

Collaborative Analytics and Brokering in Digital Humanitarian Response

Daniel Hellmann
College of IST,
Penn State University
University Park, PA USA
deh198@ist.psu.edu

Carleen Maitland
College of IST,
Penn State University
University Park, PA USA
cmaitland@ist.psu.edu

Andrea Tapia
College of IST,
Penn State University
University Park, PA USA
atapia@ist.psu.edu

ABSTRACT

During large scale humanitarian crises, relief practitioners identify data used for decision making and coordination, as critical to their operations. Implicit in this need is the required capabilities for analyzing data. Given the rapidly evolving systems of collaborative data management and analysis in digital humanitarian efforts, information scientists and practitioners alike are keen to understand the role of data analytics in response operations. Through a case study of a digital humanitarian collaborative effort, we examine the processes for big and small data analytics, specifically focusing on data development, sharing, and collaborative analytics. Informed by theories of articulation work and collaborative analytics, we analyze data from in-depth interviews with digital humanitarians. Our findings identify key practices and processes for collaborative analytics in resource constrained environments, particularly the role of brokering, and in turn generate design recommendation for collaborative analytic platforms.

Author Keywords

Collaborative Analytics; Big Data; Brokering; Articulation; Crisis Informatics; Humanitarian Aid; Ebola

ACM Classification Keywords

H.5.3 Group and Organization Interfaces

INTRODUCTION

Humanitarian crises require rapid formation of data processes. As the members of a response network deploy, they begin to identify and share data resources. These data sharing efforts often span both in-situ responders who are producing data as well as consuming analytic outputs of data analyses, and digital humanitarians, who from a distance attempt to contribute to the response. These so-called “Digital Humanitarians” represent an emerging phenomenon of technologically driven social organization, enabling

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CSCW '16, February 27–March 02, 2016, San Francisco, CA, USA

© 2016 ACM. ISBN 978-1-4503-3592-8/16/02...\$15.00

DOI: <http://dx.doi.org/10.1145/2818048.2820067>.

international scale data sharing and creating both the social and technical infrastructure for collaborative analytics [17]. Here we define collaborative analytics as the processes by which organizations share and analyze data, inputs as well as outputs, through a digital platform for communication and information transfer. Some key features of such a system include active participation of stakeholders and contributors in the analytic process and brokering activities to promote and direct participation.

While the concept of collaborative analytics bears some resemblance to crowdsourcing models, it should be noted that where crowdsourcing fundamentally relies on using participant input in an analytic process, they are generally not a party to defining the desired analysis. In contrast, participants in a collaborative analytic process constitute an ICT connected community of practice. Each member of the community contributes a valuable resource as part of a mutually beneficial collaboration.

Where resources are constrained, as is often the case in humanitarian response, there are many uncertainties surrounding the potential for collaborative analytics. While pundits extoll the virtues of big data for humanitarian relief, precisely how the data will be transformed into actionable analytic outputs is unclear.

The goal of this research is to contribute to a theory of collaborative analytics by providing insight into how the context of a crisis, including the nature of the crisis, as well as the physical and organizational characteristics, impacts how digital humanitarians engage in collaborative analytics. In particular, we investigate how data are collected, shared, and utilized through analyses. What type of work is involved? Which factors hinder or contribute positively to collaborative analytics?

In the following sections we provide background information on humanitarian response organizations, collaborative analytics and articulation work. This is followed by a description of our case study and data collection methods. Finally, we present the analytic results, discussion, and conclusion.

LITERATURE REVIEW

Organizational Factors

The ability of humanitarian organizations to effectively engage in collaborative analytics will be influenced by their network relations and the degree of change their networks can endure while either maintaining or developing stability [2]. Entering collaborative agreements can pressure organizations to change or innovate. Though this process can be temporarily disruptive, it can also serve as an investment in future stability, developed through harmonized interdependent business processes with network partners [6,9].

Research has shown NGO's and relief organizations address common information management hurdles through collaboration geared toward resource pooling and reallocation [14]. Other times, new synergies are developed through data sharing activities, which can compound the value of participating organizations' data through the sharing of analytic outcomes. A great deal of this research has centered on these data sharing practices or opportunities to capitalize on novel data sources such as social media and big data. There are, however, many hurdles. Sometimes the quality or trustworthiness of the data fails to meet organizational standards or needs [32]. Other times, there exists concern on the part of the organization that the process of leveraging big data or social media is not well understood. One hurdle with data analytics, big and small, is identifying the process to distill useful data from raw content, and more importantly, guaranteeing that process does not obscure important contextual information embedded in the data collection process [2]. Burns (2014) nonetheless advocates for pursuing these new techniques with an appropriate perspective that recognizes Big Data as more than just data, but also an all new set of analytic practices. The ultimate key is to understand those practices and tie them effectively to organizational decision making.

Data Lifecycle in Humanitarian Response

Due to the resource constraints of humanitarian response, data collection can be challenging and may involve paper-digital workflows [5]. If data is to be repurposed from existing systems, there must be clear understanding of its structure as well as the data management practices used to maintain the data [10]. Deviation from intended data management practices, particularly those hardened into systems, can make repurposing data difficult as information capture no longer adheres to expected norms. However, once the content, syntax and structure of the information is understood, strategies for combining heterogeneous data enable integration of content from a wide variety of sources can be developed [18,25,27]. For example, social media have proven to be a rich source of contextual references and time sensitive indicators of ground conditions during crisis events [21,28,31,36].

Collaborative Analytics

The concept of collaborative analytics developed here refers to an architecture of collaboration designed to bring stakeholders in a community of practice together for the purpose of sharing data and analytic results. The community will also establish and nurture the relationships necessary to further leverage analytics toward data driven decision making. We anticipate key work related concepts in such a system will include invisible work, concepts of articulation, and especially anticipation work. While similar notions have been put forth for data sharing hubs, our research sheds light on how collaborative analytics operates in a humanitarian, and hence resource constrained, environment. Concepts similar to collaborative analytics have been suggested in a somewhat cursory fashion in scholarship on participatory analytics, focusing primarily on participant led data collection and visualization [7,20]. Collaborative visual analytics covers many similar concepts in that there is a focus on data sharing and analytics with appropriate software affordances to support distributed teams. An important distinction is found in brokering activities. Whereas collaborative visual analytics manages brokering through software based agents [3,12,19], our concept relies on human based interactions.

Articulation as a Means

Articulation work, or the invisible work that happens to prepare for or enable visible work, is important to analyzing how tools fit to a particular task [26]. As collaborative analytics relies on a digital platform, the fit of the platform will influence the success of community of practice. As set forth in *Integration of Computing and Routine Work* [8], Fitting, Augmenting, and Working Around are three strategies which comprise the definition of articulation work in computing. This concept identifies work itself as a series of tasks in a lattice which describes how the tasks relate to and support each other. The articulation work comprises actions taken to adjust the lattice and is typical of any management or planning type activities. One of the primary focuses of articulation work is that it addresses contingencies and disruptions to the expected evolution of work tasks [30]. In that way it is a form of contingent adaptation or innovation that enables the continuation of computer based work. Recent interpretations of this concept have extended it toward "Anticipation Work", which is articulation done in preparation for expected technological innovations [29]. Preparation for adoption of emerging technologies is especially important for humanitarian collaborative analytics as such preparation highlights a problem or a need and anticipated solution.

The Role of Brokering

Another key facet of collaborative analytics related to articulation work is the role of brokers and the articulation work they engage in on behalf of their associates. Brokers can assist in finding solutions to misalignment of work processes by using professional connections or knowledge to address issues of undersupply or misalignment of resources.

This ability is the result of a brokers position as a boundary spanner within an organization or institution [22]. As boundary spanners, part of their functional value is as mediator of Common Ground. Common ground becomes critical when conflicting priorities between collaborators emerge [4]. Dedicated brokers who help develop and maintain common ground also provide value by handling the articulation work for collaborators, allowing collaborators more time to focus on individual priorities.

The Confluence of Articulation and Brokering

In addition to mediating work related issues through boundary objects and with an eye toward assisting establishment of common ground between collaborators, brokering offers a number of other social problem solving advantages in a crisis context. During times of war, other social conflict, or disasters, civil infrastructures are frequently damaged. This disruption causes a massive upheaval to the elements of society that depend on these civil infrastructures. The government and other parties responsible for maintaining and repairing the infrastructure may not be in a position to respond quickly, or may have lost the trust of the populace. Under such circumstances individuals will engage in articulation in order to develop alternative civil infrastructures and services to fill the gap where institutionalized services have failed [15]. As individuals find ways to work around the breakdowns in social services, they develop and rely on close knit social network of trust [24]. Through these networks, trustworthy connections can be brokered in order to secure goods, services, or reliable information. While these examples pertain mostly to the “victims” of social disruption, response personnel also must engage in a great deal of articulation work as the response effort begins. Crises are so diverse that preparing for inter-organizational collaboration can be a daunting task. In this case, the broker of knowledge about expertise and data resources enables workarounds and adjustments that keep collaborative efforts moving forward [13].

METHOD

These concepts of collaborative analytics and articulation frame our analysis of a community of digital humanitarians formed during a recent crisis. In particular, we examine a network of practitioners responding virtually to the Ebola crisis in West Africa. The goal is to extend the body of work related to how social processes, specifically brokering a collaborative analytic environment, truly work and the degree to which they are generalizable across applications.

For the past year, the African Ebola epidemic has devastated populations with approximately 27,000 human cases reported across five west African countries, Spain, and the US, resulting in over 11,000 fatalities [37]. Dozens of major charities are engaged in efforts to provide medical assistance or logistics support to the relief effort. Many developed nations and private corporations have pledged monetary as well as material support to these efforts as well. The

epidemic represents one of the most severe and unique examples of human crisis in recent history.

The duration of the epidemic is one of its most distinctive features. When one considers the basic nature of this particular crisis and other examples of disease epidemics as opposed to natural disasters and other breakdowns of civil infrastructure, disease events tend to be more protracted. A disaster may dramatically impact society, but once the event is over, response becomes more straightforward. There are of course ripples to deal with, but the event at the core of the disaster is over, which certainly simplifies assessing damage and coordinating response. Disease events, especially those with a viral growth rate, are not characterized by a semi governing of the negative impacts. The disaster continues to grow at exponential rates, making the response effort more dynamic and difficult to gauge and plan for simply because it isn't over yet. The disruption to civil normalcy is recycled through the spread of the disease.

If a silver lining can be found regarding this tragedy, it is that this duration offers the opportunity to study various aspects of the response and to develop knowledge which will improve the response in future epidemics.

For the scope of this paper, our dataset has been constrained to a subset of our overall interview subjects primarily defined as participants in the Digital Humanitarianism movement. This assures our focus is truly the perspectives of individuals involved in a global scale cooperative effort to develop and share data or analytic resources in support of the Ebola response. Amongst digital humanitarians, individual backgrounds vary greatly. While many are volunteers, some are compensated for their work in some way. Those who are compensated tend to be in coordination or leadership type roles of the participating organizations. The level of effort that these individuals apply to their activity is subsequently much higher than the typical digital humanitarian. Our research indicates that coordinators within this community tend to participate on a half to full time equivalent level, while more casual participants tend to contribute sporadically in something like 10-20 hour chunks, frequently during free time such as evenings and weekends. The result is a relatively decentralized federation of collaborators with a broad array of commitment levels and interests. Furthermore, the heavy reliance on volunteer labor drives a need for asynchronous interaction.

Data Collection

The first step of the data collection is to identify an initial list of interview subjects spanning the demographic elements related to our project as identified through a literature review. Interview subjects with demographic backgrounds such as data scientists, relief practitioners, brokers, domain researchers, or some combination of these identities and roles were sought as key informants. Subjects were recruited through the professional networks of the research team. A snowball method was used to identify and recruit through mutual contacts. The following lists provide subject

demographics and a summary of the interview guide questions.

Interview Demographics

1. High level director of global humanitarian aid brokering non-profit with 20+ years experience in technology and management
2. Data Scientist and Geographer engaged with multiple prominent data analytics nonprofits from the humanitarian aid community. Has 10+ years experience in both global and field operations.
3. Executive with a prominent technology and connectivity support startup. Has 10+ years experience in government and industry.
4. Data Scientist and Medical Doctor engaged with multiple prominent non-profits from the humanitarian aid community. Has 15+ years experience in both global and field operations.
5. Professor of crisis informatics and humanitarian aid researcher with 15+ years experience
6. Professor of crisis informatics and humanitarian aid researcher with 15+ years experience.
7. Professor of crisis informatics and humanitarian aid researcher with 15+ years experience between academia and industry.
8. Professor of Communications with 15+ years experience between academia, humanitarian aid, and environmental non-profits.
9. Epidemiologist with 5+ years experience brokering spatial epidemiology support for humanitarian aid NGO's.
10. Retired Technology Professional with 5+ years volunteer service in humanitarian aid.
11. Founder of major humanitarian aid and crisis informatics non profit
12. Open source software development executive with 5 years volunteer experience in development based non-profits
13. Technology and data science executive with 5 years volunteer experience in humanitarian volunteer non-profits
14. Entrepreneur with 10+ years experience in both government and volunteer non-profits.
15. Technology and Business Manager with 5+ years experience in volunteer non-profits
16. Entrepreneur with 5+ years experience in volunteer non-profits.

Interview Guide Summary

The questions, whose answers and relevance are likely to vary with category, are defined in harmony with the research questions. The framework consists of five components, two reflecting the users' perspectives and three reflecting the data analytic context. Underlying the framework is an interest in multi-level governance and organizational structure, as well as the work associated with big data analyses and their use in decision making.

1. Response context - Which aspect(s) or stages of the response is your organization involved in (sensitization, treatment, logistics, etc.)?
2. Decision making process - Has your organization made use of externally sourced data in its decision making? Were these raw data or analytic outputs (maps, graphs, models)? What Types of data, sources, and formats are you most interested in? How does your organizational unit (office, work group, team) deal with uncertainty in data and statistical analyses?
3. Data availability - Are the data you need largely available? If data are collected at intervals (daily, weekly, monthly), are they reliably available?
4. Analytic techniques - What types of analyses do you use, if any, to improve the Ebola response or limit the disease spread? How do you identify the audience and how often do you interact with the intended audience of analytic products? Do they provide feedback that helps shape your work? Are there analyses you need that are not available?
5. Output format - Are the outputs of your analyses (charts, graphs, maps) configurable? If so, how?
6. Organization structure – Describe your organizational structure and hierarchy. Do you collaborate internally or with external organizations? If so, can you describe those activities in terms of purpose or nature of work?

The interviews were audio (and at times video) recorded with the participant's permission. Semi-transcripts were developed for each interview by working significant timestamps, discussion threads, and illustrative quotes into interview notes during a secondary review. This process is reminiscent of the hermeneutic circle used to develop more rich understanding of text through recursive reinterpretation in light of context. Organizing thematic discussion threads within the notes and expanding those toward semi-transcript type content also bears some resemblance to Agar's language [1]. In this case thematic discussion threads are similar to strips used to organize content topics. The difference between the method employed here and prominent applications of Agars method in other qualitative research [34] is in the role of breakdown. Instead of using breakdowns in the available coding schema to extend the schema and develop understanding of the strip meaning in an existing text, divergence or recirculation of the discussion topic in the actual interview or audio recording was used to define a new note or note thread as appropriate. Adaptation of the coding schema is then handled during analysis. As a result, each note then has its own meaning in the context of the thematic thread it has been logged under, and each thread can be interpreted as the sum of meanings of its notes. This process represents a completely open form of interpretive reading based upon literal reading of the initial note set prior to actual code based analysis [16]. Once the final notes were developed, copies were sent to the interview subjects to

perform member checking and clarification of any discrepancies.

While most of the interviews conducted were one on one between the researcher and subject, one particular interview was conducted as a group, yielding 16 subjects across 14 interviews. Table 1 summarizes the metadata for the resulting interview semi-transcripts.

Interview	Time (minutes)	Semi-transcript Pages
PS-1	48	3
PS-2	35	2
PS-3	52	4
PS-4	35	4
PS-5	23	2
PS-6	88	6
PS-7	60	5
PS-8	82	2
PS-9	70	3
PS-10	60	1
PS-11	54	1
PS-12	78	2
PS-13	75	2 + video demo
PS-14	180	5 + video demo

Table 1 : Interview Results Metadata

Analysis

Once semi-transcripts were developed, the notes were loaded into a common analytic environment to perform qualitative coding. The coding schema was developed by extending a list of thematic factors generated to support the interview guide. The result was a hierarchical coding schema that reflected concepts from the interview guide as well as additional concepts emerging from a cursory reading of the interview data. The coding environment supported dynamic updates to coding decisions as well as opportunities to concurrently code single items with multiple codes. A summary worksheet compiled the raw counts of codes to a summary table from which basic statistics, diagrams, and pattern analysis could be performed. This summary dataset supports both longitudinal and cross-sectional analysis of the interviews [16]. The entire workbook design is dynamic and scalable such that revisions to the coding schema, coding analysis, and addition of new data is straightforward to implement.

FINDINGS

In summary, our findings begin with the direct implications of a lack of network connectivity. While the challenges of

limited network access are well chronicled, here we examine the specific implications for collaborative analytics, including data sharing, articulation work, and brokering.

Finding #1 – Lack of connectivity in coverage or bandwidth, inhibits widespread availability of shareable data and creates design constraints when disseminating analytic outputs.

Digital humanitarians rely heavily on internet connectivity for nearly all aspects of their work. It also has implications for their identity. The following quote from a prominent crisis informatics practitioner could not illustrate this any better.

“one of the things that hits home for many of us in the humanitarian world is pure connectivity” ... [I teach my field analysts to] ... “toggle down to the most limited bandwidth options and create redundant pathways for collaborative sharing”

“So [Social Media Company] has funded the initial phase directly with [Humanitarian IT Coordinator] to try to get some VSAT (Very Small Aperture Terminal) capacity for the NGO’s working in the area and they (Social Media Company & Humanitarian IT Coordinator) then asked us to get involved to look at more terrestrial options”

In short, the lack of quality connectivity influences nearly all aspects of their work, how data is collected and transmitted as well as how analytic outputs are formatted for dissemination. In a low connectivity environment, reports sent to the field need to communicate key facts while consuming limited bandwidth. Aside from the data volume/bandwidth issues, network accessibility is a design constraint that influences how information is presented, formatted and transmitted. For instance, bulleted facts in a text message/SMS are more accessible than a report that must be downloaded on a computer, viewed, read, and potentially interpreted.

Finding #2 – Simply the potential, rather than the actuality, of being able to share data, incentivized digital data capture and ‘driving data around.’

While for some digital data capture, as opposed to paper or whiteboard based work, presents a significant information and technology hurdle for the Ebola response, there were others who despite limited network, hardware and software resources, embarked on the arduous process of capturing, managing and transferring digital data. Simply the potential that someday, someone could potentially use the data provided an incentive to perform this work.

“The operational data at the district level was being collected and managed on a whiteboard, but if you don’t have connectivity, that’s what works”

“They had hardware and they had software and then the data was being collected and put into spreadsheets, but in most cases, they had big issues with connectivity, so they could not share that data backward”

“Most of these ETU’s (Ebola Treatment Unit) and NGO groups were sharing things on USB drives, and manually driving them from village to village in order to collect and share data. So anything that can be done with regard to Wifi coverage to help these data teams share data is something we are very interested in”

It is interesting to note the multitude of responses to a lack of connectivity when managing data. In the case of the first quote, a lack of connectivity seems to have driven analog (paper/whiteboard) type data captures. It is possible that the whiteboard approach described is meant to facilitate group work within an office. The emphasis however on “collected and managed” seems to indicate that the whiteboard really is the authoritative operational dataset. To contrast this with the first quote, the second quote indicates that computer assets were available and actively used, but in the absence of connectivity or some vehicle for sharing, the effort to maintain digital copies of the data becomes harder to justify, and the value in that effort becomes moot. This then raises the question of why the computer based work was done at all? Perhaps the anticipation of future opportunities to share, or a desire to build a historical data base for future analysis incentivized the investment in digital data management.

In the third quote, we find an example of digital data management and data sharing enabled through removable media. This is an example of improvisation, Fitting or Working Around in articulation terms, whereby collaborators change the structure or procedure of their work in order to address some deficiency or misalignment. In this case the lack of connectivity precludes network based data sharing and collaboration, so the next best alternative of thumb drives and physical transportation of data is used to share data and make arrangements for collaboration.

Hence, we observe a variety of responses to the resource scarcity, with varying implications for collaborative analytics.

Finding #3 – Collaboration is conducted mostly through the use of free and/or thin client communication and information sharing technologies.

Notions of collaboration tended to accompany discussion of data sharing and analysis, which as it turns out is deeply interconnected to connectivity. Subjects differed on their opinions as to which attributes or functionalities were most important, but lightweight and free applications (e.g. Skype) that streamline communication are common.

“It’s a nice collaborative infrastructure, but can almost kill you when you have a lot of crisis things going on” ... on the “Tyranny of Skype”

“and then when you talk about the (collaboration) solutions that are sought, it is very much the standard tools that even we are using because a lot of the connections go back into the field and around. It’s a lot of Dropbox, it’s a lot of Google docs, it’s a lot of Skype it’s a lot of Excel and office solutions and spreadsheets.”

The first quote regarding “The Tyranny of Skype” offers some insight into design principles for communication systems intended to support relief practitioners. Managing information overload becomes critical to staying focused on priorities. Systems that allow communication to be easily compartmentalized into manageable dialogs on the fly may offer some advantages to collaborators [11]. In the second quote we begin to see the types of applications commonly used. Again, basic word processing, spreadsheet, and file or document sharing functionality sees widespread adoption, but many of these applications, and specifically their key affordances, are dependent on connectivity. One key concept to note is the difference between synchronous and asynchronous interaction. Jones et al. (2008) highlights asynchronous collaboration as an affordance for managing information overload [11]. Here, in the case of the Ebola response, instead of managing information overload, asynchronous communication helps ameliorate intermittent network outages. This is a critical design feature for collaboration platforms in low resource environments, particularly where remote field work is involved.

Finding #4 –Data and resource sharing activities often require brokering actions.

Collaboration in the Ebola response is deeply interdisciplinary, and by its very nature, involves coordinating the activities of actors with different concerns and priorities. Empowering collaboration then requires some kind of facilitator willing to invest the invisible articulation work necessary to promote buy in and cooperation from the various actors. These individuals must function as boundary spanners, demonstrating insight into each individual domain of the collaborative effort in ways that build relationships, common ground, and result in collaborative action. The following quotes from one such broker of collaborative action highlight the nuance of this effort, and it’s criticality to collaborative success.

“In the Ebola response there are a number of GIS professionals helping to facilitate and map out a lot of the disease transmission, as well as where the flow of cases from one geographic region to another occurs. However there still tends to be a little bit of a divide

between the doctors, who are concerned about taking care of patients, the epidemiologists who want to get the numbers from the doctors, and the GIS professional who want the numbers from the epidemiologists. And so a lot of this is about how to best explain why this is important to other individuals.”

The follow on to the first quote illustrates the need for invisible articulation work, and a broker type activity, to assist in the development and maintenance of common ground between collaborators.

“So I spend a lot of my time explaining to GIS professionals why epidemiologists collect the data that they do, and explaining to epidemiologists how looking at their data from a spatial standpoint is important.”

Throughout the course of the research, these types of nuanced descriptions of distributed collaboration empowered through various forms of brokering activity were common. In the case illustrated, brokering took a very traditional form as an intermediary handled dialog between actors to enable cooperation. Frequently however, brokering takes a slightly different asynchronous form. Brokers in this digital humanitarian community frequently develop thematic contact lists and data inventories to help streamline their activities. These lists are shared with collaboration networks through various channels in order to promote connecting the right people to achieve a given goal. By adopting this model, the brokers disconnect themselves slightly from the brokering process while still acting as a connection between the communities they seek to empower.

DISCUSSION

In general, our findings support the notion that within a humanitarian collaborative analytics environment, articulation and anticipation work are handled by intermediaries [8,29,30] or brokers, who recruit participants and stakeholders, help solicit access to data and computing resources, facilitate and moderate communications, and connect analytic needs with data resources and participant skillsets. In a resource constrained environment, the potential for data sharing can expand the availability of digital data and systems must take into account many mechanisms for data transfer and submission. Further, as collaborative analytics are inherently interdisciplinary, brokers will play a key role in realizing the promise of ‘big data’ for humanitarian response.

Consequences of the Findings

Through interpretive qualitative analysis some key themes have been identified as useful when considering design constraints or assessing the mechanics of sociotechnical systems in the Ebola response context. First is the omnipresent role connectivity plays in the system. There does appear to be some degree of contradiction regarding whether reliable connectivity drives certain data management behaviors. Nonetheless, the prevalence of some reference to connectivity as well as the discussion

topics that coincide with references to connectivity issues matter quite a bit. First and foremost, connectivity enables network based data sharing through peer to peer data exchanges, data portals, and data services. Without reliable connectivity for all members of a collaborative network, the next best option becomes physical transfer by external media. This adjustment to a breakdown can create logistic, security, and data integrity issues that consume time and valuable resources [15,24]. The ability to easily share helps to alleviate these issues, and as illustrated through Finding #2, provides incentives to the members of the response community to capture, maintain, and share quality digital data. Furthermore, by enabling ease of data sharing and incentivizing digital data management, the groundwork for rollup of data and local analysis toward a collaborative analytics capability is established [26].

With regard to the data management practices and information behaviors potentially driven by access to reliable connectivity, the storyline presented through interpretation of the interview field seems to indicate the presence of some mitigating variable. Universally, the interview subjects agreed that connectivity is a significant part of the collaboration and information environment of the Ebola response, but the lack of connectivity resulted in a broad range of responses from field practitioners. If the ability to share incentivizes digital data capture and maintenance, and sharing is best enabled through connectivity, many might conclude that the absence of connectivity would promote the paper and whiteboard based workflows. In actuality, the interviews seemed to indicate that even in the absence of connectivity, sometimes data was still captured analog then converted to digital without any reasonable expectation that it could be shared soon. At other times, data was born digital, but due to a lack of connectivity, transferred using external media physically transferred by motor vehicles from site to site. Perhaps the notion that the ability to share in and of itself incentivizes digital data management needs to be deconstructed further. The ability to share without some benefit to the actor who must invest extra effort may not be recognized as a positive incentive. In this context then, perhaps the difference in data management behavior under circumstances of limited connectivity is also influenced in part by the degree to which that actor is actively engaged, or even interested in, digital analysis of their data and subsequent sharing and collaboration with a community of practice [8,29,30]. So while collaborative analytics depends heavily on the ability to share with ease, and ease of sharing in many cases is enabled by connectivity, for a cooperative data driven analytic environment to emerge certain social elements must co-exist.

Effective collaborative analytics requires, amongst other things, data sharing opportunities that incentivize digital data management as well as an interest in the potential benefit of participation. Interest can be promoted through brokering activities while concepts like anticipation work help to

explain why individuals and organization may or may not manage data digitally [29].

Examples of these dynamics are already emerging, with data sharing and access mitigated to a certain degree in recent years through implementation of humanitarian aid and crisis informatics data portals such as The Humanitarian Data Exchange (HDX) as well as the establishment of various digital volunteer support groups such as the Digital Humanitarians Network (DHN), Crisis Mappers, Humanitarian Open Street Map teams (HOT), Standby Task force (SBTF) and countless others. To prepare for technical aspects of collaboration these groups have developed or identified standards and provided data sharing infrastructures [29]. From a social interaction standpoint they have also enabled collaboration brokering activities across many different events and applications [23]. With that said, localized operational data sets developed within data silos have a number of issues. The unique localized datasets frequently needed for the Ebola response are still often times publicized through shared data inventory records and dropbox folders. So the data is shared via connectivity empowered architectures, but is really only openly shared as far as the collaborative network of the owner spans. Such an approach is agile, and the resources to support the IT component are low cost, making the solution attractive in resource constrained environments. The downside to these advantages is that it requires extra articulation work and brokering effort to keep accurate records of what kind of data is available and who possesses the most recent copies [35]. Metadata also becomes a factor when data development and maintenance becomes decentralized and uncoordinated. As data and results from local analysis are made available in a collaborative environment, the temporal dimension of data as well as its provenance may become important, further emphasizing the criticality of well understood data models and useful metadata.

Aside from the transparency issues of managing valuable data sources this way, there exists the issue of leveraging heterogeneous data sources. While there are techniques for addressing this issue, effectively integrating the disparate data sources requires a great deal of background knowledge, which returns us to the value of the broker. In addition to brokering knowledge about data holdings, brokering activities also occur to connect collaborators who have complementary skills or needs [35]. Social networks reflecting more semantically driven content controlled by and open to user communities might support navigating this space more effectively [33]. Skype group chats are one of the most prominent tools for such activities. In this case, group chats are established for certain communities of interest and membership is grown through a snowball recruiting method where members invite newcomers to the community into the chat session. From there introductions can be made and skills, needs, or questions can be socialized. Key members of the community emerge as brokers of knowledge and professional connections, frequently spanning across

multiple communities of interest and engaging to support numerous activities. Their value is that they bring together disparate actors with complementary skills, goals, or resources [22]. Mediation of the collaborative relationships then helps promote effective collaboration toward goals. Management of the cooperative effort may become even more important in the context of collaborative analytics, where datasets and analysis of many contributors are brought together in an analytic environment. In that sense, perhaps designing for the articulation needs of the broker may benefit the collective by empowering those who promote cooperation.

Implications for Design

Based on our current list of findings, design considerations to support a collaborative analytics environment would need to address the key features of the socio-technical system in a way that is more integrated than currently possible. The first issue to address is that of connectivity. Solutions that support asynchronous interaction between collaborators is one way to get around intermittent access. A second connectivity issue is insufficient bandwidth to support data access and sharing. Articulation, or working around the issue, is currently the only available option. In these cases the user must experiment with data access options and determine by trial and error the best strategy for data sharing. Sometimes that may mean reducing file size by reformatting or editing data. Perhaps affordances that ease this process would enable more effective workarounds. For instance, if a dataset or document needed to be shared with a group, perhaps it could be served in multiple forms and the user community messaged with the options and bandwidth requirements rather than offered a “take it or leave it when you can” option. While this paradigm could certainly become part of standard communication and collaboration etiquette, i.e. sending “small” and “large” versions, perhaps a more advanced option would work. Affordances built into ICT’s which allow users to dynamically and intelligently scope data shared with them prior to transfer might be an option.

The existence of brokering activities also present many opportunities to design useful affordances. Brokering inherently relies on the leverage of social networks and concepts such as social or professional capital and trust. While social media sites like LinkedIn or Facebook offer opportunities for collaborators to explore each other social networks, the opportunity to integrate a more semantically driven social and professional network would benefit the concept of collaborative analytics dramatically. From a semantic standpoint, users would then be interested not just in “who is connected to who” but also much deeper network information like why they are connected, what types of skills and interests can be found in a network or it’s communities. These types of information are already captured and utilized in many existing platforms, but if such platforms could be extended to also describe who has access to what kinds of data, hardware or software, or perhaps even identify

collaborative cohorts (existing or potential) with experience in a certain domain, the advantages to collaborative analytics for humanitarian aid would be significant.

CONCLUSION AND FUTURE WORK

This paper presents the initial findings of research into the data and analysis requirements of the 2014/15 Ebola response. The goal of the project is to understand the data sharing, collaboration networks, and analytic activities of data scientists and humanitarian aid personnel participating in the response effort in terms of specific operational needs. Through this understanding, the role of Big Data and analytic techniques in a collaborative environment can be clarified. Additionally, a better understanding of the collaborative systems that exists to support the response effort will help inform future design of collaborative architectures as well as identify key concepts and technologies for development and adoption in future infectious disease response efforts. The results of this research have identified internet connectivity, brokering within digital humanitarian networks as key elements to consider while designing sociotechnical systems and related applications. Connectivity empowers digital data architectures needed to support the response. Brokering socializes knowledge about data sources while promoting action between collaborators. The role of Big Data and analytics will become clearer as data architectures, methods for data collection, and collaborative processes mature and achieve widespread adoption. In the meantime, design should focus on supporting the vehicles that are flexible in affordances so user communities can continue to experiment with workarounds and collaboration methods for unique and unpredictable environments.

The nature of the environment in and of itself also has implications for the generalizability of our findings. In the Ebola context one of our key findings was centric to poor connectivity. This situation was closely connected to a need for more significant articulation work and a stronger role for brokering. If connectivity were not an issue, perhaps some of these findings would change. In that sense, findings driven heavily by the lack of connectivity may be generalizable only to similar applications. With that said, other findings have the potential to be generalizable across theory. The importance of an institutionalized brokering process to handle articulation work for the collaborative cohort seems to be deeply entrenched in the identity of digital humanitarians with most interviews indicating that across different disaster response projects, the tasks brokers took on may change, but the need for brokers was consistent.

Future research may develop additional insight through focus on local operations centric data use and the role analytics plays in response decision making. Additionally, capturing a more “in situ” perspective will likely help identify opportunities for more advanced techniques common to Big Data approaches offering significant impact to future response operations. This knowledge will ultimately help shape a vision for investment and adoption

of certain system design practices and innovations within collaborative practices throughout the humanitarian aid and crisis informatics community.

ACKNOWLEDGMENTS

The authors would like to thank all of the study participants for their time and willingness to contribute perspectives, the CSCW 2016 review committee for their comments regarding early drafts, and the National Science Foundation for funding through grant #1519023.

REFERENCES

1. Michael H. Agar. 1985. Speaking of Ethnography. *Sage University Paper series on Qualitative Research Methods* 2: 5–79. <http://doi.org/10.4135/9781412985895>
2. Ryan Burns. 2014. Rethinking big data in digital humanitarianism: practices, epistemologies, and social relations. *GeoJournal*. <http://doi.org/10.1007/s10708-014-9599-x>
3. Chee Seng Chong, Bong Zoebir, Alan Yu Shyang Tan, et al. 2012. Collaborative analytics for predicting expressway-traffic congestion. *Proceedings of the 14th Annual International Conference on Electronic Commerce - ICEC '12*: 35–38. <http://doi.org/10.1145/2346536.2346542>
4. Gregorio Convertino, Helena M. Mentis, Aleksandra Slavkovic, Mary Beth Rosson, and John M. Carroll. 2011. Supporting common ground and awareness in emergency management planning. *ACM Transactions on Computer-Human Interaction* 18, 4: 1–34. <http://doi.org/10.1145/2063231.2063236>
5. Nicola Dell, Trevor Perrier, Neha Kumar, Mitchell Lee, Rachel Powers, and Gaetano Borriello. 2015. Paper-Digital Workflows in Global Development Organizations. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, ACM Press, 1659–1669. <http://doi.org/10.1145/2675133.2675145>
6. Amydee M. Fawcett and Stanley E. Fawcett. 2013. Benchmarking the state of humanitarian aid and disaster relief: A systems design perspective and research agenda. *Benchmarking: An International Journal* 20, 5: 661–692. <http://doi.org/10.1108/BIJ-07-2011-0053>
7. Daniel Filokik, Markus Rittenbruch, and Marcus Foth. 2014. Participatory Data Analytics

- Collaborative Interfaces for Data Composition and Visualisation. *VINCI 2014*, 248–249.
8. Les Gasser. 1986. The integration of computing and routine work. *ACM Transactions on Information Systems* 4, 3: 205–225. <http://doi.org/10.1145/214427.214429>
 9. TM Harrison, Theresa Pardo, J. Ramon Gil–Garcia, Fiona Thompson, and Dubravka Juraga. 2007. Geographic information technologies, structuration theory, and the world trade center crisis. *Journal of the American Society for Information Science and Technology* 58, 1998: 2240–2254. <http://doi.org/10.1002/asi>
 10. Monika Alise Johansen, Jeremiah Scholl, Per Hasvold, Gunnar Ellingsen, and Johan Gustav Bellika. 2008. “Garbage in, garbage out.” *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*, ACM Press, 525. <http://doi.org/10.1145/1460563.1460646>
 11. Quentin Jones, Mihai Moldovan, Daphne Raban, and Brian Butler. 2008. Empirical evidence of information overload constraining chat channel community interactions. *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*, ACM Press, 323. <http://doi.org/10.1145/1460563.1460616>
 12. Paul Keel. 2006. Collaborative Visual Analytics: Inferring from the Spatial Organization and Collaborative Use of Information. *2006 IEEE Symposium On Visual Analytics And Technology*, IEEE, 137–144. <http://doi.org/10.1109/VAST.2006.261415>
 13. Benedikt Ley, Thomas Ludwig, Volkmar Pipek, Dave Randall, Christian Reuter, and Torben Wiedenhofer. 2014. Information and Expertise Sharing in Inter-Organizational Crisis Management. *Computer Supported Cooperative Work (CSCW)* 23, 4-6: 347–387. <http://doi.org/10.1007/s10606-014-9205-2>
 14. Edgar a. Maldonado, Carleen F. Maitland, and Andrea H. Tapia. 2009. Collaborative systems development in disaster relief: The impact of multi-level governance. *Information Systems Frontiers* 12, 1: 9–27. <http://doi.org/10.1007/s10796-009-9166-z>
 15. Gloria Mark and Bryan Semaan. 2008. Resilience in collaboration. *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*, ACM Press, 137. <http://doi.org/10.1145/1460563.1460585>
 16. J. Mason. 2002. *Qualitative Researching*. Sage Publishing, London.
 17. Patrick Meier. 2015. *Digital Humanitarians*. CRC Press, Boca Raton, FL.
 18. André Fabiano de Moraes and Lia Caetano Bastos. 2010. Framework of integration for collaboration and spatial data mining among heterogeneous sources in the web. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Data Mining for Geoinformatics - DMG '10*, 19–28. <http://doi.org/10.1145/1869890.1869893>
 19. Kristi Morton, Magdalena Balazinska, Dan Grossman, Robert Kosara, and Jock Mackinlay. 2014. Public Data and Visualizations : How are Many Eyes and Tableau Public Used for Collaborative Analytics ? *SIGMOD Record* 43, 2: 17–22.
 20. Michael Muller, Susanne Hupfer, Stephen Levy, Daniel Gruen, Andrew Sempere, and Reid Priedhorsky. 2011. Circles of Crowdsourcing : The Social Organization of Participatory Sensing. *MobileHCI*. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=5470364
 21. Leysia Palen and Sarah Vieweg. 2008. The emergence of online widescale interaction in unexpected events. *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*, ACM Press, 117. <http://doi.org/10.1145/1460563.1460583>
 22. Suzanne D Pawlowski and Daniel Robey. 2004. Bridging User Organizations : and the Work of Brokering Technology Knowledge Information. *MIS Quarterly* 28, 4: 645–672.
 23. Suzanne D. Pawlowski, Dan Robey, and Arjan Raven. 2000. Supporting shared information systems: boundary objects, communities, and brokering. *Proceedings of the twenty first international conference on Information systems*, 329–338. Retrieved from <http://dl.acm.org/citation.cfm?id=359759>

24. Bryan Semaan and Gloria Mark. 2011. Technology-mediated social arrangements to resolve breakdowns in infrastructure during ongoing disruption. *ACM Transactions on Computer-Human Interaction* 18, 4: 1–21. <http://doi.org/10.1145/2063231.2063235>
25. Jinkuan Song, Yang Yang, Yi Yang, Zi Huang, and Heng Tao Shen. 2013. Inter-media hashing for large-scale retrieval from heterogeneous data sources. *Proceedings of the 2013 international conference on Management of data - SIGMOD '13*, 785. <http://doi.org/10.1145/2463676.2465274>
26. Susan Leigh Star and Anselm Strauss. 1999. Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)* 8, 1-2: 9–30. <http://doi.org/10.1023/A:1008651105359>
27. Kate Starbird and Leysia Palen. 2013. Working and sustaining the virtual “Disaster Desk.” *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13*, ACM Press, 491. <http://doi.org/10.1145/2441776.2441832>
28. Anthony Stefanidis, Andrew Crooks, and Jacek Radzikowski. 2011. Harvesting ambient geospatial information from social media feeds. *GeoJournal* 78, 2: 319–338. <http://doi.org/10.1007/s10708-011-9438-2>
29. Stephanie B Steinhardt and Steven J Jackson. 2015. Anticipation Work. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, 443–453. <http://doi.org/10.1145/2675133.2675298>
30. Anselm Strauss. 1988. The Articulation of Project Work: An Organizational Process. *The Sociological Quarterly* 29, 2: 163–178. <http://doi.org/10.1111/j.1533-8525.1988.tb01249.x>
31. J Sutton, Leysia Palen, and Irina Shklovski. 2008. Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires. *Proceedings of the 5th International ISCRAM Conference*, 1–9.
32. Andrea H. Tapia and Kathleen Moore. 2014. Good Enough is Good Enough: Overcoming Disaster Response Organizations’ Slow Social Media Data Adoption. *Computer Supported Cooperative Work (CSCW)* 23, 4-6: 483–512. <http://doi.org/10.1007/s10606-014-9206-1>
33. Sebastian Tramp, Philipp Frischmuth, Timofey Ermilov, Saeedeh Shekarpour, and Sören Auer. 2012. An Architecture of a Distributed Semantic Social Network. *Semantic Web – Interoperability, Usability, Applicability* 1: 1–16.
34. E. M. Trauth and L. M. Jessup. 2000. Understanding computer mediated discussions: Positivist and interpretive analysis of group support system use. *MIS Quarterly* 24, 1: 43–79. <http://doi.org/10.2307/3250979>
35. Murray Turoff, Michael Chumer, Xiang Yao, and Bartel Van De Walle. 2004. " The Design of a Dynamic Emergency Response Management Information System (DERMIS THE DESIGN OF A DYNAMIC EMERGENCY RESPONSE MANAGEMENT INFORMATION SYSTEM (DERMIS)). 54: 1–35. Retrieved from <http://is.njit.edu/turoff>
36. Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging during two natural hazards events. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*: 1079. <http://doi.org/10.1145/1753326.1753486>
37. 2014 Ebola Outbreak in West Africa - Case Counts | Ebola Hemorrhagic Fever | CDC. Retrieved November 21, 2014 from <http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/case-counts.html>