes*. Available at: <http://www.forbes.com/sites/evapereira/2010/12/15/the-race-for-marketshare-is-on-commercial-launch-of-mobile-banking-in-haiti/> (accessed 31 December 2015).
- Sen AK (1999) *Development as Freedom*. New York: Knopf.
- Shelton T, Poorthuis A, Graham M, et al. (2014) Mapping the data shadows of hurricane sandy: Uncovering the sociospatial dimensions of ‘Big Data’. *Geoforum* 52: 167–179.
- Soden R and Palen L (2014) From crowdsourced mapping to community mapping: The post-earthquake work of Openstreetmap Haiti. In: Rossitto C, Ciolfi L, Martin D, et al. (eds) *COOP 2014 – Proceedings of the 11th International Conference on the Design of Cooperative Systems*, Nice, France, 27–30 May 2014, pp.311–326. Cham, Switzerland: Springer International Publishing.
- Stanton NA, Chambers PR and Piggott J (2001) Situational awareness and safety. *Safety Science* 39(3): 189–204.
- Streep A (2015) Nepal’s aid system is broken. So these life-savers hacked it. *Wired*. Available at: www.wired.com/2015/05/nepal-earthquake-aid/ (accessed 29 May 2016).
- Sutherlin G (2013) A voice in the crowd: Broader implications for crowdsourcing translation during crisis. *Journal of Information Science* 39(3): 397–409.
- Tuomi I (1999) Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Journal of Management Information Systems* 16(3): 103–117.
- UNESCO Institute for Statistics (2012) *Literacy world population prospects*. Available at: <http://www.uis.unesco.org/literacy/Documents/UIS-literacy-statistics-1990-2015-en.pdf> (accessed 31 December 2015).
- UNFCO (2013) *District profile: RAUTAHAT*. Available at: http://un.org.np/sites/default/files/rautahat_district_profile.pdf (accessed 31 December 2015).
- UNICEF (2013) *At a glance: Haiti*. Available at: http://www.unicef.org/infobycountry/haiti_statistics.html (accessed 24 July 2016).
- Vieweg S, Palen L, Liu S, et al. (2008) Collective intelligence in disaster: An examination of the phenomenon in the aftermath of the 2007 Virginia tech shootings. In: Frank Fiedrich and Bartel Van de Walle (eds) *Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM 2008)* Washington DC, USA, May 2008.
- World Bank (2013) *World Development Report*. Available at: <http://www.worldbank.org/en/publication/wdr/wdr-archive> (accessed 23 July 2016).
- World Bank (2015) *World Development Indicators*. Available at: <http://databank.worldbank.org/data/home.aspx> (accessed 31 December 2015).

This article is a part of special theme on Critical Data Studies. To see a full list of all articles in this special theme, please click here: <http://bds.sagepub.com/content/critical-data-studies>.