CROWDWORK, CRISIS AND CONVERGENCE: HOW THE CONNECTED CROWD ORGANIZES INFORMATION DURING MASS DISRUPTION EVENTS

by

KATE STARBIRD

B.S., Stanford University, 1997

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Technology, Media and Society

Alliance for Technology, Learning and Society (ATLAS) Institute

2012

This dissertation entitled:

Crowdwork, Crisis and Convergence: How the Connected Crowd Organizes Information during Mass Disruption Events

written by Kate Starbird
has been approved for the Alliance for Technology, Learning and Society Institute

____
Leysia Palen

____
John Bennett

____
Edwin Hutchins

____
Michele Jackson

____
Clayton Lewis

Date_______________

The final copy of this thesis has been examined by the signatories, and we
Find that both the content and the form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

IRB protocols #10-0456 and #0407.15

ABSTRACT OF THE DISSERTATION

By Kate Starbird
(Ph.D. Technology, Media and Society; ATLAS Institute)

Crowdwork, Crisis and Convergence: How the Connected Crowd Organizes Information during Mass Disruption Events

Dissertation Directed by Associate Professor Leysia Palen

Social media have experienced widespread adoption in recent years. Though designed and appropriated for a range of purposes, users are consistently turning to these platforms during times of crisis and *mass disruption*—a term used here to characterize events, including mass emergencies, natural disasters and political protests, that cause significant disruption to normal routines. Social media are playing host to new, digital forms of the social convergence behavior long known to occur in the wake of crisis events. This activity, which includes participation from local citizens, emergency responders, and global onlookers alike, produces huge volumes of data, some with potential value to affected people and responders. It also creates new challenges. Noise, misinformation, lost context and the unstructured nature of social media updates all contribute to an emerging information processing problem, with information seekers forced to "drink from the firehose" to identify the data they need.

Noting the difficulties of completely solving this problem with purely computational solutions, I address the challenge of processing social media updates into usable information from a perspective that positions the participating crowd as an asset in the effort. At the center of this inquiry is the discovery of an emerging role for remote participants during mass disruption events—that of the *digital volunteer*. This dissertation consists of four separate studies of digital volunteerism and other forms of remote participation, examining several ways members of the remote crowd help to organize information during mass disruption events. Across the different studies, I employ a mixture of methods, including qualitative and quantitative analysis of large volumes of Twitter data, interviews with digital volunteers, and participant observation within a virtual volunteer organization.

Integrating the findings from these separate studies, I introduce a new term, *crowdwork,* to describe the productive activity of remote participants during mass disruption events. Throughout, this dissertation works to unpack the popular *crowdsourcing* term, by identifying salient features of crowdwork in this context and comparing those with current understandings of crowdsourcing. Examining the larger ecosystem of digital volunteerism during mass disruption events, I describe crowdwork in this context as a multilevel filtration system, explaining how information is processed through a variety of different activities at different layers within a complex information space that includes crowdworkers, virtual organizations, and social media sites that host both the information and the information processing. This model identifies several potential "sites" of innovation where computational algorithms could both support and leverage crowdwork.

Finally, from another perspective, I examine crowdwork through the movement and transformation of information. Using the theory of distributed cognition in combination with this information-centered approach, this dissertation concludes with a holistic view of crowdwork on social media platforms as *collective intelligence* manifested within a global cognitive system.

ACKNOWLEDGMENTS

This work would not have been possible without the support of many individuals and organizations. First, I would like to thank my advisor and mentor, Leysia Palen, for all of her direction, support, encouragement, and advice during my time at the University of Colorado. There is not a chance that I would be researching in this crisis informatics space if I had not connected with her early in my graduate career, and it's questionable whether or not I would have taken my PhD studies to their conclusion without her guidance. I am extremely grateful to her for investing her time and energy into molding me into the researcher and scholar that I have become (or rather am still and always will be becoming).

Along that same vein, I would like to thank all my colleagues at Project EPIC for the conversations and collaborations that shaped this work. I would like to especially thank Sophia B. Liu, Sarah Vieweg and Amanda Hughes, whose research set the foundations for mine, and who supported me throughout my PhD studies with conversation, ideas, edits, and inspiration. I'd also like to thank peers that arrived on the scene later, especially Joanne White, who both supported my research and helped guide the development of my many projects. Certainly, the Tweak the Tweet project would never have gotten off the ground without the many hundreds and possibly thousands of tweets sent by Sophia, Sarah, Amanda and Jo.

I also cannot thank enough the other side of Project EPIC, those building the systems to support our research and staying up late at night and sometimes all night to keep those systems running. Specifically, I am grateful for the herculean efforts of Ken Anderson and Aaron Schram to collect something approaching terabytes of Twitter data to support my research, and for their Zen-like ability to weather the franticness of my collection requests which, unfortunately for the two of them, were tuned into the rhythms of disaster—rhythms that do not align at all with the natural rhythms of human life.

I would also like to acknowledge my colleagues at the ATLAS Institute, including my fellow students who contributed to my academic development, the staff who helped me over all the administrative hurdles in a PhD career, and the two Directors of ATLAS, John Bennett and Bobby

Schnabel, for their support. Though Bobby left CU and ATLAS before my first class as a PhD student, I am extremely grateful for his oh-so-slight intervention in my life 6 years ago, when he urged me to apply to ATLAS and then eventually convinced me to enroll. I am also grateful to him for continuing to mentor me from afar, and making time for me whenever he returned to Boulder.

There is one more person without whom I would not have completed this work – or even started it. Thank you, Melissa, my partner/wife, for holding my hand through my transition from my first career to my second one, for helping me decide that ATLAS was the place for me, and for sticking with me in Boulder even though you were too allergic to go outside there from April to October. When I became a hermit in my back office, tweeting, I mean, writing away, you continued to bring food to my door (and to tell me to get back to work when I was fading). Thank you!

Finally, here, I would like to thank the dozens and possibly hundreds of digital volunteers with whom I have come in contact over the last few years. Thank you for all the work you do for people in need all over the world! And thank you for supporting Tweak the Tweet and for tolerating my constant inquiries into how and why you do the things you do. Special thanks go to the many volunteers at Humanity Road.

Financial support for this work was provided through a National Science Foundation Graduate Fellowship and NSF grants #IIS-0546315 and #IIS-0910586.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION: MASS DISRUPTION AND INFORMATION CONVERGENCE

This research examines human interaction and collaboration occurring on a large-scale through social media during times of mass disruption—a phenomenon of increasing interest to researchers, news reporters, formal emergency responders, humanitarian agencies, and the public at large. The Internet has created new virtual spaces for large-scale interaction and opened up countless new channels where information flows. Ubiquitous technology (including mobile devices) now provides access to these spaces and the information produced within them from more places, geographical and virtual. Not surprisingly, social media sites and services (e.g. weblogs, social network sites, photo and video sharing sites) are playing host to new, digital forms (Palen & Liu, 2007; Hughes & Palen, 2009) of the social convergence behavior long known to occur in the wake of mass emergencies and disasters (Fritz & Mathewson, 1957; Dynes, 1970; Tierney et al. 2001; Kendra & Wachtendorf, 2003). Emergent, digital convergence activities include both information generation (i.e. citizen reporting) and information organization. Without overlooking the former, this research intends to take an in-depth look at the latter component, investigating the myriad ways in which social media users are processing a flood of data into useful information during mass disruption events. This inquiry extends from a perspective of "intelligent" crowds (Surowiecki, 2005), where users act both purposefully and "passively" (Howe, 2008) to help organize data and create useful informational resources during these events.

For this research, I define *mass disruption events* as events affecting large numbers of people that cause disruption to normal social routines. These include events like natural and man-made disasters, mass emergencies, political protests and riots.

1.1 Converging through Social Media

The sociology of disaster offers a useful perspective for examining the convergence behavior that manifests during mass disruption events. Sociologists of disaster have repeatedly shown that the first responders to disasters are rarely the formal organizations who are charged to respond, but are instead spontaneous volunteers who find themselves at the scene, or quickly converge on the scene, and begin to help (Fritz & Mathewson, 1957; Dynes, 1970; Tierney et al. 2001; Kendra & Wachtendorf, 2003; Palen & Liu, 2007). Convergence is not limited to the physical realm. Members of the public have consistently appropriated available information communication technologies (ICTs) to converge "informationally" as well.

Fritz & Mathewson (1957) define convergence as consisting of three parts: convergence of people, convergence of material goods, and informational convergence. In their description, informational convergence manifests as messages from the outside world converging on the affected community. It includes attempts to contact family or friends, inquiries about the status of people or places, and offers of assistance. At the time of that research, this form of convergence was enabled by tools like the mail system, the telephone, the telegram and the radio. It is important to note that Fritz and Mathewson described informational convergence as a problem for responders (1957).

Due to the ease of information production and transmission they enable, and through the social connections they help create and foster, social media provide opportunities for information convergence on a greatly increased scale and, I will argue, of a new kind or kinds. We now know that members of the public converge via social media during and after mass disruption events (e.g. Hagar & Haythornthwaite, 2005; Liu et al., 2008; Hughes & Palen, 2009; Palen et al. 2009). Qu et al. (2011) report that people are turning to social media during times of crisis for many of the same informational activities that Fritz and Mathewson (1957) documented in the pre-digital world, including reporting or inquiring about missing persons and requesting or offering assistance.

Several research studies describe social media being used for new types of information activities as well. Gillmor (2004) noted that the emergence of social media tools marks a significant turn in news production, giving a voice to the previously voiceless and thereby creating a "citizen journalist." Citizen journalists—also commonly referred to as citizen reporters—on the scene of mass disruption events are newly enabled to provide raw data from the ground, generating a potential new data feed for responders and other members of the public (Palen et al., 2010). Research shows that members of the public are now using social media to provide information to others during times of crisis, often from the scene of unfolding events (Palen & Liu, 2007; Qu et al. 2011). Social media are becoming places where users gather and integrate information as well (Qu at al. 2009). Discussing list-building activities that took place on Facebook groups after the Virginia Tech shootings in 2007, Palen et al. (2009) interpret the crowd's collaborative work to collect and verify the names of victims as a form of "collective intelligence." Digital convergers are also using social media to coordinate relief activities. Though self-organizing by emergent groups attempting to provide aid is an old phenomenon, often documented in the actions of physical convergers during the aftermath of disasters (Fritz & Mathewson, 1957; Dynes 1970; Wachtendorf, 2004), social media greatly increase the scale of this self-organizing behavior, allowing more people to connect to the event and join in on the relief activities.

1.2 The "Problem" of Informational Convergence

As mentioned above, Fritz and Mathewson (1957) framed informational convergence as a problem for emergency responders and local infrastructure, claiming that it overwhelmed communication facilities. Both physical infrastructure (e.g. telephone lines) and human resources (e.g. mail sorters) could be overtaxed. Fritz & Mathewson describe the problem as one of capacity, using terms like "jamming, swamping and overloading." Though issues regarding the capacity of *carrying* the messages of informational convergence (i.e. the physical infrastructure of ICT) are outside the scope of this research, the problem of the capacity related to *processing* this information is a central theme here.

One could view social media as amplifying a problem of information overload during mass disruption events, contributing to an ever-expanding flood of data that responders and (increasingly) members of the public must navigate in order to fully understand an emerging event. In the Disaster Relief 2.0 Report (Harvard Humanitarian Initiative, 2011), a report about the interaction of the humanitarian community with digital volunteer communities during the aftermath of the 2010 Haiti earthquake, representatives of humanitarian agencies claimed that data generated on social media by those affected as well as digital volunteers that converged there added to responders' existing difficulties with processing large amounts of information. Responders were neither prepared nor properly equipped to "produce useful knowledge from the flow of information and data" (p. 10). A key part of this problem stems from the *unstructured* nature of social media communications. Relief workers complained that this data required significant human processing to convert it into the uniform data formats needed for integration with their existing tools and other data sources. Though members of one digital volunteer community dispute the claim that volunteer communities amplified the problem (Meier et al, 2011), back-and-forth exchanges[1] between the established humanitarian relief agencies and the emerging digital volunteer communities were successful in highlighting that social media data change the informational landscape during mass disruption events. It is important to understand how.

Social media contribute to information processing problems in several significant ways. First, by providing new channels for information reporting, they add to the sheer volume of information available. When used by those with first-hand knowledge of an unfolding event, social media increase the volume of potentially actionable information valued by responders, locally affected individuals, and physical convergers. They also provide channels for communicating information that appeals to the broader audience, including high-level descriptions of the event, commentary, and messages of support or prayer. They are places where spam and a considerable amount of semantically similar, yet irrelevant, messages enter the information system. In Starbird & Palen (2010), we discuss the different ways that broad appeal and locally relevant information propagate through Twitter, a popular microblogging platform. Those mechanisms of information propagation allow copies of original messages to be re-broadcast through the system. All of this information flows through the same channels, contributing to volume and mixing different "signals" with "noise." With so much information available, it is hard to locate the pieces that are relevant to a given situation. For instance, an emergency response group looking to track a cyclone

[1] These exchanges occurred in the comments of the Meier et al., 2011 blog post at: http://blog.standbytaskforce.com/why-we-need-a-disaster-2-1-report/

moving across the Indian Ocean might have to sift through and discard thousands of social media messages about a victory for the Iowa State football team (whose mascot is a cyclone). As they move to respond to the event, they may have to filter further, eliminating copies and removing messages of prayers and offers of support to concentrate on posts containing immediate needs and damage reports.

Developing effective filters for separating signal from noise has been cited as *the* issue in dealing with the huge volume of data produced and available through computer-mediated communication (Shirky, 2008b). But social media adds another piece of complexity to the information puzzle that new filters will not necessarily solve: lost context. Information can lose context in different ways, during different stages of its lifecycle (i.e. during its creation, its propagation, its aggregation). A Facebook user posting a public status report about an incoming tornado may not bother to mention what city she lives in, because her message is intended for those in her Facebook social network who would already have that knowledge about her, that context for interpreting her message. In a similar way, information propagated on Twitter though the retweet mechanism may lose reference to the original time that it was posted. That same tornado warning, accurate at 4:30 in the afternoon, will most likely be inaccurate when retweeted a few hours later, though the text may still say “immediate.” This lost context can be particularly troublesome during mass disruption events where the veracity of information can be a life or death matter. Removed from the context of its creation, information as data can become ambiguous, irrelevant, untimely, and even inaccurate. Converting social media data to useful information therefore requires both filtering and “re-contextualizing,” returning the data to the context of its creation or positioning it correctly within a new context to reconstitute meaning.

Along with outdated warnings and other forms of misinformation, the issue of disinformation is another problem, one of particular concern within the domain of mass disruption. The Disaster Relief 2.0 Report (Harvard Humanitarian Initiative, 2011) recounted instances of false information posted on social media leading to wasted resources by humanitarian response agencies. That report, confirming a hypothesis which arose during our own research on social media use during the aftermath of the Haiti earthquake, suggested that individuals were reporting instances of relatives trapped but still alive in the rubble, when in fact the families were attempting to manipulate responders into helping them recover the bodies of victims not heard from in days or weeks. Though this form of manipulative disinformation is wasteful of resources but seemingly not purposefully malignant, disinformation during mass disruption events like political protests could be purposefully entered into the system in order to harm participants or compromise the political cause. Thus, the validity of social media data is a third, critical component of the problem of informational convergence.

Finally, to be useful in its aggregate form, the data must be converted into a uniform format that computer programs can process. Uniform data is much more easily collected, filtered, stored and compared than the raw information of social media. Nigel Snoad, a former representative of United Nations Disaster Assessment and Coordination (UNDAC) described the problem this way: “If you want to know where the biggest gap is, it’s the extraction of structured data from unstructured inputs” (quoted in the Disaster Relief 2.0 Report, Harvard Humanitarian Initiative, p. 22). Informational convergence via social media produces huge amounts of unstructured data that must somehow become structured to become useful to disaster responders.

1.3 The Work of Information Processing during Mass Disruption

> Dispersed pieces, tidbits of data almost, that can’t even be classified as information, were some of the less useful [information we got], because that required an intense amount of resources and coordination to turn that into actionable piece information. (Andrew Alspach, OCHA, quoted in Harvard Humanitarian Initiative, 2011, p 22)

As the above quote suggests, filtering and re-contextualizing data to generate useful informational resources requires work. And researchers in the computer science domain are working to find computational solutions to do this work. Palen et al. (2010) describe solutions in development that intend to leverage social media information for sense-making tools, aiming to equip “everyday analysts” with the means to effectively find relevant information during times of crisis. One of those approaches involves the use of natural language processing (NLP) techniques to extract and categorize situational awareness information from textual information in social media posts—in this case tweets (Verma et al., 2011). Though these approaches hold promise, we do not yet have practical and usable machine-only computational solutions for processing crisis data into resources for responders (American Red Cross, 2010b).

We are, however, witnessing the mobilization of armies of volunteers—what the Disaster Relief 2.0 report called "Virtual & Technical Communities" or V&TCs, and what I will call *digital volunteers*, digital volunteer communities, and virtual volunteer organizations—who are converging via digital media to help collect, filter, and process this data. Some of these communities are merely groups of people, connected through a social media platform or platforms, acting individually and collectively to help organize the data, both through conscious effort and by working, unintentionally, as collaborative filterers. Other communities, like the "voluntweeters" for Haiti, are effectively organizing themselves into emergent volunteer organizations. Still others (e.g. Humanity Road[2] and the Standby Task Force[3]) are defining ongoing roles for themselves in the disasters space as established, though virtual, organizations.

Digital volunteers are using social media in a multitude of ways—as sources of data to operate on, as platforms to communicate through, and as tools for connecting to other volunteers. They incorporate popular social media sites (i.e. Facebook, Twitter, Skype) from their existing daily routines into their activities as crisis volunteers, and they adopt new tools (e.g. Ushahidi) specifically for this work. Some volunteer communities create new technologies or improve existing ones to help them with their work. Many groups organize around a specific ICT or social media site, while others incorporate an array of ICTs into their work. Their communities are often born or expand during times of crisis, appear to atrophy in periods of calm, and in some cases reactivate during later events. Some communities connect to other communities in increasingly interdependent ways. Other communities splinter as internal dynamics drive subgroups in different directions.

---

[2] http://www.humanityroad.org/

[3] http://blog.standbytaskforce.com/ - The Standby Task Force supported six events during Spring, 2011 and was activated by UNOCHA during the first few weeks of the Libya uprising to monitor media and map reports of violence and displaced people.

There is a clear need for stakeholders to better understand the capacities and work products of these communities as well as their connections to both formal response and affected people. This need was brought to the surface by the Disaster Relief 2.0 report (Harvard Humanitarian Initiative, 2011), a report funded in part by the United Nations Foundation, and the subsequent replies by members of the digital volunteer communities. The original report exposed a number of unresolved issues that formal responders and humanitarian organizations face when dealing with these new digital volunteer communities in regards to the veracity and usability of crisis data. It also outlined a complex, new, and rapidly evolving landscape of disaster response, with new stakeholders (as well as old stakeholders with newly empowered voices). In a blog representing the Standby Task Force, Meier et al. (2011) refuted much of the original report and called for the joint community of disaster response and volunteers to begin work on a Disaster 2.1 report. That reply suggested that several volunteers felt their work had been misrepresented and the problem of exacerbating data overload unfairly attributed to them. Comments to Meier et al.'s blog continued the conversation and revealed both division between the humanitarian groups and virtual volunteer communities and willingness by many to figure out ways to work together better.

1.4 Crowdwork

This research aims to help conceptualize this emergent ecosystem of digital volunteerism by describing the myriad ways that digital volunteers and other members of the crowd are acting to process data into information during mass disruption events. To effectively uncover and communicate these phenomena, it will be important to unpack the popular term *crowdsourcing* (Howe 2006; 2008), which has in some ways become a catchall term for crowd-leveraging solutions within the disaster space. Examining the current conversations around crowdsourcing in the context of mass disruption reveals the term to have nebulous and contradictory definitions. This work aims to expose the underlying behaviors that constitute the popular term and to connect these behaviors to other existing frameworks for interpretation (Malone et al., 2009; Quinn & Bederson, 2011a), including collective intelligence and human computation.

A close examination of several different forms of crowd work during disaster reveals that the salient features of that work are not sufficiently highlighted within any of these existing frameworks, and will provide opportunity for exploring those interconnected efforts from other perspectives. Simply put, the vast majority of the crowd activity examined here can be thought of as either organizing information or organizing people (to organize information). This research explores both perspectives, shifting between a people-organizing focus and an information-organizing focus, and progresses towards a new, integrated descriptor, *crowdwork*, for the distributed, connected work of organizing information during mass disruption events.

The *crowd* of this work is not necessarily one of great volume; in fact, some of the activity studied here occurs in small groups of just a few people. Instead, it is the crowd implied in the *crowd*sourcing term, one that aligns with the Oxford English Dictionary's definition of "a mass of spectators, an audience" (crowd, n3, 1b) and reflects some relationship to the word *cloud*, a term that has come to mean "pertaining to or doing business on the Internet[4]." I therefore define crowdwork as the work of the Internet-connected audience, with the implied potential for—though not a requirement of—a large number of participants.

Through four studies of the connected crowd in action during mass disruption events, this research frames crowdwork as applied collective intelligence. In answering the broad question of how the crowd helps to convert information overload into useful resources during disaster, I aim to improve our understanding of both how and why these efforts "work," to reframe how we evaluate this work, and to provide design recommendations for further enabling and leveraging current and future crowdwork. An overriding aim of this research is to advance our understanding of what crowds and crowdworkers are capable of—by revealing how they self-organize and can be organized to solve real problems in the fast-paced and time- and safety-critical domain of crises and mass disruption events. I also hope that this

[4] http://dictionary.reference.com/browse/cloud

research will help formal disaster and humanitarian responders better utilize the products and capacities of crowdworkers in their response efforts.

1.5 Structure of the Dissertation

The research is based on four separate studies of crowd activity during mass disruption events, building towards a larger framework for understanding crowdwork in this context. The dissertation consists of nine chapters, the first being this Introduction. Chapter 2 provides a background on social media and mass disruption events, including a literature review that integrates existing research from the fields of human computer interaction, sociology of disaster, crowdsourcing and human computation. Chapter 3 describes the research questions addressed in this dissertation as well as the research trajectory that led to the identification of these questions. The four studies are each presented in a separate chapter (Studies 1-4 map to Chapters 4-7) that includes background, methods, and findings for that study, as well as material that pertains to the larger findings of the research. Chapter 4 (Study 1) describes the rationale behind and implementation of the Tweak the Tweet project, a proposed innovation for reporting crisis-related information via social media using a structured syntax, and relates findings from the analysis of several deployments of the syntax during crisis events in 2010-2012. Chapter 5 (Study 2) examines the self-organizing of digital volunteers—in this case remote Twitterers who called themselves "voluntweeters"—during the early aftermath of the Haiti earthquake, identifying a new role for remote participants during crisis events. Chapter 6 (Study 3) explores the organizing, work practices, tools and products of Humanity Road, a first-of-its-kind, virtual volunteer organization in the domain of crisis response. Chapter 7 (Study 4) looks at the interaction between the crowd and on-the-ground participants of a mass protest event, the 2011 Egyptian Uprising, uncovering "work" done by remote participants to support the protesters and the larger movement. Integrating findings from the four studies, Chapter 8 identifies the salient features of crowd activity during mass disruption events and outlines new perspectives for examining crowdwork in this context. Finally, Chapter 9 presents the Conclusions of this research.

1.5.1 Integration of Previously Published Work

This dissertation incorporates previously published research where the author of this work was the first author of a conference paper or book chapter that included in many cases a co-author or co-authors. Most significantly, Study 1 (Chapter 4) includes a small portion (<20%) of previously published material that involved multiple co-authors, and Studies 2 and 4 (Chapter 5 and 7) are adaptations of previously published work that had a single co-author. An adaptation of Study 3 (Chapter 6) has been submitted as a conference paper with a co-author. Chapter 8 builds on a published paper written solely by the author of this dissertation. In all cases, the text here has been adapted in places for consistency within this document and augmented with new material and interpretations that pertain to the larger theme and findings of this research. Each of these chapters contains information about how to cite material from that study, including a pointer to the original work. For Chapters 5 and 7, the previously published work should be considered the original and this version a derivative.

CHAPTER 2
BACKGROUND: SOCIAL MEDIA, MASS DISRUPTION AND CROWD WORK

This work carves out a research space at the intersection of several different domains, incorporating perspectives on computer-mediated communication (CMC) and social media, crisis and mass disruption, and what will initially be referred to here as "crowdsourcing." It integrates existing knowledge and theory from human-computer interaction (HCI) and the sociology of disaster, as well as emerging fields of human computation and collective intelligence. Though grounded in existing and in some cases long-standing research in some of these fields, this account also relies at times, especially when discussing emerging issues in social media and disaster response, on recent blogs, websites, and technical reports from the emerging community of digital volunteers.

This Background section begins by exploring the digital interaction spaces of social media and their use during mass disruption events. It then relates emerging knowledge of how people are using social media during these events to established research by sociologists of disaster, especially to the concept of informational convergence (Fritz & Mathewson, 1957). Relying on recent reports and blogs from humanitarian agencies, it will bring to light the "problem" of information overload via informational convergence in the disaster domain where, as the Disaster Relief 2.0 report claimed, responders are having to "drink from the firehose" (Harvard Humanitarian Initiative, 2011).

Accepting a proposition that computational solutions alone will not be enough to completely solve these problems, this chapter proceeds to examining the possibility of using the power of the connected crowd instead. Along this vein, the chapter begins the work of unpacking the popular "crowdsourcing" term, uncovering its roots in open source and outsourcing, and tying it in to related concepts of human computation and collective intelligence. Finally, it offers an overview of some of the crowd work configurations—emergent groups and virtual organizations—that we have begun to see in this new, developing ecosystem of digital volunteerism during mass disruption events.

2.1 Social Media

This research arises from the practical issue of converting big data into useful information in the Web 2.0 world, where new configurations of informational communication technology have spawned new possibilities for information production, broadcast, storage and retrieval. Social media are an increasingly popular form of ICT. These Internet-enabled platforms allow users to create, share and consume information from the web as well as from a variety of increasingly ubiquitous mobile devices. While social media platforms have greatly expanded the amount of data available for consumption, they have also opened up novel possibilities for processing that data into useful information. Though social media are not exactly a new phenomenon—the UseNet online discussion system was launched in 1980 and blogging first appeared in the late 1990s (Kaplan & Haenlein, 2010)—they are experiencing rapid growth and integration into modern lives.

Though the term is sometimes conflated with one subset of its members—popular social-networking sites like Facebook—social media come in many varieties, including microblogging (e.g. Twitter), photo-sharing (e.g. Flickr, Instagram), video-sharing (e.g. YouTube), location-sharing (e.g. FourSquare), and news recommendation (e.g. Digg), to name a few. Considered together, there are hundreds and possibly thousands of social media sites available, and users of these platforms are generating billions of new data points each day. Citing statistics from March 2012[5], Facebook reports that it currently has 901 million active users creating billions of pieces of content (status updates, web links, news stories, blog posts, notes, likes, photos) each day (Facebook Newsroom, 2012), while users of Twitter, a popular microblogging service, currently send 340 million messages or "tweets" per day (Twitter Blog, 2012). Instagram reports that more than 5 million photos are being shared daily on that platform (Eler, 2012).

---

[5] Usage statistics for all sites except Youtube are current for June 4, 2012. Youtube statistics were reported in a 2011 research study.

While visiting Flickr's recent photos page,[6] a dynamic script reports that 3,704 photos were uploaded in the previous minute. And in that same minute, YouTube (a video-sharing site) users uploaded about 48 hours worth of new video (Tsukayama, 2011). Clearly, social media sites are creating a steady stream, perhaps better characterized as a flood, of data.

2.2 Social Media and Mass Disruption

The first responders to disasters are rarely the formal organizations with the mandate to do so, but are instead spontaneous volunteers who find themselves at the scene, or quickly get there, and begin to help. Sociologists of disaster have noted time and again that after a disaster, people will converge onto the scene, often with intentions of providing assistance to those who have been affected (Fritz & Mathewson, 1957; Dynes, 1970; Kendra & Wachtendorf, 2003b; Palen & Liu, 2007; Hughes et al., 2008). These *convergers* are often considered a liability, another component of disasters that responders must manage. For instance, physical convergers can increase traffic on their way to the scene, crowd hospitals waiting to donate blood, and bring into an affected area unneeded food or other supplies that will have to (somehow) be removed (Fritz & Mathewson, 1957). Convergers can also be seen as a resource, improvising solutions to important problems during the response period. During the aftermath of the terrorist attacks on the World Trade Center on September 11, 2001, volunteer GIS specialists from local colleges helped to create maps for responders, and a group of chiropractors talked their way into the Ground Zero secure area and provided much-appreciated treatments to first responders (Wachtendorf, 2004).

Fritz & Mathewson (1957) define convergence as consisting of three parts: convergence of people, convergence of material goods, and *informational convergence*. In their model, informational convergence was enabled by tools like the mail system, the telephone, the telegram and the radio, and manifested as attempts to contact family or friends, inquiries about the status of people or places, and offers of assistance. Perhaps not surprisingly, we are now witnessing informational convergence via social media and other ICT (Palen & Liu, 2007; Hughes et al., 2008).

[6] Accessed: June 4, 2012 http://www.flickr.com/photos/

Social media are also enabling a new informational behavior during mass disruption, opening up channels of communication between those affected and an increasingly global audience. Gillmor (2004) noted that the emergence of social media tools marked a significant turn in news production, giving a voice to the previously voiceless and thereby creating a "citizen journalist." Citizen journalists on the scene of mass disruption events are newly enabled to provide raw data from the ground, generating a potential new data feed for responders and other members of the public (Palen et al., 2010).

2.2.1 Disasters and Mass Emergencies

Numerous research studies show people, both those affected and digital convergers, turning to ICT and social media during disaster events. Palen and Liu (2007) describe how blogs and wikis were used to help displaced people, coordinate relief efforts, and assist in locating missing people during the aftermath of Hurricane Katrina in 2005. That paper also notes the emergence of collaborative mapping efforts on visual wikis. Other studies discuss how Facebook was used in the aftermath of the Virginia Tech shootings in 2007 (Vieweg et al., 2008; Palen et al., 2009), how an online forum became a resource for many in the wake of the 2008 Sichuan earthquake (Qu et al., 2009), and how a range of ICTs, including social media (i.e. Flickr, Picasa, discussion boards, blogs, Twitter), were used to seek and provide information during the San Diego Wildfires (Sutton et al., 2008). Several research efforts have documented the use of Twitter and other microblogging services during mass disruption events, including the 2009 Red River floods and the 2009 Oklahoma grassfires (Starbird et al., 2010; Starbird & Palen, 2010; Vieweg et al., 2010), multi-victim shootings in Lakewood, WA in 2009 (Heverin & Zach, 2010), the 2010 Haiti earthquake (Starbird & Palen, 2011; Sarcevic et al., 2012), and the 2010 Yushu earthquake (Qu et al., 2011).

The 2010 Haiti earthquake marked a turn for humanitarian responders in recognizing an emerging role for social media and its users in disaster response efforts. The Disaster Relief 2.0 Report detailed the use of several social media tools, including the Ushahidi crisismapping platform, during the aftermath of the Haiti earthquake, explaining "…access to mobile and online communication enabled a kind of collective intelligence to emerge—when thousands of citizens around the world collaborated in volunteer and technical communities (V&TCs) to help make sense of a large-scale calamity and give voice to an affected population" (Harvard Humanitarian Initiaitive, 2011, p 11). Numerous articles in the popular press hailed these tools and the volunteers who worked through them as key contributors to response efforts (e.g. Mullins, 2010; New York Times, 2010a; Smith, 2010; Rosenberg, 2011; Madrigal, 2011).

After the positive press regarding its use in Haiti brought the platform to the attention of current and future digital volunteers, dozens of other Ushahidi instances were launched to support other disaster and mass disruption events in 2010 and the first half of 2011, including the Chile earthquake, the Deepwater Horizon oil spill, the Pakistan floods, several winter snowstorms in the U.S., flooding in Australia, the New Zealand earthquake, the Japan earthquake, flooding of the Red River in Canada, and political unrest and violence in Libya.

Recognizing the emergence of a new role for social media during emergency situations, the American Red Cross (ARC) conducted a survey (2010a) on how people used or intended to use social media tools during emergencies, finding that 16% of people have already used social media to get information about an emergency, nearly half say they would use social media to tell friends and family that they were okay, and more than two-thirds think emergency response organizations should be monitoring social media and responding to posts directed to their accounts. The ARC also released a report arguing for the integration of social media with formal response (2010b).

2.2.2 Political Protest

As with social media and disasters, there has been considerable discussion in the popular media concerning the role of social media during political protest. Though celebrated by western media as instrumental to the opposition protests after the Iran Election in June 2009 (Grossman, 2009), follow-up research suggests the role of social media, specifically Twitter in that case, had been over-stated. Burns & Eltham (2009) note that, for that event, the social media revolution failed to affect the change it intended, and that those tools may have been used more effectively by pro-government forces in their efforts to crush opposition protests, even putting protestors in more danger by allowing them to be more easily identified. Mungiu-Pippidi and Munteanu (2009) report that though social media helped to rally crowds during the “Twitter Revolution” in Moldova in 2009, that was not enough to change political tides in an extremely unfavorable external environment, and the revolution failed.

In October 2010, Gladwell published an essay in the New Yorker drawing attention to the overblown claims of social media relevance during the Iran Election protests and openly questioning the power of social media to affect revolutionary change (2010). That essay extended an argument by McAdam (1986) whose research examined factors that led potential volunteers to participate or withdraw from the Missisippi Freedom Summer Project of 1964, a campaign for civil rights in the American South that turned out to be extremely dangerous for volunteers. McAdam adopted Granovetter’s model of strong and weak ties (1973), noting that strong ties between volunteers were an important predictor of participation. Gladwell claimed that ties formed and maintained via social media were not strong enough to nurture participation in the kind of high-cost, high-risk activism necessary for revolution.

These claims were soon put to the test. In December 2010, Tunisia’s took to the streets of their country in a show of civil resistance to an oppressive government and unfavorable economic conditions. Acquiescing to protestors’ demands, the president of that country, Ben-Ali, stepped down and left the country on January 14, 2011. Media reported that protestors were using social media tools, including Facebook and Twitter to help organize protests (Giglio, 2011; New York Times, 2011c). On January 25, protests began in Egypt that eventually led to the ousting of President Mubarak, and again, popular media attributed a significant role in organizing efforts to social media platforms (New York Times, 2011a & 2011b). Following the relatively successful outcomes in those two countries, citizens of Syria, Yemen, Libya, and Bahrain initiated their own civil uprisings and smaller protests were held in countries across the Arab world. Though commentators would continue to question the deterministic role of social media in these events (boyd, 2011; de Vries, 2011), people were clearly using social media tools as part of their organizing efforts during political protests in 2011 (Lotan et al., 2011).

Remote audiences are also using social media to respond to political disruption. During the Iran Election protests, many Twitterers changed their profile locations to “Iran” in an effort to both show solidarity with the protesters and to prevent Twitter users who were in Iran from being identified by pro-government forces (Reinikainen, 2009; Lotan et al., 2011). Lotan et al. (2011) claim that during the 2011 revolutions in Tunisia and Egypt, Twitter served as an important source of information, coordination, and discussion for people in the Arab world and across the globe. Digital volunteers organized by the Standby Task Force used the Ushahidi platform during the 2011 political uprising in Libya to map information (much of which they found through social media posts) of violence and population displacement (Meier, 2011; Standby Task Force, 2011).

2.2.3 Seeking, Providing, and Processing Information

Social media are appropriated during mass disruption events for a variety of different informational tasks, many of which resemble the information convergence activities outlined by Fritz & Mathewson (1957). People turn to social media and other ICT to both seek and provide information during disasters (Palen & Liu, 2007; Sutton et al., 2008; Qu et al., 2009). The phenomenon of citizen-reporting is becoming an important part of the informational landscape during mass disruptions. Personal experience shared through social media can become an important source of information for others (Qu et al., 2009). Recent research shows social media being used to inquire about missing people (Vieweg et al., 2008; Palen et al., 2009; Qu et al., 2009; Qu et al., 2011) and to offer or request relief assistance (Palen & Liu, 2007; Vieweg et al., 2010; Qu et al., 2011).

Affected individuals, volunteers, participants (in the case of protests) and responders are also using social media as coordination tools. In the disaster space, Qu et al. (2009) show how social media were used to coordinate action between “netizens,” volunteer relief organizations, and officials. Study 2 in this research notes similar behavior during the aftermath of the Haiti earthquake, with digital volunteers helping to coordinate supply movements between relief organizations via Twitter (see also Sarcevic et al., 2012). Popular media accounts have suggested that social media tools, especially Twitter, have also been used to help coordinate action during political protests in the Arab world, during the Iran Election protests (Grossman, 2009), the failed Moldovan "Revolution" (Mungiu-Pippidi & Munteanu, 2009) and the Arab spring uprisings (Lohan et al., 2011, New York Times, 2011a & 2011b).

During political disruption as well as disasters, people turn to social media to look for support or offer support to others (Hagar & Haythornthwaite, 2005; Qu et al. 2009). Al-Ani et al. (2010) discuss how Iraqi citizens use blogs during a prolonged period of "extreme" disruption to engage in dialog with people outside their country, letting others know about their difficulties and garnering support for their cause. Study 4 in this research suggests other social media (i.e. Twitter) are also being used by political protestors to foster solidarity with a global audience.

Social media can also act as sites for problem solving and other information processing activities. Qu et al. (2009) report that after the Sichuan earthquake, a popular discussion forum became a place for information gathering and integration. Vieweg et al. (2008) relate how individuals came together on Facebook groups in the wake of the Virginia Tech shootings to attempt to generate lists of victims, describing this work as distributed problem solving within a collectively intelligent community. Study 2 in this work shows how self-deployed digital volunteers used Twitter after the Haiti earthquake to verify information, recommend information they thought was important or actionable, and route information to the attention of people or organizations they thought could act on it. Social media can be both a source of information and a tool for organizing that information.

2.3 Drinking from the Fire Hose – The Problem with Big Data

Social media tools are clearly amplifying, and perhaps qualitatively changing, the processing issues related to informational convergence. They enable far more voices to be heard and support entirely new information processing behaviors (i.e. citizen-reporting, remote information processing). A clear sentiment conveyed throughout the Disaster Relief 2.0 Report (Harvard Humanitarian Initiative, 2010) was that responders are being forced to "drink from the fire hose" of data, and that they are not prepared nor equipped to "produce useful knowledge from the flow of information and data" (p 11). The report explicitly claimed that newly opened communication channels for affected individuals (via social media and the Mission 4636 service[7]) as well as the activity of digital volunteer communities amplify the problem of "information overload" during emergency response.

"Information overload" is a term popularized by Alvin Toffler in his book, 'Future Shock,' first published in 1970. Toffler used the term to refer to the accelerating amount of information from print, radio and video that each person was required to process every day. According to Toffler, humans have a limited capacity to process this information and suffer a performance breakdown when they cannot keep up. Toffler was concerned by a perceived (negative) psychological toll on modern humans from the rapid and accelerating rate of information "through-put." He was afraid that humans would not be able to adapt quickly enough to processing tasks required of them. Though Toffler's fears of a psychological toll have yet to be realized, it can be argued that the humanitarian responders to the Haiti earthquake suffered a practical toll from information overload.

Examining the issues raised by the Disaster Relief 2.0 report illuminates new complexity for those concerned about information overload on the ground of disasters. Informational convergence through social media may be exacerbating the difficulties of processing all the data that is streaming into the field through an expanding number of communication channels.

> The number of messages in circulation has never been as great as it is now, but we have few instruments to filter the pertinent data, make connections on the basis of significations and needs that are still subjective, or orient ourselves within the flux of information. (Lévy, 1997, p. 9)

Considering how the technology that has enabled more communication and communication traces has outpaced technologies for processing that data, Shirky (2008b) claimed that it is not a problem of information overload, but "filter failure." In the context of disaster, we do not yet have adequate filters or other machine-based mechanisms for processing real-time, unstructured data into useful resources for responders or those affected.

[7] http://www.mission4636.org/

2.3.1 Filtering a Noisy Information Space

As discussed in the Introduction to this work, there are several pieces to the information-processing puzzle. The most obvious issue relates to the massive volume of data available and the high proportion of noise, which is itself a multi-dimensional problem. In cases of disaster monitoring and disaster response, noise can present as information wholly unrelated to the event, or information that may refer to the event but is not useful to responders. When searching social media streams, search terms like “fire” can pick up references to sports teams (“Chicago Fire”) or musicians (“Arcade Fire”) as well as information on fires that may be hundreds or thousands of miles away from the event of interest. Social media posts that are “on-topic” and refer to the event of interest may not be relevant or useful to responders or those affected. Vieweg et al. (2010) examine Twitter communication for tweets that include information that may contribute to “situational awareness” or, in other words, might be helpful to responders or those affected in forming plans for how to react to the event. They found 61% of tweets that were sent during the 2009 Red River floods by local individuals and were on-topic to the event contained what they referred to as “situational update” information, and 76% percent of on-topic tweets by similar Twitterers during the Oklahoma Grassfires in April 2009 contained that kind of actionable or otherwise useful information. Percentages for situational update information in tweets sent from people not local to the event are, most likely, considerably smaller. Additionally, posts that are both on-topic and contain situational awareness information often consist of repeated information, which can present as formal re-posts (like retweets on Twitter) or as “original” posts containing information previously published on that social media platform, on another social platform, elsewhere on the web, or on TV or radio.

2.3.2 Lost Context and the Work of Recontextualizing

A second, perhaps more vexing, problem with the data produced within social media and moved through those spaces by digital volunteers is *decontextualization*. Information posted to social media sites can lose its connection to the context of its production. As Brown and Duguid wrote (2000), “No information comes without a context, but writers and designers always face the challenge of what to leave to context, what to information" (p. 202). Constraints of social media, like the 140-character limit of tweets, may exacerbate this issue. Once data are broken down into packets, this context can be lost. Brown and Duguid were concerned about this issue, what they called the "datafication," or disaggregation of knowledge.

This problem is not specific to Twitter and disasters, though an example from this space can clearly illustrate the difficulty. When a user tweets a message on Twitter, that message is originally associated with that Twitterer's account and profile information, positioned within a stream of tweets that user has generated over time, and automatically stamped with the time of the tweet's creation. For instance, someone may post an evacuation notice for a fire in her neighborhood, saying that the Lakewood neighborhood is under a voluntary evacuation warning at this time. However, as other users retweet that message, some or all of that contextual information can fall off. The timestamp is the first to go, as retweets rarely mention the original tweet time. If the tweet is still circulating four hours later, that notice may no longer be valid, or the evacuation may now be mandatory, not voluntary. Using search features on public tweets across all accounts also contributes to lost context. Downstream Twitterers who do not follow the original account may not go back to read the author's preceding tweets or check their profile information. If several areas or several states have current fires, then someone who does not know the original tweeter may think she is talking about Lakewood, CO when the evacuation notice was for Lakewood, TX. One cause of decontextualization is the natural diffusion process that digital media and especially social media enable. Tweets, for instance, can even become disassociated with the original author as downstream Twitterers drop the creator's name to adapt the tweet to character constraints (boyd et al., 2010) or purposefully exclude the attribution and remix the text into a new tweet of their own, as we will discuss in Study 4.

Though writing in praise of the future of computer-mediated-communication (CMC), Lévy lends insight to this process of decontextualization:

> The sign no longer points toward a meaning or an object; it flows, radiates, diffuses, regenerates, and clones itself, proliferates. It is no longer a representation that has been accredited by

> transcendence, but a virus attempting to replicate itself, fighting against other viruses to occupy the media space. … Within the space of reproduction, distribution, and indefinite variation, signs no longer convoke the things they designate, nor the beings that announce them. (1997, pp. 167-168)

The other side of decontextualization is *re-contextualizing*, placing disaggregated information back into its original context or within a new context. This work of re-contextualizing data is an important aspect of how the crowd helps make sense of "big data" during disaster. Within current activities by digital volunteers, re-contextualizing can take the form of locating original sources so information can be verified, geo-locating information so it can be mapped, and compiling locations and capacities for hospitals near an impacted area into a single spreadsheet.

2.3.3 Dealing with Misinformation & Disinformation: Verifying

Veracity of social media data is also a huge issue for those trying to use that information during mass disruption events. Bad information enters the system in two ways: accidentally, as misinformation, and intentionally, as disinformation. In the social media space, the circulation and reposting of information over time can cause outdated information to appear as current, and rumors to spread quickly, often accidentally reposted by credible accounts—as was the case when multiple news agencies reported, via Twitter as well as other channels, that Gabrielle Giffords had been killed in the mass shooting at her speaking event in January 2011, when she had been injured but actually survived (Shepard, 2011).

Disinformation has been a problem in the disaster space as well. Responders to the 2010 Haiti Earthquake complained that many people sent social media messages claiming that a family member or loved one had survived but was trapped in the rubble, even though they had little reason to believe that the person was still alive and, in many cases, it appeared to responders they were merely hoping for someone to help them retrieve the body (Harvard Humanitarian Initiative, 2011).

There is evidence that the social media community can help to identify bad information. Studying the propagation of several rumors through the Twitterverse in the wake of the Chile Earthquake in 2010, Mendoza et al. (2010) found that tweets containing false information were far more likely to be challenged by other Twitterers in the space, claiming that the community works as a "collaborative filter" for verifying information.

Many of the digital volunteers that we have studied view rumor control and verification as a key activity within their volunteer work. In interviews of Twitter users who acted as digital volunteers during the early aftermath of the Haiti Earthquake—part of Study 2 in this work—we found seven of 19 participants talked about verifying or vetting information within interview responses or within their Twitter streams during that event. The Humanity Road organization, which we will describe in detail in Study 3 of the proposed research, explicitly trains its volunteers to vet information coming in through social media channels, and posts this tweet every few hours through its Twitter account:

@HumanityRoad: `Verify twice - tweet once. Rumors put lives at risk`

2.3.4 Structuring the Unstructured

The final major issue for processing a flood of data into useful information during mass disruption events is converting that data into a usable format that can be easily aggregated, categorized, searched, sorted, and merged with other existing resources. Most computational tools require standardized data, typically stored in databases or other formats (e.g. CSV, XML, RSS, KML). Informants to the Disaster 2.0 Report (Harvard Humanitarian Initiative, 2011) referred to the "unstructured" nature of social media data, suggesting that it must be structured in order to be useful. The *structuring* of information by social media users or digital volunteers would be a vital contribution to the information overload problem for disaster responders.

2.3.5 Machine-Only Computational Solutions

There are several ways to approach these issues of filtering, recontextualizing, vetting, and structuring social media data into useful informational resources during mass disruption events. However, significant to this research, we do not currently have machine-only computational solutions to these problems, and while strategies such as natural language processing and other machine learning techniques can be useful in addressing some parts of the processing problem (Vieweg at al, 2010; Corvey et al., 2010, Verma et al., 2011), issues like recontextualizing and verifying are unlikely to be completely solvable using machine-only solutions.

2.4 Crowdsourcing, Human Computation, and Collective Intelligence

One way that data are being filtered, re-contextualized, verified and structured is through the collective activity of the social media-connected crowd. The word many are using to describe this phenomenon is *crowdsourcing*, and that has become a popular and somewhat catchall term in the places where disaster response meets social media.

During the early aftermath of the 2010 Haiti earthquake, several digital volunteer groups mobilized in attempts to provide various types of assistance to those affected and those responding. The Disaster Relief 2.0 report (2011) described how some "crisis mapping" volunteers came together to create and improve publicly available maps of the affected areas, and how a group of those volunteers organized around an Ushahidi instance to collect, process, verify, and map citizen reports of damage and immediate needs[8]. In Study 2 of this research, we identified several Twitter users who "self-deployed" on that platform and attempted to help out in various capacities: by rebroadcasting what they thought was "good information," verifying information, routing information, and matching needs with offers of help. Throughout the early response and relief period, popular media and digerati praised the role of "crowdsourcing" (Biewald & Leila, 2010; Large, 2010, Mullins, 2010). The Disaster Relief 2.0 Report also used that term (a total of 28 times) to characterize the activities of digital volunteer groups during that event. Crowdsourcing efforts have continued to be credited as serving important roles in response efforts for numerous events between 2010 and the writing of this dissertation (Rosenberg, 2011; Lohr, 2011; Ross & Potts, 2011). However, the widespread use of this popular term, employed as a blanket

[8] The resulting resource is available: http://haiti.ushahidi.com/

descriptor for a variety of activities, may be obscuring the complexity of human behaviors and computational systems that support them.

Crowdsourcing, in its broadest sense, involves leveraging the capabilities of a connected crowd to complete work. The term, derived from related concepts of out-sourcing and open source, was coined by Howe (2006) who used it to describe a wide range of behaviors, tasks, and work practices that were newly enabled by Internet connectivity, including crowd brainstorming (e.g. Innoventive), crowd voting (e.g. Threadless), collectively intelligent crowd production (e.g. Wikipedia), the division of complex tasks into micro-tasks for geographically dispersed workers (e.g. Mechanical Turk), citizen science (bird counts at the Cornell Lab of Ornithology), collaborative filters (Google, Amazon, Netflix) and more (Howe, 2008). Closely related to both collective intelligence (Hiltz & Turoff, 1978; Lévy, 1997) and human computation (von Ahn, 2005; Grier, 2011), crowdsourcing can be conceptualized as an extension of open-source projects and organizing techniques to non-programmer tasks and projects. Howe provided an umbrella term to a technological turn—a turn toward collective activity mediated by the Internet and its offspring (social media). The term itself has recently become a popular one, masking much of the diversity (of activity, tasks, work practices, organization structure, etc.) that lies beneath it.

2.4.1 Outsourcing

The word crowdsourcing intentionally draws connections to related terms of *outsourcing* and *open source*. Outsourcing is a strategy used by companies to utilize work from individuals outside their employee base. Harland et al. (2005; citing Gilley and Rasheed, 2000), define outsourcing as a company going outside its internal workforce to procure something, a service or product of work, that the company either previously performed or created in-house or could have performed or created in-house.

Crowdsourcing imitates outsourcing by taking advantage of external, often remote, workforces. However, outsourced workforces are often found remote to the company utilizing them, but within another organization structure and often with co-located employees. Workers in crowdsourced projects can be dispersed across the entire globe, relying on their own resources (an Internet connection) to complete their work. Their only connections to other workers—if connected at all—are, in many cases, from within the crowdsourced project. Also, because they can involve thousands of participants in non-traditional work situations (e.g. each performing a few minutes or even seconds of work), crowdsourcing projects are not typically endeavors that could be performed in-house by an existing company, as outsourcing arrangements are.

2.4.2 Open source

Open source projects are software development efforts organized by programmers where, for many of these efforts, the end project belongs to the public domain, not to established companies or even the programmers themselves. The Linux project, established by Linus Torvald in 1992, is often cited as an ideal example of open source (Raymond, 2000; Howe, 2008). That effort, as most open source projects after, utilized the Internet to recruit programmers and organize their work.

In his essay on the effectiveness of the Linux model of software development, Raymond connects the dynamics of the open source movement to the concept of collective intelligence and the Delphi effect (2000). Contrasting the “bazaar” style of open-source development with the top-down “cathedral” building in the industry development model, he clearly favors the former approach, and argues that the crowd is a key component to debugging and the secret to why Linux worked. He also asserts that reusing code, borrowing and adapting others’ work to new problems, is a good thing. Code sharing allows one member of the crowd to leverage the intellectual work of other members for, in the case of open source projects, collective benefit.

Howe (2008) claimed that in the beginning, all code was open source. He recounted how early computer culture, constituted by the early-entrant hackers and hobbyists, favored openness and sharing among coders and called it “playful, competitive, and yet highly collaborative.” This development model, Howe believes, was disrupted by the introduction of personal computers and the proprietary software designed to be sold to owners of those systems. But the hackers and hobbyists were still tinkering beneath the surface and rose back to prominence with the Linux project, an often-cited example of the first successful open source endeavor (Howe, 2008).

Shirky (2008a) claims the success of the Linux project (in 1991) and this resurgence of open source was in part due to the Internet and the availability of a large group of connected people who could help Linus Torvald, the initiator of Linux, with his project. Shirky (2008) and Howe (2008) both insinuate that the connectivity of the Internet newly enabled or amplified the possibilities of open source projects and then crowdsourced projects of all kinds.

Ye & Kishida (2003) describe how open source communities can be examined as communities of practice (CoP) where members progress from the periphery to positions of increased skill and responsibility through legitimate peripheral participation (Lave & Wegner, 1991). Many crowdsourcing projects and communities can be viewed in this same way, as relying on a small yet dedicated core of volunteers in combination with a participation model that pulls in and trains new members.

The key difference between open source projects and crowdsourced projects is the skill set of the envisioned labor force: Open source projects primarily incorporate computer programmers (although bug finders are helpful as well), while crowdsourcing projects can take advantage of non-programmers capable of a wide range of tasks requiring human intelligence. Open sourcing can be seen as the original form of crowdsourcing, which has now recognized potential application in other domains. Open source is still a vital component of many crowdsourcing projects, including the crisis reporting Ushahidi platform, which was created and is maintained by open source coders with the goal of using people with other skills to complete other kinds of tasks in its use[9].

2.4.3 Collective Intelligence

Collective intelligence has been applied to the interpretation of a variety of phenomena in different domains. It has been offered as an explanation for seemingly intelligent behavior in large-scale interaction

[9] https://github.com/ushahidi

for both humans and non-humans (Surowiecki, 2005; Malone et al., 2009). It can refer to problems of cognition, coordination, or cooperation, and includes both conscious and instinctive—or spontaneous—collectively intelligent activity (Surowiecki, 2005). Though collectively intelligent behavior in humans can be observed offline, the connected interaction enabled by Internet technologies has opened up new ways for the phenomenon to both occur and to be observed (Surowiecki, 2005; Shirky, 2008a; Malone et al., 2009). A general view of collective intelligence, related to the Gaia concept of connected life, is that all human cognition is connected within a single system (Russell, 1983). Within this view, the rise of the Internet again acts as a catalyst for new forms of collective intelligence on a massively increased scale. In this research, I investigate collective intelligence enabled by computer-mediated communication, on-line, focusing on the ability for connected and collaborating human beings to engage in collective problem-solving activities.

Hiltz and Turoff (1978) introduced the term "collective intelligence" to the online context, deriving the concept from previous, off-line studies of the *Delphi method* in action. The Delphi method is a technique for structuring group communications to "capitalize on the strengths and minimize the weaknesses of collective problem solving" (Hiltz & Turoff, 1978, p. 45). It was designed to take advantage of the *Delphi effect*, an idea that the averaged opinion of a collection of observers is much more accurate than the opinion of a single observer (Raymond, 2000). A popularly referred to example of the Delphi effect is a contest asking participants to estimate the number of M&Ms in a jar (Surowiecki, 2005). Typically, the average of all collected estimates will be closer to the real number than any single guess. Delphi methods involve collecting opinions from multiple stakeholders during a series of "rounds," often with some discussion among participants or expert reviewers between rounds (Hiltz & Turoff, 1978; Harland et al., 2005). They can be useful when applied to problems of estimation, forecasting, and policy making (Hiltz & Turoff, 1978). Studying group interaction on early computerized conference systems, communication platforms similar to today's online forums and online chats, Hiltz & Turoff noted that the affordances of online systems provided new possibilities for structuring communication that were ideal for Delphi methods, citing, among other factors, the easy aggregation of opinions, rapid turnover between voting rounds, the ability to anonymize participation, and the potential for synchronous discussion among participants.

The first hints of machine-connected, collective intelligence—along with the computerized conference systems that Hiltz and Turoff studied—can be found in the work of Douglas Engelbart (1962, 1968, 2004). In early his work, Engelbart (1962) attempted to shift focus in the computing realm from artificial intelligence to augmenting intellect, from what machines could do to what humans could do with machines. His ideas for using machines to improve the capabilities of humans included—among many others, but of particular importance here—collaborative online workspaces and hyper-linked documents (1968). In his ideas on the potential of hypertext, Engelbart was profoundly influenced by Bush's (1945) imagined Memex machine, which proposed a knowledge system based on the associative nature of human cognition. Using the associative "trails" of hypertext, people could tap into existing knowledge in a manner more closely related to their own ways of thinking, and could take advantage of the previously enacted knowledge tree traversals of others'. In more recent work, Engelbart extends these views into a vision of connected, "collective IQ," and defines the role of computers as improving humanity's ability to solve complex problems:

> Consider a community's "Collective IQ" to represent its capability for dealing with complex, urgent problems—i.e., to understand them adequately, to unearth the best candidate solutions, to assess resources and operational capabilities and select appropriate solution commitments, to be effective in organizing and executing the selected approach, to monitor the progress and be able to adjust rapidly and appropriately to unforeseen complications, etc. (2004, p. 1).

In its recent, popular manifestation, crowdsourcing leans on a definition of collective intelligence presented by Surowiecki as "the wisdom of crowds" (2005). Surowiecki outlines three types of problems that crowds are capable of solving: cognition, coordination, and cooperation. For problems of cognition, building off an understanding of the Delphi effect, Surowiecki argues that the wisdom of crowds manifests when there is a method of aggregating disparate views from diverse members of a crowd of sufficient size. These three components—aggregation, diversity, and large number of views—are key components of many of the crowdsourcing efforts outlined by Howe (2006, 2008).

Coordination problems, which he illustrated with examples of pedestrians navigating through crowded sidewalks, are, he claims, better addressed by the self-organizing, bottom-up solutions of crowd interaction than by top-down mandates. Surowiecki borrowed Friedrich Hayek's concept of "spontaneous order" to illustrate how crowds solve coordination problems, similar to how a flock of birds navigates or an ant colony functions. Lévy (1997), however, has argued for a distinction between what he sees as (non-intelligent) instinctive "crowd" behavior (of both ants and people) and collective intelligence which, he says, implies both conscious and constructive participation. Important to the discussion here, Lévy claims that active participation in a shared knowledge space is necessary to reconstruct the meaning of decontextualized information, explaining with a techno-utopian view of web-enabled participation:

> Within the knowledge space, collective intellects reconstruct a plane of immanence of signification in which beings, signs, and things exist in a dynamic relationship of mutual participation, escaping the separation of territorial space as well as the circuits of the spectacle that characterize the commodity space (pp 168-169).

Collectively intelligent crowds actively produce and reproduce their knowledge space. Information, separated from the context of its creation, is continuously recontextualized, whether intentionally or not, through the work of the crowd.

2.4.4 Human Computation

Crowdsourcing can also be approached from a perspective of *human computation*, another emerging field, though one with roots in the pre-digital world. Human computation has been defined as "a paradigm for utilizing human processing power to solve problems that computers cannot yet solve" (von Ahn, 2005, p. 3). Grier (2011) suggests that we draw on the rich history of human computation—from the era preceding practical machine computation—as a foundation for understanding crowdsourcing. Many of the current manifestations of "crowdsourcing" in the disaster realm map well to Grier's examples of human computation throughout history, a technique often applied to solve problems that individuals alone or with assistance from available tools could not solve.

Von Ahn's work represents a return to human computation. His original approach, called "games with a purpose," was to deploy fun-to-play games that captured human processing, aggregating the actions of large numbers of game players to solve complex problems (2005, 2006). The games were designed to both entertain and to collect human responses to a uniform task, responses that could have collective meaning. One example of these efforts is the ESP game (von Ahn, 2005, 2006). The ESP game pairs remote participants together into teams of two, presents a sequence of random images, and asks players to guess labels for these images. Importantly, the two teammates are not able to communicate with each other during the game. The in-game goal is to score as many points as possible by correctly guessing the same label or labels for an image as your teammate. The goal for the human computation component of the game is to assign metadata to images to improve image search. When a certain amount of teams choose the same words to describe an image, these words are judged by the system to be good keywords to add to the metadata of the image.

The concept for "human algorithm games" (von Ahn, 2005) built upon ideas presented in the Open Mind Initiative. The Open Mind Initiative was conceived as an umbrella project, encompassing a multitude of human computation systems (Stork, 1999), but later came to concentrate on collecting "common sense" information for Open Mind Common Sense[10] (Singh et al. 2002). In that project users were volunteers, explicitly feeding information into a machine learning system. Aggregated judgments from these volunteers were used to train a machine learning classifier. The creators implemented a layered system where, on one level, "e-citizens" used their Internet connection to contribute data, and on another domain experts act as monitors. Open Mind derived some of its rationale from Open source movements with a key distinction: Open Mind included non-programmers. Designers conceived of a diverse incentive model for volunteers, relying heavily on participation driven by motivation to contribute

[10] http://www.openmind.org

to the greater good, but also suggesting designs that provided public recognition of volunteers and possible financial rewards in the forms of discounts and lotteries (Stork, 1999).

A stated goal of the Open Mind system and several of von Ahn's early human computation systems was to feed aggregated information from human-performed tasks into machine learning algorithms. Von Ahn and others eventually migrated away from solutions focused purely on using humans to train computational algorithms and began to use the collective work of humans to solve problems directly (as in the ESP game). Human computation systems are now used for a wide range of applications, including von Ahn's reCAPTCHA project, which uses word recognition tasks, required by some Internet sign-up portals to assert humanness, to do text translation (2008), and Amazon's popular Mechanical Turk platform which uses financial incentives to recruit crowd workers to tasks provided by other users. An example of the application of human computation in the crisis realm is recruiting and utilizing remote crowd workers to translate SMS messages from affected individuals to the language of responders, which occurred in the aftermath of the 2010 Haiti Earthquake in conjunction with Mission 4636 and the Ushahidi crowdmap (Harvard Humanitarian Initiative, 2011).

Though von Ahn's dissertation on human computation (2005) called attention to a new type of application domain in a connected world, the use of aggregated human computation to solve problems is older than machine computation itself (Grier, 2005; Grier, 2011; Quinn & Bederson, 2011a). Quinn and Bederson note that human computation and machine computation are inter-related and have been since the birth of the latter, citing Alan Turing's statement:

> The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer (Turing, 1950).

Grier writes that human computation, when defined as aggregating a large number of computations carried out by humans, is older than machine computation (2005), and suggests that the new fields of human computation and crowdsourcing look back on earlier human computation efforts for a theoretical foundation (2011). In his view, human computation began in the 18th century with the invention of calculus and the Industrial Revolution's introduction of division of labor. Early human computing laboratories were set up to solve computational problems in situations where there were no machines to do it, or where the machines were too rudimentary to do it alone. Human computation was used, for example, to make astronomical calculations for scientists and navigators, and ballistic trajectory calculations for military logistics. "Planners" organized computing problems into discrete tasks and "computers" worked at these tasks, often without understanding or even knowing the larger purpose of the work. Grier (2005) notes that Charles Babbage used his observations of the Royal Nautical Almanac, which employed human computation strategies, to inform the design of his computing machines. As machine computation evolved, there continued to be considerable interaction between human computation and machine computation. In another interesting example from Grier's work is perhaps the first attempt at citizen-science or citizens-as-sensors, when Joseph Henry of the Smithsonian Institute recruited volunteers to submit weather observations, first by mail and then by telegraph.

2.4.5 Back to the Future

So open source is not really new, collective intelligence is something we have recognized for awhile, and human computation is older than machine computation. What has changed? As Shirky writes, "forming groups has gotten a lot easier" (2008a, p. 18). Noting that many of these behaviors are normal human capacities, he goes on to explain, "We now have communications tools that are flexible enough to match our social capabilities, and we are witnessing the rise of new ways of coordinating action that take advantage of that change" (p. 20).

Surowiecki notes in the original text of *Wisdom of Crowds* (2004) that aggregation is the key component to collective wisdom. In the afterward of a later edition (2005), he remarks on the increased attention paid to this phenomenon in recent years and theorizes that the Internet, because it makes aggregating diverse views much, much easier, is pushing this effect into the collective conscious. It is now relatively easy to collect and compare large numbers of diverse opinions, in a multitude of different ways, and with volume that would have been nearly impossible or prohibitively time and resource-heavy to have done in the pre-Internet era. Shirky (2008a) and Howe (2008) also both insist that the affordances of the Internet and emerging dynamics of social media enable collective intelligence to manifest in ways never before possible, and make understanding this phenomenon of human behavior vastly more important. In some ways, these new views of the role of machines in collective human cognition represent a return to much earlier visions of the future of computing, outlined by Engelbart (1962, 1968) and Hiltz and Turoff (1978).

2.4.6 Overview of Different Configurations of Crowdsourcing

Currently, there are many types of web-applications, platforms and projects that fall under the "crowdsourcing" umbrella. There are also a variety of different approaches to characterizing crowd work, and significant overlap between many of them. This section provides an overview of some of the different ways that the crowd completes work, positioning them within frameworks for crowdsourcing, human computation, and collective intelligence, and highlighting features that are salient in this research.

2.4.6.1 Microtasking or "Turk-sourcing"

Amazon's Mechanical Turk platform is an often-cited example of a prototypical crowdsourcing system and in some ways has become synonymous with the crowdsourcing term. Mechanical Turk is a web-based platform that supports a microwork market, allowing project owners to distribute work across a large number of remote workers as small, paid tasks (Kittur et al., 2008). These tasks, called "human intelligence tasks" or "HITs," can normally be completed in a small amount of time (seconds to minutes), and earning a few cents to few dollars each depending upon the type and duration of the work (Ross et al., 2010).

Mechanical Turk (mTurk) supports a type of crowdsourcing that is sometimes referred to as "distributed human intelligence tasking" (Brabham, 2010) or "microwork" (Fort et al., 2011), though regarding the latter term, this research suggests that mTurk represents a specific type of microwork: *microtasking*. Microtasking platforms distribute tasks that require human cognitive abilities across a large number of people via Internet or mobile connections. Von Ahn's "Games with a Purpose," which use entertainment, rather than financial gain (or love or glory), as incentive for task completion (2005, 2006), can also be classified as microwork.

Microtasking projects utilize a top-down task-assignment strategy. The initiator of the project either designs a system specifically to collect the human intelligence actions desired, as is the case in Games with a Purpose, or defines a series of HITs for completion within an existing platform, like mTurk or Crowdflower[11]. In both cases, there is a leader who generates the project and workers who complete the tasks. This type of crowd work clearly expresses crowdsourcing's roots in outsourcing, and also maps closely to von Ahn's definition of human computation.

2.4.6.2 Collaborative Creation/Editing

Another type of crowd work is collaborative creation, which can be described as facilitating the creation or editing of a shared document. Wikipedia is a good example of this type of collective activity, one that Howe included in his original description of crowdsourcing (2006, 2008). Users of Wikipedia work together to create an Encyclopedia-like resource, proposing new topics and continuously editing existing topics. Wikipedia integrates several different types of collective activity into the creation and editing process. Malone et al. (2009) mapped Wikipedia within their genome classification model as a multi-dimensional system incorporating five different combinations of their collective intelligence building blocks, e.g. collaborative creation for generating new text, consensus decision making for deciding whether to keep the current version, etc.

Wikipedia is a specialized form of the *wiki*, a collaborative creation environment that can be easily deployed to support a range of purposes, where remote users can add and edit content (Leuf & Cunningham, 2001). Collaborative maps, increasingly important in the crisis informatics space, are extensions of wikis that allows users to add and edit features, markers, and other data to shared maps. Wikis, collaborative maps, and shared documents of other kinds are all types of collaborative creation.

[11] http://crowdflower.com/

Collaborative creation projects can also be considered as non-programming equivalents to open source movements. Members of the crowd or community put forth a project or topic that they would like others to work on, but the project then becomes a shared enterprise and the end product is not owned by any one member but collectively "owned" by the community or freely provided to the public domain. Though many wikis, including Wikipedia, assign editorial privileges to some users to control content, coordination of work within a collaborative creation environment may be self-organizing and is typically less hierarchical than microwork projects.

Soylent is a system that allows a document owner to incorporate crowd workers as collaborative editors, to check for errors, shorten paragraphs, changes verb tenses and search for citations (Bernstein at al., 2010). Illustrating an overlap between microwork and collaborative editing, that system uses paid workers incorporated into the platform through Mechanical Turk, and divides the work into microtasks.

2.4.6.3 Citizens as Sensors – Citizen Reporting

> The observers recorded their data in a coded form and telegraphed their results to Washington. Data moved toward the capital as if it were water coursing through the tributaries of a river. It flowed from farms and villages, joining other data en route and pausing at regional offices to wait for an open moment on the lines that stretched toward Washington. (Grier, 2005)

Another form of crowd work solicits information from the crowd, using humans as remote data collectors. This technique has long been employed within scientific communities to gather data. In the mid-19th century, intending to track the movement of storms in the U.S., Joseph Henry of the Smithsonian Institution recruited hundreds of volunteers to gather weather data from geographically dispersed areas and send those observations back to the Smithsonian via mail, initially, and then over telegraph wires (Grier, 2005). This was an early entrant in the crowdsourcing genre of *citizen science*. Howe (2008) described several citizen science projects that incorporated non-experts into scientific endeavors, including eBird[12], an effort launched by the Cornell Lab of Ornithology that aggregates reports of bird

---

[12] http://ebird.org/content/ebird/about

sightings so scientists can track migration habits, measuring, among other things, the impact of environmental change.

Citizen science projects are just one application among many for crowd-generated data collection. In the early 2000s, the U.S. Army adopted a motto that declared, “every soldier is a sensor” (Association of the U.S. Army, 2004), attempting to achieve “informational dominance” by taking advantage of new mobile technology and revising its view of information flow to include bottom-up data collection from soldiers in the field. That move reflected a new reality for emergency situations of all kinds: that individuals on the ground, armed with a variety of Internet-capable and often mobile devices, can and do act as citizen sensors (Goodchild, 2007; Goodchild & Glennon, 2010).

Task-assignment for citizen sensors can be both top-down and lateral in nature. The U.S. Army’s use of its soldiers to collect battlefield information and citizen science efforts to collect weather or bird info are examples of top-down crowd-collection efforts with people purposefully providing or uploading data. Data collection within these projects is a form of microwork, distributing the task of collection across a remote workforce. The crowd can also act incidentally as sensors by sharing observations and whereabouts on social media sites. Shared observations can move laterally within social networks, and can also be aggregated into a common source. This type of lateral information sharing, enabled by social media tools, is a form of citizen journalism (Gillmor, 2004; Allan & Thorsen, 2009).

Remote data reporting—whether by “citizens,” trained volunteers, or soldiers—fits multiple definitions of crowd work. Howe included several examples of citizen reporting in his definition of crowdsourcing (2008). At the same time, the data is both a product of collectively intelligent activity and a resource with the potential to contribute to collective intelligence, and the activity of identifying information to report and communicating that information falls within the definition of human computation.

2.4.6.4 Crowd-brainstorming

Howe's (2008) definition of crowdsourcing also included crowd-brainstorming activities, where the crowd is queried for ideas or solutions to a problem. Howe called this "collective intelligence in action" and used examples of both InnoCentive, a company that queries the remote crowd looking for novel answers to scientific problems and Threadless, a company that solicits t-shirt designs from the crowd and then has the crowd vote on the best designs. Both of these crowdsourcing efforts rely somewhat on the Delphi Effect, using a large, diverse crowd to find the best solution or solutions, with the latter example most closely mirroring the techniques of the Delphi Method (Hiltz & Turoff, 1978), which prescribes gathering info and ideas from a diverse group and having members vote on the best option.

Brabham (2010) differentiates between the "broadcast search approach" where a solicitor is looking for a "lone gunman" to solve a problem by throwing as large a net as possible (i.e. InnoCentive), and the "peer vetted creative production approach" where the crowd is used twice, first as the source for a multitude of ideas and second as a filter for finding the best option from among suggested ideas (i.e. Threadless). This latter approach is a form of *collaborative filtering*, a term used to describe a wide range of human-powered information filtering systems.

2.4.7 Collaborative Filters

Collaborative filtering was initially defined as people working together to perform filtering operations by annotating documents or adding metadata to documents (Goldberg et al. 1992; Resnick et al., 1994). The first collaborative filters were designed for filtering and recommendation within email systems and newsgroups (Goldberg et al. 1992; Resnick et al., 1994). They are now in use across a range of Internet technology, including Google's popular search engine, Amazon's product recommendation system, the DIGG social network, and the iTunes recommendation system, and the term now includes filtering on metadata that is not annotated or explicitly added to a document, but inferred from the behavior of participants within an information space. Malone et al. (2009) define collaborative filtering as leveraging the individual decisions of the crowd within a *social network*, where decisions are inferred from relationships between crowd members. In Starbird and Palen (2010), we describe how Twitterers use the retweet convention as a recommendation mechanism, a form of collaborative filtering.

Many collaborative filtering mechanisms are powered by *tagging*, having the crowd add metadata to digital artifacts (e.g. images, web pages, academic articles, etc.). Metadata are information, often in the form of keywords, short phrases or ratings, which give some insight into the content of the document. There are three approaches to adding metadata to documents: professionally created, author-generated, and user-generated (Mathes, 2004). Social media and other web tools support author- and user-generated metadata creation in the form of tagging. User-generated tags organize social media content within a ground-up classification system, or *folksonomy* (Mathes, 2004).

One benefit of tagging is that it makes information searchable. It allows Facebook to connect profile owners to photos posted of them, and, when a user navigates to his sister's profile page, to pull up all of the pictures of the two them together. When a researcher is looking for academic articles on collaborative filtering, author provided keywords (tags) attached to the documents as metadata enable Google Scholar to search through thousands of journal and conference papers and return only the ones in which she ia interested. Tagging is especially valuable for image search, where the original document has no textual information.

The "peer vetted creative production approach" to crowdsourcing (Brabham, 2010) incorporates purposeful tagging or recommendation into the filtering process, e.g. voting for the best t-shirt on the Threadless platform (Howe, 2008; Brabham, 2010). However, recommendation need not be explicit. Collaborative filtering can be both *active,* i.e. requiring explicit participation by users, and *passive*, i.e. leveraging normal behavior for recommendation or filtering, often without the users' knowledge (Kruk & Decker, 2005). Howe (2008) categorizes DIGG as an *active* recommendation system, where users rate online documents for the benefit of other users. Goldberg et al. (1992) note that implicit feedback, information on user behavior gleaned from normal use of the system, can also be used as feedback for collaborative filters. Considered from a framework of collective intelligence, Lévy writes, "Together, [users] organize the space, define, evaluate, color, heat, or cool it. Each one helps build and order a space of shared signification by diving in, swimming around, and simply living in it" (1997, p. 220). Howe (2008) characterizes systems that leverage activity in this manner as *passive* filters.

Active recommendation systems, like Threadless and Digg, can be viewed as microwork systems where the "task" is rating something or choosing the best option, though the product of the work is of collective interest to the workers themselves. Passive collaborative filter systems are different. Through these systems, the project owner gains value from the crowd activity, but the tasks are unnamed and the incentives neither financial nor philanthropic nor of shared interest. Instead, the incentive relies on benefits from mere participation in the information space. The crowd acts organically to organize the information within the space and a computational system attempts to extract some feature or features of that organization for application.

We have previously explored the idea that, during crises, users' "natural" behavior within an information space can help to organize it (Starbird et al. 2010; Starbird & Palen, 2010). Indeed, the ways that information moves, who it moves through, and how, can offer significant insight into its meaning.

> Rather, it is in these steps—from sources to reporters to editors and news organizations—that news is made. Without them, again, there would be no story. Nonetheless, when information takes center stage and lights dim on the periphery, it's easy to forget the necessary intermediaries. But while they may be invisible, they are not inconsequential. (Brown & Duguid, 2000, p. 6).

This is true of information circulating through both traditional media and social media. Analyzing the Twitter platform, Kwak et al. (2010) note that the speed with which information diffuses could be an indication of strength of influence of that information's author, and suggest that the number of followers a user has and the depth to which that user's messages diffuse are different measures of popularity. Suh et al. (2010) claim that "social context"—information about the Twitter author including followers, following, account inception date, etc.—can help identify the "value" of information. Study 4 (Chapter 7) of the proposed research explores the possibilities of using the "social context" of Twitter data, including how it propagates through a system, as information about that information. One consistent goal of this research, related to future application of its findings, is to identify possible methods of using both active and passive collaborative filters to generate useful informational resources from social media data.

2.4.8 Hybrid Human-Machine Computation

Quinn and Bederson (2011b) have proposed "hybrid human-machine computation" as an extension of the emerging field of human computation. Those authors are pursuing the development of systems that "tightly integrate human computation and machine resources" (2011b, p. 1), allowing humans to benefit from machine resources and machines to benefit from human capabilities. In the case of their proposed "Crowdflow" system, human work is both used directly to complete human-computation tasks and as feedback for machine learning classifiers training to complete that work in the future.

The Swiftriver system, an offshoot of the Ushahidi project, is an example of a hybrid human-machine computation system developed for use both within and beyond the crisis domain. That platform enables real-time filtering and verification of data streaming in from multiple channels, using semantic analysis and automatic classification techniques to feed human crowd workers—who then further filter, process, and verify the information. Creators describe the idea as "crowdsourcing the filter" (Hersman, 2009).

2.5 The Tools, Platforms, Communities and Organizations of Digital Volunteers

The premise for this dissertation stems from an understanding that solutions based purely on the computational power of machines will fall short of addressing the information processing problems of informational convergence, and this research attempts to shift attention to the power of distributed human computation as an alternative strategy. Considering what we know about improvisation during disaster response (Dynes, 1970; Drabek, 1986; Tierney et al., 2001; Kendra & Wachtendorf, 2003a), people working to solve these problems now are already leveraging connective technology and human-computation power to process information during mass disruption events. In some cases, self-deployed volunteers appropriate available ICT to attempt to contribute to response efforts. In others, technical volunteers have built systems to support the work of information processing during crises.

Through extensive research activity as a participant-observer during mass disruption events in 2010-2012, I have both studied and been involved in several efforts that leverage the crowd to do information processing work. This section offers an overview of several organizations and other crowd work configurations in this emerging ecosystem of virtual responders, and describes some of the ICT tools and platforms that these volunteers appropriate for their work.

2.5.1 The Social Media Resources of Digital Volunteers

Though this work considers the larger ecosystem of social media interaction and data production, the studies proposed here focus on a handful of specific social media platforms, including Twitter, Skype, collaborative Google Documents, and several platforms designed for citizen-reporting and collaborative mapping. These are platforms my research shows to have been widely adopted within the context of mass disruption. This section offers a brief overview of several of these platforms and describes how each factors in to the studies presented in this dissertation.

2.5.1.1 Twitter

Twitter is a popular microblogging platform that allows its users to broadcast short, 140-character messages. Twitter users (Twitterers) can elect to "follow" other users, which means that they receive those accounts' messages in a "friends" stream. Because connections between Twitterers are not necessarily reciprocal—Kwak et al. (2010) report that only 22.1% of following links between Twitterers are two-way—each user has separate lists of "friends" (those she has chosen to follow) and "followerers" (those who have chosen to follow her). Twitterers also maintain a profile, which consists of an account name, an image to represent their account, a name field (that may or may not differ from the account name), a short text field to describe themselves, and another text field for a self-determined location.

Twitter can be accessed through web portals and mobile devices. The platform provides a web application, a mobile application, and an SMS-to-tweet feature to its users. Third party developers have created a variety of other client applications that run on personal computers and mobile devices and incorporate other functionality, including tweet search and filtering.

One Twitter feature that holds promise in the context of citizen reporting of mass disruption is the ability to geolocate tweets. By enabling their profile and their mobile device to provide GPS coordinates or place names, Twitter users can also embed geographical information in their tweets. This data does not appear in the text content, but is attached as metadata to the tweet. Though the potential for geographical information on Twitter is high, the default setting for geo-enabling tweets is off, and currently less than one percent of tweets contain geolocation metadata (Hecht et al., 2011). The self-selected location field associated with the user profile can also be used to provide geolocation information. Approximately two-thirds of those fields contain valid (though not necessarily accurate) geographical information (Hecht et al., 2011).

Twitterers can also embed photos and other external links (URLs) in their tweets. Both of these content types appear as links within the text of the tweet and are now contained in the tweet metadata. A number of URL shortening services have come online in recent years—perhaps a reaction to the popularity and character restrictions of Twitter.

The feature that renders Twitter most useful during times of disruption is the public availability of tweets. Twitter users can choose to make their account either public or private, but the default is public and most Twitterers maintain that setting. For public profiles, all tweets are posted to a public timeline that is available for search using several application programming interfaces (APIs) provided by Twitter. This means that Twitters are not only sending messages to their followers, but are potentially broadcasting to a much larger group of listeners. Third-party developers have produced several search tools that can be accessed for free through the web, giving Twitter users and non-users alike the power to search through tweets by keywords or geographical location, and through Twitterers by name or account name. This research takes advantage of in-house software developed by Project EPIC at the University of Colorado Boulder that uses the Twitter Search, Streaming, and REST APIs to collect data (Anderson & Schram, 2011).

During the short history of Twitter, users have introduced several linguistic conventions to expand the original functionality of the platform, and these conventions have spread broadly through the Twitterverse. Honeycutt and Herring (2009) describe how the "@" symbol came to be used as a marker of addressivity, to direct a tweet to a specific Twitterer. Most Twitterers address tweets or mention other users by writing `@username` and Twitter and most client applications now explicitly support this reply/mention functionality. The *retweet* convention evolved from the @ innovation. Retweets are forwarded messages that give attribution to the original (or other upstream) authors (boyd et al., 2010). Retweets follow several different conventions; the most popular currently are `RT @`, `via @`, and `"@username: <original text>"`. Twitter began to support the retweet as a platform feature in November 2009. There now exist two types of retweets, those generated automatically through retweet buttons on Twitter or third-party applications, and those generated manually by copying and pasting and adding the attribution(s). In the Spring of 2010, approximately 11% of all tweets were retweets, and this percentage was much higher for tweets that contained hashtags (another user-driven convention, explained below) and URLs (Suh et al., 2010). For tweets that contain hashtags related to mass disruption events, retweet rates appear to be even higher; more than 50% of tweets tagged with one of the popular terms related to the Egypt protests in 2010 were retweets (Starbird & Palen, 2012). Hashtags (`#<keyword>`) are terms that users add to their tweets to tell others that they are writing about a specific topic. It is important to note that the hashtag began to diffuse widely when it was promoted and subsequently adopted for use during the San Diego wildfires in 2007 (Messina, 2007b). The convention has proven useful for searching techniques and can be extremely valuable during crisis events.

The Twitter platform serves multiple roles in this study. As described in section 2.2, we have witnessed widespread use of Twitter during mass disruption events by those affected, emergency response and relief organizations, and digital volunteers. Within this research, Twitter is a source of citizen-reported information as well as a mechanism of information propagation. It is also a site where volunteers groups self-organize and actively help to process information.

2.5.1.2 Skype Chat

The Skype platform offers a range of functionality, but for digital volunteer communities and organizations the Skype chat feature is the perhaps the most important tool. Many of the groups examined here use public and semi-public Skype chats to communicate with each other and with members of other volunteer communities. Skype chats enable synchronous textual chatting in real-time. They can be private, two-person conversations or multi-person discussions with an upper limit of 300 participants at one time. Each participant can contribute to the conversation by typing in text and pressing <return>, which sends the text to appear, preceded by its author's name and accompanied by a time stamp, in a shared, temporally ordered view of messages. Chats can be as short as a single message or can continue interminably. Skype maintains chat histories that can be recalled by any current participant back to the moment that she first entered a conversation.

Chats are rarely fully public. Chat administrators maintain the participant list and can invite or eject people from conversations. In the digital volunteer communities studied here, many chats are semi-public and administrators generally accept the requests of those who wish to take part. However, other chats are reserved for members of specific groups or volunteers that have been vetted by the group and/or trained for specific types of volunteer work.

Some volunteer communities use the voice feature of Skype to hold conference calls. The video chat feature is rarely used in this context, except to accompany screen-casts during demonstrations of new technologies.

2.5.1.3 Google Documents

Google provides a suite of tools that allow users to share and collaboratively edit different types of documents. These documents, called Google Docs, are widely used among the digital volunteer groups and are the sites of substantial information processing work. The most popular document form is the spreadsheet, which volunteers use to create schedules, maintain sign-up sheets, and collaboratively generate information resources like lists of available hospitals that include current capacities, contact information, and GPS coordinates. Google Docs allow multiple users to access documents. Administrators configure privacy settings that determine who is allowed to view and edit each document. For documents with multiple editors, users can edit simultaneously and see others' edits in near real-time.

2.5.2 Platforms for Citizen Reporting & Crisis Mapping

Crisis mapping has emerged as a popular new genre for volunteer activity during crises and mass disruption events. Generally, crisis maps are maps of impacted areas that users collectively create and edit (collaborative creation and editing). Volunteers, both local and remote, work to geo-locate pieces of information and put them onto a shared map that can be accessed by others online or through mobile devices. Crisis maps can be important after an event, an earthquake for instance, where landmarks have changed and/or locations of things such as shelters or roadblocks are in flux. Goodchild (2007) describes an explosion of participation in collaborative map-building web resources via volunteered geographic information (VGI), relating this to the idea of using citizens as sensors. Later research refers to these efforts as "crowdsourcing" and claims an important emerging role for them during crises (Goodchild & Glennon, 2010).

Volunteers for crisis mapping projects can come from a local community in response to a specific event, or from a growing pool of individuals who identify as "crisis mappers" and repeatedly participate across events, most often from outside the affected area. Some members of the latter group have formed an ongoing community, *CrisisMappers*[13], which describes itself as a network of volunteers, academics, and technology groups who work to create maps during humanitarian emergencies (Harvard Humanitarian Initiative, 2011). Other mappers from the OpenStreetMap community have created a subgroup, Humanitarian OpenStreetMap Team[14] (HOT) that responds to humanitarian crises with their specialized mapping skills and online platform. Multiple platforms and communities have arisen to support different types of crisis mapping efforts, both reflecting and establishing different organizational structures.

2.5.2.1 Google MyMaps

Google MyMaps are shared maps or "map mashups" that users can update with any number of features, marking them with graphic icons accompanied by textual explanations. A Google MyMap can be started by anyone in the crowd, and there is a low barrier for entry, as no specific skills are needed for participation, just some knowledge of the surrounding area or evolving conditions on the ground. For these reasons, Google MyMaps are often started by and used by locals during a crisis event. During the Fourmile Canyon Fire in Boulder, CO in 2010, two Google MyMaps were created by locals to document flames spotted, road closures, fire lines, donation drop-off locations and destroyed structures. The most popular received over a million views and had multiple "collaborators," or different users contributing to the map[15].

2.5.2.2 OpenStreetMap

In the wake of the Haiti earthquake in January 2010, an existing community of mapping volunteers, OpenStreetMap, utilized their open source technology to create a collaborative map for Haiti[16] that became a valuable source of information for relief efforts. OpenStreetMap is an international effort to

---

[13] http://crisismappers.net/

[14] http://wiki.openstreetmap.org/wiki/Humanitarian_OSM_Team

[15] A screenshot of an early state of that map is here: http://andrewhy.de/four-mile-canyon-fire-2010-boulderfire/

[16] http://haiti.openstreetmap.nl/

create and maintain free and publicly editable maps. With a name that reflects origins in the open source ideal, the mapping interface builds also on the wiki concept, allowing collaborative creation by a technologically-savvy community. Unlike the other mapping efforts discussed here, OpenStreetMap efforts for the Haiti earthquake and other crises are more specifically focused on geographic information than mapping changing humanitarian conditions or citizen reports. OpenStreetMap requires a more advanced skill set for participation than Google MyMaps and volunteers must be trained in specific Geographic Information System (GIS) techniques in order to participate.[17]

2.5.2.3 Ushahidi

Another example of crisis mapping is the Ushahidi platform. Ushahidi, which means "testimony" in Swahili, emerged from efforts by bloggers to support citizen journalists during a period of violence after a contested election in Kenya in December, 2007 (Okolloh, 2009). The platform was initially created to allow workers to assemble citizen reports of ethnic violence, arriving via SMS and the web, and to filter, verify, and map those reports. In its first deployment, Ushahidi represented an extremely rapid self-organizing effort by volunteers to connect via social media (initially Okolloh's blog), create a tool to support their work (Ushahidi), and develop complex work processes to maintain their ad hoc group (Okolloh, 2009).

In the early aftermath of the Haiti earthquake, an instance of the Ushahidi platform[18] was deployed, in conjunction with an SMS shortcode effort, to collect reports from affected people. Volunteers were recruited, initially through Tufts University, to identify actionable information, translate that information when necessary into English or French, geolocate that information, and structure it into a report stored within the system. The Ushahidi mapping effort continued for months, eventually transferring the responsibility for processing the reports from Tufts University students to Haitian workers coordinated through Crowdflower and Samasource (Munro, 2010).

---

[17] Personal communication with various group members, September 10, 2011.

[18] http://haiti.ushahidi.com/

After the Haiti earthquake and throughout the remainder of 2010 and during 2011, Ushahidi instances were deployed for dozens of crisis and mass disruption events, large and small, including other earthquakes in Christchurch, NZ and Japan, snowstorms in the Midwest U.S., flooding in Manitoba, CA, wildfires in Oklahoma, and Hurricane Irene's impact on the U.S. There was also a "formal" deployment connected the Standby Task Force and UN-OCHA that documented violence and the movement of people during political unrest in Libya in the Spring of 2011[19], with the goal of supporting humanitarian operations (Standby Task Force, 2011).

Ushahidi instances combine citizen reporting with collaborative creation and editing, and map deployers have experimented with using microwork to power Ushahidi instances as well (Meier, 2010b).

Though none of the studies in this dissertation focus specifically on Ushahidi, many of the crowd work communities described here, including the voluntweeters of Study 2 (Chapter 5) and Humanity Road of Study 3 (Chapter 6), have incorporated Ushahidi crowdmap support into their work practices.

2.5.3 Virtual Communities and Organizations

During mass disruption events, remote volunteers are appropriating these ICT tools and platforms to contribute to response efforts, often connecting to other volunteers and volunteer communities in their efforts. We are currently witnessing a rise of virtual volunteer organizations in the domain of disaster and humanitarian response.

Researchers associated with the Disaster Research Center at the University of Delaware developed a model of organizational types that present in the aftermath of a disaster event (Dynes, 1970). They describe organizational forms along two dimensions, one of structure (old or new) and the other of tasks (old or new). *Established* organizations have a pre-existing structure and execute pre-defined tasks. Hospitals and fire/police services are examples of established organizations. *Extending* organizations are those that have a pre-existing structure, but extend their functions to take on new tasks during a disaster

[19] http://libyacrisismap.net/

response. A restaurant that begins to serve food to victims and responders and a church that opens its doors to act as a shelter are examples of extending organizations. Organizations, like the American Red Cross or the Salvation Army, that adopt new organizational structures (e.g. take on new volunteers) to take on pre-defined tasks, are *expanding*. Finally, *emergent* organizations have no pre-existing structure and no pre-defined tasks. Emergent organizations develop when disaster convergers begin to work together to improvise solutions in the aftermath of disasters.

Peer-to-peer communication networks, including those provided by social media platforms, create new opportunities for emergent organizations (Palen & Liu, 2007), and several networks, communities and organizations of digital volunteers have arisen over the past few years. Some of these groups focus their activity around a single tool or platform, e.g. the Ushahidi crisismapping platform. Others incorporate a range of tools in a general mission of responding to disaster events. Within this latter category, there are groups who have formed to provide assistance for a single event, virtual groups created to respond to events in a specific geographical region, and organizations set on responding to events all over the world. Many of these organizations interact and collaborate with each other on a regular basis, especially during large impact events. Additionally, some volunteers shift affiliations or maintain multiple affiliations at the same time—participating simultaneously in two or more communities, sometimes during the same event.

As background for this emerging ecosystem of virtual responders, this section offers a brief description of some of these communities and organizations, including two groups that will be the subjects of some of the studies in this research.

2.5.3.1 Voluntweeters

Study 2 (Chapter 5) will focus on a group of volunteers who co-opted Twitter as a crisis communication channel to help in the Haiti earthquake response by finding actionable information, verifying it, and attempting to route it to responders. As that event progressed, many of these self-deployed volunteers began to connect and coordinate their efforts with other Twitterers who were tweeting for Haiti, eventually forming an interactive network of Twitter volunteers, or “voluntweeters” as some called themselves. Though this emergent response organization was somewhat temporary, as many voluntweeters pulled away from the network and their volunteer work after a few weeks or months, some of these digital volunteers have continued or returned to their crisis tweeting activities periodically, responding to other crisis events.

2.5.3.2 Humanity Road

Interviews with several volunteer Twitterers (for Study 2, Chapter 5) revealed that many continued to tweet for subsequent crisis events after the Haiti earthquake. A few joined up with Humanity Road[20], a virtual volunteer organization that provides informational aid during disasters. Humanity Road is the focus of Study 3 (Chapter 6) in this work. The group has roots in volunteer response to Hurricane Katrina, and came together as a formal organization during the Haiti recovery period. Many of the initial members participated in voluntweeting during the Iran Election protests in June 2009, and almost all had tweeted for Haiti. They work by appropriating available tools, including many forms of social media, to identify and distribute relevant information before, during, and after crisis events. Their work includes monitoring and filtering media and social media reports, verifying the information they find there, and integrating it into existing resources or creating new resources for affected populations and responders. The organization is almost exclusively digital, using, among other ICT tools, Twitter to recruit volunteers and distribute information, Skype and shared Google documents to coordinate volunteer efforts, and a website to display resources. During disaster events, Humanity Road accepts spontaneous volunteers and quickly trains them and incorporates them into their activities and tasks.

2.5.3.3 Standby Task Force

The Standby Task Force (SBTF) is a network of volunteers, originally organized during the 2010 International Conference on Crisis Mapping, who collaborate to provide live mapping support to

[20] http://www.humanityroad.org/

organizations involved in humanitarian response, human rights and election monitoring, and media coverage[21]. The group, which has an ongoing relationship with UNOCHA, can be "activated" by organizations needing support during mass disruption and/or violent events. During 2011, events spawning activations have included natural disasters like the earthquake in Christchurch, NZ and political disruptions like the world witnessed in Libya.

The work of the SBTF centers on the Ushahidi platform, and involves eight different teams who work together to complete four main tasks: information collection, visualization, analysis and response. While a core group of team coordinators and volunteers move from event to event, new, spontaneous volunteers are recruited during events, then quickly trained and incorporated into the various teams. The SBTF uses Skype chats to coordinate their activities in real-time, and shared Google documents to schedule volunteer shifts. During 2010 and early 2011, Humanity Road volunteers acted as informal leaders, or "coordinators" of the "media monitoring" work-group within the Standby Task Force during "activations" for several events. Many Humanity Road volunteers also worked as volunteers for the Standby Task Force during that time.

2.5.3.4 Virtual Operational Support Teams (VOSTs)

Jeff Phillips, Emergency Management Coordinator for Los Ranchos de Albuquerque, New Mexico, has proposed a method of connecting formal responders with trusted digital volunteers for remote support during events using social media: Virtual Operation Support Teams (VOSTs). Phillips first employed the term to describe and recruit volunteers to participate in an exercise intended to explore the possibility of using virtual volunteers in concert with formal responders.[22]

> This is an exercise of the concept of organizing 'trusted agents' into recognized operational structures using principles and practices of ICS to perform necessary functions in support of emergency response & recovery operations. … I have been developing 'trusted agents' and this

[21] http://blog.standbytaskforce.com/about/introducing-the-standby-task-force/

[22] Personal communication and https://docs.google.com/document/d/1cj5Ta_Ar D4ji7ZngZiirhuw_q46J3_vsNnOY_k0gg_c/edit?hl=en_US

> VOSG/T concept with the understanding that if/when we have a major emergency in my area I will NEED assistance in the social media realm. (Phillips, p. 3).

A key feature of VOST activity is social media monitoring during disaster events, using cognitive capacities of remote, human volunteers to sort and filter incoming information for local responders whose resources are taxed. Though not explicitly referenced in Phillip's initial proposal, VOSTs in some ways represent an extension of the Voluntary Organizations Active in Disaster (VOAD) structure[23] into the digital space. The VOST term and, perhaps more importantly, the ideas underneath it, have seen some traction in Twitter conversations among emergency managers interested in the problems and possibilities of incorporating social media into their operation plans[24].

In September 2011, Kris Eriksen, Public Information Officer (PIO) for the National Incident Management Organization (NIMO) Portland Team, deployed a VOST during the Shadow Lake Fire to extend her communication capacities for, among other things, monitoring social media communications (St. Denis, 2012).

2.5.3.5 Oklahoma Crisis Mappers

Another group of volunteers has formed to help identify and map information during disaster events specifically in Oklahoma. This group has deployed and maintains instances of Ushahidi crowdmaps that are continuously available to support snow/ice events, grass and wildfires, and tornados in that state. Each map is associated with a Twitter account: @okicemap, @okfires, and @oktwister. According to John Butler,[25] the creator of those mapping projects, the maps are supported by five to six core volunteers, all with backgrounds in other emergency-related, volunteer work, including ham radio operation, storm spotting, firefighting, and volunteering for the American Red Cross. Those volunteers collect citizen reports through the Ushahidi platform and process those reports onto the maps. They also generate

[23] http://www.nvoad.org/

[24] This conversation can be followed or joined by searching or using the #smem hashtag

[25] Personal communication

Ushahidi reports from information found in other media, including social media and especially Twitter. In April 2011, the group considered moving to become a formal 501c organization, but Butler expressed reservations that they might be better off remaining an "ad-hocracy." By June 2012, the Oklahoma group had branched out to other types of crisis support activities beyond crisis mapping, and had moved to advertising itself as a VOST.

2.5.3.6 The Digital Humanitarian Network

The Digital Humanitarian Network[26] was conceived in 2012 as a "network of networks" to connect representatives of the many new virtual volunteer organizations with each other as well as with representatives of traditional humanitarian response organizations. This network was designed to address the recognized challenge of coordinating activities of digital volunteers with the needs of formal responders.

2.5.4 Technology-focused volunteer networks

There are a few networks of volunteer programmers who focus their efforts on the crisis domain, but not strictly on mapping. These include the Sahana Foundation[27], a network of open source programmers who came together after the 2004 Indian Ocean earthquake and tsunami to support responders with technological assistance, Random Hacks of Kindness[28], a corporate-sponsored effort to host barcamps for "hackers," and CrisisCommons. I have attended barcamp events sponsored by Random Hacks of Kindness, where the Tweak the Tweet idea was first conceived and presented, and CrisisCommons. Through my work to deploy and support Tweak the Tweet and my volunteer activities with Humanity Road, I have interacted with CrisisCommons during numerous events, including the Haiti 2010 earthquake, the Japan earthquake in 2011, the Christchurch earthquake in 2011, U.S. snowstorms in winter 2011, and Hurricane Irene in 2011.

---

[26] http://digitalhumanitarians.com/

[27] http://sahanafoundation.org/about-us/

[28] http://www.rhok.org/about

2.5.4.1 CrisisCommons

CrisisCommons[29] is a network of technical and non-technical volunteers focused on instigating and coordinating technical or tech-centered solutions to problems within the crisis domain. Formed in 2009, the group was originally focused on hosting events, *CrisisCamps*, that brought together volunteer programmers with domain experts for weekend "barcamps" to brainstorm, design, and develop both hardware and software solutions. In the wake of the Haiti earthquake, CrisisCommons quickly organized 18 camps in cities all over the world to work on problems specific to that event. The organization has now established an ongoing structure between camps, and has helped to coordinate remote response from technical volunteers during several subsequent events, using public wikis, Skype chats, and conference calls to organize their efforts[30].

The diverse work of CrisisCommons volunteers and small full-time workforce incorporates several different crowdsourcing techniques and uses many different forms of collective intelligence. It includes open source technology development, collaborative creation of informational resources, and both lateral and hierarchical organizational structure. Within their barcamps, their chats, and their conference calls, they also use crowd-brainstorming techniques to gather in ideas for how to respond to an event and what technological projects to work on. Then they gather, in person or online, to create these tools or begin to generate the informational resources they elect to provide.

2.5.5 Incorporating Digital Volunteers into Formal Disaster Response

As these examples show, volunteers, both individuals and emergent groups, can exhibit creativity and improvise solutions which have mixed consequences for emergency managers, both aiding them and providing new challenges (Kendra & Wachtendorf, 2003a). Just as formal responders in past events have

---

[29] http://crisiscommons.org/learn-more/our-story/

[30] CrisisCommons wikis for subsequent events:
Honshu earthquake: http://wiki.crisiscommons.org/wiki/Honshu_Quake;
Christchurch earthquake: http://wiki.crisiscommons.org/wiki/CrisisCampNZ;
Hurricane Irene: http://wiki.crisiscommons.org/wiki/Hurricane_Irene

had to manage the downsides of physical and informational convergence (Fritz & Mathewson, 1957), formal responders in the digital age are struggling to incorporate the work of digital volunteers in their response and planning activities.

The Disaster Relief 2.0 Report (Harvard Humanitarian Initiative, 2011) called attention to emerging issues regarding the use of social media and the developing role of digital volunteer communities in the realm of humanitarian response. That report, commissioned in part by the United Nations Foundation, noted that the Haiti Earthquake marked a turning point in the relationship between digital volunteer communities, on one hand, and formal response and humanitarian organizations on the other. After interviewing several individuals from both groups, the authors claimed to offer a balanced account of some of the problems and potentials of these components of emergency response, and said that they intended the report to spark conversation between the two sides. A conversation did indeed ensue. A clear sentiment throughout the report was that the formal responders felt their work was complicated by the activities of digital volunteer communities. Perhaps in response to what they interpreted as an accusation, representatives of the Standby Task Force (which is in some ways an outgrowth of the Ushahidi organization and relies on their tools) posted a blog that disputed many of the report's claims (Meier et al., 2011) and numerous others left comments on that blog either defending the initial report or raising other complaints from digital volunteer communities.

Though many of these issues remain unresolved, formal responders are actively working on strategies for incorporating digital volunteer communities into their work. For instance, the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) has fostered a relationship with the Standby Task Force and worked together with that group to launch an Ushahidi mapping instance for the Libya political uprising[31]. Additionally, the VOST concept has now been employed to help formal disaster responders with the work of information processing (St. Denis et al., 2012).

[31] http://blog.ushahidi.com/index.php/2011/03/06/using-new-ushahidi-map-libya/

2.6 Conclusion

In constructing a perspective to examine crowd work during mass disruption events, this work draws on empirical findings and theoretical foundations from multiple academic fields. It weaves together established research from sociologists of disaster, recent studies on the use of social media during crisis, and an emerging literature on crowds and crowdsourcing. It exposes the roots of crowdsoucing to lie in open source and outsourcing and surfaces its relationships with the concepts of human computation and collective intelligence. This background chapter also describes many of the different configurations of connected crowd work that have manifested during and after recent mass disruption events, and enumerates some of the ICT tools and platforms that crowd workers appropriate and in some cases develop for this work.

CHAPTER 3
RESEARCH QUESTIONS & METHODS

3.1 Primary Research Question

Informational convergence in the digital age presents new challenges and new opportunities for aggregating, processing, and utilizing information during mass disruption events. The flood of information, streaming fast and furious through new channels and blending in new voices, has contributed to a perceived problem of "information overload" within the domain. Though some are working on machine computational solutions to address these issues, these alone are not yet adequate to meet these new challenges. Disaster responders and the affected public need, and are currently developing and improvising, new strategies for filtering, recontextualizing, vetting, and structuring this flood of data, to convert it from overload to resource. This research looks toward human-centered solutions to these problems, asking:

Research Question: Under conditions where the data produced are too much for a person to process manually, how does the social media-connected crowd act to organize the data moving through those platforms into useful information resources?

Though this broad research question underlies all of the research presented here, it was not the starting point for this dissertation, but was instead a destination arrived at by progressing through a series of related, though more specific questions. This dissertation consists of four separate studies designed to address these smaller research questions. The findings of these individual studies then inform a broader perspective on crowd work presented in the final chapters. As much of this work was conducted in parallel, these studies were not implemented as research stages, but alternative perspectives for examining crowd work of different kinds. However, early research questions did evolve, through discoveries made in answering them, into other questions and lead to later studies. The trajectory from one question to the next is described below.

3.2 Methods and Reporting

Because this dissertation is based on four separate studies, and because each study required different research methods—ranging from qualitative and quantitative analyses of large volumes of Twitter data collected during specific events to long-term participant observation in digital volunteer communities across events—the methods for each approach are described in the chapter devoted to that study.

3.2.1 Author Voice and Collaborative Research

This larger research project integrates smaller research efforts that include collaborations with my academic advisor, Leysia Palen, other colleagues at Project EPIC at the University of Colorado, and an external collaborator. Other aspects of this work I have completed alone. In differentiating between these two types of efforts, I use the "we" voice to refer to collective efforts within unpublished research and all previously published research that included contributions from co-authors, and the "I" pronoun for unpublished, solitary research efforts. Studies based on previous work contain footnotes with information about the original work.

3.2.2 A Note on Treatment of Data

In this dissertation research, all interview response data are anonymized. Because Twitter data are public and searchable, we have carefully disassociated the interview identities and responses from Twitter accounts and their data streams. For Twitter data reporting for interviewees, we asked each participant to choose between three kinds of anonymity: 1) use a pseudonym and altering the language of their tweets to reduce searchability; 2) use a pseudonym but maintain original tweet language; and 3) attribute original tweet content directly to their real Twitter account name. No interviewees selected the first option; most selected the third—direct attribution.

Additionally, for all other Twitterers who were not interview participants for this research but who have authored or are mentioned in tweets presented here, we have changed usernames on most accounts. However, for accounts we were able to contact through the Direct Message feature on the Twitter platform[32], we offered the above three options. Again, almost all contacted Twitterers chose direct attribution for their tweet content rather than anonymity.

3.3 Trajectory of Research Questions

The original starting point for this research examined possible solutions for automatically processing social media updates into resources for emergency responders and people affected by disasters and mass disruption events.

RQ0. How can useful information be extracted from social media updates?

This preliminary research question asked how we—researchers, technology designers, responders, affected people, etc.—could extract from social media updates, in real-time or near real-time, actionable information and other information that could contribute to situational awareness during an event. In previous work, my colleagues and I had found that large volumes of situational awareness information were shared through Twitter during disaster events, but that these were accompanied by a considerable amount of noise (Starbird et al., 2010; Vieweg et al., 2010). To be useful to affected people and responders, this information needed to be filtered—and processed in other ways as well.

Other researchers in our larger project—Project EPIC at the University of Colorado (Palen et al., 2010)—pursued machine-only computational solutions to this problem, including Verma et al. (2011) who reported some success in automatically identifying situational awareness information from Twitter messages using Natural Language Processing (NLP) strategies. However, processing solutions focused exclusively on automatically analyzing the textual content of Twitter data, in particular, suffer from several limitations related to the character constraints and informal nature of communication on that platform; traditional approaches to NLP have used textual passages from the Wall Street Journal, where language is used quite differently, to train classifiers (Gimpel et al., 2011). Though researchers are

[32] To send a direct message (DM) on Twitter, the recipient account must be following the sender account. Over the past two years, I have developed following relationships with many digital volunteers through my “Tweak the Tweet” activity and other action-research during crisis events.

exploring solutions for purely computational solutions to processing social media updates (e.g. Corvey et al., 2010; Gimpel et al., 2011; Verma et al., 2011; Corvey et al., 2012), we do not yet have working solutions for filtering and classifying actionable and situational awareness information at the high levels of accuracy and precision necessary to be relied upon exclusively in the high-stakes domain of disaster response and mass disruption events. Additionally, other complexities of social media data like the ones noted in the introductory chapters of this work—e.g. lost context, language ambiguity, misinformation and disinformation—may mean that solutions based on machine computation will never achieve this alone.

Accepting this as a constraint led to the formation of a new research question, as I imagined a solution that shifted some of the burden of processing this information onto the communicators (people), asking them to format their messages in such a way that machines could easily identify and extract key information. This question became the inspiration for the "Tweak the Tweet" microsyntax (Starbird & Stamberger, 2010) and the basis for Study 1 (Chapter 4) of the proposed research:

RQ1. Can we shift the burden of interpretation onto the users and teach people to make their social media communications machine-readable? How?

Addressing RQ1, the Tweak the Tweet idea was initially conceived during a Random Hacks of Kindness barcamp[33] in November 2009 as a tool for affected people to use during an unfolding event to communicate their needs or first-hand knowledge of the situation in a machine-readable format. If used "correctly" by social media users, the proposed microsyntax would render their messages simply parse-able. Study 1 (Chapter 4) outlines the rationale and eventual implementation of Tweak the Tweet and describes in detail how the innovation was deployed during multiple events in 2010 and 2011. Early findings demonstrated considerable difficulty in teaching affected people, especially during times of impact, to format their social media communications in a standardized format. However, a significant discovery during our first deployment opened up the possibility of another approach: adjusting the user

[33] http://www.rhok.org/

scenario for Tweak the Tweet from an affected person to a remote volunteer. RQ2 encapsulates that discovery and the new research direction that resulted from it:

RQ2. Whom can we teach to make their social media communications machine-readable? How?

In our initial deployment of Tweak the Tweet, described in Studies 1 and 2 (Chapters 4 and 5), we were surprised to find that the microsyntax was adopted not by those on the ground during the event, but by remote individuals—digital volunteers—who were using the innovation to help process information by translating actionable information into TtT syntax. Those self-deployed, remote volunteers were also taking part in a wide range of other ICT-enabled activities, attempting to help affected people in Haiti by filtering, routing, verifying, and mapping information. This discovery of a new kind of work and a new kind of remote, crowd worker shifted the focus of this overall research towards understanding who these people are, what their goals are, and how they work in various capacities during mass disruption events. RQ3 asks how they organize themselves to do this work:

RQ3. How do remote individuals use social media to organize themselves to process information during mass disruption events?

Studies 2 and 3 (Chapters 5 and 6) in this work were designed to address RQ3 by examining different configurations of digital volunteers. Study 2 (Chapter 5) describes how the digital volunteers we observed in Haiti began to act in concert with others in the space, forming what we characterize as an emergent organization (Dynes, 1970; Kreps & Bosworth, 1994). Other, more formal organizations are now working to recruit and train remote volunteers for repeated deployments over time and across events. I will examine the self-organizing activities of one of these virtual volunteer organizations, Humanity Road, in Study 3 (Chapter 6).

There are many different kinds of digital volunteers and different forms of crowd work. Many of the digital volunteers that we observed tweeting during the Haiti event were intentionally participating in the work of information organizing, actively working together to process information. However, it is also possible to view every action by every member of the crowd tweeting about an event as tiny movements towards organizing the space. During a mass disruption event, micro-actions within social media platforms such as routing information, recommending information and users, friending and following can provide meta-information about the information flowing through the virtual space. This observation leads to the following question, addressed in Study 4 (Chapter 7) of this proposal:

RQ4. How does the larger crowd act to organize information through individual actions within the social media space? How can we derive meaning from the collective "work" of the crowd?

In answering the larger research questions presented here, addressed in part through these sub-questions, my aim is to develop a broad understanding of how the crowd works to process information during mass disruption events. The questions addressed here lead to an illumination of the capacities, motivations, and salient features of crowdwork in this context, showing it to be rich, diverse, collaborative and interactive. The answers to these questions also help inform strategies for leveraging and supporting crowdwork in this context and beyond.

CHAPTER 4
STUDY 1. TWEAK THE TWEET: SELF-ORGANIZED HUMAN COMPUTATION THOUGH SOCIAL MEDIA[34]

Addressing Research Question 0, strategies for processing information in social media updates, and Research Question 1, shifting the burden of this processing from machines to people[35], this study discusses the rationale, deployment, and lessons learned from the "Tweak the Tweet" project. Tweak the Tweet (Starbird & Stamberger, 2010) was conceived as a solution for the Twitter platform for the problem of processing disaster information. TtT asks users to structure their tweets in such a way as to render them machine-readable, making it possible for remote computers to aggregate, filter and categorize tweeted information in real-time. In the first part of this study, I will explain the rationale behind the original Tweak the Tweet idea, and how it was informed by previous research on the use of Twitter during the Red River flooding and Oklahoma grassfires in Spring 2009 (Starbird et al, 2010; Vieweg et al, 2010). Next, I will describe the work we—my colleagues at Project EPIC and I—did to design, implement, deploy and support Tweak the Tweet (TtT) during multiple crisis events in 2010 and 2011. Featuring data collected during deployment efforts for the Haiti earthquake as well as several subsequent events, I evaluate deployment strategies, assess many of the assumptions underlying the Tweak the Tweet concept, and derive lessons learned. These latter sections address Research Question 3 of this larger research: how to teach people to make their social media communications machine-readable. Finally, I explain how the Tweak the Tweet project evolved into the larger dissertation research discussed here.

---

[34] Significant sections of this Chapter related to early deployments of Tweak the Tweet (those that took place in 2010) have been adapted from an earlier work:

Starbird, Kate, Leysia Palen, Sophia B. Liu, Sarah Vieweg, Amanda Hughes, Aaron Schram, Kenneth Mark Anderson, Mossaab Bagdouri, Joanne White, Casey McTaggart, & Chris Schenk. (2012). Promoting Structured Data in Citizen Communications During Disaster Response: An Account of Strategies for Diffusion of the 'Tweak the Tweet' Syntax. In Chris Hagar (Ed.), *Crisis Information Management: Communication and Technologies*, Cambridge, UK: Woodhead Publishing Limited. ISBN: 978-1-84334-647-0.

To cite material from this Chapter, please cite this original work as well as this dissertation.

[35] See Chapter 3 for an explication of research questions and their mapping to the separate studies of this work.

4.1 Tweak the Tweet Background: Extracting Situational Awareness Information from Tweets

Remarking on what he called the “rise of the citizen journalists,” Gillmor (2004) asserted that the coordinated terrorist attacks of September 11, 2011 marked a turn in news production, as ordinary people were newly enabled to produce and share information, including valuable context about the unfolding event, to a global audience through web-based technologies—i.e. email, mailing lists, and blogs. Not long after, social media platforms—which began to see large-scale adoption around the world about a half-decade later—and ubiquitous mobile technology created an information ecosystem where more information could be shared with more people from more places, both geographical and virtual, than ever before. Twitter, a social media platform that supports microblogging, launched in 2006 and was soon used for citizen journalism during disaster events, including the 2007 Southern California Wildfires (Messina, 2007b; Sutton et al., 2008).

In research studies of two concurrent, natural hazard events in March and April of 2009, we found that local people were using Twitter to share information about the event, including information that could contribute to situational awareness (Starbird et al., 2010; Vieweg et al., 2010). Situational awareness, a term employed in several safety-critical domains that essentially means understanding what is happening during an activity or event, involves taking into account all of the available information in order to make the most-informed decisions possible (Zhang et al., 2002). Vieweg et al. (2010) reported that a majority of tweets that were sent by locals and were “on topic” to the disaster event—61% during the 2009 Red River Floods and 76% during the 2009 Oklahoma Grass Fires—contained information that could contribute to situational awareness.

Local users were not the only ones turning to the Twitter platform to seek and share information or comment about the event. During the 2009 Red River Floods, the majority of Twitterers who tweeted about the event were located outside the affected area (Starbird et al., 2010). Twitter had become a site of informational convergence (Fritz & Mathewson, 1957; Hughes & Palen, 2009), where a global audience met up after an event to participate in a range of activities, including information seeking and sharing

(Palen & Liu, 2007), helping to coordinate response (Starbird, 2010; Qu et al., 2011, Sarcevic et al., 2012), offering support often in the form of prayers, and sharing photos and other high level accounts of the event (Starbird & Palen, 2010). Though this convergence was often well intentioned, it contributed to a large and rising volume of information, much of it noise, flowing through Twitter during disaster events. Additionally, a large majority (90% during the 2009 Red River Floods) of tweets could be considered "derivative," meaning that the information was already available elsewhere on Twitter or the surrounding online information space. Researchers and practitioners in the domain of crisis response began to ask how we could build tools to extract, from this growing and noisy information space, actionable information and other information that could contribute to situational awareness for responders and affected members of the public (Palen et al., 2010; Vieweg et al., 2010). This question served as the preliminary research question for this dissertation work:

> ***RQ0.*** *How can we extract useful information from social media updates?*

Researchers in the field of natural language processing have been experimenting with using machine learning algorithms to infer meaning from the content of tweets (Vieweg et al., 2010; Verma et al., 2011). Though early research in this area shows some modest results, NLP classifiers for situational awareness information require hand-coding to train and are not portable across events—i.e. classifiers trained on data from a flood do not work well classifying data from an earthquake (Verma et al., 2011). Twitter may provide ongoing difficulties due to its short message format and non-standard language use. Additionally, the issue of lost context and the resulting ambiguity of individual messages, discussed in the background section of this work, may create permanent difficulties for extracting useful information even when classifiers can identify tweets as being potentially useful.

Recognizing the complexity of this problem, I began to consider an alternative strategy, which became the Research Question 1 of this dissertation:

> ***RQ1.*** *Can we shift the burden of interpretation onto the users and teach people to make their social media communications machine-readable?*

4.2 Tweak the Tweet: A Microsyntax for Crisis Reporting

Reflecting on this question at a Random Hacks of Kindness (RHOK) barcamp[36] in November 2009, in collaboration with another researcher, I proposed the Tweak the Tweet (TtT) concept. Tweak the Tweet (TtT) is a crisis-reporting microsyntax that encourages users of the Twitter platform to format their event-related tweets in a specific, standardized way (Starbird & Stamberger, 2010). The Disaster Relief 2.0 report cited the "unstructured" nature of social media communications as a problem for those trying to process and make sense of the information coming through those channels in the wake of the 2010 Haiti Earthquake (Harvard Humanitarian Report, 2010). Tweak the Tweet was designed to alleviate that problem, offering a way for people to "structure" their Twitter updates, making them machine-readable by a simple computer algorithm.

4.2.1 The Mechanics of Tweak the Tweet

Tweak the Tweet instructs users to place certain hashtags in front of different pieces of information within their tweets to tell computers *what* their tweets are about and *where* in the tweets to find certain things. For example, they are asked to place a `#loc` hashtag in front of any location information. A parsing algorithm can then search for the #loc string inside the text of each tweet and pull out the text that comes after that tag, storing it in a database field for tweet location.

```
#haiti #need food and h20 #name Villa Manrese #loc Haut Turgeau #info
1000+ ppl no aid 18 days #contact JL 555-5555
```

In this example, the `#haiti` tag tells the remotely-located process that captures tweaked tweets that this tweet is referring to the Haiti earthquake event. The presence of the `#need` and `#loc` tags indicates that the tweet is using the TtT format. After the computer program identifies this tweet as TtT, it will create a 'need' record to correspond to the tweet's "primary" tag. The need will be recorded as "`food and h2o`"; the name of the entity in need will be "`Villa Manrese`"; the location will be "`Haut Turgeau`"; and the contact information will be "`JL at 555-5555`".

[36] http://www.rhok.org/

Incorporating minor alterations and clarifications to the original proposed syntax (Starbird & Stamberger, 2010), TtT now asks users to consider four separate components for each tweet-report:

1. TtT Twitterers are asked to use an event tag to tell the collection software that their tweet relates to a specific event.

    `#haiti` for the 2010 Haiti Earthquake

2. Users are asked to choose from a list of disaster-related report categories—e.g. `#damage`, `#shelter`, `#need`, etc.—and to include one and only one of these tags, followed by the report details.

    `#need medical supplies and water`

3. Location information is extremely important for reporting of disaster-related information, and TtT Twitterers are instructed to include it, where applicable, using the `#loc` tag to mark its place within the tweet.

    `#loc 123 Main Street in Clarksville`

4. Finally, users can choose from a range of other hashtags to mark up other pieces of information within the tweet—e.g. `#contact`, `#time`, `#source` or `#src`, etc.

4.2.2 Why Use Twitter for a Crisis Reporting Innovation?

TtT is designed to take advantage of several features of social media in general and affordances of the Twitter platform specifically, some of which have already contributed to the adoption and appropriation of Twitter during times of crisis. These include the short message length, social network organization, and the ability to broadcast to a global audience. Additionally, Twitter has a history of user-driven, linguistic adaptations that set a precedent for an innovation like Tweak the Tweet.

Tweak the Tweet relies on the public broadcast and searchability of tweets—the fact that the vast majority of tweets are public and all public tweets can be searched using a variety of APIs provided by

Twitter, as well as many end-user search tools. This means that all TtT tweets are available for both manual search and for tools built to access the Twitter stream and process the tweets. Importantly, anyone can build tools to access TtT tweets, and these tools can be deployed from any place with an Internet connection. Because Tweak the Tweet uses an existing social media site to carry its messages, it significantly reduces the development overhead for an Internet-based, crisis-reporting application.

This latter feature also means that people do not need to download and learn to use a whole new tool during an event. When the idea was conceived, we knew from previous research (Starbird et al., 2010; Vieweg et al., 2010) that people were already using the Twitter platform to broadcast actionable and situational awareness information during crisis, so TtT represented a significant change in *how* the information would need to be communicated, but did not require the user to adopt a whole new technology. Importantly, when using the innovation, users' crisis reports would not be siloed within a single crisis-reporting application. TtT users could broadcast to their social networks, reach a global audience, and have their tweets collected as crisis reports all at the same time.

This relationship between TtT and an existing social media site carries other advantages, stemming from the connection between social media updates and account profiles. Twitter supports not anonymity but pseudonymity (Pfitzman & Köhntopp, 2001), in that every tweet can be traced back to an account and each account has an ongoing history as well as a profile that its owner maintains. TtT reports can therefore be connected to accounts, and these accounts can be assessed for credibility. This relationship between account and tweet activity may also create motivation for contributing information, as users can get credit in the form of social and symbolic capital (Bourdieu, 1984; Bourdieu, 1986; Putnam, 2000) for reporting information. It may also be true that in a situation of pseudonymity, users are less likely to knowingly introduce false information. These latter assumptions, important components of the original rationale for Tweak the Tweet, will be examined within this chapter.

Finally, by building off the hashtag convention, Tweak the Tweet seeks to leverage an ongoing phenomenon within the Twitter platform whereby users—as opposed to Twitter developers—introduce and occasionally adopt on a massive scale linguistic conventions that enhance the existing functionality of the platform. Addressed tweets or "@mentions" (Honeycutt & Herring, 2009), retweets (boyd et al., 2010), and hashtags (Messina, 2007a) are all examples of these user-driven adaptations. Originally introduced and advocated by Messina as a way to improve searchability and group formation, hashtags are a way for users to mark up a tweet with specific keywords (preceded by the # symbol) to tell others what their tweet's topic or topics are. Hashtags experienced a first surge in visibility, due in part to Messina's advocacy, during the Southern California Wildfires in 2007 (Messina, 2007b) and had achieved widespread adoption by late 2009, when TtT was first conceived. The original Tweak the Tweet concept was designed to capitalize on this widespread adoption, acknowledging that many users already understood and were accustomed to seeing and using hashtags in tweets. Again, it was thought that syntax use would fit well within users' existing social media behavior, which included adopting new linguistic conventions and enforcing new norms.

4.2.3 An Analog to Tweak the Tweet: Citizen Science

From the beginning, a major challenge for the Tweak the Tweet concept was always: *How will Twitterers learn how to use the TtT syntax?*

Tweak the Tweet was designed to reduce the computational complexity of interpreting social media messages. However, in doing this, it effectively shifts the burden of communication from the technical architecture of the system onto the people sending these messages. To use the syntax, Twitterers have to learn a new way of communicating—adding specific hashtags to their tweets in specific places, adjusting word order, and still managing to fit their message within the 140-character constraint of the platform. As we developed the idea, one question that continually surfaced was: how could we teach the syntax to potential users?

The initial design of Tweak the Tweet sought to answer this teaching question by incorporating the strategies of citizen science (Starbird & Stamberger, 2010), a correlate of citizen reporting. Citizen science projects use distributed workforces of amateur scientists to contribute to large scientific endeavors, typically through data collection. These projects enable scientists to leverage a cheap workforce and to gather data over a large geographic space and over a long period of time (Cohn, 2008). Citizen science considerably predates the digital age—the Audubon Society's annual Christmas bird count is over a century old, and in the mid-19th century, Joseph Henry of the Smithsonian Institute recruited hundreds of volunteers to help track the movement of storms by gathering weather data from geographically dispersed areas and sending those observations back to the Smithsonian (Grier, 2005).

Citizen science efforts have typically relied on the latest ICTs; Henry's weather project through the Smithsonian Institute accepted reports via mail, initially, and then over telegraph wires. Today, existing citizen science projects are incorporating online tools and web-based reporting into their efforts, and new ICTs are enabling new types of projects. In his book introducing the *crowdsourcing* concept, Howe (2008) described several new citizen science projects, including eBird[37], an effort launched by the Cornell Lab of Ornithology that aggregates reports of bird sightings so scientists can track migration habits, measuring among other things the impact of environmental change.

Distributed weather reporting is a close relation to both citizen science and crisis data reporting. The National Weather Service (NWS) currently solicits citizen reports of weather information through two programs: SKYWARN and Cooperative Observers[38]. The SKYWARN project, established in the 1970s receives information through trained storm spotters via telephone or HAM radio reports. The Cooperative Observer program, created in 1890, incorporates general climate data measured and submitted by over 11,000 volunteers using mail, the telephone, or a computer. In a similar effort, the FAA has been collecting voluntary reports of actual weather conditions from pilots for decades, in the form of pilot weather reports or PIREPS. Casner (2010) describes the PIREP program as "an attempt to recruit the population of pilots and use them as a corps of trained amateur weather observers" (p. 348).

[37] http://ebird.org/content/ebird/about

[38] www.weather.gov/om/brochures/Citizen_Scientist.pdf

Citizen science in the form of weather reporting has experimented with a concept like Tweak the Tweet. In 2010 the NWS deployed a Twitter microsyntax for citizen reporting of weather information[39], asking users to set apart location information in tweets using:

```
WW <location information> WW
```

One key component of citizen science efforts is training. Volunteers must be educated in how to submit the right data in the right format (Cohn, 2008). The Christmas Bird Count requires volunteer groups, or "counting circles," to include at least one experienced birder who helps guide the group in the correct counting and reporting protocols.[40] PIREPS weather reports require a standardized format, and most pilot training programs include instructions on how to make these reports (Casner, 2010). For their weather reporting Twitter microsyntax, the NWS provided a web page with instructions.

4.2.4 A Protocol for Teaching TtT Syntax: Prescriptive Tweeting

Similar to the citizen science and weather reporting programs described here, Tweak the Tweet requires volunteer information reporters to learn what needs to be reported and how it needs to be reported. Within the initial Tweak the Tweet concept was a proposed protocol for teaching the syntax using *prescriptive tweets,* or tweets that explained and demonstrated the "correct" syntax (Starbird & Stamberger, 2010). It was hoped that prescriptive tweets would be distributed by a formal response agency that would also monitor the incoming tweets, and that affected people would see these prescriptive tweets and be able to adapt their information to the format.

Listed below are several examples of prescriptive tweets from the initial TtT presentation (Starbird & Stamberger, 2010):

```
2:02pm @RedCrossLW: #lakewoodfire #lwfire use these tags for reporting,
#fireline, #line_down, #addy, #city, #X_streets
```

---

[39] http://www.nws.noaa.gov/stormreports/

[40] http://birds.audubon.org/get-involved-christmas-bird-count

```
2:04pm @RedCrossLW: #lakewoodfire #lwfire more tag reporting: #wind,
#visibility, #road_close, #injury, #num, #comment

2:06pm @RedCrossLW: #lakewoodfire #lwfire use this format: #fire #city
[city] #addy [address or cross streets] #floor [floor]

2:07pm @RedCrossLW: #lakewoodfire #lwfire Twitterers, please make sure
that you are your families are safe before uploading fire data!

2:08pm @RedCrossLW: #lakewoodfire #lwfire use this format: #imok [name]
#city [city] #location [place] #addy [address or cross streets]

2:10pm @RedCrossLW: #lakewoodfire #lwfire For those of you in Lakewood,
TN, use the #lwfire hashtag to report info.
```

It is important to note that the syntax changed quite significantly from this initial presentation to its current version, through adjustments made during different deployments. This evolution will be described below.

4.3 Tweak the Tweet Implementation, Resources, and Instructions

I implemented Tweak the Tweet in stages, adding new pieces during and between deployments. This section describes the technical infrastructure of Tweak the Tweet, including the resources and instructions that have been made available to the public. Figure 1 gives an overview of the separate components of the TtT system and demonstrates how these pieces interact. Figure 2 is a timeline that contains the TtT deployment windows for each of the events featured in this chapter, as well as markers for when different components of the TtT went online.

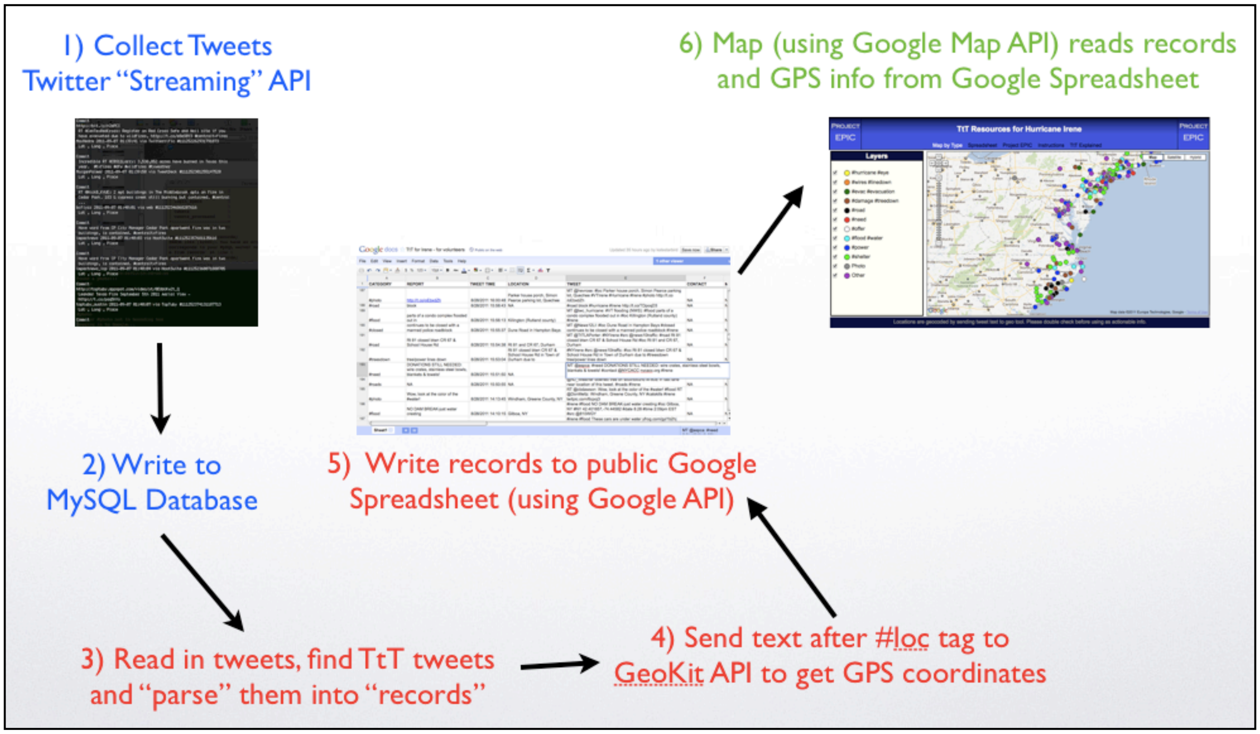


Figure 1. TtT Implementation Diagram

4.3.1 Collection

The collection strategy for Tweak the Tweet evolved from piggybacking on Project EPIC's data collection architecture that had been designed primarily for research (Anderson & Schram, 2011) to using a specific collection solely for TtT support. The TtT collection script is a Python script that relies on the Tweepy library,[41] a free and publicly available library, to access Twitter data through the Twitter Streaming API. To use the Streaming API, one designates a list of terms and Twitter returns tweets[42] that contain those terms as they are posted the public timeline—this is sometimes referred to as "filtering" and is a forward-in-time or real-time search as opposed to the backward-in-time search functionality of Twitter's Search API. For the TtT collection, the term list is a collection of hashtags identified as being specific to the crisis event. As the event evolves and new terms begin to gain traction, new search terms are manually added to the designated list, and the collection script is restarted.

---

[41] http://code.google.com/p/tweepy/

[42] Twitter returns some subset of tweets that contain these terms. They do not report exactly how that subset is determined. (cite boyd's big data paper)

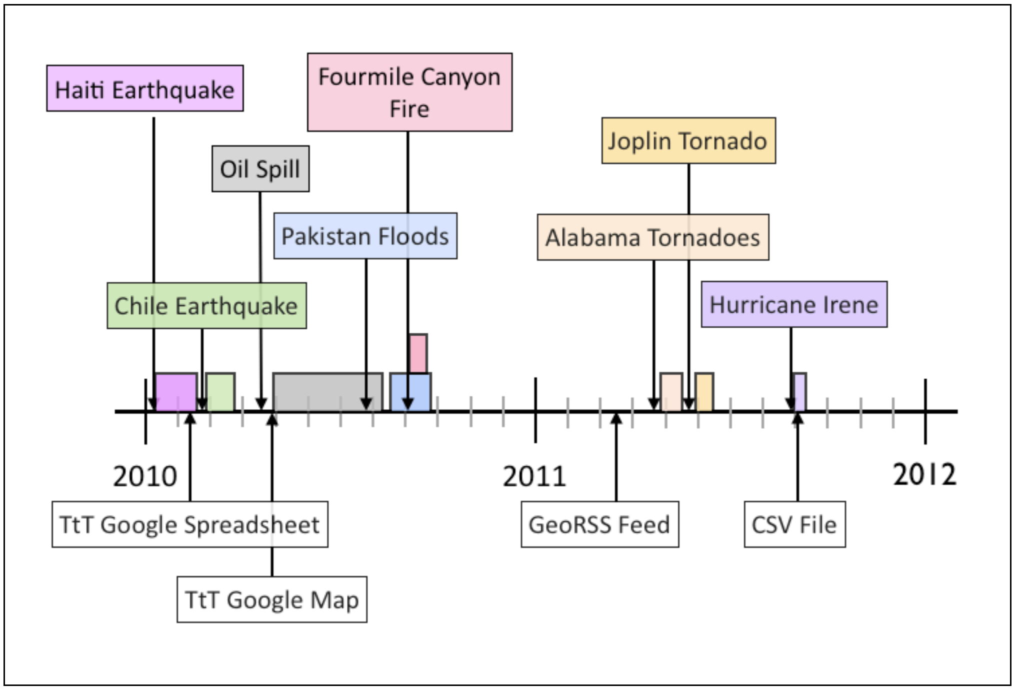


Figure 2. Timeline for Featured TtT Deployments and Adding Public Resources

4.3.2 Temporary Storage

During a large-scale crisis event, tweets arrive continuously and in high volume. Originally, I experimented with processing TtT tweets in real-time from within the collection script, but that strategy caused data loss, because the parsing functionality could not keep up with the incoming tweet stream. To deal with this issue, I now push incoming tweets to temporary storage, where a separate parsing script accesses them in batches. Initially, I used a text file as temporary storage, but I later turned to using MySQL to store both the tweets and the processed records.

4.3.3 Parsing

The parsing script, implemented in Ruby, is based on a series of regular expressions. Here, I explain the general flow and highlight important aspects of that program.

CONFIG file: Because the salient categories for disaster reports vary across event type, impact size, duration, and geographical features of the affected area, the TtT syntax is designed to be dynamic and the parsing script is meant to support this flexibility. I currently accommodate this need for flexibility with a configuration file for each TtT deploy that can be quickly changed within any text editor to adjust certain features of the syntax.

Parsing Intervals: As mentioned above, to deal with the speed of incoming tweets, the parsing script runs separately from and concurrently with the collection script. It runs in intervals, pulling all tweets that have arrived in the system since the last time the process ran.

Tweet Parsing: For each tweet, the script executes the following process to create, if possible, a TtT record from the tweet:

1. *Is the tweet TtT?* The script compares the textual content of the tweet against a list of TtT terms to confirm that the tweet is in TtT format. If not, it exits the process for the current tweet and moves to the next collected tweet.
2. *What is the Report Category of the tweet?* The script compares the tweet content to a list of report categories, ordered by specificity to generality of those terms. For instance, `#damage` would be identified as the report category before `#fire` or `#info`, which are more general terms. The script then sets the main category of the report to this tag and takes the text immediately following this tag (ending in the next #) as the textual "report."
3. *What is the Location of the tweet?* I describe this process in more detail below, in the Geolocating section. If available, the record is assigned GPS coordinates.
4. *What other pieces of information are present?* Next, the process compares each tweet against the entire list of possible data tags (`#contact`, `#time`, `#photo`). Where any of these tags is found, the script parses the information following that tag into the corresponding field for that tag within the record.
5. *Is the tweet REALLY TtT?* To reduce noise in the set, some main categories, especially those with general terms like `#fire`, must have location information associated with them in order to be classified and stored as reports. For records that do not have location information, the script then

compares the report category against a list of general terms, and if it falls in one of those categories, it is discarded.

6. *Have we already processed this tweet record?* The program compares this record against all of the other records stored in the database. If the record is a close match to another record sent within a designated period of time, then it is discarded. If not discarded, the TtT record is now stored in the database as a record.

At the end of each processing cycle, all new records are then posted to a variety of resources, described below.

4.3.4 Geolocating

Geolocating tweets is the most complex part of the TtT processing infrastructure, and I continue to improve the algorithm for location disambiguation. For geolocating TtT data, the following steps are taken for each tweet that passes a first line test for TtT format:

1. *Check for Metadata Location*. A small portion of tweets (~1%) contain location information embedded as GPS coordinates, bounding boxes, and/or place names in the metadata of the tweet[43]. If the tweet has GPS coordinates or a bounding box in its metadata, the program initially stores those coordinates as the tweet report's location. However, if the tweet also has valid location data in the textual content of the tweet (2, below), the program preferences the textual location over the metadata location. This preferencing is done for a simple reason: it allows people, including remote volunteers and affected people who may be moving, to report information about a location at which they are currently not.

[43] Analysis of several disaster events in 2011 and 2012 indicates that the percentage of tweets with GPS metadata is typically near or below 1%, and that this number varies according to event type, location, and the amount of global participation. For example: Egypt Protests = 0.56%; Hurricane Irene = 1.02%; Texas Wildfires in Sept 2011 = 1.32%; and SE Tornados on March 2, 2012 = 1.02%.

2. *Check for User Designated Textual Location.* The tweet content is then searched for the `#loc` or `#location` string, and if found, the algorithm takes all of the text after the tag before the next # and stores that as the textual location in the tweet report. If a textual location is not found, the algorithm returns an empty textual location and either the metadata GPS coordinates, if available, or empty GPS coordinates for the tweet record.

3. *Send Textual Location for Geolocating.* If a textual location is found, the algorithm sends the location using the GeoKit library[44] to find GPS coordinates by cross-referencing several different geolocating tools.

4. *Check GPS Coordinates against Bounding Box.* Within the config file for each event, a bounding box for the event is designated. After receiving GPS coordinates for the textual location from GeoKit, the algorithm checks to see if these coordinates are within the bounding box for the event. If they are, the coordinates are stored in the tweet record and geolocating process ends.

5. *Supplement Location Information.* To improve location disambiguation, the config file also allows a system administrator for an event to associate certain place names with specific event tags. For instance, during the flooding in Vermont after Hurricane Irene, `#vtfloods` became associated with Vermont and Burlington, VT. If Step 4 does not produce valid GPS coordinates within the bounding box, the algorithm takes all of the lists of places from the config file that are associated with hashtags found within the tweet text, and, one at a time, adds these terms to the end of the textual location and sends those to the GeoKit for geolocating. For instance, if a tweet contained `#vtfloods` and `#loc 123 Main Street`, the program sends "123 Main Street, Vermont" to the GeoKit, and if that does not return valid GPS coordinates within the box, the program then sends "123 Main Street, Burlington, VT" to the GeoKit.

[44] http://geokit.rubyforge.org/

6. *Set Tweet Record Textual Location and GPS Coordinating.* As soon as valid GPS coordinates are found within the bounding box, the GPS coordinates for the tweet report are set to these coordinates—overriding any metadata coordinates. If no GPS coordinates within the bounding box are found, the algorithm returns the location text as the tweet record's textual location and either the metadata GPS coordinates, if available, or empty GPS coordinates.

4.3.5 Resources

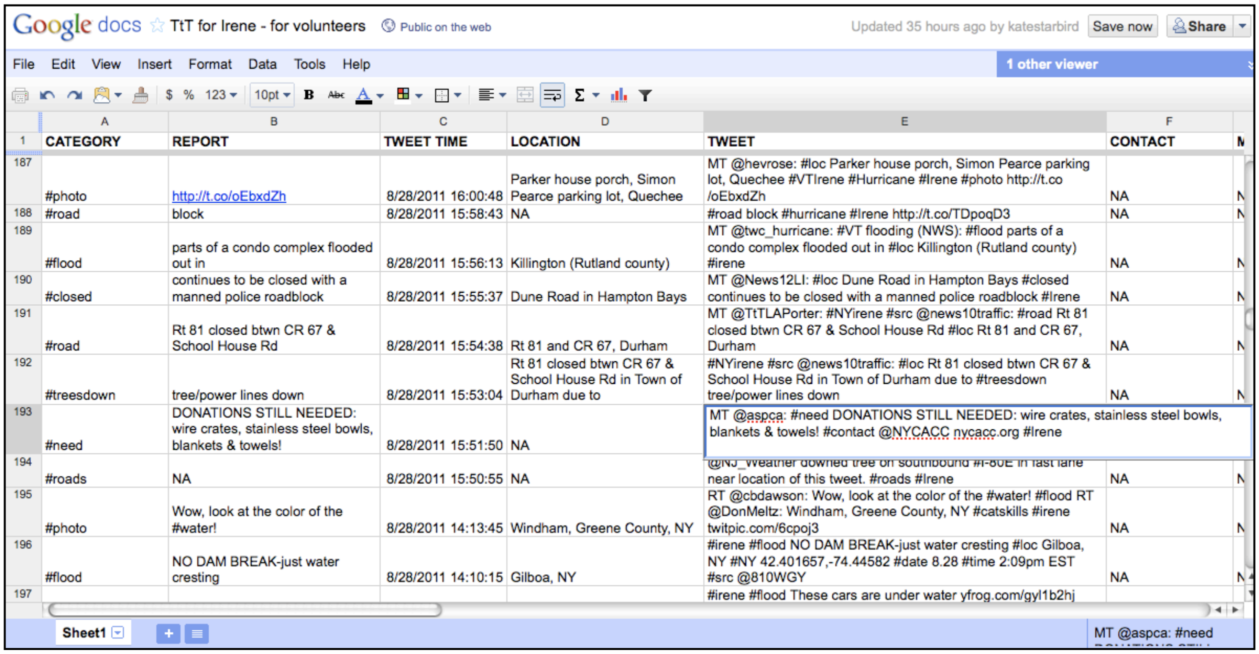
Tweak the Tweet data is meant to be public. To make the data available to the affected public and responders, to show users both *that* the data is being collected and *how* the data is being collected, and to demonstrate proof of concept, I have introduced several public resources that showcase TtT-reported data.

4.3.5.1 TtT Google Spreadsheet

For each TtT deployment, beginning in February 2010 for the Haiti Earthquake response, I provide a public Google Spreadsheet containing records parsed from TtT tweets (e.g. Figure 3). The TtT reports are uploaded to the Google Spreadsheet from within the Ruby processing script, using the Google Spreadsheet API.

The spreadsheets serve several purposes at once. They create live resources from TtT tweets that can and have been accessed by affected people, remote volunteers, and others during several events—e.g. during Fourmile Canyon Fire the Google Doc environment recorded hundreds of people accessing the spreadsheet at the same time. The spreadsheets also act as a learning resource for TtT Twitterers, demonstrating how TtT tweets are parsed into records.

I have experimented with making these spreadsheets editable, allowing users to update fields, especially GPS locations and a "CONFIRMED" field. Initially, I used an extra routine in the script to update the internal TtT database to reflect spreadsheet user changes. However, reading and writing to a Google Spreadsheet while it is being edited often caused the data to become out of synch, and these errors would then propagate to the database. To avoid this problem, spreadsheets are now editable, but edits are not updated in the MySQL database. An ideal interface between TtT data and a second layer of digital volunteer activity would better support verifying, editing, and updating existing records, while keeping the current versions of the records visible to the public.



	CATEGORY	REPORT	TWEET TIME	LOCATION	TWEET	CONTACT
187	#photo	http://t.co/oEbxdZh	8/28/2011 16:00:48	Parker house porch, Simon Pearce parking lot, Quechee	MT @hevrose: #loc Parker house porch, Simon Pearce parking lot, Quechee #VTIrene #Hurricane #Irene #photo http://t.co/oEbxdZh	NA
188	#road	block	8/28/2011 15:58:43	NA	#road block #hurricane #Irene http://t.co/TDpoqD3	NA
189	#flood	parts of a condo complex flooded out in	8/28/2011 15:56:13	Killington (Rutland county)	MT @twc_hurricane: #VT flooding (NWS): #flood parts of a condo complex flooded out in #loc Killington (Rutland county) #irene	NA
190	#closed	continues to be closed with a manned police roadblock	8/28/2011 15:55:37	Dune Road in Hampton Bays	MT @News12LI: #loc Dune Road in Hampton Bays #closed continues to be closed with a manned police roadblock #Irene	NA
191	#road	Rt 81 closed btwn CR 67 & School House Rd	8/28/2011 15:54:38	Rt 81 and CR 67, Durham	MT @TtTLAPorter: #NYirene #src @news10traffic: #road Rt 81 closed btwn CR 67 & School House Rd #loc Rt 81 and CR 67, Durham	NA
192	#treesdown	tree/power lines down	8/28/2011 15:53:04	Rt 81 closed btwn CR 67 & School House Rd in Town of Durham due to	#NYirene #src @news10traffic: #loc Rt 81 closed btwn CR 67 & School House Rd in Town of Durham due to #treesdown tree/power lines down	NA
193	#need	DONATIONS STILL NEEDED: wire crates, stainless steel bowls, blankets & towels!	8/28/2011 15:51:50	NA	MT @aspca: #need DONATIONS STILL NEEDED: wire crates, stainless steel bowls, blankets & towels! #contact @NYCACC nycacc.org #Irene	
194	#roads	NA	8/28/2011 15:50:55	NA	near location of this tweet. #roads #Irene	NA
195	#photo	Wow, look at the color of the #water!	8/28/2011 14:13:45	Windham, Greene County, NY	RT @cbdawson: Wow, look at the color of the #water! #flood RT @DonMeltz: Windham, Greene County, NY #catskills #irene twitpic.com/6cpoj3	NA
196	#flood	NO DAM BREAK-just water cresting	8/28/2011 14:10:15	Gilboa, NY	#irene #flood NO DAM BREAK-just water cresting #loc Gilboa, NY #NY 42.401657,-74.44582 #date 8.28 #time 2:09pm EST #src @810WGY	NA
197					#irene #flood These cars are under water yfrog.com/gyl1b2hj	

Figure 3. TtT Spreadsheet for Hurricane Irene

4.3.5.2 TtT Google Map

I have also provided a TtT Google Map with each instance, beginning in May 2010 with the deployment for the Deepwater Horizon Oil Spill. The TtT Google Map is an HTML and Javascript web application that uses the Google Map API to render the map and the Google Spreadsheet API to read the data for the map from the public TtT Google Spreadsheet. The map uses the GPS coordinates in the processed tweet records to geolocate the icon for each report, and assigns a color to the icon associated with the accompanying report category.

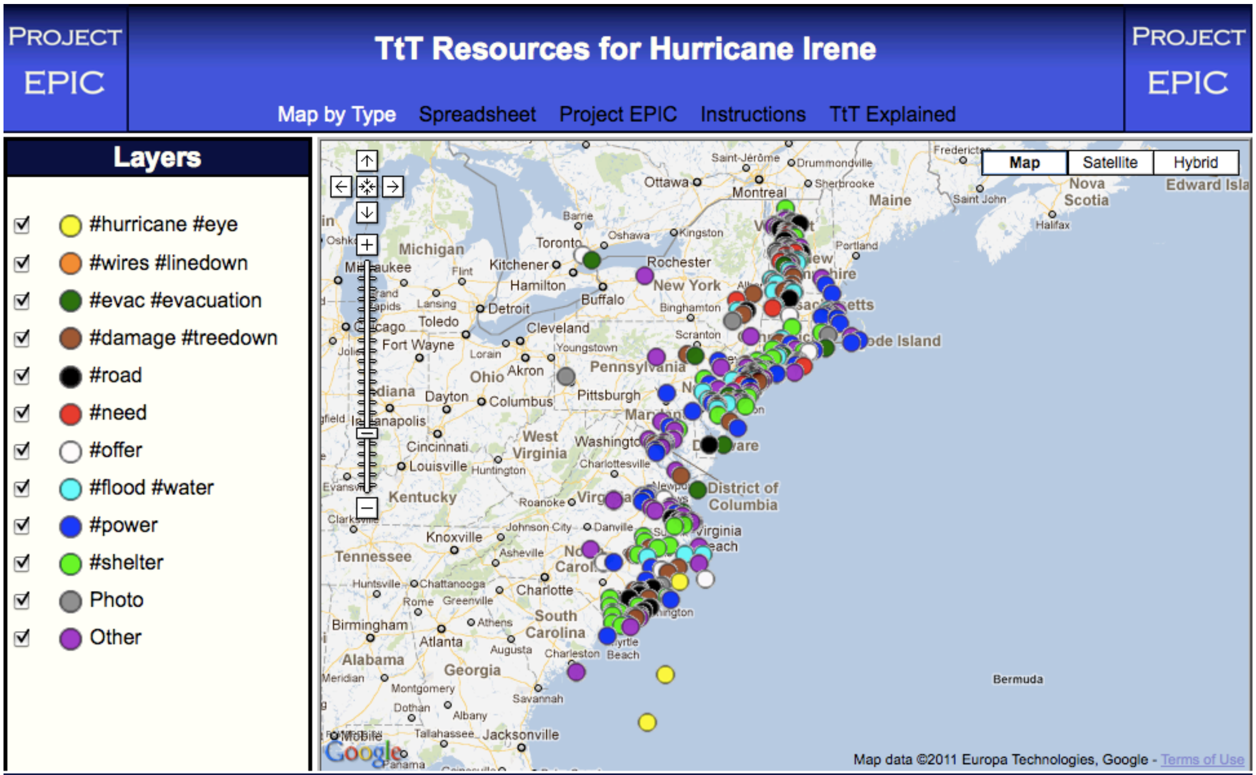


Figure 4. TtT Google Map for Hurricane Irene

4.3.5.3 GeoRSS Feed and CSV File

To create TtT data feeds that are compatible with and can be easily incorporated into other resources, for certain events I have provided a GeoRSS Feed and a public CSV file with all of the TtT records from the internal database, those these feeds have not reflected edits to the public Google Spreadsheet.

4.4 Deploying Tweak the Tweet

With help from my colleagues at Project EPIC, I led deployments for Tweak the Tweet for dozens of events in 2010 and 2011. Table 1 lists 20 of these deployments, ones that witnessed measurable participation from people outside our research group. In the following sections of this chapter, I first discuss in detail the mechanics of the TtT deployment for the 2010 Haiti Earthquake and the lessons learned from that initial deployment about syntax adoption and use. Later, I describe significant features and findings from deployments for four other events (those bolded in Table 1): the 2011 Chile

Earthquake; the Deepwater Horizon Oil Spill in 2010; the 2010 Fourmile Canyon Fire near Boulder Colorado; and the Joplin tornado in May 2011.

By juxtaposing event details and deployment strategies with analysis of Twitter data collected during these events, this reporting assesses many of the underlying assumptions of the syntax. Though some remain open questions, I will present findings from these deployments that offer evidence both supporting and challenging the following statements, core assumptions that constituted the early rationale for Tweak the Tweet:

1. Because TtT tweets are public, anyone can and therefore *someone will* build tools to process TtT tweets.
2. Prescriptive tweets will be an effective way of teaching the syntax and driving adoption.
3. Syntax users will be affected people.
    a. Twitterers will be able to learn and then use the syntax during an event.
4. Leveraging an existing social media platform will have multiple benefits.
    a. It will be easy to build the infrastructure for collecting and processing TtT tweets.
    b. Users will be motivated to share information in TtT by mechanisms of social and symbolic capital.

4.4.1 TtT for the Haiti Earthquake: Examining Underlying Assumptions of TtT

On January 12, 2010, at 16:53 local time (EST), a 7.0 magnitude earthquake shook the country of Haiti, causing catastrophic damage. Hundreds of thousands of lives were lost with many more casualties and an estimated 1.5 million people were displaced (New York Times, 3 September 2010). In the early aftermath of the quake, with the country's already-vulnerable infrastructure almost entirely destroyed, thousands of people were reportedly trapped in the rubble of collapsed structures, electricity and phone services, where available before, were down, and while foreign rescue teams attempted to reach trapped victims, relief agencies struggled to meet basic needs such as food, water, shelter, and medical care for other affected and displaced people. The New York Times (3 September 2010) referred to the relief effort as a “logistical nightmare” due to the severity of need and the wide-scale destruction of the existing infrastructure.

Event	Onset Date	TtT Collection Periods
Haiti Earthquake	**Jan 12, 2010**	TtT: Feb 12 – Feb 25
		Research: Jan 14 – Feb 1
Chile Earthquake	Feb 27, 2010	TtT: Mar 1 – Mar 25
		Research: Feb 26 – Mar 15
Eyjafjallajokull Volcanic Eruption	Apr 2010	TtT: Apr 17 – Apr 22
Deepwater Horizon Oil Spill	**Apr 20, 2010**	TtT: May 31 – Aug 9
Pakistan Floods	Late July 2010	TtT: Aug 15 – Sep 19
Fourmile Canyon Wildfire	**Sept 6, 2010**	TtT: Sept 6 – Sept 16
Hurricane Tomás	Oct 31, 2010	TtT: Oct 30 – Nov 8
San Francisco World Series Riots	Nov 1, 2010	TtT: Nov 1 – Nov 2
California Flooding	Dec 2010	TtT: Dec 17 – Dec 23
Astoria Fire	Dec 16, 2010	TtT: Dec 19 – Dec 21
Extreme Weather in Winter 2010-2011	Dec 2010 – Mar 2011	Several Windows
Queensland Floods	Jan 2011	TtT: Jan 12 – Jan 25
Cyclone Yasi	Feb 3, 2011	TtT: Feb 1 – Feb 7
Christchurch Earthquake	Feb 22, 2011	TtT: Feb 22 – Feb
Golden CO Fire	Mar 20, 2011	
Alabama Tornadoes	Apr 25, 2011	TtT: Apr 28 – May 12
Joplin Tornado	**May 22, 2011**	TtT: May 29 – Jun 13
Hurricane Irene	Aug 27, 2011	TtT: Aug 24 – Sept 5
Tropical Storm Lee	Sept 2, 2011	TtT: Sept 2 – Sept 10
Texas Fires	Sept 5, 2011	TtT: Sept 6 – Sept 16

Table 1. Tweak the Tweet Deployments

Though Project EPIC conducts research in disaster studies year-round, our move to assist in the Haiti event was based on a desire to help through the one means we had available—Tweak the Tweet—not with the intent to use the event to conduct research on the proposed structured data format. In the end, we learned a great deal about the deployment of a TtT instance as well as the underlying rationale. Within hours of the earthquake, CrisisCommons[45], an emergent organization comprised of a coalition of largely technology-oriented individual, governmental and NGO volunteers aimed at assisting in crisis response,

[45] http://crisiscommons.org/

began to organize conference calls. The group invited our research team to participate, and during one of those calls, on January 14, someone listening in suggested the possibility of using Tweak the Tweet. Afterwards, our research team met locally to consider the moral and practical implications of operationalizing the nascent idea to assist in the Haiti relief efforts.

Our team was concerned about the readiness of the TtT syntax. The syntax had yet to go through a single round of usability testing. Our only use scenario was based on an affected person tweeting from the ground; we did not know how easy it would be to learn the syntax, especially for a person experiencing the stress of a crisis event. Also troubling was at that time, there was no infrastructure in place to *process* the tweets, though we were hopeful that the CrisisCamps[46] efforts planned for later that week would produce tools for digesting the data. ICT access was another issue. Even prior to the earthquake, very few people on the ground in Haiti were connected to Twitter, and use of other networked computing services was also already limited. Would TtT be privileging the needs of people who had the means to access social media to the detriment of others who could not promote their needs in this way? And finally, as researchers, we struggled with the idea of taking an active role in an event and that an intervention of this kind would change the social landscape of the phenomena we were trying to study, and without certainty that it would be for the better.

After some debate, we concluded that the potential gains for the affected area outweighed the risks. We rationalized that the 'users' would probably not be affected Haitians communicating their needs, but international relief workers who could use the format to coordinate response efforts. We did not know if it would work, but we felt that if TtT could possibly help make the situation better, even for a single person, then it would be worth the effort to 'deploy' the syntax. At 19:51 EST on January 14, researchers at University of Colorado's Project EPIC[47] officially deployed Tweak the Tweet for the Haiti earthquake relief efforts, in collaboration with parallel efforts at CrisisCamps. A notable result of this decision was

[46] CrisisCamps are bar-camp style events organized by CrisisCommons to improve disaster response by connecting responders and other experts with volunteer programmers.

[47] Project EPIC: Empowering the Public with Information in Crisis. http://epic.cs.colorado.edu/

that the deployment itself was a multi-faceted and -staged project that required significant time commitment by seven researchers at Project EPIC, working in a variety of roles.

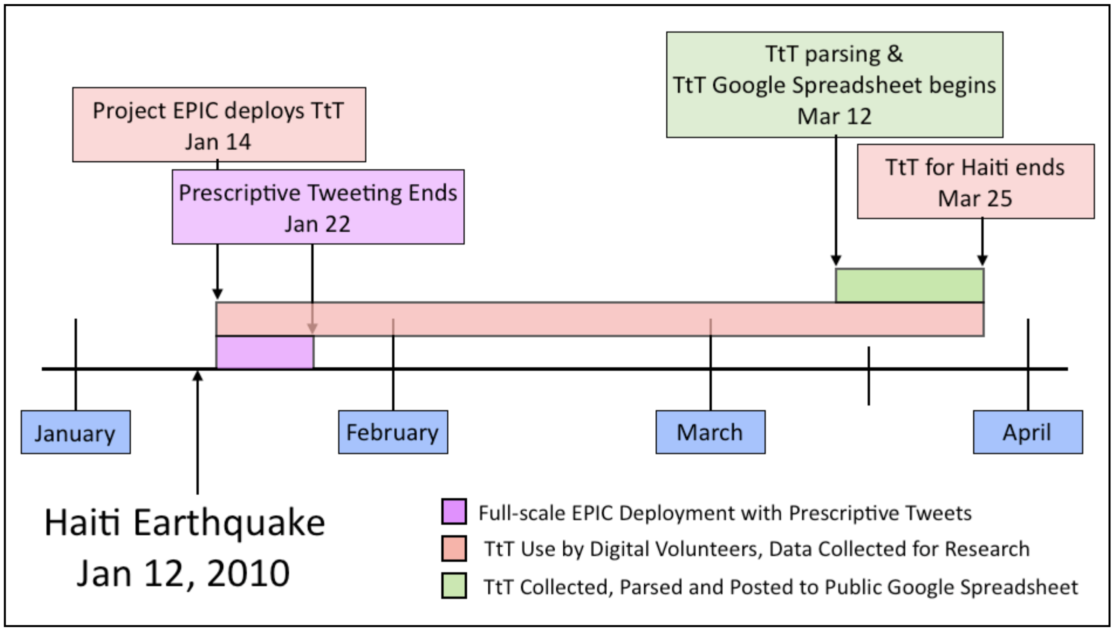


Figure 5. Timeline for TtT Deployment for Haiti EQ

4.4.1.1 Haiti Earthquake Deployment Strategy

This section details the multi-faceted deployment strategy developed and carried out by Project EPIC researchers during the early aftermath of the Haiti Earthquake.

Prescriptive Tweeting

Following the initial concept as outlined by Starbird and Stamberger (2010), we began our TtT deployment for the Haiti Earthquake by tweeting out *prescriptive* tweets that modeled the syntax:

```
#haiti pls tweet in format: #haiti #offering [list offers] #loc [location]
#num [amount] #contact [@ or #]

#haiti use 1 main hashtag per twt: #imok, #ruok, #need, #offer or #have,
#open [road, store or other], #close [road, store or other]
```

To distribute the prescriptive tweets, we activated a Project EPIC Twitter account (@epiccolorado) and had researchers use their own personal accounts as well to bootstrap the creation of an audience.

From the @epiccolorado account, automated scripts broadcast the prescriptive tweets at regular intervals—rotating through a sequence of tweets, sending one every ten minutes. To reach as broad an audience as possible, and following the initial TtT suggestion for using formal response and mainstream media to distribute the messages, we direct-addressed the accounts of journalists and media outlets, response agencies, and other influential Twitterers, requesting that they retweet the prescriptive tweets. Table 2 shows the number of prescriptive tweets and retweets, as well as the percentage of prescriptive tweets distributed by Project EPIC researchers, and Figures 6-8 illustrate the distributions of prescriptive tweets over time in comparison to other types of TtT tweets.

Tweeting "Example" Tweets

On the second day of the deployment, January 15, several Project EPIC researchers began to tweet out "example" tweets to demonstrate the format in use. The tweets contained actionable information found in other places online, including other tweets, blogs from NGOs on the ground in Haiti, and email listservs. These example tweets would end up having a significant effect on both TtT adoption and TtT use. We later changed our terminology for this behavior, shifting from calling them *example* tweets to *translated* tweets. Table 2 and Figures 6-8 show the distributions of translated tweets sent by Project EPIC researchers and other Twitterers, over the whole set and as they varied over time.

Garnering Media Attention

In a second component of our deployment strategy, we attempted to garner media attention for the syntax to increase its visibility, encourage adoption, and let potential technical volunteers know that we needed help to build the infrastructure to support TtT processing. We sent targeted prescriptive tweets, informational tweets about the syntax, and tweets with links to our online resources (described below) to media outlets, journalists, and humanitarian response agencies. The University of Colorado also released a press release about our deployment efforts.

Collection Stats	
Jan 14 – Feb 1 (Research Collection Period)	Feb 12 – Feb 25 (Spreadsheet Support)

TtT Tweets							
Feature	**Total**	**RTs**	**RT %**	**Feature**	**Total**	**RTs**	**RT %**
TtT Tweets	5016	2030	40.5%	TtT Tweets	77	14	18.2%
Prescriptive Tweets	1061	154	14.5%	Prescriptive Tweets	2	0	0%
Translated	3668	1807	49.3%	Translated	61	14	23.0%
Original	99	53	53.5%	Original	0	0	na
Altered Form of TtT	188	16	8.5%				

Twitterers					
Feature	**Total**	**# EPIC**	**Feature**	**Total**	**# EPIC**
TtT Twitterers	295	6	TtT Twitterers	19	1
Prescriptive**	21	5	Prescriptive**	1	1
Original**	10	3	Original**	0	0
Translators**	79	5	Translators**	11	1
RTers**	255	3	RTers**	8	0

Reports			
***	**Feature**	**Total**	**# EPIC**
	TtT Reports	25	1
	Original	0	0
	Translated	25	1

Feature	**Total**	**% EPIC**	**Feature**	**Total**	**% EPIC**
TtT Tweets	5016	38.5%	TtT Tweets	77	3.9%
Prescriptive Tweets	1061	83.8%	Prescriptive Tweets	2	100.0%
Reports			Reports	25	4.0%

Table 2. Haiti Earthquake 2010, TtT Statistics

*Altered form of TtT syntax.
**Some Twitterers tweeted multiple types of reports – so they are counted in each place.
***TtT infrastructure not yet set up to create reports from TtT tweets.

4.4.1.2 TtT Resources for the Haiti Earthquake

Researchers at Project EPIC created and made available several resources to support Tweak the Tweet during the Haiti Earthquake deployment, and other technical volunteers, who connected to the project through CrisisCommons or learned about it through our efforts to promote it, developed applications that collected and processed TtT tweets as well. This section describes some of these available resources, divided into those that helped teach the syntax or enabled syntax use and those that processed and published TtT tweets.

Teaching the TtT Syntax

When we deployed the syntax, we wanted it to be available for use by all volunteers and responders, and this required translating the hashtags into several different languages, most importantly French and Haitian Creole. Using Twitter, we recruited several volunteers to help us translate the syntax, the prescriptive tweets, and a page of explanation and instructions, using a public wiki to facilitate this work. Soon, the wiki became a place where multiple volunteers collaborated, both to assist in translating activities and to help clarify the instructions across all languages. Wiki users also made suggestions about which hashtags should be included as main categories and data tags. This wiki became the primary set of instructions outside of our prescriptive and example tweets. Later, as we recognized that a core group of Twitterers were participating in a different type of TtT "translating" activity—described in much more detail below—Project EPIC researchers recorded and posted an instructional YouTube video demonstrating how to translate a regular tweet into a TtT tweet[48]. Project EPIC researchers also designed and developed a Tweak the Tweet Editor, a web-based syntax editor that helped users create TtT tweets. Using a form structure, that application guided the user in selecting a report category and filling in dynamically provided secondary data fields, and then generated a TtT tweet from the form.

Processing TtT Tweets

Several tools that supported TtT use were brought online during the deployment. The first tool, developed by volunteers at a CrisisCamp on January 16, was a Twitter account that automatically retweeted every TtT tweet. Unfortunately, this tool suffered from consistently hitting the Twitter rate limit for tweet volume over time and was repeatedly blocked as spam. During the first week after the event, Project EPIC researchers developed software to collect TtT tweets and post them to an RSS feed that they provided to the Sahana Foundation[49], a humanitarian response organization with a technical focus. Long after the acute emergency period, but at a time when relief activities were ongoing in mid-March, we

[48] http://www.youtube.com/watch?v=RQhWMzG7d9w
[49] http://sahanafoundation.org/

launched the first version of the TtT processing infrastructure (described in Section 4.3), which posted processed tweet records to a public TtT Google Spreadsheet[50].

4.4.1.3 Data Collection for TtT Use during Haiti Earthquake

Twitter Data

We used two separate collection strategies for the Haiti Earthquake event: a Project EPIC collection during the emergency period (Jan 14 – Feb 1) served the dual purpose of capturing TtT tweets for the RSS feed (described in the section above) and generating the research data set; and we deployed a separate collection for TtT processing to the Google Spreadsheet from March 12 – March 25.

Our Project EPIC data collection strategy employed a two-part collection protocol derived from our study on Twitter use during the 2009 Red River Floods (Starbird et al., 2010). First, we ran a collection script to capture tweets. During the active deployment of Tweak the Tweet (Jan 14 – Jan 24), we used the Twitter Search API to identify and collect all TtT tweets—i.e. tweets that contained both a Haiti event tag (`#haiti`) and one of the specialized TtT hashtags (e.g. `#loc`, `#contact`, `#need`, etc.). In the second part of our collection protocol, we identified each account that had contributed a tweet to the keyword-based collection, and then executed a back-in-time search using the Twitter REST API to capture their entire user stream (their *contextual stream*). On February 1, we collected the contextual streams for the *research collection period* (Jan 10 – Feb 1) for every Twitterer who sent a TtT tweet during our active deployment. This collection, the *Haiti TtT Contextual Streams*, contains 339 Twitterers and 292,928 tweets.

During a first round of analysis, we manually classified each tweet that appeared to be in TtT syntax as one of the following:

- *original* – sent by a user to describe his/her own needs
- *translated* – information available elsewhere on the Internet that was translated by this user into TtT

[50] http://bit.ly/Haiti-TtTweets

- *altered form TtT* – user appeared to be trying to use TtT or something like it, but is not strictly following the syntax rules
- *not TtT* – tweet contained a TtT tag, like `#need`, but is otherwise not in TtT

We also coded each Twitterer according to their overall use of TtT syntax:

- *original TtTer* – any Twitterer who sent one or more original TtT tweet
- *TtT translator* – any Twitterer who created and sent the first version of at least one translated TtT tweet
- *TtT retweeter* – an account that only retweeted TtT tweets, but did not create any TtT tweet of its own
- *not a TtTer* – an accounts for whom the collected TtT tweets were all *altered form TtT* or *not TtT*

Table 2 contains the distributions of these tweets and Twitterers across these categories.

Interviews

Through this early analysis of the Twitter data collected for TtT users, we identified 84 Twitterers (outside our research group) who created and sent TtT tweets—ten of these Twitterers were *original* tweeters who had crafted TtT tweets containing their own needs and 74 were *TtT translators* who had translated information they found elsewhere into TtT format. Attempting to learn more about syntax use as well as the broader digital volunteer behavior we were seeing (discussed in detail in Study 2, Chapter 5), we elected to do a follow up interview study with *TtT translators,* Twitterers who were creating new tweets using the syntax but who had not been directly affected by the earthquake.

We successfully contacted and interviewed 20 of these TtT translators during July and August of 2010. The interviews were conducted via email for all but participant, with whom we completed a phone interview. For the email interviews, each participant was sent a message with the same set of open-ended questions. For the phone interview, these same open-ended questions guided an hour-long verbal exchange. Though this interview instrument consisted of six sections, the reporting in this chapter will focus on the responses from the one section that focused exclusively on TtT use. Study 2 (Chapter 5) will examine the whole of the interview data, which covers a much larger range of digital volunteer activities.

4.4.1.4 Analyzing Deployment Strategies

The original Tweak the Tweet idea positioned prescriptive tweeting as the primary mechanism for diffusing the microsyntax. Most of our interview respondents, users who used TtT at least once in a *translated* tweet, did assert that they were first exposed to TtT in the content of someone else's tweet. However, it is not clear whether prescriptive tweeting was the primary route of TtT diffusion. Tracking the movement of prescriptive tweets and adoption rates in comparison to some of our other deployment strategies challenges this foundational assumption in multiple ways.

Tracking Diffusion of Prescriptive Tweets

Using a combination of algorithmic and manual analysis, I traced the tweet-retweet chain for all 1061 prescriptive tweets identified in the set, and found that prescriptive tweets did see moderate diffusion within the Twitterverse. Project EPIC researchers sent 854 prescriptive tweets and retweets and these were retweeted 131 times. This indicates about one retweet for every 6.5 prescriptive tweets. Interestingly, Twitterers who were not EPIC researchers sent 43 tweets and received 17 retweets, about 1 retweet for every 2.5 prescriptive tweets. This shows that prescriptive tweets by other Twitterers were much more likely to be retweeted than prescriptive tweets by EPIC researchers.

This effect could relate to several issues, including differences in network strength and visibility for EPIC accounts. It is possible that EPIC researchers had fewer followers during this event than other Twitterers who were retweeted for prescriptive tweets. Importantly, the @epiccolorado account sent its first tweets (ever) out *after* the event, and had fewer than ten followers at the time of its first Haiti-related tweet. However, EPIC had a retweet rate that was slightly higher than its individual researchers who, on average, had higher and in some cases much higher follower rates at the beginning of the event. This suggests that network strength was not the only factor in retweet rates for prescriptive tweets.

Another possibility is that the bot-like nature of the Project EPIC Twitter account and the high volume of EPIC prescriptive tweets had a negative impact on the likelihood of those tweets being retweeted. Accounts that repeatedly tweet out similar information are often referred to as bots, because many of them are operated in whole or part by computational algorithms (Singel, 2011). These accounts can develop a low reputation on Twitter and even be designated as “low quality” or “spam” by the Twitter platform and subsequently removed from search results. While the latter possibility here (blocking by Twiiter) did not occur, during the TtT deployment @epiccolorado, the account that generated over three quarters of *all* prescriptive tweets, was using a bot-like mechanism to distribute them and this could have discouraged retweets of their prescriptive tweets over time. However, Figure 6, which plots EPIC tweets vs. prescriptive tweets temporally, shows a relatively stable rate of retweets for tweets over time, indicating that the lower rate for EPIC was not a case of diminishing returns.

The evidence here is not enough to be conclusive, but suggests that network effects, tweeting patterns and other measures of influence affect diffusion of prescriptive tweets.

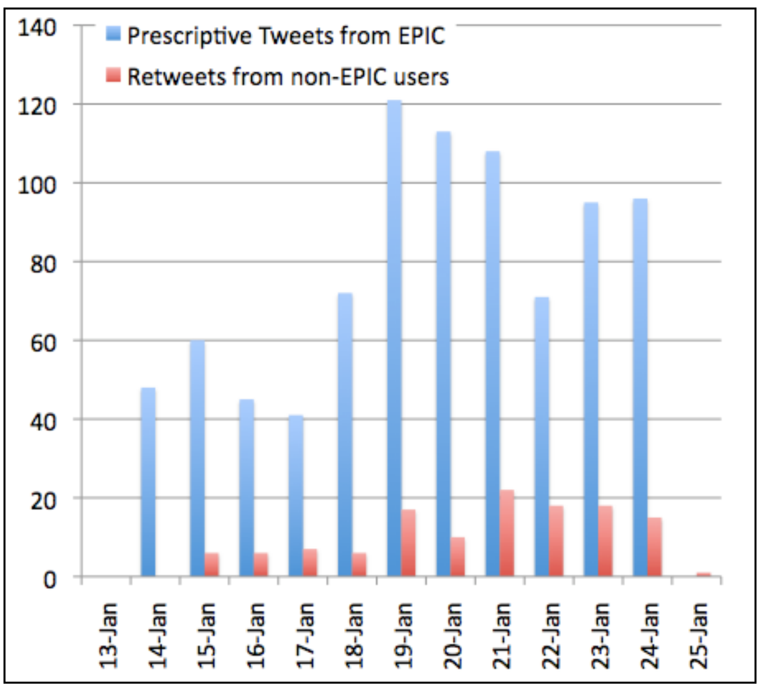


Figure 6. Diffusion of Prescriptive Tweets: EPIC Prescriptive Tweets vs Crowd Retweets

Connecting Prescriptive Tweeting to TtT Use

Another way to measure of the effectiveness of prescriptive tweeting is to analyze its relationship to TtT use—i.e. original and translated TtT tweets. If TtT use rises and falls at the same time or slightly after prescriptive tweeting, it would follow that prescriptive tweeting likely had an impact on TtT use, especially if other factors can be eliminated. Figure 7 shows the impact of two different tweet strategies over time, comparing the volume of prescriptive tweets and EPIC translated tweets to TtT tweets (not retweets) sent by Twitterers who were not EPIC researchers, and indicates that prescriptive tweeting is not a singular factor in TtT adoption and use.

On January 15, the second day of prescriptive tweeting, users outside the EPIC research groups sent their first tweets using the TtT syntax. Both of these were original, `#ruok` tweets sent by individuals outside of Haiti looking for information on missing persons.

> **@smiley424:** #haiti #ruok #name Esther Bonhomme #loc PORT-AU-PRINCE #contact @smiley424

> **@vasselli:** #haiti #ruok #name JeannD'arc Noralus #loc St François de Salle #num 36215306

TtT use increased slightly on January 16 to 13 TtT tweets—one original and twelve translated. The most dramatic increase in TtT use occurred the next day, January 17, when 155 TtT tweets were sent. Of these, all but one were translated tweets. TtT use then fell back from January 17's highs, and began to fluctuate around an average of 96 TtT tweets per day between January 18 and January 22. After this, TtT use dropped to what might be considered maintenance levels and continued between 20 and 60 TtT tweets per day until the end of our research collection on February 1.

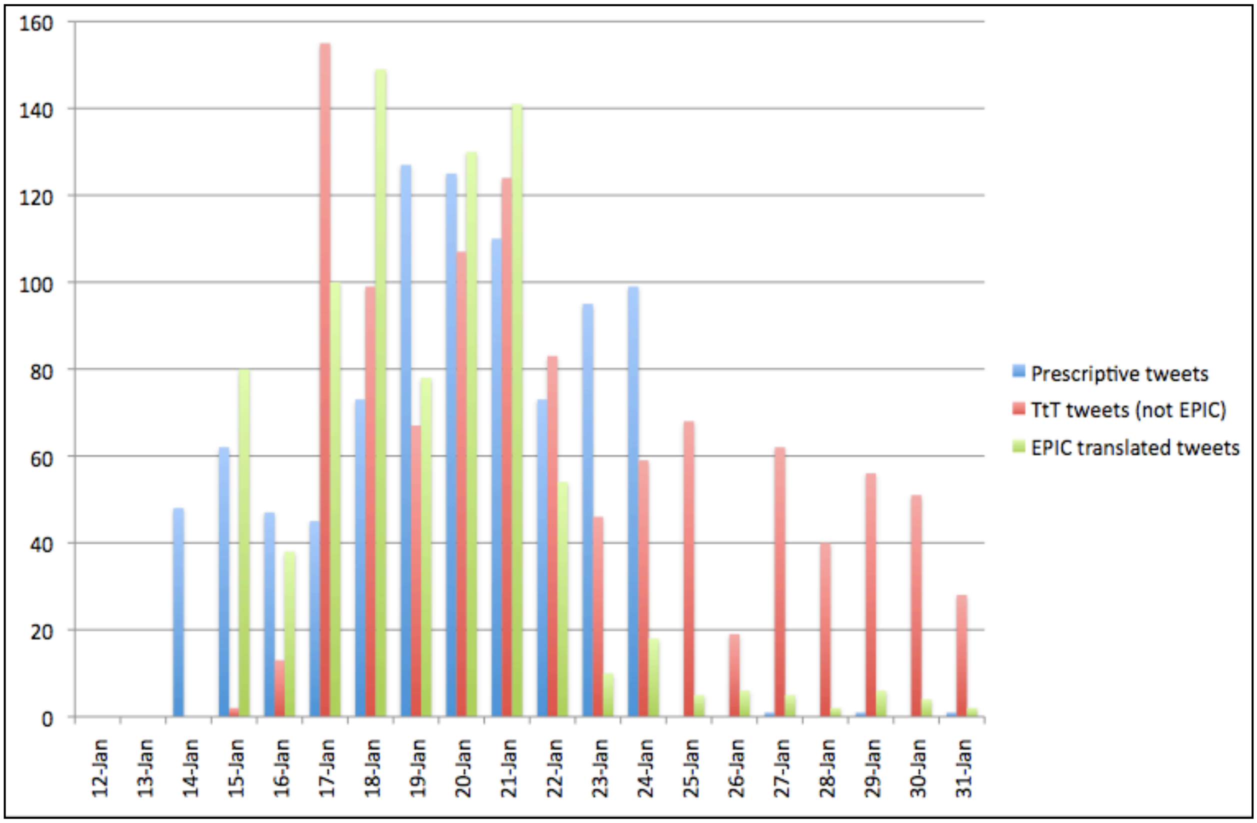


Figure 7. Volume of Prescriptive, TtT, and EPIC Translated Tweets Over Time

Comparing the increases and decreases of TtT use against prescriptive tweeting reveals an incomplete correlation between the two. For example, the rise in TtT tweeting that takes place on January 17 occurs while prescriptive tweeting remained steady at just over 40 prescriptive tweets per day. Conversely, a permanent decrease in TtT use occurred on January 22 and January 23, at a time when prescriptive tweeting had fallen some, but before prescriptive tweeting stopped completely after January 24. The complete cessation of prescriptive tweeting had no immediate impact on TtT tweeting volume.

This analysis does not show uniform correlation between prescriptive tweeting and TtT use, indicating that prescriptive tweeting may not be the single force behind TtT adoption during the Haiti TtT deployment.

Connecting Translated Tweeting to TtT Use

TtT adoption and use by Twitterers outside of EPIC corresponds much more closely to changes in another EPIC deployment strategy, that of *TtT translating*. On January 15, EPIC researchers began to send out example tweets of actionable information translated into TtT syntax. Translated tweets by EPIC researchers and others then diffused via the retweet through the Twitterverse, doing so at a much higher rate than prescriptive tweets—49.3% of all translated tweets are retweets, about one retweet for each translated tweet. Original tweets were retweeted at an even higher rate—one retweet for every 0.87 original tweet.

The correspondence between the drop in EPIC translating activity and overall TtT use is the clearest indicator of a relationship between the two. On January 22 Project EPIC greatly reduced the number of translated tweets that we sent out, from 141 the day before to 54, and on January 23 researchers almost entirely ceased translation activity for the remainder of the research collection period[51]. These drops correlate closely to the decrease in TtT use by other Twitterers between January 21 and January 23, from over 120 to less than 50

Figure 7 suggests that EPIC translation activity may also have driven early adoption of the syntax. When TtT use was experiencing rapid adoption, between January 16 and January 18, prescriptive tweets remained relatively steady, but EPIC translation increased rapidly. However, this link is weaker than the correlation between the drop in EPIC translation and overall translation during the latter part of our research collection window.

Perhaps a better measure of adoption is the number of Twitterers adopting or using the syntax each day. Figure 8 shows the number of different Twitterers (including EPIC researchers) who were using the syntax each day and the number who started using the syntax that day, compared to the number of prescriptive tweets and EPIC translated tweets sent.

---

[51] After that point, some researchers sent a few translated tweets with new actionable information as part of individual volunteer activities and not part of the TtT deployment strategy.

TtT use does not start at all until Project EPIC researchers begin to send translated tweets on January 15. After that date, the number of TtT Twitterers rises and falls in a very similar pattern to the amount of EPIC translated tweets sent the day before, with some overall growth between January 15 and January 22. Note that after January 17, the number of new adopters does not consistently increase the number of total users, indicating that some Twitterers use the syntax one day and not the next.

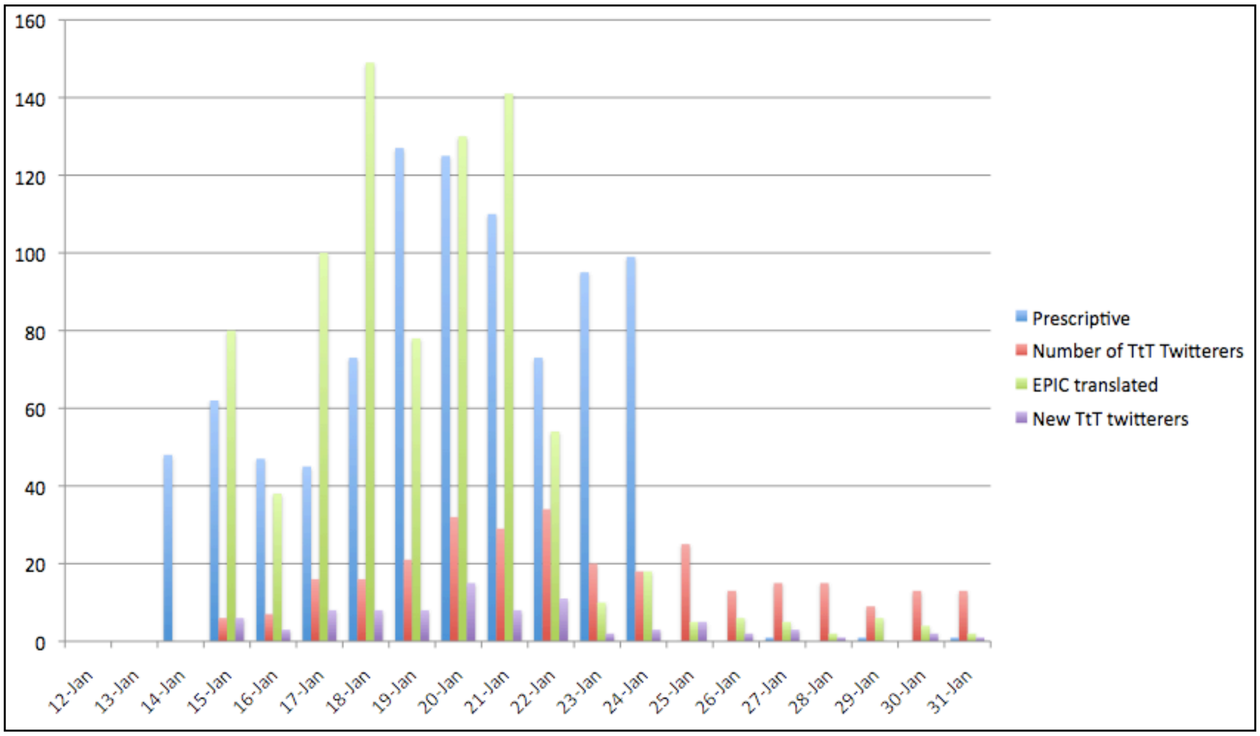


Figure 8. Prescriptive Tweeting and EPIC Translating to Number of TtT Twitterers Over Time

Though the number of Twitterers using the syntax each day appears to be correlated to both EPIC translating and prescriptive tweeting, the number of new adopters of the syntax, like the overall volume of TtT tweets, is tied strongly to EPIC translation activity. When EPIC translating stops on January 22, new user adoption falls permanently from 8-15 for the six previous days to less than 5 for every day after. Significantly, this drop precedes the end of prescriptive tweeting.

Eventually, the number of overall TtT users stabilizes around a low, but consistent number. On and after January 23, there appears to be a core of 10-15 Twitterers who had learned the syntax, incorporated it into their tweeting practice for the Haiti Earthquake response, and continued to use it even though the original advocates of the syntax (EPIC researchers) had stopped promoting it. There are a variety of possible explanations for why other Twitterers tried the syntax out for a short period of time and then shifted away from using it.

Summary of Deployment Strategies

Together, Figures 6-8 indicate that TtT tweet volume is more likely related to translation activity than to prescriptive tweeting. Since the vast majority of TtT activity by those outside of Project EPIC was TtT translating (see Table 2), it is not surprising that the modeling of translating behavior by EPIC researchers and the diffusion of this behavior through the Twitterverse had a greater effect than prescriptive tweeting. It is also not surprising that the innovation was more likely to spread—via the retweet mechanism—when tweets contained actionable information embedded in the format instead of abstract instructions. These findings suggest that a deployment strategy incorporating TtT use in context via translated tweets with actionable information is more viable than a strategy based wholly on prescriptive tweets. However, these findings also indicate that there are more factors at play in determining TtT adoption and TtT use than merely the tweeting actions of EPIC researchers. Importantly, diffusion of the syntax may rely on network effects and TtT user reputations, and it also may be affected by tweet volume over time.

4.4.1.5 Assessing The Field of Dreams Implementation Plan

The original Tweak the Tweet idea relied in part upon a hypothesis that since anyone *could* build applications to process TtT data—because it would be public and available through Twitter's APIs—that someone *would*. Our research on Twitter use during the Red River Floods in 2009 discovered several programmers who were assisted in distributing information about the flood using algorithmically-controlled accounts that tweeted flood heights at regular intervals (Starbird et al., 2010). When we developed the Tweak the Tweet concept, we theorized that similar *innovators* would rapidly develop tools to process TtT tweets once people started incorporating the microsyntax into their messages. In retrospect, this idea could be referred to as the "field of dreams" implementation plan (in reverse)—i.e. if someone comes, they will build it.

Shortly after deploying TtT for Haiti, we recognized a serious flaw in this plan. Without the infrastructure in place to process the microsyntax, it was difficult to convince others to use and promote it. On the other side of that equation, without evidence that people could and would use the microsyntax, it was hard to motivate innovators to build the tools. Though multiple CrisisCamps provided a potential volunteer workforce for development, recruiting programmers to work on the TtT project turned out to be a tough task—there were many other available technology-based projects vying for attention among CrisisCampers.

This problem had two pieces. There was no proof that anyone would use the syntax, and so efforts to build the tools could be in vain. This was a hard prospect for volunteers, and they shared this feedback with me during post-project debriefs. Additionally, the microsyntax itself was very much in flux even as it was launched on January 14. It would go through significant user-driven evolution both in use and through our wiki (see 4.4.1.2, *Teaching the Syntax*), and EPIC researchers would dramatically alter TtT ourselves on January 18, simplifying the structure and removing the amount of hashtags required to make TtT tweets look more "normal" to Twitterers. With a changing syntax, computational tools would need to be adjusted over time, something that did not fit well with the weekend warrior volunteer arrangement of CrisisCamps. In fact, the most sophisticated application for processing TtT tweets[52] came from a technical volunteer outside of the CrisisCamp community, and that developer worked over the course of several days to improve and adjust his processing architecture (personal communication with Simon Twigger).

And so our field of dreams implementation plan fell victim to the chicken and egg paradox, as it relied on a group of initial users that were not yet extant during the early phase of TtT deployment, when volunteers with technical skills were actively seeking opportunities to help. However, some users did

[52] http://tweetneed.org/

experiment with using TtT, and the syntax did experience some adoption, even without the infrastructure to process it and create resources from it.

4.4.1.6 Discovering the TtT Translator

Findings from the initial deployment of TtT for the 2010 Haiti Earthquake also call into question the proposed use scenario of Tweak the Tweet—that of an affected person using the syntax to communicate her own needs during an emergency event. Though dozens of people created and sent out TtT tweets during the event, only ten people used the syntax to report their own needs and of those, only four were reporting from Haiti[53]. The syntax was far more likely to be used by people who were not directly affected by the disaster, but who wanted to help out. These remote volunteers worked to locate actionable information from a variety of sources, then translated this information into TtT syntax and tweeted it out to their followers and the broader Twitter public. This discovery of the *TtT Translator* triggered a re-conceptualization of the TtT use scenario, shifting the primary user persona from an affected person to a digital volunteer. It also led to the discovery of a community of digital volunteers who participated in a range of information processing activities, of which TtT was only a small part.

Though Study 2 (Chapter 5) focuses exclusively on the diverse activities of digital volunteers, the following sections in this chapter (4.4.1.7 and 4.4.1.8) address their motivations for TtT adoption and their strategies for using the syntax to help process information during the response efforts.

4.4.1.7 Investigating Why Digital Volunteers Adopted and Used TtT

Considering that there were few visible resources that demonstrated TtT tweets were being collected and processed, and no indication or suggestion that TtT tweets and the resources derived from them were being directly incorporated into response efforts, it is somewhat surprising that the microsyntax was used at all, and yet it was—Table 2 indicates that TtT had 89 users during the research collection period. Some users even continued to use it after the active EPIC deployment ended, into February and March of 2010.

[53] Others were reporting needs for volunteers or donations in the U.S. and Canada.

Table 2 shows that 19 Twitterers used TtT between February 12 and February 25, a period where TtT was supported with collection, parsing, and a public Google Spreadsheet listing TtT reports.

Our interviews with 20 TtT translators reveal a variety of different motivations that users had for incorporating TtT in their tweets.

> Joey: After reading the website and info, I thought it may be helpful to those reading and sorting information to identify needs. It only made sense!

Like Joey, several respondents claimed that Tweak the Tweet just made sense, that they believed the idea as described within the prescriptive tweets and on the EPIC Website would help make their tweeted information more useful.

> Cindy [describing her reasons for using TtT]: Fast data transfer, analysis and mapping was crucial during this time. Anything that sped up information sharing was great!

> Meg: I completely got the need to create a system that would allow for searchable, categorizable information. I was using Ushahidi 24/7 so I understood the benefit of being able to sort through the MASSIVE amounts of data.

Meg's comment reveals an understanding from personal experience of the difficulty of sorting through the "massive" amount of data moving through different information sources at the time. During the emergency period of the Haiti earthquake, she was participating in a wide range of volunteer activities (these will be discussed in greater detail in Study 2, Chapter 5), and TtT was only a small part of this work. She mentioned in her interview feeling overwhelmed with the amount of data that she needed to process, writing that it was "difficult to keep up with the stream at times." Meg understood the need for filtering and categorization, and therefore Tweak the Tweet made sense to her as a possible solution to a problem she was actively experiencing.

> Karen: I realised other people would be using the Twitter stream in different ways to collate information and so a standardised format made sense

Karen also understood the rationale behind Tweak the Tweet, and mentions in her interview response another aspect of TtT that is echoed by several other participants: Tweak the Tweet provided standardization which they either thought to be useful as described in our materials or perceived to have value of other kinds.

> Mary: Because it standardized everything. It gave a SOP/ Standard Operating Procedure or MO/ Modus Operandi of doing tweets. I thought it made perfect sense. Tweets to me are made of 2 kinds of information. The emotional and factual. Tweets often mix the two to get them to be read and RT. But, so many people were tweeting and so many tweets were duplicated. TtT system made everything look the same. It deleted the emotional. Went for the factual only. Made it systematic.

> Sarah: I choose to tweet with the syntax because uniformity in messaging is essential in high pressure, emergency situations and it was easy to use. I also noticed that people (some volunteers with Ushahidi even) spent a lot of time trying to decipher messages with missing info and what the source of the info was. I felt that if it had been vetted by Ushahidi already or other aid groups then those questions had been answered (or would be) if the messages were uniform and where composed with TtT. In short the time spent looking for missing parts to a message could be avoided by using TtT. I found it useful

> Tara: I realized my impression of those using the syntax was to take them more seriously, so once I had confirmed information it was translated into syntax. It also greatly aided in data entry into Ushahidi and updating the constant stream of information.

In the first interview excerpt above, Mary notes that one reason she liked the TtT format was, she claims, because it removed the emotion from tweets, and left only factual, standardized information. Sarah (above, second excerpt) also cites the uniformity of TtT tweets as an advantage during emergency situations. She notes that the format had a secondary benefit of encouraging users to put all of the relevant information inside a single tweet, and suggests that she used the microsyntax to mark her messages as being "vetted." In the final example above, Tara also indicates that she used TtT as a marker for her own verified information. Her comment suggests that she perceived TtT use as an indication of expertise in the volunteer tweeting space. In the second part of Tara's response, she reveals that she was manually searching, sorting and processing tweets and that TtT made this work easier. Note that none of these three interviewees remark on the usefulness of TtT resulting from the automatic processing that TtT was conceived to enable; all three describe their TtT use as motivated by beneficial, though unintended consequences of TtT use that were not outlined in the initial Tweak the Tweet description.

Interviewees also discussed what they perceived to be limitations to TtT adoption and use. One huge concern was whether or not their messages were being heard and utilized by those in a position to help people in Haiti. In her interview response, Meg commented, "I think ppl need to see how it's used on the other side in order to fully adopt it." Another interviewee echoed that sentiment, stating that in her overall opinion, the syntax was, "Terrific in emergencies as long as it's being monitored; we had no way of knowing if it really was." TtT users and presumably other Twitterers who did not adopt the syntax wanted to know that an agency or agencies were tuning in to this data stream. Like the technical volunteers who voiced concerned over building a system that no one would use, volunteer TtT translators wanted to know that their effort was not wasted.

4.4.1.8 Examining How Twitterers Used TtT during the Haiti Earthquake Response

As discussed above, the vast majority of TtT users during the Haiti earthquake response were TtT translators. Exploring how these Twitterers did this translation work reveals dynamics of connected crowd work at the micro level, in the individual transformations of tweets towards states of increased structure.

The most common form of translation involved monitoring the Twitter stream, identifying actionable information that was not in TtT format, and creating a tweet in TtT format using this information. In many cases, translators were tracking the Twitter stream using web applications and Twitter clients to manually search for tweets that contained the #haiti hashtag.

> Cindy: *Many people* tweeted the #Haiti hashtag, I monitored about 15 hashtags during the height of the crisis, via the TwitBird iPhone application… Hashtags are very important for breaking outside your Twitter walled garden. Without those, everything is closed to your existing contacts.

Some refined these searches to limit the volume of tweets.

> Ellen: I followed the tweets coming out of Haiti and then started doing a twitter search for tweets within 50 miles of Port-au-Prince and continued to search for anything with the hashmark #Haiti.

Others began to identify Twitter accounts that were on the ground and others that they trusted, and used those social networks to find actionable information. In some cases, other Twitterers would identify voluntweeters as potentially useful broadcast points for their information, and sent them requests directly.

> Tara [discussing the sources she used for information she tweeted using TtT]: From unfiltered tweets and Facebook posts, direct requests in mentions to my attention from Twitterers noticing my efforts and emails. Once people saw through my feeds that I was getting results, the information and requests bombarded me. It was then a matter of confirming the information through direct communication, or other sources.

The micro-work of tweet-to-TtT-tweet translation took on a variety of forms. In the example below, voluntweeter @barbaraslavin creates a TtT tweet by moving pieces of information from an existing tweet into a new tweet, and arranging this information to follow and therefore accompany appropriate TtT hashtags, including a main category tag and four data tags.

(Original tweet) @**jeanclaude:** Jocelyne Sampeur-Michaud @ Institut Franse, Ruelle Robin PaP needs food and water for children toddlers and pregnant woman. 2516-9046 #Haiti

(TtT translated tweet) @**barbaraslavin:** #haiti **#needs** food H20 **#name** Jocelyne Sampeur-Michaud **#loc** Institut Franse, Ruelle Robin P-a-P **#info** 2516-9046 **#source** @jeanclaude

* **Bold** added for emphasis of tweet author, **Blue** added to indicate added tags for TtT tweets

In this example, @barbaraslavin changes the order of information from the original tweet and in some places significantly shortens content by abbreviating and eliminating words. Attribution to upstream authors and original sources, along with the addition of TtT tags within the tweet content, add considerably to the character length of tweets, and this forced TtT users to be very concise and in some cases make decisions about what content to eliminate and what to keep to meet the 140-character tweet restriction.

@barbaraslavin credits the source of this information by adding the TtT `#source` hashtag, followed by the Twitter handle of the upstream Twitterer. This is one of two strategies for creating TtT translated tweets. Other TtT translators created *tag-added retweets*, using the RT @ convention to credit the upstream Twitterer, and keeping much of the initial word choice and order intact, but adding tags that the original user had not included in the original tweet, and in some cases truncating content and even moving content around.[54] For example:

> (Original tweet) **@jeanclaude:** I need to get wismond and his family: clothes, food, water and a tent! Call me 34955555 or 36555555

> (Tag-added RT, TtT) **@TamiH68:** RT @jeanclaude **#haiti #need** : Wismond & family: clothes, food, H20 & tent. **#con** [**JC** at 34955555 or 36555555]

In the above example, @TamiH68 creates a tag-added retweet from @jeanclaude's message seeking help for a family in need. To translate the tweet into TtT syntax, @TamiH68 first carries over the text from the initial tweet, possibly by using the retweet function within Twitter or another application. Next, she adds the `#haiti` event tag to the beginning of the tweet, important for having the tweet show up in TtT collections and in general searches for Haiti tweets. Next, she adds a # in front of the word "need" in the original tweet text, and removes a few extra words—"I," "to get," and "his"—to make the tweet shorter and have it conform better to TtT. She also shortens the word "and" to "&." Finally, she replaces the "call me" text from the original message with the `#con` hashtag, which is a shortened TtT hashtag meant to designate contact information. It is now positioned in front of the phone numbers @jeanclaude listed in his original tweet. Finally, she adds "KJJ," a shortening of @jeanclaude's name, in front of the phone numbers to indicate that listeners should contact him to follow up. @TamiH68 also changes some of the punctuation and spacing in the tweet.

There were also several examples where users merely added a single tag (like `#need`) to a tweet in a format that was not fully TtT—and perhaps not even inspired by TtT—but nonetheless helped to mark

[54] Retweets that modify content in this way are now sometimes called "modified tweets," indicated by using "MT @" instead of "RT @"

that tweet as belonging to a particular category. In the example below, @Box_Timer retweets a non-TtT tweet by @paulriner, adding the `#need` tag before the retweeted text to indicate that the tweet has been designated as containing information about an actionable need.

> **@Box_Timer:** **#need** RT @paulrhiner: Sis of St Joseph de Cluny need food/water/med supplies Urgently at Ste Rose de Lima sch on Ave John Brown. #HaitiRT

Though this user was not classified as a TtT translator, because the account owner did not fully incorporate TtT into his or her tweets, @Box_Timer added tags to six different retweets—marking five tweets as `#need` and one tweet as `#help`.

Some TtT translators created synthesized TtT tweets by collating pieces of information from multiple tweets or other sources. At times, they used direct messages or other routes of communication to trace down all of the necessary components for creating a TtT tweet—i.e. report, location, contact information, and source. In some cases, like the example below, they used mapping services and consulted with people who had local knowledge to find GPS coordinates for locations, and added the coordinates to the `#loc` (location) field within the text.

> **@francest:** my name is Orel Ducasse. i am dehydrated and in need of food and water. Please:Rue lamarre impasse Mouzin next of Cine Capi ...

> @Mission_Peace: #rescuemehaiti #Loc Latitude: 18.54298 / Longitude: -72.343102 next to Cine in Port Au Prince #haiti #contact Orel Ducasse #need food, h20

Like Tara, who is quoted in the section above (4.4.1.7) as saying that she used the syntax as a marker for confirmed information, many TtT translators worked to verify information before tweeting it out using the syntax.

> **@coffeegal** (2010-01-20 11:24): @Ayiti12345 Are these 100 people still trapped? or stranded sans med/water/food?

@coffeegal (2010-01-20 11:48): #haiti #stilltrapped #loc Eglise ST. Antoine, Nazon. #num 100 #src @Ayiti12345 via Signal FM Radio #date 20 11amEST

In the example above, a TtT translator attempts to confirm a report of 100 people still trapped. This information request did not receive a response in the *Haiti TtT Contextual Streams,* though it is possible that @Ayiti12345 or another Twitterer outside that dataset responded, or that someone responded through a DM or another channel. In any case, a few minutes later, @coffeegal decides to tweet the report out in TtT format (though the #stilltrapped tag was not a standard TtT tag).

Some voluntweeters who knew about TtT could not always find time to use the format correctly and began to rely on others to translate their tweets of actionable info into TtT format. Meg, commenting on this, "There were some ppl that were REALLY good at it and they would catch many of our forgotten tweets and redo them. Sometimes we were going so fast that it was hard to remember to use it." To some extent, a layered system with some division of labor arose, with some Twitterers identifying and tweeting actionable information and others, in a second step, helping to structure that information.

As Mark shares below, Twitter was not the only source of information for TtT translation:

> Mark: In the first days after the earthquake, we were trying to find tools, blogs, sites, Twitter users, who had any information that would help our students follow what was happening on the ground in Haiti.

Additionally, the information in many TtT tweets can be traced to reports in the Ushahidi platform. Some TtT users created all or almost all of their TtT syntax tweets using information from Ushahidi: 90% of Meg's TtT tweets were originally Ushahidi reports. Conversely, many tweets went into TtT format and then became Ushahidi reports when another volunteer entered the tweeted information in the web portal for that platform.

@MijaFlores (before 2010-01-20 15:29:58): @Jim_P_Edwards have to report another forgotten orphanage BBDF 38 Joseph GUIDE Clercine 22 Tabarre food meds water RT pls

There were nine retweets of this tweet in the *Haiti TtT Contextual Streams* between 15:29 and 16:00. A little over an hour after the initial tweet, this information was translated into TtT format. A few hours after that, it appears within a tweet that links to a related Ushahidi report.

> **@lwestinmark** (2010-01-20 16:49:39): #haiti #need food meds water #src @MijaFlores #info another forgotten orphanage #loc BBDF 38 Joseph GUIDE Clercine 22 Tabarre

> **@rojosalvo** (2010-01-21 02:50:06): #haiti Bon Berger de Fatima Orphanage: BBDF 38 Joseph GUIDE Clercine 22 Tabarre EVENA DESHOMMES. TELEPHONE: 3470-4803... http://dlvr.it/fzP

TtT tweets also contained information originally found in blogs, media reports, Facebook posts, and NGO websites, as well as information learned from personal communication between the TtT Twitterer and an affected person.

4.4.1.9 TtT as Self-Organized Human Computation

This description of TtT translation activity during the aftermath of the Haiti Earthquake shows digital volunteers appropriating the tool as a structuring mechanism to improve the quality of their communications, and as a marker for their own expertise in the crisis tweeting space. Individually, these Twitterers were using the syntax to amplify, structure and route packets of actionable information. Collectively, the TtT translators performed as a largely self-organized, human computation system that worked to transform dispersed and disaggregated pieces of first hand and other actionable information into a state of greater organization where information could be easily aggregated, synthesized and even mapped by simple computer algorithms.

One way to approach this collective activity is through the framework of distributed cognition (Hutchins, 1995), which conceives of cognition as distributed across physical objects and social networks. In this framework, cognition happens as transformations of information— in the movements and changes in how information is represented. Examining the digital volunteer behavior described here using a distributed cognition lens positions the network of TtT translators as a singular cognitive system that “thinks” through the actions of structuring tweets—e.g. adding hashtags, reorganizing word order, then rebroadcasting information in the standardized format. This translation work is distributed across the network of connected individuals as well as the tools and resources that these people incorporate in their work processes. Though no single node of the system, and no single user or data packet, contains a complete picture of the entire system, collectively the individual transformations of information representations act to move the system to increased states of organization. Conceivably, TtT processing tools could be seen as components of this cognitive system as well.

TtT use by digital volunteers connected through a social media platform provides a clear example of collective intelligence as distributed cognition. Evidence from other activities of digital volunteers presented later in this work (Studies 2-4, Chapters 5-7) offers further support for using the distributed cognition framework to characterize connected crowd work.

4.4.1.10 Haiti Summary: Lessons Learned from the Haiti Earthquake Deployment

The TtT deployment for the 2010 Haiti Earthquake response provided several important lessons about the viability and usability of Tweak the Tweet (or another microsyntax for crisis reporting) as well as insight into the activities of an emerging crowd of digital volunteers. In regards to our proposed user scenario of an affected person, it turns out to be difficult to get the message out to and to teach a new syntax or tool to those affected while an event is unfolding. A better strategy for an information structuring innovation is to target and train the converging volunteer workforce to use the tool.

These findings also suggest that the broader Twitterverse is more interested in the information distributed within a crisis reporting microsyntax than the rules of the syntax. Adoption, in this case, appears to be less about following instructions than copying the behavior of others in the space. Though sending prescriptive tweets may better explain the rationale and rules for the syntax, tweets with actionable information diffuse further and are more likely to increase adoption.

Another factor that could have limited early adoption of the syntax was the lack of a robust infrastructure for processing the tweets during the early stages of the deployment. Prospective users of the syntax may have been more likely to use the syntax if they could see that their tweets were being processed. Along those same lines, TtT users that responded to our interview requests claimed that an established connection between TtT tweets and response organizations would also have increased their motivation for using the syntax.

4.4.2 Chile Earthquake: Conceptualizing the Deployment as a Campaign

Less than two months after the Haiti quake on February 27, 2010, an earthquake of magnitude 8.8 struck the country of Chile. Though an earthquake-prepared Chilean infrastructure prevented the kind of widespread and catastrophic damage that occurred in Haiti, hundreds of lives were lost not only from structural collapse but also from a tsunami that hit several coastal towns. Immediately following the earthquake, with more experience now about how to help propel such an effort, our team again went to work distributing messages about the syntax, asking for Chilean Spanish language translation help by hosting publicly-editable wikis, and deploying new tools that we had recently implemented for supporting the format in the Haiti recovery efforts. On March 1, four days after the initial quake, we had a working TtT infrastructure that collected TtT tweet reports, in both English and Chilean Spanish, and published the records parsed from those tweets onto a public Google spreadsheet. Spreadsheet columns included report type, report content, time, location, contact, status, and 'more info' (See Figure 3). On March 25 when we ended that collection effort, we had recorded 228 TtT reports, which were mostly about missing people (`#sebusca`).

Type	Time	Name or Need	Location	Contact	More Info
#sebusca	3/1 18:34:54	carlos dominguez torres	parral	@martes	trabaja en peluquería morales pza de armas
#sebusca	3/1 10:19:35	silvia riveria	concepció n	@marcom arco	
#sebusca	3/1 10:05:22	marisa cordoba	los angeles	9 55555555	profesora historia liceo de niñas 57 años
…					

Table 3. Examples of missing person records from tweets sent after the Chile Earthquake. The titles in the first row are translated here from Spanish to English

We felt that TtT deployment for the Chile earthquake was substantially more effective and efficient than for Haiti. We were now practiced at launching such an effort, and understood the deployment to be a part of a socio-technical “campaign.” And critically, Twitter use in early 2010 was much higher in Chile than in Haiti. Additionally, as a result of the Haiti effort, we had publicly available tools (i.e., the spreadsheet) that demonstrated the added value for prospective users in the Chile response. However, the importance of a local advocate who can localize such a socio-technical effort cannot be underestimated: the influence of a single Twitterer, @Clandrea[55], was significant. @Clandrea is woman living in Chile who had just survived the earthquake and who had been added to many Twitter Lists of popular media outlets during the event, including NPRNews, YahooNews, NYTimes, HuffingtonPost, and CBSNews. Twitter Lists are curated lists of users that another Twitter user can make, typically recommended sources, often associated with a specific topic. We identified @Clandrea as someone who had both international and local influence during the event and began to communicate with her through Twitter, asking her to use TtT and to retweet some of our prescriptive and example tweets. Analyses of tweet distribution showed that her activity was a significant factor in the adoption of TtT for this event. Therefore, an important component may not be how many prescriptive or example tweets are sent, but rather *who* sends them. Ideal messengers of the TtT format seem to be account owners who are recognized as having local authority during an emerging event. Though originally hypothesized as response agencies, effective TtT prescriptive tweeters may include influential local citizens as well.

4.4.3 Deepwater Horizon Oil Spill

By May 2009 we had developed the interactive mapping component of the TtT infrastructure, a public Google Map that marked those TtT tweets that contained location information. Markers were color-coded according to the type of the report; users could click on markers to see the tweets and the report generated by that tweet.

[55] Though other user handles in this paper are anonymized, @Clandrea’s real name is used here with her permission.

We launched this tool for the first time on May 9, during the third week of the Deepwater Horizon Oil Spill disaster, in concert with another Tweak the Tweet campaign to allow citizen reports of the oil spill via Twitter. Due to uncertain health safety issues for citizen reporters, we were originally reluctant to encourage people to investigate and report oil impact, so we decided to deploy TtT initially to aggregate only reports of bad smells or fumes near the impacted areas. Working again with CrisisComomons, we created a special Twitter account, @oilreport, to interact with locals and send out prescriptive tweets, initially focused solely on the `#smell` hashtag. Later, after seeing many tweets with oil impact information and other efforts to allow for citizen reporting of oil through other channels, we added and promoted the ability for Twitterers to report oil impact on the shore, affected wildlife, response activities, volunteer opportunities, and area closures. Using TtT in concert with some behind-the-scenes, manual work to add location information to tweets, we collected and mapped over 800 oil impact reports between May 9 and August 9, 2010[56].

4.4.4 Fourmile Canyon Fire in Boulder, CO: Unexpected Local Authority

On September 6, 2010, an emergency hit close to home for us at the University of Colorado, directly affecting our own community including many of our friends and colleagues. The Fourmile Canyon Fire began burning at about 10am on Labor Day morning and would continue to burn for ten more days. 3500 people were evacuated from their homes during the initial days of the fire, and approximately 9000 more were put on evacuation notice later in the week due to a forecast of high winds (Eliot & Banda, 2010). By the time the fires were put out, 169 homes had been lost, though thankfully there were no serious injuries.

Seeing the smoke first hand, we began tweeting about the fire within the second hour of the event, and at 12:11pm, while the community was still settling on a standardized hashtag to report the event (eventually `#boulderfire`), we deployed an instance to collect and map TtT tweets to an interactive, Google map[57], a tool we had added to our suite during the Deepwater Horizon oil spill. The deployment

[56] The spreadsheet for the Oil Spill is available at http://bit.ly/oiltweetmap

[57] Map for the Fourmile Fire: http://www.cs.colorado.edu/~starbird/boulderfire_map.html

for the Fourmile Canyon Fire was our most successful to date. We collected 811 unique TtT tweet reports in ten days and 161 different authors contributed a TtT tweet to our records. Locals tweeted photos of the smoke plumes, as well as reports of need and offers of help for those who evacuated during the fire. Twitterers outside Boulder who had volunteered and used TtT during previous events began to help from afar. Several people contacted us through Twitter and were added as editors to the public Google spreadsheet we created, which became a collaborative space where GPS locations were added and records were verified.

Hundreds of people accessed this spreadsheet and other emerging volunteer groups asked to tap the data to add to their own information resources. Our interactive map was linked to by several other websites and some media outlets.

We felt that this comparatively successful deployment of TtT stemmed from a perceived authority of Project EPIC by others especially with respect to this local event, both in our community and afar. Though we were not the most important or influential individual Twitterers during this fire—those titles more likely belong to other Twitterers, namely @fishnette and @laurasrecipes who picked up more followerers and were more highly retweeted during the event—our collective effort to deploy the TtT syntax and its growing base of associated tools during the Fourmile Canyon Fire generated information resources that were accessed by people wishing to gain more information about the event. This relatively successful effort was a culmination of many of the things we had learned in other deployments, and offered further evidence towards some of our emerging hypotheses about how to communicate TtT during a crisis event.

4.4.5 The Joplin Tornado: A Volunteer-Led Deployment

The 2011 tornado season was a dramatic one. In late April, a series of major tornadoes struck areas across the U.S. Southeast, and about a month later, a similar series of tornadoes hit the U.S. Midwest, damaging communities across several states and devastating the city of Joplin, Missouri. The tornado that struck there was later estimated as an EF5 (NWS, 2011), the highest classification for tornados. That massive tornado hit in the early evening of May 22, 2011, causing 161 fatalities (Reuters, 2011b), over 900 injuries, and damage to 8000 structures (Reuters, 2011a). Afterwards, digital volunteers from both remote locations and nearby cities rallied to help Joplin, attempting to use social media to support their volunteer activities and organizing efforts. As the example tweets below demonstrate, volunteers worked in a variety of capacities: e.g. helping to promote the locations of shelters and other services, aiding the coordination of relief efforts, and connecting people with their lost belongings.

> `@whysomuch` (May 23, 2011): RT #Shelter is open: Campus of MO SO State Univ. at the Legget and Platt Athletic Center - 3950 Newman Rd #Joplin, Mo

> **@JoplinRelief** (May 25, 2011): The City of Purdy is looking for a location to setup and cook 1000 burgers and hotdogs. Also have pallets of pet food. #joplin

> **@fitzcarmen** (May 29, 2011): Found this clearing debris 22nd & Jackson. wedding pillow. Nuestra Boda? Now @ #Joplin lost found. Recognize names? http://t.co/OLSgzaW

On the Twitter platform, several event-specific hashtags emerged, including `#MOhaves` and `#MOneeds`. These tags were introduced just hours after the tornadoes hit Joplin in a tweet sent by a veteran of previous voluntweeting efforts—the tornados that hit Alabama and the month before.

> **@jsandford:** Encouraging #joplin to use #MOHaves and #MONeeds to classify where people have or need supplies. #mowx #WeAreAlabama

Twitterers assisting in the Alabama tornado relief efforts developed the convention during that previous event to assist them (`#ALNeeds`, `#ALhaves`). These hybrid hashtags saved volunteers tweet real estate by combining two types of categorization in a single hashtag term, and helped them to differentiate between needs and offers in Alabama and those in other affected areas. This is in example of a user-driven structuring mechanism being introduced into digital volunteer activity by the volunteers themselves. The suggested hashtags experienced some adoption for the Joplin event—more than 1000 Joplin-related tweets contained the `#MOneeds` tag.

We deployed Tweak the Tweet for the Joplin tornadoes shortly after midnight on April 23, about seven hours after the initial impact, and supported syntax use until June 13, more than three weeks later. Over the course of the 22-day deployment, we processed 356 unique TtT reports from 45 contributors. From the early moments of this deployment, we provided a public spreadsheet[58], a Google Map[59], and online instructions for how to use TtT for Joplin (Table 4), all standard practice for deployments by this time. However, there were several significant differences between our deployment for this event and those we had done for previous events.

One eventual change in our strategy was a move to a grid format for displaying instructions on a webpage, dividing syntax use into four clear components.[60]

Tweak the Tweet
Instructions for TtT for Joplin, MO Tornado

Event Tag (choose 1 or more)	Main Category Tag (choose only 1)	Location Tag (always include if you have location info)	Other Data Tags (choose as many as you need and can fit)

[58] https://docs.google.com/spreadsheet/ccc?key=0AkuhimfFYZrOdERPWVY1TWNlY2h2R0xjd01Va0ZNLXc&hl =en_US#gid=0

[59] http://www.cs.colorado.edu/~starbird/TtT_Tornados_map_byEvent.html

[60] This adjustment in communicating TtT instructions was suggested by digital volunteer and Humanity Road volunteer/leader, Catherine Graham.

#joplin #tornado	#need #moneeds #offer #mohaves #donation #damage #shelter #road #medical #closed #open #missing #imok #photo After main category tag, write what is needed/offered or damaged, etc.	#loc location info # To add location, type #loc, then write your location information, then end with another tag symbol OR add location anywhere in tweet in lat, long form: 46.8771863,-96.7898034 OR enable geo-location on your phone and Twitter client settings for locating the tweet where you are.	#contact #con #call #more #details #info #time #source #src After data tags, write the info that goes with these tags… i.e. #contact Mary 555-5555 #time 3pm #source @CNN

Include an Event Tag

2. Include **one** main category tag
3. Include location info like this… #loc location info #
4. Include as many other data tags as you need and can fit.

Table 4. Instructions for TtT for Joplin Tornados

Another difference between this deployment and those described earlier in this chapter was the dynamic adjustment to the user-created, hybrid tags for event plus need (`#MOneeds`) and event plus offer (`#MOhaves`). Though the instructions above do not reflect this, we added functionality to parse tweets with those tags as both part of the Joplin collection and placed into the respective main category. To reduce noise, we required that tweets of this kind include the `#loc` tag to ensure that they were using the TtT syntax.

Along with the inclusion of the improved grid instructions and the hybrid event-category tags, this deployment diverged significantly from previous TtT efforts in another important way: it was the first time that a TtT effort was driven primarily by the crowd and not led by an EPIC researcher. In the hours after the impact, several veteran digital volunteers recognized a need for TtT support in the emerging relief efforts and contacted the TtT-specific Twitter account (@kate30_dev) to request a TtT deployment for the Joplin tornados. While I worked to bring the TtT resources online, digital volunteers took charge of the campaign, spreading the word about how to use TtT and linking to the resources—i.e. the map, spreadsheet and instructions for its use. As the event progressed, new volunteers who were drawn to the response by personal concern and physical proximity to Joplin, began to use TtT and then took the lead in using and teaching the syntax.

A table of the top ten Twitterers in terms of volume of unique TtT reports attributed to their accounts (Table 5) demonstrates that the one active EPIC account during the event (@kate30_dev) was not a majority contributor to the effort—as had been the case in all previous TtT deployments. Of the high-volume TtT users during the Joplin tornado relief efforts, four were veteran digital volunteers and five new users who adopted the syntax for the first time during this event.

The data we collected to support this event indicated that veteran digital volunteers had a direct influence on newly mobilized, local and peripheral volunteers in spreading TtT use. For instance, @MichaelDB is a Twitterer from Springfield, MO, a city about an hour and a half away from the affected area. His first four tweets that use TtT syntax are retweets (like the one below) from the account of @rqskye, the most prolific TtT tweeter during this event and a long-time digital volunteer and TtT user.

@MichaelDB (May29 11:55am): RT @rqskye: #MOHaves #Donations Accepting clothing; plz bring hangers #Loc Wildwood Baptist Church 4827 East 20th St #Joplin #Cont 555 5 ...

Minutes after sending out four of these retweets in quick succession, @MichaelDB begins to use the syntax himself:

@MichaelDB (May 29 12:00pm): #need URGENT Swamp fans to cool overheated pets
#loc 140 E Emperor Ln
#cont #Joplin Humane Soc 555-555-5555
#src Annie@JHS 12:20
#MONeeds

Another important figure in TtT diffusion during Joplin was @goMarielle, a Twitterer who learned about TtT a few days after the event (May 27) from a tweet sent by @CrisisMappers, another veteran digital volunteer and a TtT promoter. After learning about the syntax and how it worked, @goMarielle, also a resident of Missouri, began use the syntax herself, to tweet out links to the TtT resources and to encourage other Twitterers to use the syntax, often using @mentions to make personalized requests.

> **@goMarielle** (May 29): @8NewsWeather Thanks for spreading the word: Tweak the Tweet efforts for #Joplin #relief tinyurl.com/JoplinMappers needs VOLUNTEERS

> **@goMarielle** (May 31): @CrisisCamp NEED #TtT volunteers to #TAG tweets for #joplin #tornado relief efforts. on.fb.me/iNAWOp #JoplinMappers PLS RT

Through her TtT promotion activity, @goMarielle connected with other Twitterers who were from Missouri, and went on to mentor later adopters in incorporating TtT into their volunteer tweeting activity. Interestingly, some of the later adopters initiated accounts that were solely for the purpose of TtT tweeting for these events and even incorporated TtT into their account names.

User	TtT Tweets	Affiliation
@rqskye	78	Voluntweeter
@MichaelDB	61	Springfield MO – 1 and 1/2 hours away
@loisannporter	29	Voluntweeter
@goMarielle	31	Columbia MO – 4 hours away
@TtT4J*	20	Springfield MO – 1 and 1/2 hours away
@kate30_dev	20	Researcher
@MarkMyWords*	17	Missouri
@TheFireTracker2	11	Voluntweeter
@ChevyTtT*	10	Columbia MO – 4 hours away
@oktwister	8	Voluntweeter

Table 5. High Volume TtT Twitterers during the Joplin Relief Efforts
*** new accounts started after the event to tweet primarily for Joplin relief efforts (all are silent now)**

More than any other event in our two-year effort to deploy and support TtT use, the Joplin event suggests the possibility of a sustainable future of TtT, whereby veteran digital volunteers deploy Tweak the Tweet, promote the format, and then identify, recruit and train new volunteers in syntax use during the response period. This suggests an important design direction: hosting Tweak the Tweet resources on a public website that allows volunteers to launch an instance and to tailor that instance to the current event by dynamically adding event tags and main category tags for the evolving event.

4.5 Summary: Lessons Learned from Deploying Tweak the Tweet

Through deployments for more than 20 events in 2011 and 2012, across a range of crisis event types, Tweak the Tweet has provided a wealth of insight along a number of dimensions. Findings shed light on the usability and viability of the TtT syntax specifically as well as design considerations for crisis-reporting microsyntaxes in general. This project has also increased our understanding of digital volunteer communities and offers design possibilities for encouraging and shaping digital volunteer activity in the future.

Among the significant findings is the refutation of our "field of dreams" implementation strategy. During the Haiti response period, we quickly learned that we could not count on spontaneous digital volunteers to build and maintain a complex software infrastructure to support an innovation of this kind. Though we could count on some assistance from technically-able volunteers brought together at barcamps or self-deployed in the social media space, the core infrastructure for Tweak the Tweet and any similar effort may require some organizational support to keep it in place and at the very least a self-appointed leader responsible for maintaining the infrastructure and recruiting assistance. This is true both in maintaining the technical components during and between events and managing the non-technical aspects of the campaign and syntax structure during events, but these two sides of Tweak the Tweet can be organizationally separate. For instance, with a usable interface and robust infrastructure, digital volunteers could deploy and maintain a TtT instance, updating the geographical region and syntax and promoting the automatically-generated resources. However, the technical infrastructure underlying those efforts needs to be functional, usable and sustainable.

Another place where our initial rationale fell short was in how we proposed to teach the format. Prescriptive tweeting turned out to be a less effective method of driving adoption than other strategies. Actual use of the TtT syntax during real events and incorporating real, actionable information was the most effective means of spreading the syntax. Resources that demonstrated how TtT was processed, like the spreadsheet and Google Map, also served to educate users on how the syntax worked. Online instructions that included some of the rationale behind TtT use were better teaching instruments than periodic prescriptive tweets. In events where we did not provide website instructions, TtT users often requested them. Over time, we improved the method of communicating these instructions, taking into account feedback from digital volunteers. At several points in time, we also experimented with providing online applications to support TtT tweet creation and translating activity. Though these tools showed promising results, this is an area where more research and development could improve the usability and teaching methods of Tweak the Tweet.

An important conclusion of this project is that leveraging existing social media platforms for crowd work has multiple benefits. The first advantage involves infrastructure development and maintenance. By deploying this innovation through the Twitter platform, we were able to piggy-back on the existing infrastructure of a commercial product with hundreds of millions of users, a platform designed to support huge volumes of messages and accustomed to adjusting to capacity issues around spurts of activity—e.g. the burst of activity after the announced assassination of Bin Laden produced 3440 tweets per second[61]. This meant that developers of the TtT infrastructure were not charged with supporting the technically challenging aspect of carrying the expectedly high volume of communications during crisis events, and could quickly bootstrap a functional crisis-reporting system with a fraction of development cycles required to build a similar reporting system from scratch.

Conducting TtT reports through a social media platform also enabled a sort of natural diffusion for the innovation, whereby potential new users were exposed to the innovation through its use. Because TtT tweeting is public and visible in a social media environment, its use can be self-promoting, helping it to diffuse through emerging networks of digital volunteers. Motivations for TtT use involve the development of social capital in the form of expanding networks (Putnam, 2000; Ellison et al., 2007) and

[61] This statistic was reported by Twitter from their @twittercomms account in a status update on May 2, 2011. https://twitter.com/twittercomms/status/65133603398483970

symbolic capital (Bourdeau, 1984), or what Malone et al. would call "glory." For some syntax users during the Haiti response efforts, TtT was one component of efforts to create a social media brand as a *crisis tweeter*. The public visibility of their volunteer work differs significantly from systems (e.g. Ushahidi) where digital volunteer work occurs in a closed environment—in those cases social and symbolic capital gains are almost entirely limited to that which can be received from others already participating in the group. Motivations for these different types of digital volunteer activities may vary accordingly.

4.5.1 Re-conceptualizing Tweak the Tweet as Distributed Human Computation

Perhaps the most pivotal discovery of this research was the early observation that TtT was not experiencing substantial adoption by affected individuals, but was instead appropriated by remote, geographically-dispersed, digital volunteers in their efforts to assist in response activities. This has led to a revision in rationale for how and why Tweak the Tweet should be used—remote volunteers are now the targeted user group for the syntax. It has also led to a re-conceptualization of the project.

The syntax was initially conceived as a data collection mechanism using the crowd power of distributed sensing in the form of trained citizen reporters. However, its use during the Haiti response efforts manifested as self-organizing, distributed human computation by remote volunteers. This form of crowd work could be classified as a form of microwork, whereby users are tasked with searching media spaces to identify actionable information and then translating or structuring that information into a prescribed format. From this perspective, the crowd work of TtT translation has a low barrier to entry, in that potential volunteers need only to learn how to identify the right kinds of information and then how to use the syntax to structure it. Considering its place within the now larger domain of crowd work during disaster, this distributed structuring activity could be one component of a multi-staged crowd-filtering effort, which I will explain later in this work.

However, understanding the phenomenon of digital volunteerism calls attention to a new question: In a world where a large group of remote information-processers are available to take on the task of TtT translation, is it needed? With the existence of Ushahidi and other similar platforms—with, currently, much higher participation rates during large-scale events than TtT—would it not be better to have volunteers skip the step of TtT translation and simply take information from the original sources and enter it directly into reports for Ushahidi or another platform? Or to have volunteers synthesize this information directly into resources like the ones Humanity Road (Study 3, Chapter 6) creates? Considering these alternate routes of information structuring, perhaps the best argument for TtT use is the difference in activity visibility, which this research suggests may have a positive influence on the motivations for use, the rate of diffusion in a newly activated volunteer population, and, as a result of those two things, the potential scale of participation.

4.5.2 From Citizen Reporting to Human Computation to Crowd Work

The preliminary direction of this research was inspired by the problem of extracting meaning from social media updates in the crisis context. Aligned with this trajectory, RQ1 addressed the viability of a user-side intervention for processing social media updates—the Tweak the Tweet project. The discovery, made during the early efforts to deploy Tweak the Tweet, of TtT translation and the encompassing phenomenon of digital volunteerism shifted the overall focus of this research from the goal of extracting information to a slightly altered purpose of understanding the dynamics of connected crowd work during disaster. In this way, the findings of Study 1 (this chapter) provided a springboard for the subsequent research studies described in the remainder of this dissertation. The digital volunteers discovered during our action research to deploy Tweak the Tweet during the Haiti earthquake response became the participants of Study 2 (Chapter 5) and constitute a large portion of the volunteers whose behavior is examined in Study 3 (Chapter 6). This progression served to redefine the broader thrust of this research as an investigation of how the crowd organizes and is organized to help process information during not just crisis events but mass disruption events of other kinds as well.

CHAPTER 5
STUDY 2. “VOLUNTWEETERS”: SELF-ORGANIZING BY DIGITAL VOLUNTEERS IN TIMES OF CRISIS[62]

The following study grew out of our efforts to deploy Tweak the Tweet during the aftermath of the 2010 Haiti earthquake. This study begins by addressing Research Question 2, to whom and how can we teach a crisis reporting syntax for social media, and later shifts to answering Research Question 3, how remote individuals use social media to organize during mass disruption events, in this case a catastrophic natural disaster[63]. The investigation examines how remote, digital volunteers worked in various capacities to filter, process and move information during that event, incorporating a tweet-level analysis of how they organized information along with an empirical and theoretical examination of how they organized themselves to do this work.

5.1 Introduction

The new behaviors of mass interaction that information and communication technology (ICT) enable are affecting the way we seek and provide information, as well as the way we imagine our roles and responsibilities in such matters. Here we empirically consider the emerging role of the “digital volunteer” as an element of the phenomenon popularly known as “crowdsourcing.” We relate these behaviors to the social science theory of collective behavior (Dynes, 1970; Kreps & Bosworth, 1994) and show how microblogging platforms serve as a new arena for self-organizing.

We examine this phenomenon in the context of disasters—specifically, the January 12, 2010 Haiti earthquake—when onset of the presence of such volunteers was rapid and their activities often

---

[62] This work is an adaptation of an earlier work:

Kate Starbird and Leysia Palen. (2011). "Voluntweeters": self-organizing by digital volunteers in times of crisis. In Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11). ACM, New York, NY, USA, 1071-1080. DOI=10.1145/1978942.1979102 http://doi.acm.org/10.1145/1978942.1979102

To cite material from this Chapter, please cite this original work as well as this dissertation.

[63] See Chapter 3 for an explication of research questions and their mapping to the separate studies of this work.

remarkable. Previous research and development efforts have emphasized the potential for "crowdsourcing" via social media to increase situational awareness during crisis events (Okolloh, 2009; Liu et al., 2010). However, use of the popular umbrella term risks both obscuring the underlying behaviors that constitute "crowdsourcing," and erroneously casting them as novel by-products of new media. Spontaneous volunteerism is not a new feature of crisis events (Fritz & Mathewson, 1957; Dynes, 1970; Tierney et al., 2001; Kendra & Wachtendorf, 2003b)—disaster events in the pre-ICT era were places where such large-scale self-organizing phenomena could previously be seen. This paper attempts to unpack "crowdsourcing" in crisis response by applying an existing framework of self-organizing in disaster settings (Kreps & Bosworth, 1994) to new digital volunteer behaviors. In so doing, it reveals new forms of volunteerism that were not previously possible.

This work extends previous research on the use of social media during crisis, e.g. (Hughes et al., 2008; Qu et al., 2009; Guy et al., 2010; Starbird et al., 2010; Vieweg et al., 2010). It proceeds from our efforts to deploy the Tweak the Tweet microsyntax, described in detail in Study 1 (Chapter 4), for the 2010 Haiti earthquake. Though the initiative was not widely used at that time, the innovation nevertheless served as an opening to seeing and subsequently studying the broader sphere of digital volunteerism. A recap of TtT deployment efforts for that event sets the stage for the description that follows of the multi-phased empirical study of digital volunteerism and self-organizing behavior.

5.2 Deploying Tweak the Tweet for the 2010 Haiti Earthquake

On January 12, 2010, a 7.0 magnitude earthquake struck Haiti near its Port-au-Prince capitol. Due to the earthquake's intensity and shallow depth, as well as Haiti's vulnerable infrastructure, the destruction was catastrophic. Economic damage has been estimated at up to $13.9 billion (Cavallo et al., 2010) with 200,000 to 250,000 lives lost (New York Times, 2010b). In the aftermath, thousands were trapped beneath collapsed structures, hundreds of thousands were injured, and nearly 1.5 million people had been displaced. Relief agencies struggled to meet basic needs in this catastrophic event.

With a desire to assist the response in Haiti during a time of overwhelming need, and in collaboration with the CrisisCamp initiative (crisiscommons.org), our research group deployed the Tweak the Tweet (TtT) syntax on January 14, 2010 at 19:51 EST.[64] We activated a Twitter account to distribute prescriptive tweets at regular, automated intervals (hourly), in both English and French. At the same time, several team members began to tweet information from various sources, including other tweets as well as email messages from disaster-related lists, using TtT syntax. We used the dedicated account as well as personal accounts to maximize audience reach. We updated our project website to explain how the syntax worked, developed a web-based editor that helped people generate tweets in the syntax, and created how-to screencasts posted to YouTube. We issued a press release to encourage broad attention to and adoption of the syntax.

Though the effort to support syntax use continued for several weeks, we ended our active campaign and sent our last prescriptive tweet on January 24, due to the recognition that the syntax was hard to use from the ground, and that we could not assure users that TtT tweets were being channeled to people and agencies who could respond.

5.2.1 Unexpected Activity of Tweak the Tweet Translation

The TtT syntax was originally intended for use during a disaster by those directly affected by the event. We deployed the idea during the aftermath of the Haiti earthquake knowing that only a small percentage of the population of Haiti were Twitter users, and that using a syntax could be difficult, but hoping that the structured data format could nevertheless be useful in some fashion. Not surprisingly, few people used the syntax from "the ground."

However, what was unexpected were the many Twitterers located around the world who emerged as "translators"—those who translated information from multiple sources into the syntax and tweeted it out to their followers. Previous research indicates that "synthetic" information activity—where users modify

[64] All times are reported in local Haiti time, which was Eastern Standard Time (EST) at the time of the quake.

and synthesize information from multiple sources, including Twitter itself—is an important form of information production on Twitter (Starbird et al., 2010). We recognized this synthetic information generation in the activity of translation from raw, unstructured information to structured data forms as the basis for a new form of the volunteerism that pervades disaster response (Fritz & Mathewson, 1957; Dynes, 1970; Tierney et al., 2001; Kendra & Wachtendorf, 2003a & 2003b; Palen & Liu, 2007). We initiated this study to investigate the behavior and motivations of the TtT translators.

Note that for this paper, *translating* and *translator* specifically refer to TtT syntax translation, though many digital volunteers contributed by offering classic language translation to assist in this international response.

5.3 Method

We employed a multi-step process for data collection, beginning with large-scale tweet data collection using Twitter APIs, then a preliminary tweet content analysis to identify the target study population, and finally email-based interviews with a sample of the population to understand their motivations and experiences while tweeting during the immediate aftermath of the Haiti earthquake.

5.3.1 Tweet Collection and Analysis

5.3.1.1 Twitter Data Collection

During the active TtT campaign period (January 14-24), we used the Twitter Search API to collect every tweet that contained both the `#haiti` hashtag and at least one specialized TtT hashtag (e.g., `#need, #offer, #loc`). On February 1, we identified every Twitterer who had at least one tweet in that set and, using the Twitter REST API, captured their entire Twitter status update streams (*contextual* streams) from January 10–February 1. These streams contain *all* of a user's public tweets, including public tweets addressed to other users. This initial data set consists of 339 Twitterers, six of whom were from our research group, and a total 292,928 tweets.[65]

5.3.1.2 Qualitative Analysis and Coding of Tweets

We conducted an analysis of tweet content in the contextual streams to better understand the behavior of Twitterers in our data set, and to identify how each was using TtT syntax. We manually coded every tweet that contained a TtT hashtag, excluding prescriptive tweets, as: 1) *original tweet* (a user putting her own offers or needs into the syntax); 2) *TtT retweet* (of another's TtT syntax tweet); 3) *translation tweet* (a user translating information from another source into TtT syntax); and, 4) *non-TtT tweet* (a tweet that contained one of these tags, but was not in the TtT format). Discerning between TtT translations and retweets could not be done computationally because some retweets are not "properly" credited using one of the known conventions, and because many of the Twitterers in our sample generated *tag-added retweets*, meaning that they began with a conventional retweet, including the `RT @` marker, but then modified it by inserting TtT tags. Table 6 shows the number of tweets in each category, excluding those sent by our team. Of 2911 total tweets in the syntax, 1040 were translations of information found in other sources, including other, non-TtT tweets.

Total TtT tweets (not including from our team)	2911
Translated (including *tag-added* retweets)	1040
Original (from affected people & remote organizations)	39
Retweets of TtT tweets	1732
Unknown – cannot classify as RT or Translated	100

Table 6. Coding for TtT tweets

5.3.1.3 Identifying the Translators

This tweet coding differentiated people who were using TtT on their own behalf, those who were translating other information into TtT, and those who were only retweeting messages already in the

[65] The Twitter API allows us to collect only the most recent 3200 tweets from each user. For some high-volume users (20 of the total and 2 from our interview set), we could not go back as far as Jan 10 for their contextual streams.

syntax. Table 7 shows the frequency of the different TtT user types. Though only ten Twitterers (four in Haiti) used TtT syntax for directly tweeting their own needs, 74 of the 333 Twitterers were found to have translated at least one tweet into TtT syntax.

Twitterer type	Number of Twitterers of Type
Original Twitterers - tweeting on their own behalf	10
Original Twitterers located in Haiti	4
Original Twitterers not located in Haiti	6
Retweeters	215
Translators	74
Our Own Research Team Members	6
Twitterers whose #hashtag use was not TtT	38

Table 7. Number of Twitterers by TtT user type

5.3.1.4 Other Remote Digital Volunteers

Though non-translators were not included in the interview portion of the data collection (described next), qualitative analysis of their tweet streams suggests that many exhibited similar remote volunteer behaviors as those in the primary participant pool. Many were members of the volunteer networks that we describe below and acted as remote operators during the event, directing the flow of critical information. It was these consistent, Twitter-abetted, volunteering behaviors that became the ultimate subject of this research, with the TtT syntax serving as one vehicle with which these new behaviors were exhibited.

5.3.2 Interview Data Collection

5.3.2.1 Participants

The 74 translators were identified as candidates for the interview phase of this study, designed to better understand volunteer translating behavior. We attempted to contact participants through their Twitter accounts, the only available point of access, using "direct messages" ("DMs" are private person to person twitter messages) to those who were "following" our account (which is required for DMs), and publicly addressed tweets to those who were not. Between the Haiti earthquake and the commencement of the interview study in the summer, two accounts were deleted, and several others became inactive, making it likely that many of our addressed tweets went unread. 37 of the 74 Twitterers responded; 27 of these consented to participate and 19 completed the full interview. Respondents were widely distributed across volume of TtT tweets, from one TtT tweet sent to 171 TtT tweets sent. The response rate among those who were reachable was 51%. Eight initial respondents did not complete the interview for reasons including ongoing involvement in Haiti relief efforts, illness, and others not specified.

5.3.2.2 Email Interview Instrument

The interviews were conducted via email for all participants, though participants were given the option of a verbal interview. Each participant was sent a message with the same set of open-ended questions.

There were six sections to the interview (which were untitled in the instrument): 1) general Twitter use; 2) Twitter use during Haiti; 3) use of TtT syntax; 4) awareness of Twitter followers; 5) other volunteer experience; and 6) further comment and suggestions for future use of social media during crises. The section themes and the questions within them were derived from the prior qualitative analysis of the contextual tweet streams.

Because our analysis began to reveal that Twitter connections served critical functions in volunteering behavior, respondents whose tweet streams showed interactivity with other translators were sent a follow up email inquiring about their twitter-related social connections. For several participants, this second round led to an ongoing exchange with unscripted follow-up questions, just as in a face-to-face open interview. Due to the back-and-forth nature of most exchanges, we felt the term “survey” was not representative of the data collection method and instead use the term “email interview.”

5.4 Findings

5.4.1 The Participants

We asked gender and age questions and used tweet analysis to capture basic demographics for our interview group and to make comparisons to the larger group of TtT translators.

A majority of Twitterers who translated information into TtT syntax tweets are women. In our interview pool, there are 17 females and two males. To determine if this breakdown is representative of all translators and not due to other factors, we coded each Twitterer for gender by analyzing the name and photo associated with the Twitter profile and the content of the account owner's tweets. For the entire 74 TtT translators, we found 46 females, 16 males, one account that was operated by two people (one male and one female), and 11 accounts where gender could not be determined. Where we could determine and assign gender in the entire translator group, 74% are female.

Though most of the interviewees were located in North America during the event (nine in the US and six in Canada), our interview group includes one Twitterer each from Turkey, Australia, Switzerland, and the UK. The average age is 40.1 years.

5.4.2 Previous Twitter Use

Figure 9 shows the date that each study participant joined Twitter. Three participants created their accounts after the earthquake, though one operates another account started nine months earlier. Two others joined Twitter earlier, but they did not send their first tweet until after the event. The remaining 15 interviewees were actively tweeting from their current accounts before January 12, 2010**.**

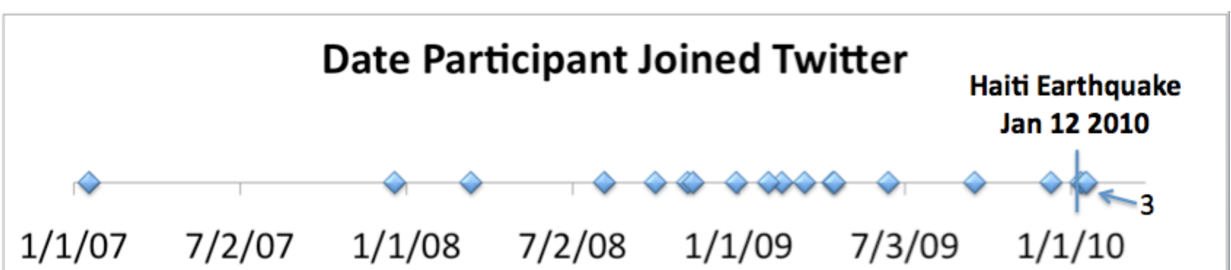


Figure 9. Twitter Join Date for Interview Respondents

The experienced Twitterers in the study offer a variety of reasons for joining Twitter, including promoting a cause with which they are personally connected, and keeping up on pop culture and politics. Three report tweeting during previous disasters or crises and two of those specifically mention the political protests in Iran (June–December 2009) as their entry point to "crisis tweeting."

5.4.3 Personal Motivations for Haiti Tweeting

For those who began tweeting after the earthquake, all three cite personal connections to Haiti during the earthquake as motivation for creating accounts:

> Emily: My brother, a volunteer teacher in Port-au-Prince, had arrived in Haiti two days before the EQ. We… were not able to contact him and not sure where he was staying... As you know, the news was horrific and we were beyond worried. I joined Twitter approximately 12 hours after the EQ in order to try and communicate with surviving students from his school to find out if they had any news on him.

Three other participants cite connections to people in or from Haiti as reasons for starting to tweet. Five of the six personally connected Twitterers began with a single cause—a person or place with specific needs—and later branched out to communicating about other needs and issues.

Among those without personal connections, interview responses and tweet streams reveal the singular motivation of trying to help out, in any way possible:

> Alena: Why I did it? has no other explanation other than I had to. One part of the world was in pain and I could not sit back watch others do something when I had a little chance to send some drinking water to people if I could.

> Maria: I think that's when I went on Twitter and started tweeting. Then I discovered a whole bunch of people tweeting for Haiti and started doing it myself and building up connections as much as I could in order to try to save some lives if possible. ... As you'll see some of us tweeted 16 hours a day or more... I just hoped what I was doing was helping. I'll never know if my tweets actually helped but that's ok as well.

As these excerpts suggest, many of these people began tweeting for Haiti as a way of volunteering their time—sometimes entire waking days—from areas far removed from the event. Most, like Maria, entered the space with little direction or knowledge about how to help. Maria mentions that one of her first volunteer activities was to build up connections. We will return to these connections at length below to explain how volunteers were able to begin with little more than a Twitter account, yet end up helping in remarkable ways.

5.4.4 Using Tweak the Tweet Syntax

We asked participants where they first saw or heard about TtT syntax, why they decided to use it, and where they got the information that they translated into TtT tweets.

5.4.4.1 Discovering Tweak the Tweet

The majority of respondents learned about the TtT syntax from a tweet—many of these were prescriptive tweets from our team members' accounts found through hashtag searches, or seen in an RT from another Twitterer they follow. Five participants report being directed to the syntax instructions by other Twitterers in their networks and two mention receiving help from another Twitterer who "explained" the format to them, or "taught" them how to use it. A few interviewees appear to have learned the format by mimicking its use in others' tweets. One of these did not realize she was using TtT syntax, though her tweet streams show that she began to use the `#location` and `#contact` data tags. Another explains that she discovered TtT when one of her tweets was retweeted in the format:

> Emily: Another Haiti volunteer on Twitter retweeted my post with the correct syntax, and I went to the "Tweak the Tweet" website, read up on it, and started using it for all my Haiti tweets and retweets except personal responses to people looking for information or offering help to [Haitian village] directly.

It is interesting that this respondent refers to the retweet of her information as being in the "correct syntax." This statement implies that the syntax was somehow more correct or more official than her original tweet. This theme shows up in several of the responses about TtT use.

5.4.4.2 Reasons for Adopting TtT Syntax

We asked all interviewees why they chose to use TtT, giving examples to each from their own Twitter stream. Several replied that it just "made sense":

> Susan: Because it standardized everything. It gave a SOP/ Standard Operating Procedure or MO/ Modus Operandi of doing tweets. I thought it made perfect sense.

The statement above suggests that this Twitterer welcomed a standardized method for her tweeting and it echoes Emily's comment about "correct syntax." When these Twitterers say that TtT made sense, they may not be talking about the rules of the syntax, per se. In fact, many of the same report that it was quite hard to use. What seems to appeal to them was the idea of standardization. The syntax itself became an object of authority—an organizing feature in the socio-technical milieu of Twitter activity. This is an idea we return to in the Discussion.

Some participants saw the syntax as a way to manually filter and identify key information for themselves, from within an inherently noisy information space:

> Claire: Myself having been searching using hashtags and keywords thought this was an excellent way to filter out (most likely) credible sources directly involved in the effort from possible citizens of Haiti or family members just sharing information and sorting through the chatter.

Some felt that using the syntax marked their own tweets as authoritative and helped their accounts rise above the fray:

> Claire: I realized my impression of those using the syntax was to take them more seriously, so once I had confirmed information it was translated into syntax.

A few noted that the format could be used as a way to make individual tweets more useful, regardless of back-end tools for filtering. By formatting or "syntaxing" their tweets in this way (as some came to call it), they made sure tweets had the necessary pieces of information. They could use the syntax as a way to instruct others to do the same:

> Jenny: I choose to tweet with the syntax because uniformity in messaging is essential in high pressure, emergency situations ... I also noticed that people … spent a lot of time trying to

> decipher messages with missing info and what the source of the info was…. In short the time spent looking for missing parts to a message could be avoided by using TtT.

These responses suggest that some Haiti Twitterers adopted TtT as another resource in their own volunteer work. The syntax helped them format tweets to include all the necessary pieces of information, such as contact number, location, and source. It was a signal they could give to say that the information in their tweets had been verified by someone who had experience in the space. ("Experience" might only amount to days worth, which can be enough to warrant credibility and even develop mastery relative to the newness of the volunteering tweeting phenomenon). TtT also helped some do manual filtering by allowing search on certain hashtags to identify important, actionable information, a topic we return to later.

5.4.4.3 Source Material for TtT Tweets

For their own translated TtT tweets, interviewees reformatted and passed on information from a variety of sources. Most report using other tweets as a key source. They first located tweets with key information that were not initially in TtT format, then added the TtT tags and retweeted them. Several mention verifying the content before passing it along. The example below shows how one Twitterer took the information from a tweet she saw and translated it into TtT syntax, noting her source.

> (Original) **@Delmon:** Marie Girard @ Institut FR, Ruelle Robin PaP needs food water 4 children, toddlers + pregnant woman 5555-5555

> (Translation) **@barbaraslavin:** #haiti #needs food H20 #name Marie Girard #loc Institut Franse, Ruelle Robin P-a-P #info 5555-5555 #source @Delmon TwT

Participants also listed public blogs and Facebook posts, as well as DMs, emails and phone calls from contacts (some on the ground in Haiti), as sources for their TtT tweets.

Another key source for TtT tweets noted by six of our interviewees was Ushahidi. Ushahidi is a collaborative reporting environment that aggregates and maps information provided by citizens (Okolloh, 2009). Originally developed during violent political unrest in Kenya in 2008, Ushahidi was used during the Haiti crisis in conjunction with an SMS shortcode to collect incident reports from the ground. Ushahidi organized and encouraged volunteers to help process the raw SMS messages into structured reports (Liu et al., 2010). Some of our respondents monitored Ushahidi incident reports, identified new or newly verified information, and then tweeted it in TtT syntax. A few also directed these tweets to users they felt could act on the information.

5.4.5 Remote Operators: Mechanisms of Emergence

Analysis of interview responses revealed two types of digital volunteers within our translator group. The first was the type we identified at the onset of this research study, digital volunteers whose primary activity was the translation of other information into TtT tweets. For the other type, TtT was merely another resource adopted as part of a diverse tool set. In this second group, Twitterers were not simply translating information, but instead were acting as *remote operators*, moving information between many sources using a variety of tools. TtT was only a small part of that activity. Many exerted considerable effort into building and leveraging connections to move information— and in some cases supplies— between affected people on the ground, response agencies in Haiti and abroad, and other volunteer crisis workers all over the world.

None of the Twitterers in the second category were able to simply log on to their Twitter account and start acting like remote operators. Instead, they progressed from simple activities (e.g., retweeting or translating tweets) to more complex ones (e.g., verifying or routing of information).

Ushahidi was a key bridge between information-based activities (similar to TtT translation) and the more complex volunteer work of remote operators. Over time, some volunteers shifted from taking information from Ushahidi incident reports and tweeting it (sometimes in TtT syntax) to other activities, including verifying details, especially the origin date, and following up on contact email addresses and phone numbers to confirm reports. Others began to monitor Twitter by searching for certain hashtags, including TtT tags, and then used that information to create new Ushahidi reports or update existing ones.

This shift between translating and pure data entry to information verification is one we see in several interview responses.

Though TtT translation and Ushahidi data entry were the primary entrance points to crisis tweeting for many, others began with different types of tweet-driven activities. One started her Haiti volunteer work by adding minutes to Haiti residents' cell phones as requested by other Twitterers who were in Haiti or had connections to people there.

@MelyMello (Jan 15 24:26): @ayitiJo we can top up your phone, can't we? add more minutes to it for you? Just need your phone #

@MelyMello (Jan 16 17:38): @jean123 please let 50955555555 know that 630HTG were just added

@deJacmel (Jan 16 ~19:00): @MelyMello Please Add min 2 ths cell numbers for me. They R helping Amer families to contact their haiti relatives.

@MelyMello (Jan 17 13:11): @janeSM want to help? Help me add minutes to a WACK of phone numbers I have been sent!

@MelyMello (Jan 17 14:32): @janeSM - please send any urgent request to help locals to @deJacmel - we're adding $ to cells of his local contacts to do just that

@MelyMello (Jan 17 20:33): @Meira_Davi just emailed you the updated list - Paypal is almost empty again - have two more numbers to do - then we're at $0 :(

@MelyMello (Jan 17 21:00): @deJacmel - can UR ppl help? At Brochette (Carrefour), 1500 ppl have nothing to eat or drink since Tues evening. Call 509.337-53-2154

As the sequence above describes, @MelyMello enlisted the help of other remote Twitterers in her effort. She created a PayPal account and requested donations. She established digital relationships with other Twitterers (@deJacmel) who had connections to people on the ground in Haiti. Later, she would use these contacts on the ground to get information on ground conditions, verify reports, and even request help for other needs she identified.

The interview excerpt below shows the evolution of isolated activity to a set of network connections and expanded set of activities:

> Linda: In the beginning, I worked alone ... I started recognizing ppl who seemed to have good info and we would support each other, RT each other and help to find info for each other.

The exchange below, taken from the contextual streams of one of our translators and another Twitterer in our initial 339 (the latter was not coded as a translator), demonstrates how some volunteers came to use connections and information on Twitter in combination with contacts on the ground to act as remote emergency operators.

@CarolB (Jan 18 10:53): @IstanbulTWSTVL who needs the truck? What type of truck? Where in clercine?

@IstanbulTWSTVL (Jan 18 11:04): @CarolB @janeSM Needs a truck by Clercine and help with & UN DRS that need ride at rue de l'enterrement #xx

@CarolB (Jan 18 11:07): Okay made contact with someone should be sending truck.

@IstanbulTWSTVL (Jan 18 11:21): @CarolB Second situation was stevePs people on the way to Clercine had no phone contact he has posted number @steveP is this solved

@CarolB (Jan 18 11:27): @IstanbulTWSTVL It has been arranged. Truck has been arranged.

In this example, @IstanbulTWSTVL routes information gleaned from Twitter sources to @CarolB, who has a connection to people on the ground. Volunteer Twitterers like @IstanbulTWSTVL would find information that had not yet been acted on through Twitter and Ushahidi reports, then act on it themselves in one of two ways. Some would follow up with contact numbers directly, calling affected people or responders on the ground in Haiti to confirm and update their reports. Others tried to find another Twitterer to whom they could direct the information, getting it closer to those who could physically act on it.

Maria: It didn't matter if I actually helped directly; the important thing was figuring out who to send the help requests to and depending who was online at the time…

5.4.6 Examining Collective Work to Process Information at the Tweet Level

In the extended example that follows, we see an interactive process between several Twitterers to react and collectively process a report of human trafficking. The tweets below are selected examples from over 200 tweets and retweets sent about this incident, and illustrate, at the tweet-level, several types of work that these volunteers completed, including identifying, amplifying, geo-locating, verifying and structuring information.

Figure 10, an eDataViewer image, shows the Twitter conversation surrounding this report, distributed across Twitterers and over time. In this graphic, lines of horizontal spheres are the tweets from contextual streams of individual users. Tweets are placed along the horizontal axis according to the time of the tweet—time progresses from left to right, and the vertical white lines are day markers for (midnight on) January 26, January 27, and January 28[66]. Tweets that refer to the report are colored blue and given a larger size for emphasis. As the image shows, the discussion begins on January 26, and during its first four hours several TtT Twitterers send a tweet—and in some cases many tweets—referring to the report. The conversation goes quiet during the early morning hours on the 27th and then picks up again somewhat that next afternoon as Twitterers follow up on the previous day's activity. An analysis of how the information moved and changed over time lends insight into how voluntweeters worked.

[66] Though the days markers are shifted two hours to the left due to differences in Haiti time and collection time. The final colored tweet occurs just after midnight on the 28th.

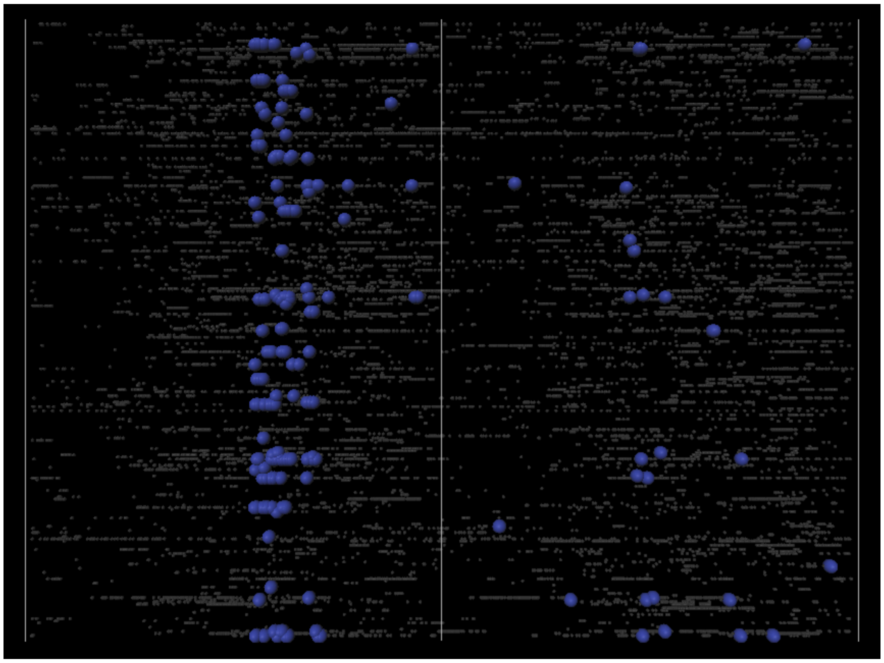


Figure 10. Twitter Conversation about a Child Trafficking Report at Hopital Espoir

5.4.6.1 Identifying, Amplifying and Routing Actionable Information

The discussion began with the following tweet, sent by @SergeGilles, a Haitian located in Haiti at the time of the earthquake.

> **@SergeGilles** (2010-01-26 15:03): I Heard that there is a Human Traffic of childen in Hopital Espoir, Delmas 75. Can @UNICEF Check this out?

The tweet broadcast information that @SergeGilles had heard about human trafficking of children, a widespread concern for many humanitarian agencies during the aftermath of the quake (Evans, 2010; Gupta & Agrawal, 2010). Within this tweet, @SergeGilles also made a public request to UNICEF, through an @mention to their account, that they follow up and “check out” the report.

@SergeGilles's original tweet was retweeted 16 times within the *Contextual Streams of TtT Twitterers*[67]—most of those were sent within the first hour and all but one went out within the first two hours. A few minutes after the original tweet, @MarkJones retweeted the information out to his own followers, using different language but crediting @SergeGilles as the source:

> `**@MarkJones** (2010-01-26 15:09): My friend @SergeGilles is getting reports of kids being trafficked @ Hopital Espoir, Delmas 75. Please contact authorities NOW!`

And in less than an hour's time, this second tweet version of the information was retweeted eleven times by TtT users. These 27 total retweets demonstrate volunteer Twitterers working to identify and then amplify actionable information. Several of these Twitterers also attempted to route the information to particular accounts that they felt might be able to act on the information. For example, within minutes of @MarkJones's tweet, @RedLion began tweeting out his own tweets with the information, addressed to humanitarian response agencies, sending a total of seventeen addressed tweets about this incident, using either of the two versions presented below, and changing the mentioned account for each (bolded here for emphasis):

> **@RedLion** (2010-01-26 15:33): URGENT **@navynews** Confirmed reports of Human Trafficing of childen in Hopital Espoir, Delmas 75. Can you check this out please?

> **@RedLion** (2010-01-26 15:43): rt NOW **@usarmy** RT @MarkJones 100% sure of source. It may be too late by the time we get there. Kids being sold @Hopital Espoir, Delmas 75

In parallel information routing efforts, several other users sent similar tweets addressed to organizations that they perceived to have people on the ground at Haiti at the time. Interviews with TtT translators revealed that some voluntweeters felt that they could help by routing actionable information to

[67] This data set of TtT users is only a subset of all Haiti Twitterers and therefore the retweet numbers in this example represent a subset of the total number.

people or organizations that could act on it. These tweets demonstrate how voluntweeters did this using addressed tweets in their public Twitter streams. Some may have also been using private direct messages and other platforms to attempt to move actionable information into the hands of people who could help.

5.4.6.2 Geolocating information

In a separate effort, remote Twitterers also worked, together with a Twitterer on the ground, to find GPS coordinates for the child trafficking report. Within an hour of the initial report, @MarkJones requested help from his followers in finding GPS coordinates for the hospital (first tweet below). @mallori50 quickly replies with coordinates—in a tweet that was retweeted by another of the TtT Twitterers— and @MarkJones rebroadcast those coordinates in his own tweet (third tweet below). The source of the coordinates provided by @mallori50 is not completely clear, the text appears to be copied from another resource, which was published on "Jan 24, 2010" and also listed the hospital as "open."

> **@MarkJones** (2010-01-26 15:58): Help me get the GPS Coordinates RIGHT NOW for Hopital Espoir, Delmas 75
>
> **@mallori50**[68] (2010-01-26 ~16:02): @MarkJones: Jan 24, 2010 ... GPS Coordinates: Coordinates: 18°33′0″N 72°18′0″W ... Hopital Espoir (Hope Hospital) is open ...
>
> **@MarkJones** (2010-01-26 16:08) PM GPS Coordinates appear to be 18°33′0″N 72°18′0″W ... Hopital Espoir (Hope Hospital) is open It is off 75

After sending out those preliminary coordinates, @MarkJones then noticed a tweeted reply from @SergeGilles[69] that came in seconds before with a slightly different set of coordinates. He quickly replied

---

[68] @mallori50 is not a TtT Twitterer and this account's tweets are not in the *Contextual Streams of TtT Twitterers.* This tweet is found as a retweet within the tweet stream of another TtT Twitterer, and for this reason, the tweet ending has been truncated and the time is approximate.

[69] Neither @SergeGilles now @MarkJones were TtT users during the research collection window and therefore neither of their data streams are contained within the *Contextual Streams for TtT Twitterers* data set and neither are pictures in any of the EDV images. However, on the same day that we collected the streams for the TtT users, EPIC researchers also collected user streams for Twitterers we noticed, during

to @SergeGilles and posted another tweet updating the GPS coordinates for the Hospital to those recommended by @SergeGilles.

@SergeGilles (2010-01-26 16:07): @MarkJones Hopital Espoir +18° 32' 28.44", -72° 16' 47.87"

@MarkJones (2010-01-26 16:09): @SergeGilles OK. I will repeat those now.

@MarkJones (2010-01-26 16:09): Better coordinates: Hopital Espoir +18° 32' 28.44", -72° 16' 47.87"

5.4.6.3 Making Composite Tweets: Adding Context

@MarkJones had by then provided two sets of GPS coordinates for the location of the original report of child trafficking. However, none of his tweets yet combined all three components: the report, the location, and the GPS coordinates. This could have become a problem for downstream Twitterers who might have been able to act on the information, including those receiving routed tweets from other voluntweeters, who may not have the time or other resources to follow the ongoing effort to geolocate the information. Soon, several Twitterers began to fill this gap by sending out composite reports including both the initial information and the newly found GPS coordinates. The two tweets below are examples of these composite tweets:

@MLBaxter (2010-01-26 16:14): @NavyNews Per @MarkJones Hosp Espoir +18° 32' 28.44", -72° 16' 47.87" - Grd rpt of Child Trafficing - pls see his twts - Notify authorities

@greenbrookmom (2010-01-26 16:19): @MSF_USA need your help, right now!!! child trafficing delmas,! Kids are sold from Hosp Espoir +18° 32' 28.44", -72° 16' 47.87" Notify...

28 tweets in the *Contextual Streams of TtT Twitterers* are composite reports of the incident that include GPS coordinates. Figure 11 shows an image from the eDataViewer of the crowd activity to geolocate the child trafficking report, displaying requests for GPS coordinates in aqua, tweets with

---

our TtT deployment, as having impact in some capacity. The tweets in these examples from these two users were captured during that "high impact user" collection.

coordinates for the hospital (but without the original report) in pink, and composite tweets that include the GPS coordinates in purple. The image shows a quick progression from the original report to composite tweets, and demonstrates how the work took place across multiple voluntweeters accounts.

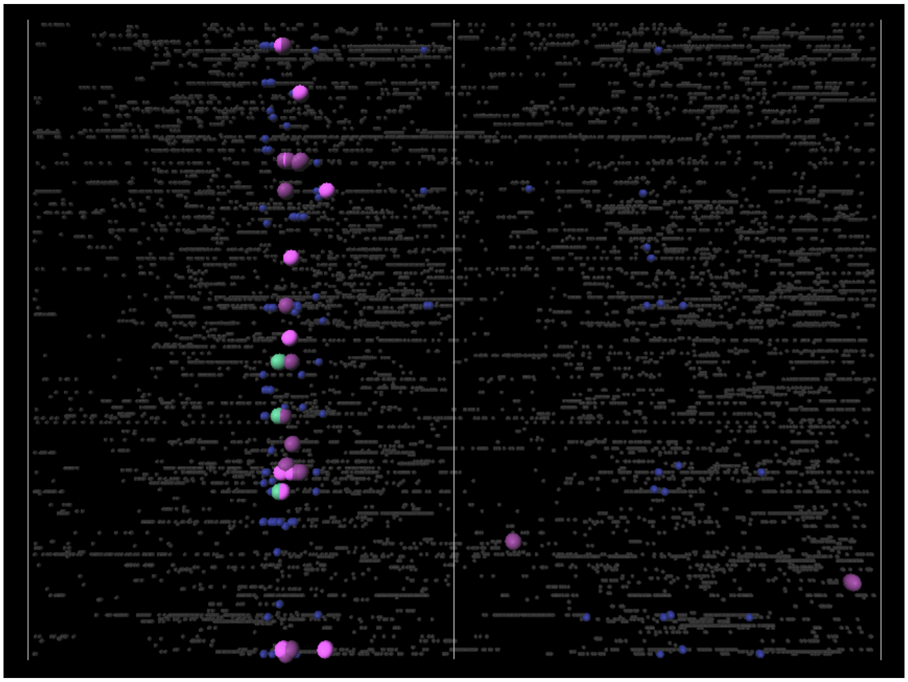


Figure 11. EDV Image of Effort by Twitterers to Geolocate Report of Child Trafficking

5.4.6.4 Structuring Information into Reports

Before the first composite tweets went out, but after the efforts to geolocate the information, someone[70] created an Ushahidi report from the information, mapped at the GPS coordinates provided by @SergeGilles. @rojosalvo tweets out a link to that report, along with a synopsis of the reported incident:

> **`@rojosalvo`** `(2010-01-26 16:12): #haiti CHILDREN being trafficked, please help NOW: @ Hopital Espoir, Delmas 75. Please HELP` **`http://dlvr.it/sJF`**

Two Twitterers quickly created TtT tweets about this child trafficking incident. @forjo1 generated the first TtT tweet (first below). She marked the tweet as a `#need` tweet and included the `#contact` tag,

[70] Ushahidi reports are anonymous, so it is unclear who submitted the report for this incident.

designating @MarkJones as the contact person. She also included the GPS coordinates, though without marking them with the TtT `#loc` tag.

> **@forjo1** (2010-01-26 16:41): @Unicef #Haiti #SOS Hopital Espoir +18° 32' 28.44", -72° 16' 47.87" **#contact** @MarkJones **#need** security

> **@lawndude** (2010-01-26 16:45): #haiti #need Security #loc 18.55000 -72.29995 #info ongoing human trafficing

@lawndude quickly created another, similar TtT tweet from the information. Like @forjo1, he designated the tweet as a "#need" tweet and reported that "security" was needed. This, along with the tweet time—only a few minutes after @forjo1's—suggests that this second TtT tweet was informed by the first. @lawndude added a #loc before the GPS coordinates and changed these to a decimal approximation of the original degree-minutes-seconds format. He also added an #info tag followed by "ongoing human trafficking" to give more context to the report of security needed. The relationship between these two TtT tweets may indicate that the second TtT tweet was intended to be a correction of the first, meant to both add context and bring the tweet closer to standard TtT format.

5.4.6.5 Verifying Information

Another activity that was clearly taking place within the Twitter streams of digital volunteers was an effort to verify the human trafficking report. In the quote below, Linda relates how an orientation towards verifying information that reflects lessons learned from voluntweeting during Haiti:

> Linda: "Crisis tweeting" is an art really. Accuracy can mean the difference between life and death for the people directly affected. That is why we ask that ppl not RT info that they havent confirmed.... Sometimes, during Haiti, ppl would tweet info or urgent needs (like someone trapped) and it was several days old. Ppl want to help but dont appreciate that they may cause harm.

Verifying, for the voluntweeters during this event, took place in several stages. In the first few hours, as the information was being routed to news media and response agencies, @MarkJones made several assertions concerning the credibility of the information and the existence of multiple sources.

@MarkJones (2010-01-26 15:17): @susanlee @erhj2 @UNICEF -I have a credible name of a nurse in this hospital who is giving us an SOS for human trafficking! HELP!

@MarkJones (2010-01-26 16:07): Confirmed local POLICE have used this hospital for trafficking before. Please pray. Hopital Espoir means Hospital of Hope.

@MarkJones (2010-01-26 16:20): @LargeSoda I have received a direct eyewitness report from a trusted source inside of this hospital.

@MarkJones (2010-01-26 16:32): I have confirmed eyewitness inside reports of kids being sold right now @ this hospital. Contact military or aid orgs now & urge help

Though we do not have evidence within the *Contextual Streams of TtT Twitterers*, these repeated assertions of the credibility of his source suggest that he was fielding inquiries from others, possibly newspersons or response agencies, who were attempting to confirm the information—his contextual stream contains several tweets addressed to news agencies and humanitarian groups during this time, as well as retweets of responses he received from some of those accounts. His tweets also suggest that much of this conversation was occurring on Twitter's backchannel, through direct messages. In his public stream, he repeatedly instructed other Twitterers to "DM him" for more information.

Later that afternoon and evening, @MarkJones sent out a string of tweets stating that he had been in contact with military and other response agencies and that they are following up on the report. His tweets suggest that some of these sources eventually reported back to him that they could not, however, find evidence of human trafficking at the hospital. In the following tweet, he addressed two Twitterers, one of whom is a TtT translator, perhaps in reply to inquiries from them about the status or even the validity of the report:

@MarkJones (2010-01-27 00:10): @friends4ever @loveandlight just wrote an update for you all. No great news. Not sure what I expected, but no bust yet. Maybe tomorrow?

At this point, some Twitterers had begun to question this report and other variations of the information that began to diffuse within the set.

In the first tweet in this section on *Verifying Information*, from very early in the event, @MarkJones states that he had contact information for a nurse inside the hospital who confirmed that trafficking was occurring there. A few hours later, another Twitterer sent out a tweet that appeared to be a misrepresentation of @MarkJones's tweet. This tweet, sent by @JaneSM[71], stated that a nurse was *involved* in the trafficking:

> **@JaneSM** (2010-01-26 ~18:03): CHILD TRAFFICKING BY A NURSE @ Hospital Espoir, DELMAS 75 contact #SOS #RESCUEHAITI contact @MarkJones CHILDREN BEING SOLD!

The tweet was followed by at least two variations from the same account that then directed the information about the trafficking nurse to the @UN and @USAID_Haiti accounts. Over the next 25 minutes, these tweets were retweeted by ten accounts within the *Contextual Streams of TtT Twitterers*, spreading the misinformation. Before that diffusion of misinformation ended, another volunteer Twitterer corrected @JaneSM, tweeting:

> **@sam123** (2010-01-26 18:06:43): @JaneSM check again, think info from a nurse not traffic by...

> **@sam123** (2010-01-26 18:14:22): @JaneSM Mark named nurse who reported it not who was guilty of it. Go back & see.

Other voluntweeters in the *Contextual Streams of TtT Twitterers* jumped into the conversation to tell @JaneSM that the information was already being acted on and that the AP and military were already "on it." Some accused her of overreacting. Many requested DMs from her so they could switch to conversing

[71] @JaneSM was not a TtT Twitterer and I do not have access to her full contextual stream from this time period. There is evidence within the *Contextual Streams of TtT Twitterers* of her tweets and interactions through retweets of and addressed tweets to her account by other TtT Twitterers. Because of this, all of the timestamps for @JaneSM's tweets are approximate times, interpreted from how the tweet diffused within the contextual streams of others.

through the Twitter backchannel, away from the public eye. In interviews with TtT translators, a few said that they would use or attempt to use direct messages to confront people about misinformation.

Soon, @JaneSM changed her strategy and began to question the credibility of the whole report, asking:

> **@JaneSM** (2010-01-26 ~18:35): Either we get to the bottom of it NOW and find out what is up with that.....or we get to the bottom of people making false statements

> **@JaneSM** (2010-01-26 ~18:40): Okay original source was NOT SERGE and NOT Mark they followed up on source and situation...going to find it now.

These tweets are two of a long stream where @JaneSM reacted strongly, perhaps to criticism of her accusation of the nurse being involved, to the report, addressing both the sad possibility of human trafficking and the rumor-like nature of Twitter. In the second tweet above, @JaneSM demonstrated that she had been following up on the report and searching for its original source. She tracked back through the tweets that were sent referring to the reported information, and retweeted several of the intermediate versions to demonstrate which Twitterers passed on what information.

She and a small group of others continued to question the human trafficking report, though their voices did not diffuse far. Each of the following tweets received less than three retweets in the TtT Twitterers data set.

> **@radioto** (2010-01-27 00:15): Is this story is true or not ? "...Heard that there is a Human Traffic of childen in Hopital Espoir, Delmas 75..." PLEASE OFFICIAL SOURCE

> **@henson234** (2010-01-27 06:15:46): @JaneSM re: Hopital Espoir, only MarkJones knows? Is he in Haiti or Atlanta? If in Atlanta how does he know?

Less than 24 hours after the initial report, @SergeGilles, the author of the first tweet that made the accusation of child trafficking at the Hopital Espoir, tweeted that he followed up on the information and that he could not confirm it:

@SergeGilles (2010-01-27 12:37): I investigated the child trafficking case @ hopital espoir. It was a rumor. However It's still a pending thread in other areas #haiti

This declaration of the report to be merely a rumor, by its original source, was retweeted 16 times, and re-sourced in tweets using other text six times. Altogether, this shows significantly less diffusion than the original report and follow-up composite tweets, suggesting that though a well-meaning crowd does act to self-regulate, misinformation travels further than corrections of that misinformation.

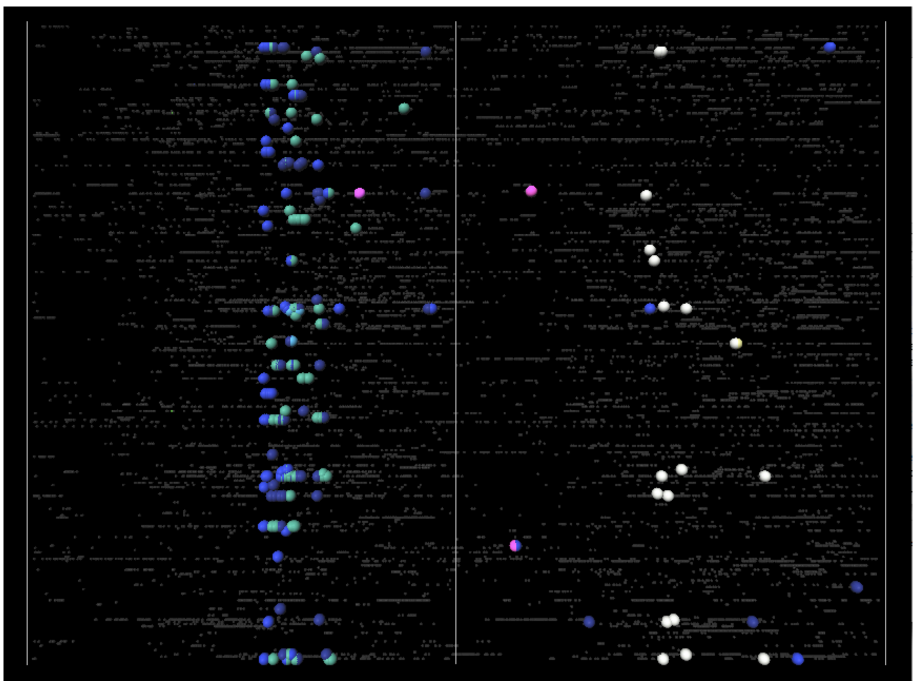


Figure 12. Effort by Twitterers to Verify Report of Child Trafficking

Figure 12, another eDataViewer graphic, shows how Twitterers worked collectively to verify the child trafficking report. All colored spheres contain information from the original report. The brighter blue spheres contain a reference to the original source, @SergeGilles. Light green spheres are tweets that also make an assertion about the credibility of the source or say that the information has been confirmed. This behavior occurs slightly after initial reporting. Pink spheres are tweets openly questioning the validity of the report. This confronting behavior only shows up in the public stream after the initial flurry of activity surrounding the report. Finally, white spheres are tweets stating that the information is now deemed by its original source to be a rumor. Interestingly, the diffusion of corrected information is less concentrated and significantly lower in volume than the initial efforts to amplify the provocative report.

The image also shows that reports with uncorrected information from the original report were still moving through the system after the correction tweet fully diffused. One of these tweets, sent hours after the child trafficking report was determined by the crowd and the original source to be a rumor, is a third TtT tweet about the incident:

> **@shortcutt** (2010-01-28 00:42): #haiti #need HELP CHILD TRAFFICKING #loc +18° 33' 0.00", -72° 18' 0.00" Hopital Espoir, Delmas 75

Demonstrating anew the problem with lost context in the Twitter space, this last tweet appears to have been created using an earlier composite tweet or possibly multiple tweets as sources for the information. However, the author has not followed the entire conversation surrounding the report and does not recognize that the report is both outdated and potentially false.

This excerpt shows verifying behavior occurring in several different ways. @MarkJones's assertions of his sources and the information credibility, and other Twitterers' retweets of these assertions, demonstrate an effort to make a claim of validity for the report. When @JaneSM misinterpreted the information and made a blatantly incorrect claim, many Twitterers confronted and corrected her. Reacting to this confrontation, she then began to question the validity of the whole report, and with the help of (and with resistance by) several other Twitterers, she began to track back through the Twitter stream to trace the original source, and sent addressed and perhaps also directed messages to accounts that acted to diffuse the information, seeking to get more background on the report. Meanwhile, @MarkJones had effectively sparked a large response effort and several media and response agencies attempted to follow up on the information. Perhaps in reaction to one of these two things, the public questioning of the information's validity or the on-the-ground response effort that has been set into motion, the original source of the information went himself to the hospital to attempt to confirm the information and reported back that it is just a rumor.

5.4.6.6 Information Processing as Collective Intelligence

This long excerpt regarding the information movement of a single report of child trafficking shows how incremental acts by voluntweeters to amplify, route, verify, and structure reports functioned collectively to organize information and make it more useful to others in the space. Later in this dissertation, I will frame this group information processing behavior as collective intelligence, using the framework of distributed cognition to demonstrate how the network of voluntweeters worked to process information.

5.4.7 Connections Between Translators

The ability to leverage connections was a powerful part of the work process for the digital volunteers in this study. When we examine their contextual tweet streams, we see that the TtT translators were a highly interconnected group. Figure 13 shows a visualization of the entire network of translators (80), including our own research group members (in black). Edge thicknesses represent the (natural log of the) amount of interactions between each pair of Twitterers as determined by *addressed tweets* or *mentions* within the contextual tweet streams. This is not a measure of Twitter *followers* and *followings*, but rather of who was publicly "talking" to whom during the January 10-February 1 time window. Four translators were *unconnected* to the network, translating tweets without direct interaction with other volunteer Twitterers. Nine others were connected to one other Twitterer within the group. The rest had direct interactions with multiple members of the network. The network is dense, with each translator connecting with, on average, 7.7 other translator-volunteers (excluding our researchers), and some of these connections are leveraged repeatedly, with over 40 mentions.

Our Project EPIC accounts (in black) are visible on the top right of the dense part of the network cluster. Note that we are not central to the network, and that the density exists elsewhere (in the lower left quadrant). In fact, when we create another graph removing all of our researchers from the network, the core of the network remains intact.

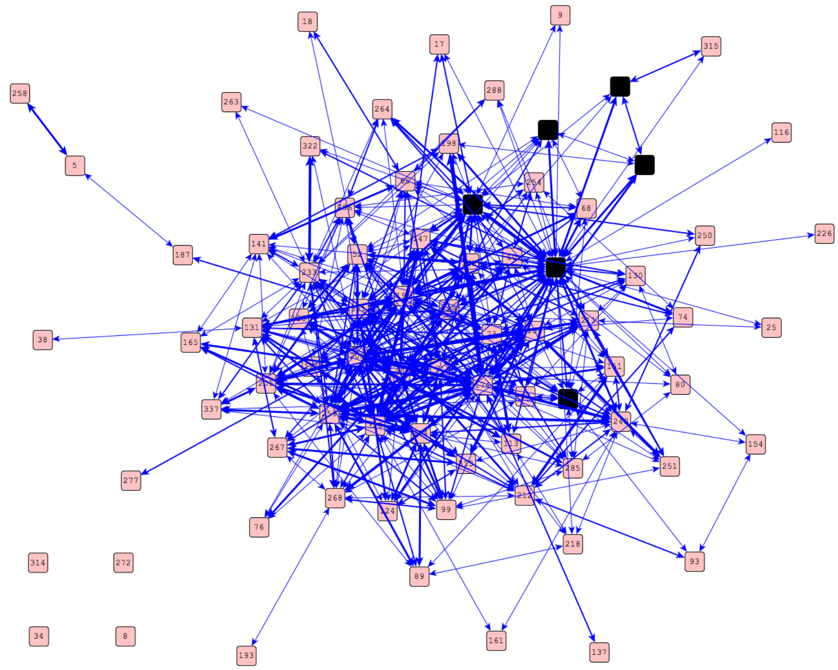


Figure 13. Connections between TtT Translators

To understand when and how these connections were formed, we sent each interviewee a list of all the other translators that they addressed a tweet to during the study time window in a follow-up email. We asked whether or not they knew each Twitterer in real life or through Twitter prior to the earthquake. Though a few listed one or two previous connections, a clear majority replied that they had never previously connected with these other Twitterers.

> Julie: We all came together by circumstance.
>
> Karen: All of these people I had never spoken with, prior to the Earthquake. I would have found all of them via the #Haiti #HelpHaiti or other Haiti hashtags, or occasionally a retweet from someone already in my Haiti network.

In the latter excerpt, Karen explains how she created her volunteer network, using hashtag searches to identify other Twitterers with whom she wanted to connect, and sometimes used established connections to find new ones.

> Karen: I met many people through them whom I also came to trust. We worked on Twitter, Skype, Google Wave, Ushahidi & email - and during one emergency, on the phone when Skype failed us.

5.4.8 Features of Ad-hoc Communications Infrastructures

Participants we identified as remote operators incorporated a variety of social media and communication tools into their work practice, including email, Twitter, Skype, Google Wave, and Ushahidi. Within the Twitter platform, they used different affordances for different communication needs: public updates for broadcasting or seeking a wide range of help; addressed tweets for making connections, directing information, and challenging misinformation; DMs for moving resources, exchanging other contact info, and confronting possible hoaxers. Several cited the need to verify and manage misinformation as rationale for moving to tools with more access control, like Google Wave.

> Meg: It was difficult to keep up with the stream at times, and some of us tried to branch [off] a "core" team to Google Wave, so we could have conversations without strangers butting in and confusing matters.

Without previous training, crisis tweeters cobbled together sophisticated, ad-hoc, digital infrastructures by appropriating existing tools.

> Linda: Social media will evolve to meet the needs that present during a crisis. ... Ppl say a crisis isn't the time to learn a new technology. That may be true but that's when solutions are inspired. Like meatball surgery in a war zone, you make do with what you have and along the way, you might invent a better way of doing something

5.4.9 Volunteering through Tweeting: Role Identity, Momentum & Consequences

For many respondents, crisis tweeting during the Haiti event became and continues to be a significant part of their Twitter identities. Over half of interview participants report or have evidence in their recent tweet streams indicating continued volunteer tweeting during crises. Five still appear active in their Haiti tweeting (Aug 15-Sept 10) and seven have tweeted for other, more recent, crisis events. In their profile descriptions, @rachellehoude now refers to herself as a *voluntweeter*, @Meira_Davi claims to be a *crisis tweeter*, @MelyMello states that she is "heavily involved in Haiti relief via working w/ an amazing, effective, dedicated group of volunteers," and @cait proclaims that she "won't forget Haiti." Their crisis tweeting—a term we have adopted from the practitioners to describe this behavior—has come, in some ways, to define their Twitter personas.

5.4.9.1 Self-Identifying as a "Voluntweeter"

Between January 21 and January 23, Twitterers throughout our translator dataset began to tweet or retweet updates with the terms "voluntwitter" and later, "voluntweeter."

> **@janeSM** (Jan 21): Am stunned-have gotten sups in-saved people fr rubble-brought them drs-we have best team! We R voluntwitters!

> **@rachellehoude** (Jan 21): I HEART VOLUNTWITTERS!

These tweets show a growing self-identification with the role of crisis tweeting as crisis tweeters, or voluntweeters. For those who embraced this role, the connection between their publicly enacted crisis tweeting behavior and their Twitter persona seems to contribute motivation for continued participation. A few participants displayed another layer of awareness, that these digital volunteer activities represented the birth of a new phenomenon.

> @rachellehoude, from her blog: "Everyone over the past few weeks has felt that buzz, that collective energy flowing through each one of us, inspiring action and involvement. It's been incredible to be part of it and connect with like-minded people. It's been awesome to witness the birth social media's soul."

Long-standing research on the sociology of disaster suggests that the desire to help in times of crisis is age-old, and in fact is a behavior in disaster response that is critical to response and recovery (Dynes, 1970; Fischer, 1998; Tierney et al., 2001). Large numbers of people are known to converge onto the site in the wake of a disaster event to observe or to help (Fritz & Mathewson, 1957; Kendra & Wachtendorf, 2003b). Currently, novel forms of this phenomenon result from the affordances of social media tools, which allow more and remote people to contribute in ways not previously possible (Palen & Liu, 2007).

> Kelly: I don't have a lot of money and I'm not brave enough… to travel to Haiti, but this was an opportunity to get involved and, with the use of twitter, in a big way! I was already impressed with the reach twitter had, but seeing what it could do during a crisis like Haiti BLEW my mind away.

5.4.9.2 Emotional Impact

In their initial responses, few interviewees mentioned the emotional impact from their digital volunteering. Follow up comments were far more likely to touch on this, even though we did not ask specifically about these experiences. Some remarked about rescue attempts that arrived too late and calls for help that went unanswered:

> Maria: It was horrid to hear all these cries for help; it haunts me to this day and I still have nightmares about it.

One respondent told us that she had not continued crisis tweeting and remarked that she was not sure that she could do it again, because the work was too "emotionally draining." Another respondent, one who had participated in crisis tweeting during the political unrest in Iran, talked about how she backed away from her account after that event, before starting all over again for Haiti.

> Linda: During Iran, I got very caught up in it and was on twitter constantly. I dropped off after that when I changed my [Twitter] name. I kept my follower count low, stayed away from politics and emotional issues.

Though we could not reach the people who had ceased tweeting from their Haiti-focused accounts, we can speculate that some, like Carla, had abandoned their Haiti-tweeting identities purposefully, suffering from emotional and perhaps also physical exhaustion.

5.4.10 Frustration with Formal Response

Another theme that emerged is a frustration that some volunteers began to feel towards the formal response effort:

> **@rachellehoude** (Jan 21): This is getting frustrating. I keep seeing the same calls for help, day after day. Can the NGOs at least HIRE us so we can do this for them?

After over a week in the space, coordinating small-scale response using ad-hoc networks and over-the-counter digital infrastructures, this voluntweeter felt unappreciated and even obstructed when the "formal" response moved into place. Another explains how the new "organization" of response prevented them from accessing the resources that they had been using for their "informal" response activities.

> Meg: …after that stage, it became clear that the larger NGOs were locking down streams of access for individuals. UN Cluster meetings were mandatory in order to obtain aid from the storage facilities. It became harder to help and our teams started to crumble due to feelings of powerlessness. It was a very frustrating time and most of my efforts turned more to shedding light on the broken system than trying to fight against it to get little done.

In follow up comments, many participants said they hoped the formal response would learn how to better coordinate with social media volunteers in the future.

5.5 Discussion

The activities of the crisis Twitterers are a new form of previously recognized organizing behavior made possible through the availability of new media. Sociologists of disaster, including Dynes (1970), call organizations that did not exist prior to a major disruption of the social order *emergent organizations*—groups of people that previously had no standing structure or defined tasks. (The other organizational types—*established*, *expanding*, and *extending*—have varying degrees of pre-existing structure and task definition.)

Emergent groups are usually self-organizing, and come about in disaster settings to meet some unmet need. *How* they come together is the subject of Kreps and Bosworth (1994), who discuss a "theoretical tool" to explain collective behavior and organizing through a progressive manifestation of key features: Domains (D), Activities (A), Resources (R), and Tasks (T). Organizations can arise out of any sequential permutation of these mechanisms (though some arrangements are more common than others), as Kreps and Bosworth (1994) empirically describe in a historical survey of accounts of emergent organizations in numerous disasters. Employing their framework to the behavior that many have come to call "crowdsourcing" we believe helps better explain and identify the underlying phenomena behind this term, including what makes it possible in disaster, and perhaps in other aspects of social life as well.

The temporal sequence that we believe best explains the behavior of the emergent voluntweeter population we studied is the $R \rightarrow A \rightarrow T \rightarrow D$ configuration of self-organization, with the Resources and Activities mechanisms of self-organizing being most developed, and with some indications of progression to the Tasks and Domain mechanistic stages. Groups that begin with Resources or Activities appear less often in Kreps and Bosworth's empirical taxonomy; this might be a type we see arising more often from today's ICT-supported environments. Note that not all emergent organizations endure long enough or organize completely enough to manifest all four attributes in their lifespans.

5.5.1 Resources (R) as Mechanisms of Self-Organization

For the "voluntweeters" in crisis events as described here, the means by which they began to work as volunteers in the Haiti event was the medium of Twitter. In other words, their initiation into the space was through the accessibility of *resources,* which are the "individual capacities and collective technologies of human populations" (Kreps & Bosworth, 1994, p. 25).

Twitter was the gateway into an information space that itself contained resourceful features that helped would-be volunteers navigate within it, and begin to self-organize with others in an evolving practice of crisis tweeting.

5.5.1.1 Hashtags

Hashtags were used by many as an entrance into the space—to find key information and identify the people participating in relevant conversations. Hashtags can be used to filter information to help narrow the focus of search. They were an important basis for the eventual organizing of activities (stage 2 in Kreps and Bosworth's organizing model). For example, `#rescuemehaiti` signaled a particular kind of help request, and became a tag around which people organized:

> Alena: One of them…proposed an idea; to use #rescuemehaiti hashtag. Promptly me and him made sure people asking for help would tag this word and within half hour everyone was following and using this tag. He said one day later officials contacted him and said this tag was very helpful so we should continue using it.

Increasingly, it appears that the hashtag is employed as a mechanism for identifying useful social connections.

5.5.1.2 Syntax

Among our population, the TtT syntax also served as an organizing feature (though not the primary one) within the network of crisis tweeters. TtT added structure to tweet information; even when the syntax was not used in full, it helped voluntweeters to remember to include—and teach others to include—the necessary pieces of information into a single tweet to make it complete.

There was also indication that some saw the syntax as imparting a kind of authority because of its perceived rigor and assumptions about attention to accuracy by those using it. The manner in which some interviewees referred to the TtT syntax suggests an important structuring effect. One participant speaks about "syntaxing" as a task; another refers to the syntax as a kind of language, as in "translated into syntax." Though TtT syntax users may not fully understand what a syntax is, this co-opting of the term into language about voluntweeting suggests the syntax's role as an organizing feature.

5.5.1.3 Data Entry & Data Movement

"Individual capacities" are another resource that can initiate self-organizing, particularly among social action types of activities. Here, stepping into the Twitter space as well as the disaster volunteer role (typically for the first time), the basic activities of retweeting and entering data into a syntactical form using TtT, or entering data into Ushahidi's records, were the origins of a crisis tweeting practice.

5.5.2 *Activities (A) as Mechanisms of Self-Organization*

Activities are defined as "conjoined actions of individuals and social units" (Kreps & Bosworth, 1994, p. 25), and we find this to be the second organizing mechanism in a large sector of the crisis tweeting network. It is in this stage when those drawn to Twitter, having discovered and used initial resources, begin to coordinate activities.

5.5.2.1 Data, Translation, Filtering, Verification, Cross-Referencing

The activities described in this paper—information filtering, amplification, verification, and synthesis, tweet translation, and cross-referencing data records between Ushahidi and TtT-formatted information—mark a graduation from data entry to coordinated activity with other people and groups.

5.5.2.2 Action Work

In addition, the hybrid virtual-physical "action" work of @Melymello, for example, who leveraged her Twitter social connections to obtain goods and services of direct benefit to those in Haiti, is an instance of conjoined action.

5.5.2.3 Conjoining Activities with other Organizations

A few notable organizations were entities to which the emergent band of voluntweeters began to connect. These included Ushahidi, CrisisCommons and Shaun King, an Atlanta, Georgia pastor of a faith-based organization. Shaun's Twitter account was the most mentioned in our entire dataset. Thirteen of the 19 interview respondents mention or address him in their tweets, and two talk about him in their interviews. His activities began during the early aftermath of the quake and are ongoing in September 2010, with his aHomeInHaiti.org website serving as a place to coordinate donations and provide shelter for Haitians.

The self-organizing mechanisms of resources (R) and activities (A) seem to be clear in this emergence of a coherent, productive group of voluntweeters. In some cases, we begin to see introduction of division of labor—which in part describes tasks (T)—and some instances of movement to more formal collectives of domains (D).

5.5.3 Tasks (T) as Mechanisms of Self-Organization

Tasks are defined as "collective representations of a division of labor for the enactment of human activities" (Kreps & Bosworth, 1994, p. 25). As a core set of crisis voluntweeters gained experience with this emergent practice, some broke off to have extended and more private conversations elsewhere. These interviewees note that Skype and Google Wave became the means by which coordinated activity was discussed. Conversational excerpts provided earlier, including that between @CarolB and @IstanbulTWSTVL, demonstrate some division of labor, as does @MelyMello's activity of adding minutes to cell phones.

Some establishment of norms, especially with respect to how and when to retweet and verify information, began to occur with the newly experienced, acting as mentors, pointing out to less experienced people what constitutes the most helpful tweeting behavior. Volunteers also began to challenge possible hoaxers, a Twitter behavior that Mendoza et al. recognize as occurring during the aftermath of the February 27, 2010 earthquake in Chile as well (Mendoza et al., 2010).

5.5.4 Domains (D) as Mechanisms of Self-Organization

Finally, for Kreps and Bosworth, domains are "collective representations of bounded units and their reasons for being" (1994, p. 25). For emergent, self-organizing groups, "bounded social collections" may or may not come about. Our investigation reveals some mix of consequences in social sub-groups of the broader crisis Twitterer population.

There are volunteers who did a good amount of work for Haiti, but then fell away after a time. With Haiti as the sole focus of their Twitter use, their accounts go dormant or are even deleted entirely after their event participation ceases.

Still, others started reflecting on what it meant to be a good "crisis tweeter," and publicly identify as such. Though some of these multi-event crisis tweeters continued to work without affiliating with an organization, others found emerging organizations like *Humanity Road* as digital "places" to focus their volunteer activities. Humanity Road (humanityroad.org) is a virtual organization that was sparked by digital volunteering activities during the 2009 political unrest in Iran, but then formalized during the Haiti event (personal communication, Chris Thompson). Its members, two of whom were interview respondents, describe themselves as performing crisis tweeting during disaster events, as well as mentoring crisis tweeters between events. The formation of Humanity Road represents the full culmination, as described by Kreps and Bosworth, of the graduation from resources and actions (for them, during the Iran unrest) to "domain"-driven features of organization (during the Haiti event).

5.6 Conclusion

The empirical examination of the work practice, products and motivations of crisis tweeters who emerged in the aftermath of the 2010 Haiti Earthquake reveals important features of self-organizing in a highly networked world. The broad attribution of social media-related behaviors to "crowdsourcing" can be more finely understood, in this case, as a collection of resources, capacities and a progression to increasingly more defined tasks and even organizational identity. In the matter of sudden and tragic events, the desire that some feel to help is newly enabled by resources like Twitter, where assistance can be provided remotely. Information creation and movement as the basis for social connection and subsequent collective action is at the core of these operations. The emergent ICT-abetted behaviors we have documented here and their consonance with knowledge about existing self-organizing mechanisms suggest that the digital volunteer will become a common and likely influential feature of social life.

CHAPTER 6
STUDY 3. AT WORK AT A "DISASTER DESK": ENACTING & SUSTAINING A VIRTUAL VOLUNTEER ORGANIZATION[72]

In this study, which offers additional insight towards Research Question 3, we continue to explore the organizing efforts of digital volunteers—in this case shifting our attention to a nascent virtual organization of volunteers who constituted a subgroup of the Haiti voluntweeters described in Study 2 (Chapter 5). Looking beyond that emergent-network response to a single event, this study investigates the ongoing work of a virtual, volunteer-based, disaster response organization to develop and sustain its activity over time and across events.

6.1 Introduction

During the response and relief efforts following the 2010 Haiti earthquake, a small group of geographically dispersed volunteers, many of them voluntweeters (from Study 2, Chapter 5), came together using online communication tools and formalized as a non-profit organization called Humanity Road, with the intention of continuing to respond to the Haiti disaster and then on to other crises and humanitarian events around the world using ICT. Humanity Road, a 501(c)(3) based in the United States but incorporating volunteers from all over the world, became a first-of-its-kind, non-profit, virtual volunteer organization operating in the domain of disaster relief.[73] The group assists in response efforts by helping to verify and process information from impacted areas and providing public safety information.

Related to the theme of this broader work, this chapter focuses on the organizing, work practices, tools, and products of Humanity Road, exploring the relationships between ICT, structure and action in an emerging new genre of virtual volunteer organizations.

---

[72] A shortened version of this chapter will appear as:

Starbird, Kate and Leysia Palen. (2013). Working & Sustaining the Virtual "Disaster Desk." *ACM 2012 Conference on Computer Supported Cooperative Work* (CSCW 2013).

[73] An announcement on Humanity Road's website from July 2010 asserts this first-of-its-kind claim: http://www.humanityroad.org/announcements/humanity-road-inc-sm-first-of-its-kind-digital-disaster-relief?A=SearchResult&SearchID=1519674&ObjectID=27990&ObjectType=7

6.1.1 Virtual Organizations

Though a formal virtual organization for disaster volunteerism was a novel concept in 2010, virtual organizing has a history as long as technologically supported communication. In their early investigation of computer-mediated communication, highlighting the features and potentials of computer conferencing and predicting a networked future in social and organizational life, Hiltz and Turoff (1978) described how technology opened up new channels for organizing, enabling group members to work together across time and space, and enabling new manifestations of collective intelligence. Thirty years later, Shirky (2008a) celebrated many of these same themes, asserting that technology was aiding the formation of new kinds of groups, enabling new ways of coordinating, and occasioning new possibilities for collective action. Both works spoke broadly about new possibilities and configurations for groups in a connected world. Though neither used this term, others have discussed the dynamics of online groups as *virtual organizations*.

During the 1990s, as email communication began to see widespread use both in formal organizations and informal workgroups, distributed teams and virtual organizations became the foci of significant research inquiry (e.g. Finholt & Sproull, 1990; Finholt et al., 1990; Handy, 1995; Ahuju & Carley, 1998; DeSanctis & Monge, 1999). Ahuju & Carley (1998) define a virtual organization as "a geographically distributed organization whose members are bound by a long-term common interest or goal, and who communicate and coordinate their work through information technology." This research in the domain of mass disruption suggests a slight alteration to this definition, a removal of the "long-term" distinction, as some virtual organizations of digital volunteers, like the voluntweeters described in Study 2 (Chapter 5), organize quickly and in some cases temporarily to respond to specific events.

Virtual organizations (VOs) are enabled and structured by the technological tools that members use to connect, communicate, and organize. Several researchers have noted that therefore virtual organizations are significantly different from traditional organizations that rely on face-to-face communication (e.g. Hiltz & Turoff, 1978; Finholt et al., 1990; Ahuju & Carley, 1998). Before computer-mediated communication (CMC) technologies had achieved widespread adoption, Hiltz & Turoff (1978) claimed that, due to the affordances of the technologies used for communicating, VOs would be more lateral, less hierarchical, and less centralized than traditional organizations. These predictions echo in commentary about virtual organizations that continues to highlight presupposed egalitarianism. Finholt et al. (1990) hypothesized that groups relying on CMC would experience less "spoke and hub" or centralized communication where messages are mediated through certain individuals, and more "circle" or de-centralized communication where messages move directly between all members of the group. However, that study found that email use in ad hoc groups actually developed into patterns that fit both kinds of communication models, centralized communication that aided coordination and de-centralized communication that was good for information sharing.

Ahuju and Carley (1998) report that VOs are characterized by informal interaction that results in emergent structure, generated as members perform their organizational tasks. DeSanctis and Monge (1999) counter that virtual organizations actually require *more* formal interaction, arguing that because the volume of communication is higher in CMC interactions, there may be "pressure to make some communication more formal or programmed in order to gain efficiencies and bring routine to otherwise customized work." Those authors highlight several features of communication specific to CMC, asserting that as communication volume increases, efficiency decreases; that as comprehension decreases, establishing, supporting, and communicating a shared social context becomes critical; that lateral communication reduces hierarchy and domination and increases the diversity of participation; that norms of technology use emerge and then develop as the organization evolves; and that interpersonal relationships develop and deepen as another evolutionary effect. Mapping these features of CMC to the dynamics of virtual organizations, DeSanctis and Monge (1999) argue that VOs will be characterized by highly dynamic processes, reconfigurable structures, and edgeless, permeable boundaries that may occasion increased conflict, especially when norms are not compatible across cultures.

It is important to note that early work on virtual organizations was often based not on "pure" virtual forms, but on virtual groups formed by leveraging previously established relationships between members, often within or between existing entities, and groups whose members still relied to some extent on face-to-face communication. Focusing on that latter characteristic, Fiol and O'Conner (2005) classify virtualness by the amount of face-to-face contact among team members, distinguishing between *pure virtual*, *hybrid*, and *face-to-face* teams. This study focuses on the former distinction, whether the organization has or has had a structure that pre-exists its virtual form, assigning *pure virtual* to organizations that arise originally in virtual form and have no significant pre-existing structure in physical space. Using this definition, both the voluntweeters and Humanity Road are pure or nearly pure virtual organizations, emerging from interactions mediated by connective technology.

6.1.2 Social Computing in a Virtual Volunteer Organization

Exploring relationships between traditional volunteer (non-profit) organizations and social computing, Voida et al. (2012) report that these organizations, contrary to the perceptions of many, do not need help recruiting new volunteers—a perceived application of social media tools. Volunteer coordinators feel that the "episodic" volunteering encouraged by public, "all call" social media broadcasts is not useful to them. Instead, they see potential for social computing to help them meet other priorities, including fostering community among current volunteers and supporting a deep engagement with these volunteers. Voida et al. (2012) describe this potential as yet unrealized in traditional non-profit organizations.

But what of the purely virtual, volunteer organization? Humanity Road, an early member of the emerging ecosystem of virtual humanitarian response groups, provides an opportunity for exploring virtual volunteerism within a domain for which it may be particularly well suited. Pure virtual organizations rely entirely upon social media and other ICT for their organizing, and for Humanity Road, these tools are tightly woven into the social structure of the organization—shaping who they are, what they do and how they do it. Additionally, disaster events often act as catalysts for the *convergence* of people, information and resources (Fritz & Mathewson, 1957). During these events, many physical convergers shift into roles of helping (Fritz & Mathewson, 1957; Dynes, 1970; Kendra & Wachtendorf, 2003b) and we see this phenomenon in the online arena as well (Hughes & Palen, 2009). Humanity Road, in fact, emerged from digital, spontaneous volunteerism (see Study 2, Chapter 5), and that phenomenon remains a constitutive element of the organization. Their continued incorporation of spontaneous volunteers intentionally differs from how traditional volunteer organizations, specifically the American Red Cross, function.

This chapter follows the trajectory of Humanity Road from an emergent group to a formal non-profit, considering how its work, its articulation of work, and the ICT it adopts are interrelated, and how they together express the virtual organization it has become. This account maintains a particular focus on how the organization sustains itself over time in a domain, both virtual and event-driven, that is twice-over ephemeral.

6.2 Research: Analytical Frame and Method

6.2.1 Evolution of an Organization: Structuration as Interpretive Frame

Structuration is a social theory originally posed by Giddens (1979; 1984) that claims that institutional social structure and human action are mutually and recursively enacted and defined. A post-positivist theory that rejects strictly deterministic explanations of social behavior, structuration views human action as limited and guided by the social structure within which it is enacted, and yet also functioning to constitute that same structure in an ongoing process of mutual constitution.

In describing structuration, Giddens differentiates between *structures*, defined as rules and resources or transformational relations, and *social systems*, which are "reproduced relations between actors or collectivities, organized as regular social practice" (1984, p 25). Social systems, in his view, exhibit structural properties, but structure only exists in its "instantiations" as reproduced social practices and as "as memory traces orienting the conduct of knowledgeable human agents" (1984, p 17). Giddens argues that structure is both the "medium and the outcome" of these reproduced social practices (1979, p. 25). Using a concept he refers to as the *duality of structure*, Giddens explains that "the rules and resources drawn upon in the production and reproduction of social action are at the same time the means of system reproduction" (1984, p 19). Structuration theory therefore views social action as both shaping and shaped by these structures. Giddens also stresses that structure both constrains and enables action, an important feature that maps well to studies of how technology shapes human action.

Orlikowski (1992) provides a framework for applying structuration theory to research on the role of technology within organizational change, explaining that technology is both the product of human action and a medium of human action, functioning to enable and constrain it. Drawing from the recursive relationship described by Giddens' concept of the *duality of structure,* Orlikowski proposes the *duality of technology* and suggests this phenomenon of mutual constitution exists at both an individual and institutional level.

Desanctis and Poole (1994) offer another extension of Giddens' theory to address how ICTs affect organizational change. Adaptive structuration theory (AST) attempts to account for the fact that the same technology deployed in similar environments can lead to different structural outcomes. AST outlines a method of uncovering the recursive relationship between technology and action that involves enumerating the *structural features* of an existing organization and identifying the *structural moves* at the micro-level of enactment. Marking a significant difference from the research methods described by Orlikowski (1992), AST prescribes a substantial focus at the micro-level, on individual instances of appropriations of technology, the visible actions that give evidence for how a rule or resource is being used. Desanctis and Poole elaborate, "New social structures emerge in group interaction as the rules and resources of an (advanced information technology) are appropriated in a given context and then reproduced in group interaction over time" (1994, p 129).

6.2.2 Examining Mechanisms of Organizing in Virtual Organization through Digital Traces of Communication

Weick (1976; 1995) asserts that the organization, as a noun and a singular thing, is a myth, and argues for a shift towards looking at organiz*ing*, the process. For virtual organizations, much of the process of organizing unfolds within the digital communications between, and digital artifacts created by, members of the group. Because a large portion of this communication is recorded digitally as it occurs, virtual organizations provide a unique opportunity for studying organizing through the digital traces of communication. DeSanctis & Monge (1999) elaborate, “The virtual organization provides a metaphor for considering an organization design that is held together, literally, by communication.”

This statement reflects a CCO view of organizing, whereby *communication constitutes organization*, i.e. the two co-produce each other and develop together (Putnam et al., 2009). Putnam et al. explain, “In this perspective, communication exists prior to the organization and shapes the context in which structural forms emerge” (2009, p. 8). In the CCO perspective, an extension of structuration theory, organizing happens through communicative acts that shape the structures of the organization-in-progress. Extending this perspective, McPhee and Zaug (2009) propose four different “flows” through which communication constitutes organization: *membership negotiation*, *organizational self-structuring*, *activity coordination*, and *institutional positioning in the social order of institutions*. This study of virtual organizing relies on the CCO perspective, positing a constitutive role for communication within the organization and seeking to describe the organizing of Humanity Road through the digital traces of their communication. The work also demonstrates how the four flows of McPhee and Zaug’s (2009) model develop and move within a virtual organization, and exposes a fifth communicative flow within this virtual organization that is literally the work that they do.

6.3 Methods: Participant Observation within a Virtual Organization

This investigation of the Humanity Road (HR) organization is based primarily upon participant observation by the author of this research, who spent more than 16 months as a volunteer and participant observer within the Humanity Road organization, taking part in several efforts to respond to disaster events in 2011 and 2012.

6.3.1 Humanity Road: Overview and History

Humanity Road is virtual organization of digital volunteers that responds to disaster events across the globe using social media and other online tools to help process and distribute information. The organization is incorporated as a 501(c)(3) in the United States, though it recruits and incorporates volunteers from all over the world. They respond to a variety of crisis events, including natural and man-made disasters, by virtually deploying as a remote, volunteer workforce that collects, verifies, and routes information to people and organizations who need it.

Though HR incorporated as a formal non-profit in April 2010, their organizing began several years earlier in 2005, when sisters Catherine Graham and Chris Thompson envisioned and then deployed an "Internet Café" for disaster victims of Hurricanes Katrina (they later consulted on a similar effort for Ike) (Humanity Road, 2011). A few years later, in the summer of 2009, Graham and Thompson were among a small but growing group of online volunteers, at that time helping to provide first aid information to protestors involved in the Iran Election protests. Two of the nineteen voluntweeters whom we interviewed for Study 2 (Chapter 5) reported that their entry point to what they called "crisis tweeting" began with activity around those protests[74]. Less than a year later, the 2010 Haiti Earthquake acted as a catalyst for bringing together a core group of online volunteers, many of them voluntweeters, who would later become the original members of Humanity Road.

HR's activities to support the Haiti Earthquake response began on January 14th when volunteers created a website (helphaitiheal.wordpress.com). Chris Thompson explains that, early in their efforts, the group published a list of "Twitter Commandments" on their website (Figure 14), giving other crisis

[74] Though Graham and Thompson are considered part of the larger voluntweeting effort for Haiti, they were not interviewed for Study 2 because neither were TtT translators. However, both retweeted TtT tweets and were part of the larger sample analyzed for that study.

Twitterers guidance for how to sort facts from rumor and tweet responsibly during disasters (personal communication, September 2010). Through their digital volunteer activities on Twitter and Skype, relying to some extent on the emergent structure of the *voluntweeters*, the original members of HR began to meet and recruit others. Thompson claims that the group's early presence in the voluntweeting space and the structure they offered for how to tweet helped them meet and recruit new members (personal communication, September 2010). The group filed to become a formal non-profit organization in late February 2010, and celebrate their birthday as coinciding with Earth Day of that year—April 22, 2010.

Twitter Commandments

1. **Do No Harm –** It is safer to share no news than to share inaccurate news. Rumors put lives at risk.
2. **Do not Panic** – You do not need to know it all
3. **Take a Deep Breath –** Do not be distracted by noise and confusion
4. **Verify Source.** If you do not know for sure the source is reliable – do not retweet the information. Use 2 reliable independent sources for major news
5. **Verify Facts.** Get facts, locations, address, specific need, number of people impacted, population at risk, dig deep into details, the more the better.
6. **Listen to the officials and experts**. Use caution and reason and follow those who appear to have a 'handle' on how to respond in these situations.
7. **Use Tweak the Tweet** – We recommend formatting your tweet using Tweak the Tweet (http://epic.cs.colorado.edu/TtT/editors/pakistan/).
8. **Not sure where to start?** Pick one topic and stick with it. Become an expert on fielding researching facts and providing help on specific topic.
9. **Repeat the first 3** Twitter Commandments as needed.

Figure 14. Humanity Road's Twitter Commandments

They selected their name, Humanity Road, after a road in New Orleans that received attention after Hurricane Katrina. That name reflects a connection between the original core group of volunteers and the southeast United States. However, the organization recruits, trains, and deploys (virtually) volunteers from all over the world. Over their history, they have had volunteers reporting in from the U.S. and Canada as well as countries in Africa, Asia, the Caribbean, Europe, the Middle East, and South America,

In their first year as a formal organization (2010), HR volunteers responded to 72 events in 53 countries and participated in a collaborative exercise related to social media and disaster response. During 2011, they virtually deployed for 132 crisis events, including earthquakes, fires, volcanoes, and civil unrest, and took part in ten exercises. Through their response work and within the exercises they participate in, they have formed partnerships with several other organizations in the disaster response domain. As of March 2012, HR continues to follow its core mission of collecting, verifying and providing information during disaster events to the emergency response officials and affected people who need it.

6.3.2 Becoming a Humanity Road Volunteer

During the response period following the January 2010 Haiti Earthquake and in subsequent disaster events that year, I had considerable contact with HR members through my efforts to deploy the Tweak the Tweet—many HR volunteers were also TtT translators, TtT retweeters and/or voluntweeters. Most of my early interaction with HR members occurred through the Twitter platform, and over time I developed followed/following connections with many of HR's volunteers. One of my colleagues at Project EPIC took part in some of the group's early organizing efforts in March 2010. However, I did not become a HR volunteer until late 2010.

I was first invited into the virtual work environment of the organization in November 2010, during a second wave of digital volunteer activity to support humanitarian efforts in Haiti. This was at a time when the Haitian people were still coping with the massive casualties, economic upheaval and physical displacement from the catastrophe of the earthquake itself, were struggling to combat the first wave of a cholera epidemic that would infect nearly half a million people in its first year (Centers for Disease Control and Prevention, 2011), and were simultaneously facing an approaching hurricane, Tomás, that was likely to strike the island country the following morning (Kurczy, 2010).

On the night of November 4, I was attempting to help other digital volunteers, several of whom were HR members, to use Tweak the Tweet to create a shelter map for Haiti. One of the organization's volunteers, Amanda[75], whom I had previously connected with over Twitter and with whom I often exchanged both public @mentions and DMs (direct messages which are private tweets between two

[75] All names of Humanity Road volunteers are anonymized, except leaders Catherine Graham and Chris Thompson.

users), suggested I join the group's Skype chats to better coordinate our efforts. HR uses the Skype platform's synchronous, textual chats for much of their real-time coordination during disaster events. Amanda "invited" me, creating access for me to participate, into several Skype chats that evening, including HR's *Café Window*, a backchannel for their volunteers, and a *Haiti Cholera Window* where volunteers from HR were working together with other digital volunteers and organizations to coordinate digital response efforts to the cholera epidemic. Earlier that day, volunteers had begun to use the Haiti Cholera Window to work together to support Haitians during Hurricane Tomás as well.

I spent several hours in the HR chatrooms that night and through the next day, until it became clear that the impact from the hurricane had been relatively minor and there was no longer a need for us to support a shelter map. Over the next few days, I continued to interact with HR volunteers and leaders in their public chatrooms and in one-on-one Skype chats, learning more about their efforts to help collect and process information related to the cholera outbreak, and gathering more information about the organization in general. After being encouraged to do so by its leaders, I became a formal member of HR a few weeks later, by filling out an online form provided by their website.

I began my research efforts as a participant observer in late January 2011. In this capacity, I both act as a digital volunteer within the HR organization and as a researcher of the organization. When interacting within the group's Skype windows, I make this dual role explicit in my handle, which appears as "Kate Starbird – student/researcher" for everyone following the conversation.

6.3.3 Studying a Virtual Organization through its Digital Traces

The primary data for this research are the digital traces of communication that HR members create and then leave behind in the social media platforms and other online tools that enable their interactions. These tools include Skype, Google Docs, Twitter, and the organization's webpage. The first two in that list of platforms maintain a digital record of all the interactions between users of the tool and, in the case of Google Docs, the documents they create. Skype saves the entire history for every textual chat (the feature used most by HR volunteers), and allows every participant of a chat to access the chat history spanning from the moment she entered that chat window to the current time. Unless purposefully deleted by their creators, Google Docs are saved by that platform in their final version and also maintain a history of edits, showing which user edited what content in what order, and those histories can be recalled by anyone with document privileges at any time in the future. For HR's Skype chats and the Google Docs we incorporate into our work, my research draws on the digital traces of the interaction histories maintained by the platforms themselves.

Twitter is also an important tool for HR, used for a variety of tasks including coordination, communication, and recruiting. To capture the digital traces of Twitter communication, the research relies on Twitter data collection software developed and maintained by my colleagues at Project EPIC (Anderson & Schram, 2011). These tools use a combination of Twitter APIs to collect data pertaining to HR and to the disasters to which the organization responds. For the entire period of my participant observation within HR, Project EPIC has used the Twitter REST API to collect the complete public timelines of those HR volunteers who have provided their Twitter handles to the organization. Using Twitter's Search and Streaming APIs, Project EPIC also collected tweets that contained certain event-related search terms or hashtags during many of the disaster events covered by HR. Additionally, Project EPIC collected several data sets during the emergency period of the Haiti earthquake, including the contextual streams of voluntweeters described in Study 2 (Chapter 5), and these Haiti data provide evidence of some of the early organizing efforts of future HR volunteers.

Another important record of HR's activity is their website (http://www.humanityroad.org), which provides general information about the organization, tools for volunteers, information for collaborative partners, recruiting materials for prospective new members, and resources for the public during crisis events. Through this website and the public-facing, digital material published and stored there, the organization communicates its messages for the outside world. The website is therefore a place both for doing the work of the organization—i.e. providing resources to responders and the affected public, and for establishing the organization's institutional positioning within the social order of other institutions (McPhee & Zaug, 2009).

This research examines other digital documents—e.g. Word, Excel and PDF files—created by the organization's members and shared through Twitter, Skype, and email. It also relies on field notes of observations of our collective activity, recorded in a digital document during and immediately following those event responses, team trainings, and team meetings.

Drawing from this extensive, digital record of the diverse activities of HR, captured through the ICT tools with which we communicate, and using snapshots of the digital resources we use and create, this study illustrates how this virtual volunteer organization works—what we do, as well as why and how we do it. Significantly, the fourteen-month period of my participant observation overlaps with a very early stage in the organization's history, and offers a window to examining virtual organizing in a group where structure is still nascent and extremely fluid. Incorporating a CCO view of organizing, this study looks through the ICT-mediated and digitally stored communication at the organization-in-process, with particular focus on the relationship between the tools that digital volunteers use, the actions they take in a time critical and high stakes environment, their developing work practices, and the evolving organizing structure.

6.4 Observations: Work Practices, Tools and Products of a Virtual Volunteer Organization

Now approaching the end of its second year in action as a formal organization, HR has developed into a robust organization with a multi-faceted mission and continues to establish a role for itself in the domain of humanitarian response. Since its 2010 incorporation, HR membership has shifted from three to two leaders, with adjustments in the regular membership makeup as well. Today it relies of a small but steady group of the two leaders and six now-seasoned volunteers who devote considerable time and resources to HR activities both during and between disaster events.[76] A goal of the organization is to be "24 always on, ready to go," an aim supported, in some part, by their global workforce which positions volunteers across time zones. In direct contrast to strategies employed by the traditional non-profits described in Voida et al. (2012), to meet increased resource needs during response efforts, HR also incorporates a larger workforce of intermittent and spontaneous volunteers who activate during specific disaster events. After events HR will ask new volunteers to formally register as members. Some members persist across events and participate in monthly planning events, and others activate only episodically. The number has gradually risen to about 20 active members during any four-month period.

The overall work is quite diverse, including a continued focus on monitoring social media during disaster events and creating resources from this information. Volunteers also work between events to help educate the public on how to prepare and respond to events ranging from small emergencies to large-scale disasters. Its messages range from commonly broadcasted warnings on how not to get caught in floodwaters to advice focused on pet and animal care during various types of crises. The organization's mission has expanded and branched out in new directions—for instance, the goal of educating the public has broadened through efforts by HR's leaders and other experienced volunteers to train traditional responders and others on how to incorporate social media into their response plans. Throughout these first two years, the organization has worked to create connections to other disaster response organizations, both traditional and digital, and now regularly collaborates with many of these groups for real-time coordination of digital response efforts and while participating in disaster response exercises and simulations.

Activities and roles within the organization are just as diverse as the range of its broader mission. Among other things, volunteers work to monitor multiple information sources during disasters, update content on websites and Facebook pages, tweet preparedness and response messages, translate materials into multiple languages, create press releases and respond to media, connect with representatives of other

[76] During 2011, in addition to full-time activity by HR's two leaders, four volunteers responders to more than 30 events each, and three of those reported over 1000 hours of service.

organizations, put out a monthly newsletter, help to develop and maintain the technological components of the website, and help to raise funds for the organization.

The diverse work practice and expanding mission of HR offer vast opportunity for research inquiry. This study focuses on three distinct, but related aspects of the virtual volunteer organization. First, it offers an overview of the core functions of HR, of who they are and how they view their role in disaster and humanitarian response. In describing these functions, this account shifts between two views, providing both a snapshot of a nascent organization at a single point in time, and a short history of the becoming of this organization. To this latter point, I also aim to illustrate how the group has come to define its goals, its products, its role, and its work practices. This is the second aim of this chapter, examining how a pure virtual organization enacts the work of organizing, highlighting in particular the relationship between ICT, action, and structure. The third goal of this chapter is to address the issue of sustainability for the organization: How does HR, an organization that emerged from spontaneous volunteerism in the wake of highly publicized crisis events, carve out a role for itself in a developing ecosystem of virtual volunteer organizations and maintain its membership and activities over time and across events?

Though the findings in this study draw from participant observation that spanned over fourteen months and more than 100 hours of interaction, a significant portion of this study concentrates on a single event response to a large earthquake in Peru that occurred in October 2011, mid-way through the second year of HR's existence. This one event demonstrates several significant aspects of HR's work, offering a view of organizing-in-progress through an extended excerpt of the organization-in-action. The excerpt incorporates digital traces of interaction mediated by several different platforms during a three-hour response to that event. To address other goals of this study, e.g. to examine how volunteers work between events, and to explore the role of HR in the larger realm of disaster response, this study also draws on smaller excerpts from other events and analyzes digital traces of volunteer activity between disaster events.

6.4.1 HR's Mission: Driven by Need, Led By Experience, Powered by Volunteers

Upon incorporation in the spring of 2010, HR declared their mission as:

> We are a global citizen's action team of experienced and focused individuals dedicated to educating the public by providing accurate, critical recovery information, before, after and during a catastrophic disaster. We are a non-governmental not for profit organization aligned with the United Nations global disaster response clusters as well as US state federal and local disaster response groups. We pledge to act responsibly and quickly in identifying, facilitating, collecting and disseminating required information and solutions to the public affected by the event

This statement strongly echoes the activities HR volunteers were engaged with at the time: response in an immersive information context of rapid social media posting. It also reflects a strong orientation towards positioning the organization within current disaster response and humanitarian relief structures. However, there is no clear allusion to the virtual nature of the organization and its work. In the fall of 2011, the mission statement was "reduced … and [made] easier for folks and volunteers to remember" (Graham, personal communication). This is HR's current Mission Statement, available on its website, which describes the organization's current focus and perceived role in disaster response:

> *Driven by Need, Led By Experience, Powered by Volunteers.* Humanity Road's mission is to educate the public before, during and after disasters on how to survive, sustain and reunite with loved ones. Humanity Road volunteers are trained and equipped to use Internet and mobile communications technology to collect, verify and route information online during sudden onset disaster. Using the Internet, they provide public safety information as well as directing the public to governmental and aid agencies that are providing assistance for the disaster. (Humanity Road's Mission Statement[77])

This statement, like the original, hints at the organization's roots in volunteer activity preceding its incorporation in April 2010. The first sentence of the statement, an assertion that "Humanity Road's mission is to educate the public … on how to … reunite with loved ones," resonates closely with earlier experiences of HR leaders Thompson and Graham during Hurricanes Katrina and Ike (years before HR became an organization) where they worked to set up Internet Cafés to assist affected people in

[77] The first sentence is broadcast on their home page: http://www.humanityroad.org. The whole statement is available via a link from their home page to http://www.humanityroad.org/AboutUs.htm

connecting with loved ones. The second sentence asserts that volunteers are trained to "collect, verify, and route information" during disasters. These core activities align closely to earlier efforts by HR volunteers and other voluntw*eeters* to assist the humanitarian response during the 2010 Haiti earthquake and the protests in Iran after 2009 Election. Unlike the first mission statement, this version clearly scopes out the organization's focus on using ICT in their response efforts, reporting that volunteers "use Internet and mobile communications technology" to do their work.

These connections between HR's mission statement and previous volunteer efforts reflect how the organization developed atop previous structure laid down by volunteers' activities during earlier events. Because many of the original volunteers had already been active in an emerging space of digital volunteerism, it is also probable that the actions of other volunteers in the space had influence on early digital volunteerism efforts of HR members, and therefore had a shaping effect on HR as a whole. Additionally, two of the founders of HR had previous experience working with the American Red Cross (ARC) and part of the mission and the early "design" of the organization hints to this relationship—as a concrete example, the early rationale for the Internet Cafés grew out of Graham's work with the ARC during the aftermath of Hurricanes Katrina and Ike, where she recognized a gap between informational needs of the affected population and information distribution strategies.[78]

The mission statement itself acts as a structuring mechanism within the organization, contributing to volunteers' understanding of what the organization is and what it does. This is an example of organization's *self-structuring*, one of the four communicative flows described by McPhee and Zaug (2009). The statement serves a second communicative purpose of *institutional positioning*, being a message to the outside world that carves out HR's role during disaster response efforts.

Following that mission statement, volunteer work at HR centers on educating and informing the public before, after and during disaster events on how to, in their words, "survive, sustain, and reunite

---

[78] Personal communication with Catherine Graham and YouTube video posted by Chris Thompson. In the video, Thompson describes the relationship between Graham's ARC work and the Internet Cafés, explicitly noting a connection between those efforts and Humanity Road.
http://www.youtube.com/watch?v=1oZx58zDQDM

with loved ones." In a recent *All Hands Conference Call*, a monthly meeting that takes place on Skype through a combination of multi-person voice chat and textual chats, group members discussed their "core business functions" as being threefold: Disaster Preparedness and Education, Disaster Response, and Process Improvement.

[Excerpt: Skype Textual Chat – 3/24/12]

> **HR Catherine Graham** (3/24/12 9:36am MST):
> Core Business Functions
> In what line of work is the Nonprofit involved?
> Disaster Preparedness
> Disaster Response
> Process Improvement

6.4.1.1 Disaster Prevention and Preparedness: Educating the Public Before Events

The first of HR's three core functions is *Disaster Preparedness and Education*. This long-standing role for the organization comes through in the mission statement within the larger goal of "educat[ing] the public before, during and after disasters." HR devotes a considerable portion of volunteer and technical resources to preparedness and education, using its website, its @HumanityRoad Twitter account, and the Twitter accounts of many volunteers to communicate a multitude of diverse messages to the public. The HR website (www.humanityroad.org) maintains a variety of disaster education resources, including a "whole community" page with advice on how to organize a local response group, and several blog pages that contain entries with preparedness and response information. For example, a January 2012 entry on the *HR Top Picks* blog page provides information about the "Ready, Set, Go!" program, which emphasizes preparedness and early evacuation during wildfires, while the *RedCrossDog* blog and the *Animals in Disaster Digest* blog both offer pet and animal care information. These latter two are part of a larger effort within HR to address the welfare of pets and other animals during disaster events.

Though the resources on HR's website are publicly available, volunteers use Twitter to broadcast disaster preparedness and education messages. The following are examples sent from HR accounts:

@HumanityRoad (Mar 15, 2012): After #Tornado: Use the telephone only for emergency calls. ow.ly/9m93S #MIwx #hmrd

@HumanityRoad (Mar 15, 2012): After #Tornado: Use battery-powered flashlights when examining buildings – do NOT use candles. ow.ly/9m93S #MIwx #hmrd

@AlexanderB (Mar 15, 2012): #Tornado: Watch for flying debris. Flying debris causes most injuries/fatalities. http://1.usa.gov/9kqUn3 #hmrd #miwx #ohwx

---

@RedCrossDog (Mar 16, 2012): Woof! #Flood: If you can't evacuate #livestock, move them to higher ground. bit.ly/gq6bMi #rcdog

@RedCrossDog (Mar 16, 2012): Woof! #Winter tip: #Puppies don't tolerate cold like adult dogs. You may need to housebreak on paper during the winter. bit.ly.i4YiGu

@HumanityRoad (Mar 27, 2012): Are you Ready? Functional needs vary. Be sure to make provisions for medicines that require refrigeration. 1.use.gov/elqL4M #hmrd

@HumanityRoad (Mar 27, 2012): Today's #Sustain tip > Fresh clean water is critical to surviving in disaster. Add or rotate a fresh gallon in your supply kit today. #hmrd

For outgoing messages from HR volunteers, the tweet content aligns to the operational strategy of the group at the time, i.e. whether the group is actively responding to an event or not. The top three tweets from the above example were sent *during* and immediately *after* a March 2012 series of tornado events. Tweets sent by the organization during event impact and response periods are designed for two purposes: first, to give information about what to do in the moment to those immediately affected; and second, to educate others about how to respond to a certain type of event at a time when HR believes they are most receptive to learning—that is, when an event of a type likely to affect them is happening to someone else.

The four tweets in the second part of this example were sent during times when volunteers were not actively responding to disasters. In the organization' mission statement, between time is referred to as "*before* periods." For HR, a small organization with a limited amount of volunteers available at any given time, these between periods are times to change focus from the high-action, real-time information processing activities of live event responses, to education messaging concentrating on prevention and preparation advice. The shift in operational strategy visible in their outgoing tweets and the language used in their mission statement to describe their functional stages correspond to a disaster life-cycle model where phases of *mitigation* and *preparation* precede disaster *response* and *recovery* (Drabek, 1986). The underlying structure of HR's operational strategy reflects knowledge of the domain of disaster response brought into the organization by HR leaders Thompson and Graham, both of whom had previous experience as ARC volunteers.

HR has developed a way to reduce the workload of and standardize their preparedness and educational messaging through *tweetables*[79], pre-fabricated messages that offer information about how to prepare or respond to specific kinds of disasters. The organization recommends its volunteers broadcast these messages between and during events, using personal accounts or, for those who have access, the official @HumanityRoad account. During events, volunteers can choose from a variety of tweetables to select ones that relate to the conditions of the current event.

A large percentage of HR's outgoing Twitter communications, especially those sent between events, are tweetables. Though they are not representative of all of the communications from HR and its volunteers, all of the tweets in the example at the beginning of this section are tweetables. To extend the

[79] Listed here on the original Humanity Road website: http://humanityroad.wordpress.com/virtual-emergency-operations-center/tweetables/. Currently, Humanity Road maintains their tweetables in a Google Doc with access limited to volunteers with privileges to tweet from the @HumanityRoad account. This limited access is designed to prevent accidental data deletion from the document. Volunteers can request @HR tweeting privileges.

breath of their message across different times zones and keep the account active when volunteers are not, HR uses the Twuffer tool to schedule tweets for future broadcast. The last two tweets in the above example were scheduled through Twuffer, which appears as the "source" of the tweet in its meta-data.

Several distinct tasks have emerged to support the dissemination of tweetables: identifying tweetable material, crafting a tweetable message using 140 characters, translating tweetables into multiple languages including rapid translation into the language of an affected area when needed, tweeting and retweeting relevant tweetables during disasters, and scheduling non event-specific tweetables for the HR account—or "feeding the Twuffer," as some volunteers call it. The following excerpts from Skype chats show HR leaders discussing tasks related to tweetables.

[Excerpt: Skype Textual Chat – *HR Useful Links* window – 11/9/11]

> **HR Catherine Graham** (11/9/11 7:11am MST):
>
> How about adding something like this to our power outage tweetables and anytime tweetables
>
> In high risk weather or power outage? Establish check-in times with loved ones & set phone to optimum settings or turn off between use

[Excerpt: Skype Textual Chat – *HR Useful Links* window – 10/1/11]

> **HR Chris Thompson** (10/1/11 11:46am MST)
>
> @FEMA We teamed up with @FCC with a full list of tips on using your technology to stay in touch after a disaster http://go.usa.gov/8YY
>
> that's great tweetable stuff and twuffer food

Tweetables can be seen as a form of organizational *self-structuring*. These ready-made messages offer structure to volunteer's activities, giving them something tangible to do during events. Even an inexperienced volunteer can jump in by retweeting tweetables as they are broadcast by the @HumanityRoad account, or she can access the list of tweetables herself and choose to tweet one that relates to an emerging event. These messages also work to keep the outgoing public communications from HR volunteers and the @HumanityRoad account aligned with the organization's overall message.

As tweets are added to the list of tweetables, usually by seasoned volunteers, volunteers continue to define what that overall message is and how it appears to the outside world.

6.4.1.2 Disaster Response: Educating the Public *During* and *After* Disaster Events

The second core business function of the HR is *Disaster Response*, which primarily manifests as informing the public *during* disaster events. This part of the mission overlaps with work to educate the public on how to respond to a current disaster. It also includes the collective efforts by volunteers to process information and create publicly available resources during disaster events. HR's mission statement sums this up by saying that they "collect, verify, and route information," but they also filter, synthesize, and structure it, among other things.

Information collection and processing activities occur for two different purposes and in two slightly different, yet overlapping phases. In the initial period after a disaster event has been identified, collection efforts focus on improving the situational awareness of active volunteers so they can make decisions about how the organization should respond to an event—how much effort should they expend, should they shift resources from other HR activities, should they attempt to activate offline volunteers, how many resources should they create and of what kind, when should they stand down, etc.

As volunteers work to process information for their own situational awareness, and as questions about whether and how to respond are answered through information-seeking and collective interpreting activities, information-collating efforts begin to serve the second goal of creating resources for the affected public and responding agencies. These resources are published and continually updated (if needed) on the organization's website, linked to from outgoing tweets by @HumanityRoad and other volunteer accounts, and are routed to other organizations—including organizations on the ground of disaster events and other digital response organizations—through tweets, emails, conference calls, and for the latter group, in text-based Skype chats that bridge virtual volunteer communities and organizations.

Hurricane Jova

Catherine Graham - Monday, October 10, 2011

Hurricane Jova is targeting the Pacific Coast of Mexico with anticipated broad impacts to several Mexican states. Stay informed with weather updates from your local weather service such as The Weather Channel - Hurricane Central and to see current wave heights with Hurricane Jova, visit ocean weather. Be sure to follow local official guidance regarding preparations and evacuations. Below are helpful links for some local official sources.

Local Officials
Red Cross Mexico @Sedenacional
Mexico Meteorological Service @huracanconagua
Mexico Meteorologist Alberto Unzon @chaac_tlaloc
State of Colima, Mexico Civil Protection @PC_Colima
State of Colima, Mexico Governor @gobernador_mam
Mexico Seismological @sismologicoMX

Public Instructions:
(Colima) Los fenómenos más comunes que provocan desastres en nuestro país son incendios, huracanes, sismos e inundaciones. Este plan es una medida preventiva que requiere la participación de todos los miembros de la familia; incluso los niños y los ancianos. Se trata que todos los integrantes de la familia sepan qué hacer antes, durante y después de un desastre. Plan familiar de Protección Civillan
(Manzanillo) Manzanillo, Colima.- Ante la cercanía del **huracán "Jova"** a las costas de Manzanillo, el Sistema Municipal de Protección Civil pide a la población en general observar una serie de medidas precautorias. ¿Qué hacer antes, durante y después de un huracán?

Media
@MetMex
@Ntelevisa_Com
Manzanillo TV Online
The Weather Channel - Jova
@climamexico

Major Hashtags: #Jova, #huracn #hurricane
Micro hashtags: #Colima #Vallarta #Michoacan #Jalisco #Nayarit

Figure 15. Event Diary (Top Section Only) for Hurricane Jova, October 10, 2011

Figure 15 is a screenshot from the *Event Diary* blog on the HR website, containing a small subsection of a single blog entry related to Hurricane Jova, which was approaching Mexico on October 10, 2011. During large-impact or potentially large-impact events like this one that are determined to warrant a response effort by the organization, a volunteer or HR leader will create a blog entry within the "Event Diary" for the current event. This entry represents one publicly visible product of the group's efforts to process information during the event, containing a range of potentially useful information including a high-level account of the event, lists of relevant websites and Twitter accounts, warning information broadcast by official sources, hashtags to follow, etc. Entries for events post-impact typically include damage reports, nearest airport and hospital information, links to maps and other situational awareness reports, shelter information, the status of power and communication infrastructure, and a list of social media accounts that are providing first-hand information about the event.

Volunteers also turn to Twitter to send information about unfolding events, using that platform to route information to specific people and organizations and also to broadcast information to a general audience that may include affected individuals. These outbound tweets are closely connected to the information processing activities of the group. Selected tweets from HR volunteers during a series of tornados that struck the southeast United States in early March 2012 provide examples of some of this information processing work and the resulting Twitter activity.

> **@AlexanderB** (Mar 2, 2012): RT @foxnewsradio: #Tornado warning sirens are sounding in #Harrisburg, IL

> **@HumanityRoad** (Mar 2, 2012): RT @femaregion4: FEMA Blog- Preparing for Continued Severe Weather http://t.co/EQJQ9rsG #alwx #tnwx #kywx #gawx #tornadoes #hmrd

> **@SallyB** (Mar 2, 2012) #alwx RT @JimCantoreAL -descending funnel cloud reported just NW of Jeff and Nick Davis Rd .. valid now

The above tweets show HR volunteers retweeting accounts of official sources, amplifying the messages of those accounts. By monitoring their friend/follower networks and executing multiple hashtag searches, volunteers work to filter the public stream of Twitter messages, identifying trusted sources, Twitterers who are on-the-ground in an event, and relevant hashtags. In the third tweet, @SallyB adds an event-related hashtag (`#alwx`) to the beginning of @JimCantoreAL's original tweet, structuring the tweet to extend its reach by pushing that message into public searches for the popular hashtag. Conversely, HR volunteers also help instruct people on what hashtags to track for information relevant to the event or, as in the case of these tornado events that crossed geographical borders, for information about how the event is affecting a specific location, as the example below shows.

> **@AlexanderB** (Mar 2, 2012): RT @citizen_corps: Tags for following today's #tornado alerts and news: #alwx #gawx #kswx #ilwx #tnwx #kywx #inwx #alabama   cc: #smemchat

Though in this case @AlexanderB retweets a non-HR account (@citizen_corps) to inform the social media crowd on what hashtags to follow, HR volunteers will often create these lists themselves, then post them on the HR website and tweet them. Volunteers also create lists and tweet out account names of official sources and people to follow who are on the ground in the event. The following tweet demonstrates HR directing people in Tennessee to an official account for their area:

@HumanityRoad (Mar 3, 2012) #Tennessee #TN Follow @T_E_M_A and get update on situation at their website http://t.co/GRUM7iSK #tornado #hmrd

Though much of their messaging is meant for a broad audience that may include affected people and other volunteers, HR's volunteers also respond directly to people in the impact zone of disasters, using @mentions and private direct messages (DMs) to provide preparedness information and relay information about current conditions.

In the tweet excerpt below, after a relatively large earthquake in Turkey, a HR volunteer monitoring Twitter sees messages from two people talking about the trouble they are having contacting loved ones in the affected area. @HumanityRoad sends a public tweet directed at two of these Twitterers (the fourth tweet below) suggesting that they use text messaging instead of a voice call. Later, one of those Twitterers thanks @HumanityRoad and lets the volunteers know that they followed HR's advice, and it helped.

@aloaurora (May 18, 2011): Apparently I'm not crazy; a magnitude 6.0 earthquake hit Kutahya, Wst Turkey a little while ago…some phones don't seem to be working

@aloaurora (May 18, 2011): @johnsmyname Tried calling some friends, no signal or else recorded messages. Got through to couple others, not sure if lines are jammed

@johnsmyname (May 18, 2011): effing Christ. seems like everybody is calling Turkey. no way of getting through

@HumanityRoad (May 18, 2011): @aloaurora @johnsmyname if they have cell phones – trying sending a text messages – keep it short – just say RUOK?

@aloaurora (May 18, 2011): Thanx for the tip, tried it & it worked :)

@HumanityRoad (May 18, 2011): @aloaurora Glad to hear it worked! Text messaging frequently works when voice networks are too congested during disaster.

In the following tweet, @SallyB, an experienced HR volunteer, sends a tweet to a person in an area affected by floods, advising him to evacuate:

@SallyB (Jan 26, 2011): @samsun You'd best get yourself south of the river--roads are not getting any better!

Direct contact with affected people fits within their goal to help people who are sheltering in place or are otherwise immediately affected by the disaster at hand. This is another intentional difference between their operational strategy and that of the ARC, who focuses on the "sustain" phase of disaster, but does not provide help during the "survive" or impact phase (personal communication, Thompson). This strategy, which HR calls "just in time" messaging, is enabled by the social media tools that allow volunteers to identify and communicate with people in need.

In addition to the organization's website and Twitter communications, HR volunteers also create and contribute to resources available on other platforms. Volunteers have worked in Google Docs and other spreadsheet environments to develop and maintain hospital and shelter lists that include multiple and sometimes changing fields, like capacities and current occupancies. They also collaborate with outside groups and other volunteers to feed and maintain resources during events. During the cholera epidemic in Haiti in the Fall of 2010, HR collaborated with volunteers and researchers of the HealthMap effort to integrate clinic and case information about the emerging epidemic into a shared tool (Brownstein, 2011).

Though the examples given in this section are not an exhaustive description of all activities and resources created by HR during disaster events, they offer an overview of the types of tasks in which volunteers take part and the work products that they generate in their Disaster Response efforts. This chapter will delve further into *how* volunteers accomplish this work in Section 6.4.2.

These communicative acts of HR volunteers, created during their response activities and broadcast through available ICT, are difficult to classify within McPhee and Zaug's model of four communicative flows (2009). Though there may be elements of activity coordination, self-structuring and institutional positioning within these communications, those are not the primary motives or functions of these acts. For HR, a virtual organization operating both on and within an interactive environment constituted by information-communication, these outgoing communications constitute the information-processing activities that are literally the work of the organization.

6.4.1.3 Process Improvement: Extending our Mission to Educate

Another significant component of HR's work is *Process Improvement*, though this organizational function is not explicitly scoped out in the mission statement and has only recently been cited as a primary focus of the organization. One way that HR representatives fulfill this newly recognized role is by helping humanitarian responders, emergency managers, digital volunteers and others understand how to best leverage social media and digital volunteer communities to assist in response efforts. In a recent Skype chat within the HR Café window, Graham described Process Improvement as "contributing in ways that would help improve the tools/technology of disaster response" and rationalized that it does fit within HR's mission to "educate the public." HR activities that align with this increasingly important aspect of the organization's work include social media trainings for professional responders, disaster response exercises, consulting arrangements, conference presentations, and working with researchers to contribute to an increased understanding of the role of social media during disaster response.

HR's original foray into Process Improvement occurred during the 2010 Haiti Earthquake response, prior to the incorporation of the organization, when the original volunteers published their "Twitter Commandments" on the HelpHaitiHeal website[80]. Figure 14 contains a version of those commandments, extracted from HR's first website.

---

[80] http://helphaitiheal.wordpress.com/event-summary/about-2/

The commandments, a *self-structuring* device that pre-existed the organization, were designed to educate other volunteers on how best to use social media, especially Twitter, to assist in response efforts. Through these commandments, the original HR volunteers were attempting to improve how digital disaster work was done. This foundational document of HR activity shows that *Process Improvement* was a core function long before the organization identified the term and began to add it to their conversations about what their mission is.

Currently, HR leaders and seasoned volunteers are working in a variety of capacities towards improving disaster response. One way they are doing this is by leading formal trainings for professional responders, transferring what they have learned about media monitoring and social media crisis communications. In her role as a HR leader and volunteer, Catherine Graham has been consulting and training United States military personnel around their use of the QuickNets platform[81], a rapid response tool for situational awareness that incorporates social media feeds (personal communication). Graham, Thompson and other experienced volunteers also attend and present material at conferences related to disaster response, social media, and crisismapping.

Though their consulting efforts and conference presentations take place outside the virtual interaction spaces used as material for this research, HR volunteers do participate digitally in scheduled exercises which leave digital traces of their Process Improvement activities—HR took part in eleven exercises in 2010 and 2011, including multiple Camp Roberts Relief exercises, Exercise 24 and Exercise 24 Europe, and the Samoa Cyclone Simulation. During many of these exercises, HR volunteers helped to simulate social media activity during disaster on one hand, and worked to process the simulated social media information on the other. For HR volunteers, exercises are less about practice, as volunteers already get substantial "practice" during real events. Instead, they are seen as opportunities to connect with other response organizations and to demonstrate the capacities of their remote workforce to partners and potential partners in the emergency response domain.

[81] http://quick-nets.org/

6.4.2 How Humanity Road Organizes during Disaster Events

The preceding description of HR's mission offers insight into what the organization does, what their goals and work products are, and what they see their role to be within the domain of disaster response. This next section provides an in depth view of *how* they do this work, specifically during the *Disaster Response* phase of their efforts, the aspect of the organization's activity that aligns most closely with the larger theme of this research.

6.4.2.1 Responding to a Large Earthquake in Peru in October 2011

This section is focused around a single, extended excerpt of HR activities during a response to a large earthquake in Peru on October 28, 2011. The 6.9 magnitude earthquake struck Peru in the early afternoon, at 1:56pm[82], near the town of Ica. This region of Peru had suffered an 8.0 magnitude quake in 2007 that caused hundreds of fatalities and considerable damage to nearby cities (Puertas & Elsen, 2007). Though later assessment showed the impact of this event to be relatively low, as there were few casualties and little damage, the initial report of earthquake location and magnitude and the history of impact in the area led HR volunteers to deploy (virtually) for the event. The response included eight volunteers, including myself, and took place over approximately three hours, beginning informally just minutes after the earthquake, at 2:06pm, and ending officially when Chris Thompson announced that we were "standing down" at 5:11pm, though many volunteers deactivated some time earlier.

This extracted passage from the response activities, which relies heavily on the text-based Skype chats during this time, offers a snapshot of the organization at a single point in time. It also demonstrates the fluidity of its organizing—how change, sometimes significant change, occurs in the midst of responding to events. In some cases, that change later becomes a fixture of the ensuing organizational structures, embedded within their commonly understood work practices. Importantly, this structuring effect is often assisted by the digital record of the organizing-change-in-action.

---

[82] Timestamps for the digital communications presented here are adjusted to be consistent across all volunteer accounts and are normalized around PET, the time zone of the affected area in Peru.

6.4.2.2 Using Skype Chats as Virtual Workrooms for the Organization

A significant portion of organizing and response efforts at HR occurs on the Skype platform, through text-based, synchronous chats. Volunteers participate in several Skype chats at once, with each taking place in its own window. Different windows host different kinds of conversations: an *Urgent Events* window hosts real-time response activities; the *Useful Links* window is used both between and during events to post links to websites and tools that volunteers feel could help in current or future efforts; volunteers are encouraged to go to the *Work Diary* window to report their recent activities; and the *HR Café* is used as a backchannel for lighter conversation. Occasionally, groups leaders or veteran volunteers may create a new window for a specific event, or special windows that allow them to invite volunteers from outside HR to collaborate on a response effort. These windows serve to separate the organization's work areas. Skype chats can be viewed as virtual workrooms, digital places where volunteers go to do their work. Though they are sites for several communicative flows within the organization, at the beginning of this excerpt these chats serve as the primary location for *activity coordination*:

[Peru Excerpt, Part 1: Skype Textual Chat – *HR Urgent Events* window – 10/28/11] [83]

1 **Deepak** (12:08:25 pm PET): hello all

2 **HR Chris Thompson** (12:18:19 pm PET): (wave)

3 **sam2011** (1:47:20 pm PET): Good evening.

These three greetings appeared in the *HR Urgent Events* window on October 28, in the hours leading up to the earthquake. Here, Deepak enters the conversation, saying "hello all." HR Chris Thompson responds about ten minutes later with an animated emoticon that appears in the Skype window as a cartoon face waving its hand. About an hour and a half later another volunteer, who has several Skype handles including sam2011, enters "Good evening" into the shared window.

[83] This is the beginning of a long excerpt revealed in sections throughout this chapter/paper. Line numbers indicate progression of the conversation within the HR Urgent Events window during this extended excerpt.

Within the virtual work environment of these Skype chats, volunteers come and go throughout the day, occasionally dropping a few words into one or several windows. These greetings serve multiple purposes. Though Skype icons for "status" can be manipulated by users to describe their availability and are visible to other users, these are generally designed for someone's entire Skype audience. Volunteers who use the tool for other purposes in addition to their HR activity may have a broad range of connections and may at times want to be seen as available to some contacts (like other volunteers) while unavailable to others. Therefore, many volunteers do not use the visible Skype status to tell others that they are available to work that day. Instead, they leave greetings in the *HR Café* or *HR Urgent Events* windows to signal to other volunteers that they are following the conversations there and are available to help if an event occurs.

Volunteers also leave greetings in the window to initiate conversation with other volunteers, and not always on the subjects of disaster events, social media, or HR. The *HR Café* window especially is used for backchannel conversations, to exchange pleasantries, to talk about everyday occurrences, and to inquire into or share details about a volunteer's personal life. These interactions contribute to relationship building among the volunteers, and they are typically encouraged, except during active events when they can be distracting. The following excerpt from the *HR Café* window, which also occurred in the moments leading up to the Peru earthquake, shows volunteers essentially hanging out in the virtual chatroom, mixing disaster talk and availability updates with light conversation:

[Peru Excerpt Part 2: Skype Textual Chat – *HR Café* window – 10/28/11][84]

1 **sam2011** (1:47:46 pm PET): Good evening, Hope everyone is well

2 **HR Chris Thompson** (1:48:15 pm PET): doing well thanks Sam

3 **HR Chris Thompson** (1:48:21 pm PET): hope you are too

4 **marc123** (1:48:35 pm PET): evening?! where art thou?

5 **sam2011** (1:49:04 pm PET): Thou art on Hayling Island, UK

[84] Numbering begins at 0 for this conversation, which takes place in a different window from the other sections of this excerpt.

6 **HR Chris Thompson** (1:49:18 pm PET): the 5.4 kermadec didnt produce any effects, the Bangkok flooding is still bad - but with language barriers - we have not had much success in that area of the world

7 **marc123** (1:49:27 pm PET): aha. i skipped your kingdom this summer

8 **sam2011** (1:50:21 pm PET): Shame!! My kingdom of Hayling Island is wonderful!

9 **sam2011** (1:50:43 pm PET): Hi Chris, I am great thank you looking forward to Geneva!

10 **kaitlin_rice** (1:52:59 pm PET): Thanks for the update, Chris! Hi to Sam and Marc!

11 **marc123** (1:53:21 pm PET): hi!

12 **kaitlin_rice** (1:54:16 pm PET): (wave).

13 **sam2011** (1:54:31 pm PET): Hi Kaitlin

14 **kaitlin_rice** (1:57:17 pm PET): Hi Sam. Nice to C U virtually! Sorry I can't chat much now - am working - just made a quick stop at the Cafe - I hope you have a nice weekend!

15 **sam2011** (1:58:15 pm PET): I am working at laptop but have been watching window all day. On and off this evening, Yell if you need anything

16 **kaitlin_rice** (1:59:07 pm PET): TY! Good to know!

6.4.2.3 Activating the Disaster Desk

The next section of the extended excerpt contains the two subsequent messages posted in the HR Urgent Events window, sent immediately after, in the linear progress of the chat, the greetings by Deepak, HR Chris Thompson and sam2011 (Peru Excerpt Part 1), and just minutes after the most recent activity in the *HR Café Window* (Peru Excerpt Part 2). In these posts, HR Sally Bridges shares information about a recent earthquake with other volunteers tuned into the Skype windows, calling attention to the possibility for response.

[Peru Excerpt Part 3: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

4 **HR Sally Bridges** (2:06:30 pm PET):

5 http://earthquake.usgs.gov/earthquakes/recenteqsww/Quakes/at00ltshuy.html

6 **HR Sally Bridges** (2:06:51 pm PET):

7 2011 October 28 18:54:32 UTC

8 Details

Maps

Tsunami

Earthquake Details

This event has been reviewed by a seismologist.

Magnitude

6.9 (Preliminary magnitude — update expected within 15 minutes)

Date-Time

Friday, October 28, 2011 at 18:54:32 UTC

Friday, October 28, 2011 at 01:54:32 PM at epicenter

Location

14.500°S, 75.800°W

Depth

15 km (9.3 miles) set by location program

Region

NEAR THE COAST OF CENTRAL PERU

Distances

52 km (32 miles) S (180°) from Ica, Peru

122 km (76 miles) SSE (163°) from Chincha Alta, Peru

183 km (114 miles) W (278°) from Puquio, Peru

299 km (186 miles) SSE (153°) from LIMA, Peru

Location Uncertainty

Error estimate not available

Parameters

NST= 17, Nph= 17, Dmin=831.7 km, Rmss=0.87 sec, Gp=148°,

M-type="moment" magnitude from initial P wave (tsuboi method) (Mi/Mwp), Version=1

Source

West Coast and Alaska Tsunami Warning Center/NOAA/NWS

Event ID

at00ltshuy

Sally has been online, tweeting about the happy ending to a recent missing person's case, and monitoring her friends' Twitter feeds as she interacts with other others about that event. Perhaps within her casual Twitter activity or maybe by intentionally checking another resource (it is unclear from her Twitter timeline and Skype communications), at some time between 2:03pm and 2:06pm PET, Sally recognizes that there has been a recent earthquake. Her first post in the *HR Urgent Events* window (lines 4-5), which occurs about twelve minutes after the earthquake, is a link to the United States Geological Survey (USGS) website page created for this event, and the second post (lines 6-38), sent just seconds after the first, is an earthquake report copied and pasted from that USGS site, containing the earthquake magnitude, location, depth, nearest cities, and several other important pieces of information. This report, whose content and structure was determined by the USGS and has been digitally transferred into a HR workroom, becomes the starting point for HR's work around the quake.

The conversation continues:

[Peru Excerpt Part 4: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

39 **HR Catherine Graham** (2:07:28 pm PET): ooh

40 **HR Sally Bridges** (2:08:00 pm PET): http://wcatwc.arh.noaa.gov/

41 **HR Sally Bridges** (2:08:33 pm PET): No tsunami warning, watch or advisory so far

42 **HR Chris Thompson** (2:08:43 pm PET]: wow

43 **HR Sally Bridges** (2:09:03 pm PET): https://twitter.com/#!/search/peru

44 **HR Catherine Graham** (2:10:12 PM): no dart buoys near it - none triggered http://www.ndbc.noaa.gov/?lat=20.000000&lon=-120.000000&zoom=1&type=h&status=r&pgm=&op=&ls=false

Catherine Graham, a HR leader, is the first to respond to Sally's post. Her short comment (line 39), "ooh," demonstrates to Sally and others that Catherine has now seen this information and is taking it into account. Catherine's extension of the word to include two o's may also be intended to convey an early interpretation of the report—i.e. 6.9 is a large magnitude and this may be a big event. HR Chris

Thompson's "wow" remark (line 42), posted about a minute later, imparts the same effect, that this event may require humanitarian response. For HR, this means a possible activation of the *Disaster Desk*.

Though the term Disaster Desk takes on several slightly different meanings, the common understanding is that it is a set of tasks that volunteers complete during events, such as monitoring social media for information, verifying and analyzing information, and using social media and the group's website to communicate situational awareness and response information. The Disaster Desk incorporates many tools, but is operated from the Skype chatrooms by one or multiple members. It is essentially the active state of the volunteers during Disaster Response efforts, but has also come to imply the monitoring of possible events as well—there has been a recent shift from having the Desk be something that volunteers "activate" to something that a single volunteer can "have" or "monitor" as well. This transformation is part of an effort by HR leaders to empower volunteers to take over monitoring responsibilities for set periods of time, freeing up leaders and seasoned volunteers to work on other tasks or just go offline for awhile.

The four other posts in Part 4 (lines 40-41 and 43-44), which all occurred within four minutes of HR Sally Bridges' first post about the earthquake and only twenty minutes after the initial event, indicate that HR's Disaster Desk has been activated—i.e. volunteers are now performing tasks linked to the Disaster Response functionality of the organization. While Catherine and Chris are beginning to tune into the event and are perhaps absorbing the information in the USGS earthquake report, Sally moves quickly on to other tasks, first checking an NOAA website to assess the possibility of a tsunami, then posting a link to a report on that website along with a summary (lines 40-41)—"no tsunami warning so far." Half a minute later Sally posts a link to a Twitter search she has created for the event (line 43). And not long after that, HR Catherine Graham posts a link to another NOAA page (line 44) that signals if any dart buoys have been triggered, which would indicate a possible tsunami.

Once an event has been reported and the Disaster Desk has been activated, volunteers move quickly to gather information from multiple social media and Internet sources, reporting on what they find and posting links to the original source in the *Urgent Events* window. Following a routine for earthquake events that has developed over time, volunteers first search for information about the event's location and impact, then look for reports of possible tsunamis, and after that begin to look for on-the-ground and official sources of information about the event and the affected area. Initially, the information is shared with the intention of helping other volunteers understand what is going on, to contribute to the situational awareness of the group.

6.4.2.4 Dividing Labor and Tool Use

In the next section of the excerpt, volunteers begin to coordinate their activity, explicitly articulating their intentions to take on specific tasks using specific tools.

[Peru Excerpt Part 5: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

45 **HR Catherine Graham** (10/28/11 2:10:28 PM): I"m in Tweet tracker so i'll take a look there

46 **HR Sally Bridges** (10/28/11 2:10:28 PM): (thumbs up icon)

47 **HR Chris Thompson** (10/28/11 2:10:41 PM): I am @HR and will begin tweeting

48 **kaitlin_rice** (10/28/11 2:13:38 PM): I can RT from kaitstweets for a few

HR Catherine Graham tells the group that she is using the Tweet Tracker tool (line 45), a special tool that maps tweets containing geolocation metadata, allowing her to home in on tweets coming from the impacted area of events. HR Sally Bridges gives a thumb's up using a Skype icon here (line 46), but the timing of this post indicates that it is a response to an earlier part of the conversation. HR Chris Thompson follows by stating her intention to use the @HumanityRoad twitter account to begin tweeting outbound messages (line 47). Her tweets will loosely follow another well-rehearsed routine, tweeting out a "Did you feel it?" message linking to the USGS website for the project by that name, and then sending specific tweetables that inform affected people about how to best respond. She broadcasts the earthquake tweetables in English and Spanish (since the affected geographical area is a Spanish-language country) and with hashtags added specific to the current event.

@HumanityRoad (10/28/11 2:11 PM): Did you feel it? M6.9 earthquake; near the Coast of Central Peru http://t.co/ImaT7JM5 #terremoto #sismo #hmrd

@HumanityRoad (10/28/11 2:13 PM):In a disaster, voice lines are often congested, try sending a brief text message - just say RUOK or IMOK #peru #hmrd #terremoto #sismo

@HumanityRoad (10/28/11 2:19 PM): After a #quake – Expect aftershocks & stay away from damaged areas. #Peru #terremoto #sismo #tremblor #hmrd

@HumanityRoad (10/28/11 2:20 PM): Después de un #terremoto – esperar que las réplicas y se mantenga alejado de las áreas dañadas #tremblor #sismo #hmrd

Kaitlin_rice also explicitly states her intentions for how she will help (line 48), offering to use her own Twitter account to retweet Chris's tweets from the @HumanityRoad account, extending the reach of their messaging. Due to technological constraints of their virtual work environment (their Skype chats) and the ad hoc nature of their work practice, these volunteers use very explicit language to tell others what they are doing and/or what they plan to do. Because they do not share a physical workspace, volunteers cannot literally see what others are doing or even who is in the room working during a given event. Additionally, though roles have certainly developed, there is no set procedure dictating exactly who should be doing what, so activity must be coordinated in the moment. Thus, *activity coordination* during events is often very clear in the Skype chats, because it has to be.

Part 5 of the Peru excerpt also indicates a relationship between tools and division of labor, where tools structure the kinds of tasks that take place as well as who does what. HR volunteers use many different tools during their response efforts. The short passage demonstrates volunteers turning to the Twitter platform to both gather information and to disseminate it. For the former task, there are several different tools available that can be used to mine Twitter data in different ways—e.g. hashtag search tools, geographical search tools, and tools that translate incoming tweets automatically. However, there is no single tool that HR volunteers can use to do all of their Twitter monitoring—no tool that can filter by multiple terms and hashtags, filter by geographic range, translate, and geolocate tweets all in the same place. For this reason, volunteers select different tools to use, often tools with which they are uniquely proficient, as Catherine is with Tweet Tracker[85]. The division of labor is therefore done as much by tool as by task. Through the selection of these tools and tasks, human capacity, tool capacity, and the functions of the organization become mutually structuring. The affordances of available tools enable the organization to take on certain types of tasks. Over time, by selecting or being assigned to certain types of tasks that require certain types of tools, members develop expertise in different areas, and these collective expertise come to shape the overall work practices and work products of the organization.

6.4.2.5 Identifying On-the-Ground and Official Sources

Some of the first tasks that volunteers take on during disaster events involve identifying on-the-ground and official sources of information. Finding first-hand and trusted sources fits within the larger goal of increasing situational awareness—for themselves, affected people and responders. This material will later become the raw material for the resources that volunteers generate and present to the public. The following passage from the extended excerpt shows volunteers moving to identify these sources, verify them, and pull information from them into their shared workspace.

[Peru Excerpt Part 6: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

HR Catherine Graham (2:14:19 pm PET): This Author appears to be in the area of the quake
Author: danielmunoz
Location:Lima
Date-Time:2011-10-28 12:04:35
Tweet:Joder, magnitud 7 en Ica #Terremoto #Temblor
Latitude:-12.0459739
Longitude:-77.0306154
HR Sally Bridges (2:14:43 pm PET): http://www.citypopulation.de/Peru-Agglo.html
HR Catherine Graham (2:14:54 pm PET): Author: alfredolm6

[85] Tweet Tracker is a beta version tool developed by researchers at the Data Mining and Machine Learning Lab at Arizona State University. http://tweettracker.fulton.asu.edu/

58 Location:Peru

59 Date-Time:2011-10-28 12:03:52

60 Tweet:ULTIMA NOTICIA: TERREMOTO EN ICA. #terremoto #temblor

61 Latitude:-6.8699697

62 Longitude:-75.0458515

63 **HR Sally Bridges** (2:16:13 pm PET): Ica (32 mi), Chincha Alta (76 mi) , Puquio (114 mi) and Lima *186 mi) are nearest towns/cities--seeing tweets that buildings in Lima shook.

64 **HR Catherine Graham** (2:17:42 pm PET): Sorry Danielmunoz is in LIma - the tweet was located in Peru my mistake

65 **HR Chris Thompson** (2:17:45 pm PET): anybody got an official source?

66 **HR Chris Thompson** (2:17:59 pm PET): s'ok - I am seeing reports that the quake was felt in Lima

67 **HR Catherine Graham** (2:18:43 pm PET): Author: vagm1992

68 Location:Ica - Perú

69 Date-Time:2011-10-28 12:07:06

70 Tweet:Temblor en Ica.. Tirando para #terremoto .. No es broma.. ESTOY TEMBLANDO!

71 Translation: Earthquake in Ica .. Pulling for # earthquake .. No joke .. I'm shaking!

72 Latitude:-14.0690032

73 Longitude:-75.7370612

HR Catherine Graham's first (lines 49-55), second (lines 57-62) and final posts (lines 67-73) in this section contain information in the form of tweet records that she found using the Tweet Tracker tool in her quest to locate on-the-ground sources. The format of the data posted here, which includes the author name, the location, the timestamp, the tweet text, and the latitude and longitude, indicates that Catherine has copied and pasted data directly from Tweet Tracker into the HR Urgent Events window. Finding on-the-ground sources is an important task for volunteers, a way for them to directly contribute to response by bringing first hand information out into the collective knowledge or situational awareness of the event.

On line 63 of this section, HR Sally Bridges indicates that she is also searching for information coming from the ground, using the Twitter search she had initiated previously. Though her first search, shared on line 43, was merely a search on the term "Peru," Sally often uses an Advanced Twitter Search that enables her to set geographical boundaries for the search and therefore pick up tweets likely to have originated in the affected area. Both Tweet Tracker and the Advanced Twitter search have a few drawbacks, giving incomplete and sometimes inaccurate results. Tweet Tracker is limited to tweets that contain GPS metadata, typically only around 1% of tweets for disaster events.[86] Though Twitter's Advanced Search does also assess the self-reported location in the Twitterer's profile, this is also neither an exhaustive search nor a totally accurate filter, since only 66% of accounts have valid geographic location information (Hecht et al., 2011) and this information can be incorrect.

These tools have limitations, but volunteers do not simply accept the results. Instead, they appropriate these tools not to constitute, but to assist them in their work, using them in conjunction with their own fact-finding and verifying skills to make determinations about on-the-ground sources. Sometimes this work is individual, but at other times it becomes a collective effort to determine the proximity of a source to the event. Returning to the Skype conversation (Peru Excerpt Part 6), after posting a link to an official website with information about the population of the affected area (line 56), HR Sally Bridges works to try to confirm HR Catherine Graham's assertion (line 49) that the author of the first tweet she found was in the affected area. At this point, the three active volunteers hold a short discussion about the geographical boundaries of the impacted area (lines 63-66), attempting to determine if Catherine has accurately identified an on-the-ground Twitterer. On line 63, HR Sally Bridges reposts the names of the possibly affected cities and their distances to the epicenter of the quake, and then summarizes reports that she has seen through her Twitter search, indicating that the earthquake was felt in Lima. HR Catherine Graham attempts to self-correct (line 64), thinking that Lima is outside the affected area, but HR Chris Thompson and HR Sally Bridges offer evidence that her initial claim was indeed accurate (lines 65-66).

---

[86] Analysis of several disaster events in 2011 and 2012 indicates that the percentage of tweets with GPS metadata is typically near or below 1%, and that this number varies according to event type, location, and the amount of global participation. For example: Egypt Protests = 0.56%; Hurricane Irene = 1.02%; Texas Wildfires in Sept 2011 = 1.32%; and SE Tornados on March 2, 2012 = 1.02%.

6.4.2.6 Verifying: A Foundational Task for Humanity Road

This work to find and confirm on-the-ground and official sources falls into the umbrella task of verifying, perhaps the foundational task of the HR organization. Verifying information and teaching others how to verify information are intimately connected to who HR is as an organization. Verifying appears in their Mission Statement and it is the subject of their most tweeted tweetable:

@HumanityRoad: Verify twice - tweet once - rumors put lives at risk

This mantra is repeated over and over again in volunteer trainings and reiterated periodically within their active chatrooms. Instructions for verifying information also appear on their checklist, which I explain in detail later in Section 6.4.2.8. Verifying was also the main focus of the Twitter Commandments (Figure 14), which suggests its roots lie in the digital volunteer efforts that its members took part in before HR became a formal organization, i.e. the Iran Election protests and the early weeks of the Haiti Earthquake response.

Thompson has described how the Twitter Commandments helped early HR volunteers recruit new members to their group (personal communication). The work of verifying was an original organizing point for the group, a center of gravity, something that members organized around. In the initial days of HR, the Twitter Commandments codified a work practice that the early voluntweeters had been developing, and over time a considerable about of HR's organizational effort towards Process Improvement has focused on encouraging others in the social media space to verify the information they find before passing it on.

6.4.2.7 Exercising Transactive Memory

As the volunteers continue to work to gather information, HR Sally Bridges decides to generate a Twitter search that automatically translates from one language to another (line 75).

[Peru Excerpt Part 7: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

74	**HR Chris Thompson** (2:20:29 pm PET): Posted the checklist in the useful links window

75 **HR Sally Bridges** (2:22:25 pm PET): https://twitter.com/#!/Peru21pe Trying to translate. Need Kate's link that we used for Turkey and can't find it. :(

76 **HR Chris Thompson** (2:25:15 pm PET): Bettie - here - http://t.co/RcMpOliy

77 **HR Sally Bridges** (2:25:47 pm PET): Thanks, I bookmarked it (I thought)

78 **HR Sally Bridges** (2:26:35 pm PET): Okay, updated for Peru: http://www.xlhit.com/s?sv=tw&q=peru&tl=es&ml=en

Sally has used this tool before in our response to another event, and though she has forgotten how to find the tool that offers this functionality, she remembers that it was I who first shared the tool with the group. Instead of taking the time to track back through all of the Skype conversations to find where I had posted the link during a previous event, she instead attempts to contact me to ask if I remember where to find it.

> **@SallyB** (10/28/11 2:25 PM PET) @katestarbird What is the link for the tweet translation site that you used for Turkey? Need it for Peru. In Skype Urg window now.

Sally is attempting to use a cognitive strategy of *transactive memory*, a model proposed by Wegner (1986) that defines group memory as a combination of the cognitive capacities of individual members with individual knowledge about the other members' areas of expertise. Sarcevic et al. (2008) have examined transactive memory in a physically co-present team of trauma resuscitators, demonstrating how group members relied on collective memory to store and recall information. This excerpt from HR's response efforts shows transactive memory in action within a virtual team.

Using this at its basest level as an information retrieval strategy, Sally does not have to remember the information as long as she remembers where she can find it, and in this case which other group member can find it for her. Some volunteers also keep track of what group members are skilled at certain kinds of activities, and will sometimes request another volunteer take over some part of the task they are working on. For example, much later in this event response, HR Samantha Walker will post a request for assistance with translating in the Skype window (line 79) and two other volunteers will quickly reply, one with the translation (line 81) and another with a link to a tool where the volunteer could do her own translating (line 86):

[Peru Excerpt – Transactive Memory: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

…

79 **HR Samantha Walker** (2:56:00 pm PET): Can anyone translate this please

80 **HR Samantha Walker** (2:56:06 pm PET): Peru Alerta naranja de tsunami en costas de Ica información VHF desde radios aficionados...gente autoevacua por precaución

81 **HR Alexander Breuer** (2:56:32 pm PET): Peru tsunami alert orange coast of Ica information from VHF amateurs ... people as a precaution autoevacua

82 **HR Sam Walker** (2:56:46 pm PET): I will get an online translator to get a rough idea of info I am finding

83 **HR Alexander Breuer** (2:56:46 pm PET): = google translate

84 **HR Samantha Walker** (2:57:01 pm PET): thank you (Hug)!

85 **HR Alexander Breuer** (2:57:07 pm PET): :$

86 **HR Sally Bridges** (2:57:12 pm PET): http://translate.google.com

87 **HR Sally Bridges** (2:57:23 pm PET): (wave)

Transactive memory can be examined through the lens of distributed cognition (Theiner, 2010), which implies that remembered and rehearsed connections between each group member and the resources that other group members provide are all part of a larger cognitive system. This perspective of viewing the work of digital volunteers through the framework of distributed cognition will be explored at greater length in Chapter 8.

6.4.2.8 The Recursive Relationship between Action and Routines

Returning to our temporal position within the extended excerpt, at this stage of the event response, about thirty minutes after the earthquake, there are three volunteers working at the Disaster Desk, collecting information and sharing it with each other in the *Urgent Events* window. At 2:30pm[87] HR Chris

[87] Up until this point, the excerpt included every post in the Urgent Events window, but it now skips ahead to salient moments.

Thompson suggests to Catherine, her co-leader of HR, that they initiate an Event Diary (a blog post entry about the event, see Figure 15) for the earthquake:

88 **HR Chris Thompson** (2:30:27 pm PET): Catherine - event diary page?

A few minutes later, HR Catherine Graham posts to say that she agrees with that determination (line 89). HR Chris Thompson then suggests that Catherine start the diary and that the other participating volunteers help feed her information for what will be a public resource about the event (line 90).

[Peru Excerpt Part 8: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

89 **HR Catherine Graham** (2:32:45 pm PET): Yes Chris, I think this should be an Event Diary

90 **HR Chris Thompson** (2:34:09 pm PET): if you can start it we can feed you

91 **HR Catherine Graham** (2:34:36 pm PET): k

92 **HR Chris Thompson** (2:34:54 pm PET): Good map to put on Event Page if possible http://www.iris.edu/seismon/zoom/?view=eveday&lon=-77&lat=-13

HR Chris Thompson quickly begins this "feeding" process, posting a link to a map that would be good to put in the Event Diary blog (line 92). At about this time, several volunteers check in to the Skype windows and offer help.

[Peru Excerpt Part 9: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

93 **HR Samantha Walker** (2:35:04 pm PET): what would u like me to do?

94 **HR Samantha Walker** (2:35:21 pm PET): monitor?

95 **Deepak** (2:36:22 pm PET): Can I help in any way

96 **Deepak** (2:36:23 pm PET): ?

97 **Kate Starbird - student/researcher** (2:36:56 pm PET): Let me know if I can help as well.

In the initial part of the response, through Part 7 of the extended excerpt, the effort was largely self-organized, coordinated laterally by volunteers who self-selected their current or intended tasks and then articulated them to others. These volunteers followed what seemed to be established patterns for the group in terms of the types of tasks and the order of tasks that were undertaken. However, as more volunteers, some of them inexperienced, join the chat and offer to help in the response, a need arises for a more explicit and directed division of labor. Seasoned volunteers and leaders begin proposing tasks to the group at large (at first) and then assigning them to specific volunteers to facilitate creation of the Event Diary page.

[Peru Excerpt Part 10: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

98 **HR Chris Thompson** (2:37:30 pm PET): we need the official weblinks for national and regional - the FEMA like links for Peru

99 **HR Chris Thompson** (2:38:13 pm PET): we are creating an event page

100 **HR Chris Thompson** (2:39:17 pm PET): Use this page as a template of the types of links to collect - http://www.humanityroad.org/EarthquakeTurkey

101 **HR Chris Thompson** (2:39:37 pm PET): Also the useful links window contains our official checklist of what we do on an emerging event

102 **HR Chris Thompson** (2:40:31 pm PET): We also need someone to monitor for urgent needs

103 **HR Chris Thompson** (2:40:40 pm PET): so just tell us which piece you are working on

…

104 **HR Chris Thompson** (2:42:43 pm PET): Deepak - can you monitor Tweet tracker for urgent needs? Kate-can you Google search for information on impacted hospitals, buildings, people

105 **Kate Starbird - student/researcher** (2:43:19 pm PET): yes - hospitals, buildings, people using Google... that's what I'll do now

106 **HR Kaitlin** (2:43:21 pm PET): Please repeat nearest cities to epicenter

107 **HR Chris Thompson** (2:43:21 pm PET): Sam - can you collect the best twitter ids to monitor - some are already in this window

108 **HR Samantha Walker** (2:43:32 pm PET): yes NP

At first, HR Chris Thompson offers suggestions about the kinds of things that need to be done (lines 98-102). She lists a few potential tasks and asks volunteers to choose what they want to work on (line 103). This strategy aligns with the organization's goal of working in self-directed teams. But soon, perhaps unwilling to wait for volunteers to figure it out themselves or unsure that they will select the most appropriate tasks for the moment, HR Chris Thompson begins to assign specific tasks to volunteers (lines 104-108). This passage demonstrates a significant shift in how activity coordination is happening for this event. In the early stages of the group's response, a few experienced volunteers appeared to be self-organizing in a lateral way, self-selecting tasks and articulating their own actions and intentions to let others know. Though not explicitly communicated during this time, that small group's efforts were largely shaped by an underlying structure of pre-established routines. These seasoned volunteers already knew that after an earthquake, they first check for tsunamis, and then they start to monitor media and look for on-the-ground and official sources, etc. Some of the volunteers that are joining the efforts now have less experience and less knowledge of the protocol that Sally, Catherine and Chris have been following. This requires a shift from a leaderless, self-organizing group to somewhat of a top-down, hierarchical one. Chris takes control and starts to lead the less experienced volunteers.

Attempting to both instruct and educate, HR Chris Thompson points us to two digital documents that contain information about how the organization "works" during disasters. The first is an Event Diary page (like the one in Figure 15) that had been created for a previous event. HR Chris Thompson suggests that volunteers use this as a template to give us an idea of the kinds of information that need to be collected. In this instance, Chris is using that previously created resource, captured and preserved digitally, to intentionally structure current/future action.

HR Chris Thompson also directs volunteers to refer to the "official checklist" which she has posted in the *HR Useful Links* window:

[Peru Excerpt – Checklist – *HR Useful Links* window – 10/28/11]

HR Chris Thompson (2:16:40 pm PET): Checklist

HR Chris Thompson (2:16:48 pm PET): What happened, did it really happen, where, when, details

If yes – share information on texting

If no – share information on verifying before tweeting

Who is the Event Official Source

Examples: Hurricane, Tornado, Flood, Earthquake, Wildfire, Health

What is the potential impact to the population

Examples: collapsed buildings, approaching fire, tornado, storm

Examples: search & rescue, evacuation, shelter in place, preparedness

How big is the potentially impacted population

Emergency, Disaster, Catastrophic Disaster

What do we know about the population – language, geography, vulnerability

Who is the governmental official source

Examples: Rescue, Evacuation, Shelter, Find Hospitals, Quarantine

Is the official source providing guidance to the public?

If yes – share

if no – share standardized guidance

Identify first responders (vs) official responders

Website, blog, social media user, email

Identify Sources: Official, Unofficial, Trusted, Unknown, Untrusted

Use 2 separate sources to verify if no official source has reported event

Identify collaborative partners that can assist

Summarize the Event (see sample below)

Post event summary in Urgent Window and Work Diary Window in Skype

The checklist, an organizational document used to structure volunteer activities, serves multiple organizing purposes, including activity coordination and organizational self-structuring. The routine prescribed here encapsulates a significant cluster of organizational know-how, including pre-existing domain knowledge and lessons learned from experience by the organization's members. Looking back to the early part of this Peru Earthquake excerpt, though not explicitly referenced until later, the routine delineated in this checklist can be seen as shaping the activity already occurring in the Skype windows, structuring the volunteer's initial collation efforts.

Leaders of HR initially created the checklist as part of new volunteer training, and they continue to use it educationally as a formal part of what is now called "Disaster Desk training." In October 2011, just before the event covered here, Chris began to use the "scrape and paste" method of distributing the checklist during events, invoking it both as a reference and as an instruction set for current activities (personal communication, Chris Thompson)

The checklist is a dynamic document, built through our media monitoring experiences as digital volunteers, and continuously evolving as volunteers incorporate lessons learned and leaders seek to clarify and streamline the inscribed work process. Existing as a digital document that can be copied and pasted into current activity windows or electronically distributed as a PDF file, the checklist is a lightweight resource that leaders and experienced volunteers use both internally as a reference for their own work practice and as an intentional structuring mechanism that allows them to both pass on their previous knowledge and to help coordinate current action. The checklist encodes a selected collection of previous, ad hoc activity coordination as established routine for future work. Digital documentation of action facilitates the transfer—*what we just did* becomes *what we do*. In this way, action and routine have a recursive relationship, with action shaping routine and routine shaping action.

6.4.2.9 Moving to a Living Event Diary

This next section of the extended excerpt further explores the relationship between action and structure, and highlights a third component of structuration within the virtual organization, demonstrating how technological tools shape the virtual organization's work practices. In this part of the example, a problem with the current work practice becomes clear to several volunteers: the multiple, simultaneous use of the *HR Urgent Events* window as a place for activity coordination, information sharing and resource building, as well as the staggered arrival of new volunteers onto the virtual scene of the response efforts, results in a complex and confusing work environment. While some tasks are being repeated, others are forgotten, and even though volunteers make an effort to articulate their intentions and their actions, it is hard for us to keep track of what everyone is doing. Addressing this problem, one of HR's leaders introduces a new tool to the group, and this tool quickly and permanently changes how the organization responds to disaster when multiple volunteers are active.

At the end of Part 10 of the extended excerpt above, HR Kaitlin Rice returns to the *HR Urgent Events* window (line 106) after a long period of absence[88], since the initial moments of the group's reaction to the earthquake. Though it is unclear what task she is taking on, she posts an undirected request, not specifically to any one volunteer, for information about what cities were nearest to the earthquake. Following that request, Sally drops her current task and goes to retrieve that information for Kaitlin. As she is posting the distances to the window for the second time (lines 110-112; 114-120), another volunteer joins the conversation:

[Peru Excerpt Part 11: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

Jiro (2:43:34 pm PET): whats up guys!

HR Sally Bridges (1:43:54 pm PET): [3:16:09 PM] HR Sally Bridges: Ica (32 mi), Chincha Alta (76 mi) , Puquio (114 mi) and Lima *186 mi) are nearest towns/cities--seeing tweets that buildings in Lima shook.

<<<

HR Kaitlin Rice (2:44:02 pm PET): TY

HR Sally Bridges (2:44:30 pm PET): [3:06:42 PM] HR Sally Bridges: Distances

52 km (32 miles) S (180°) from Ica, Peru

122 km (76 miles) SSE (163°) from Chincha Alta, Peru

183 km (114 miles) W (278°) from Puquio, Peru

299 km (186 miles) SSE (153°) from LIMA, Peru

<<<

[88] Kaitlin uses a different Skype account when she returns. Earlier in the excerpt, her Skype handle was kaitlin_rice. For the remainder of the excerpt she is logged in as HR Kaitlin Rice.

121 **HR Chris Thompson** (2:44:38 pm PET): hello Jiro

122 **Kate Starbird - student/researcher** (2:45:07 pm PET): no damage immediately reported ... http://www.bloomberg.com/news/2011-10-28/peru-hit-by-magnitude-7-quake-tsunami-center.html

123 **Jiro** (2:45:13 pm PET): hello Chris

124 **Jiro** (2:45:31 pm PET): i have been offline for a while because of low current here in my area

125 **HR Chris Thompson** (2:46:10 pm PET): (thumbs up)

In answering Kaitlin's request, Sally tracks back to her earlier posts in the Urgent Events window, then copies and pastes them into the window at the current time. Copied-and-pasted elements from earlier Skype conversations begin with the original timestamp and author and end with a line break followed by the "<<<" characters. When copied and pasted, the timestamp appears in the time zone of the person pasting the clip and does not update according to the time zone of the other participants. In this case, those copied stamps say 3:16 PM and 3:06 PM, which means the timestamps are now EDT, but Skype does not specify the time zone.[89] Because volunteers are spread across time zones, the pasted time may not align with the original post time in others' windows. Though this is not an issue here, copying and pasting previous Skype communications into the window without editing the timestamp could cause problems when the original time of the information is important.

The interaction grows more confusing. In this next section, HR Sally Bridges and HR Kaitlin Rice continue to identify and share information related to their respective collection tasks (lines 126; 128-140); Deepak responds (line 127) to HR Chris Thompson's earlier directive asking him to monitor tweet tracker (Part 10, line 104, above); another volunteer, HR Alexander Breuer, arrives in the window and states his availability to help (line 141); and I post information about nearest cities to the earthquake (line 142).

[Peru Excerpt Part 12: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

126 **HR Sally Bridges** (2:46:11 pm PET): URGENT: Witnesses will point to local press that the sea in Pisco has withdrawn about 10 meters. posted by @@info_emergencia

[89] The PET is added here to make these times clear to the reader of this document. The Skype window does not show the time zone.

Deepak (2:46:32 pm PET): will do

HR Kaitlin Rice (2:47:05 pm PET): For AiD Veterinarians in Lima >> Lima:

Bertchi Veterinary Clinic
Jorge Dintilhac 675
San Miguel L-32, Lima, PERU

Tel: +51 (1) 464-3342
Fax: +51 (1) 451-8077
EMail Bertchi Veterinary Clinic

Dr. Javier Chávez
Dr. Fernando Chávez
Dr. Ricardo Matínez

HR Alexander Breuer (2:47:07 pm PET): Hello all - I am also available

Kate Starbird - student/researcher (2:47:43 pm PET): Nearest city seems to be Ica - EQ was 32 miles S of Ica

Coordination and communication problems begin to take a toll on the group's efficiency. Though these activity coordination efforts already reflect some adaptation to the affordances of Skype—for example, how volunteers explicitly articulate availability and current tasks—this response effort is bringing to light some constraints of the ICT that we have not yet managed to overcome. My post (line 142) is a response to HR Kaitlin Rice's request for nearest city information (line 106), suggesting that Sally and I were working on the same task at the same time, and that I kept working even after she had completed it because I had not noticed her posts. Though Chris's request to him went out only four minutes earlier, Deepak's response (line 127) is out of order in the continuing conversation, and it may not be clear to all, especially the newly arrived, what he is agreeing to do. HR Kaitlin Rice's post about veterinary clinics (lines 128-140) addresses a significant area of our data collection efforts (pet and animal care), but its length forces the window to scroll down, leaving earlier activity coordination efforts out of view.

This coordination complexity continues:

[Peru Excerpt Part 13: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

143 **HR Sally Bridges** (2:47:51 pm PET): Alexander, I have to step away for about an hour to show a couple of homes

144 **HR Sally Bridges** (2:48:08 pm PET): Will you keep an eye on this? http://www.xlhit.com/s?sv=tw&q=peru&tl=es&ml=en

145 **HR Kaitlin Rice** (2:48:19 pm PET): for AiD list >   Veterinaria Animaniac

146 Jr. casma 295 Urb. Los Pinares - Los Olivos

147 Lima39, Peru

148

149 Tel: 521-4752

150 EMail Veterinaria Animaniac

151

152 Dr. Luis Morales Montejo

153 Dr. Josue Paz Palacios

154

155 Information: Animales menores

156 **HR Alexander Breuer** (2:49:11 pm PET): (thumbs up)

157 **HR Samantha Doyle** (2:49:22 pm PET): Chancellor of Peru Twitter account - @CANCILLERIAPERU

158 **HR Alexander Breuer** (2:50:01 pm PET): there are many RTs about the water going 10 m into the sea

159 **HR Sally Bridges** (2:50:16 pm PET): Watching @AlertaNews24 and @info_emergencia Haven't found Red Cross yet.

160 **HR Alexander Breuer** (2:50:37 pm PET): Chris, did you already send "after quake expect tsunmai" tweets?

161 **HR Chris Thompson** (2:51:13 pm PET): < insert process improvement on data collection here >

HR Chris Thompson's comment about process improvement at the end of this section of the excerpt (line 161) explicitly acknowledges the coordination difficulties we are encountering. We have a problem keeping up with everything that has already been done, who is available to work, who is doing what, and what remains to be done. Though most volunteers are making an effort to articulate their availability, intentions, and current tasks, and our leaders are trying to direct inexperienced volunteers to needed tasks, it is still hard to keep up with everything. Volunteers are duplicating work in places, and its unclear what tasks remain.

Under the stress of the situation, with too many volunteers doing too many things all at once and in the same place, the socio-technical infrastructure that underlies our work practice is breaking down. Star and Ruhleder claim that infrastructure becomes visible only at these points of breakdown (1996). But for this virtual organization, even at breakdown, it is not the technical infrastructure that is first questioned, but the organization's work practices (line 161). The technical infrastructure remains a given, taken for granted by HR leaders and volunteers who have built their organization atop robust, commercial platforms developed for mass adoption like Skype, Twitter and Google. In this case, the volunteers do not question the technical platforms upon which their organization runs, as they are powerless to change those systems. Instead, they initially direct their attention to the social and organizational configurations of the group, the "process" of HR Chris Thompson's comment, which is inextricably blended already into the technical infrastructure. Chris is suggesting that they find a way to better use the functionality of the tools that they have. Later, Catherine will address this issue of *process improvement* at the level of the socio-technical infrastructure, by appropriating another of these widely used tools—Google Documents, which they have used before for other kinds of work—to help coordinate their activity. Though they cannot change the functionality of the tools that are available to them, they can change how they incorporate these tools into their activities.

At 2:57pm PET, not quite an hour after HR began its response to the Peru Earthquake and about six minutes after HR Chris Thompson remarked about the need for process improvement, HR Catherine Graham posts a link in the HR Urgent Events window introducing this new tool.

162 **HR Catherine Graham** (2:57:56 pm PET): https://docs.google.com/document/d/1WG8HOC-ryO0pxwOeeBtIB-8ptuWQLfwsqNVA8V1dE1I/edit?authkey=CL3swe0L#

163 **HR Catherine Graham** (2:58:54 pm PET): I have created a document to collect the items we need for publishing i.e. put hospitals links in the links column. I'll start collecting whats alreadyhere in the urgent window for it

As she explains, Catherine has just created a shared document for the volunteers to use to help coordinate their information collation activities (Figure 16). The document is a shared spreadsheet hosted by the Google Document platform, which allows multiple users to edit digital documents simultaneously. Initially, Catherine sets the privileges on the document to public so volunteers (and anyone else with the link) can all access it and contribute to it.

Humanity Road Event Monitoring
6.9 Chile Earthquake
Volunteers reporting in:

National and Regional Links	
Emergency Numbers	
Nearest Airports	
Other Airports	
News and Social Media Twitter links	
Hashtags	
Facebook	
Maps and Situational awareness reports	
Video link (Embed code please)	
What happened?	
Survive	

Hospitals	
Sustain	
Shelters	
Reunite	
Red Cross	
Animals in Disaster links and information	
Urgent Needs Information	
General Information	
Transportation, Infrastructure and Road Conditions:	
Picture Links	
Video Link	

Figure 16. Google Document for Event Monitoring – Initial State, 2:56 pm PET

Catherine's Google Doc is about to permanently change the way HR volunteers work during multi-volunteer response efforts. The document itself is much more than an artifact; it is a tool, serving both as a platform for interaction and a record of that interaction. Though it will soon become a structuring mechanism, in its initial instantiation the Event Monitoring Google Doc is a manifestation of existing structure, dually shaped in content and functionality by the prior organizing efforts of HR volunteers and the technological affordances of the platform in which it was created. Catherine pulls the document's initial content (the information categories) directly from the current format of the Event Diary (Figure 15). This content is also closely related to the Disaster Desk Checklist (shown above in the Peru Excerpt – Checklist). By turning to Google Docs and configuring the document in this way, Catherine attempts to recreate an early work practice of collaboratively creating information resources—this practice was previously abandoned when volunteers moved from a website tool that enabled multiple editors to one that did not (personal communication, Catherine Graham).

The creation of the Event Monitoring Google Doc demonstrates again how structure emerges and moves through an organization. This introduction of a new tool also provides an opportunity to examine how ICT influences organizing. The format and function of the new Event Monitoring tool are greatly influenced by the affordances of the Google Doc spreadsheet tool, which provides the table format and enables the multi-user interaction in the workspace. These relationships between the characteristics of this new tool, previous action and underlying technological affordances soon become connected to future action and resulting structure in the organization.

Within minutes of Catherine's post sharing the link to the document, several volunteers pull up the document and begin to work there, using the categories provided in the first column to guide us in our collection efforts. As she promised, Catherine begins by tracking back through the Skype chats and copy-pasting previously posted information from the chats to the Google Doc. She first drops a section of Sally's initial earthquake report (itself copied and pasted from the USGS site) into the field next to "What Happened?"

By 3pm PET, two other volunteers have moved information into the fields corresponding to "News and Social Media Twitter Links" and "Hashtags." One of those volunteers edits "6.9 Chile Earthquake," a mistake in the header of the initial document, to "6.9 Peru Earthquake." More hashtags are added, along with more Twitter handles. I add information I collected earlier into the "Nearest Airports" row. Someone drops in the information about nearby cities, and someone else—probably Kaitlin who is the leader of the Animals in Disaster cluster—adds the information she has been posting about veterinary clinics. At 3:06 pm PET the document appears as it is in Figure 17.

Although we can see when volunteers add new content, the Google Document platform does not allow us to see who is working in what area. We all appear as "anonymous" to each other. To enable us to know which other volunteers are responding, Catherine's original document contained an area in the header where volunteers could assert that they were "reporting in." By this time, ten minutes after Catherine urges us to turn to this new document, seven of us have reported in for work there.

Humanity Road Event Monitoring
6.9 Peru Earthquake
Volunteers reporting in:
Alexander, Kate, Sally, Chris, Cat, Kaitlin, Sam

National and Regional Links	@AlertaNews24 @info_emergencia
Emergency Numbers	
Nearest Airports	Pisco Airport (PIO) GPS: 13° 44' 41" S, 76° 13' 13" W
Other Airports	
News and Social Media Twitter links	@Info_emergencia @SubTVChile @terratvperu @news_in_peru @AlertaNews24 @perutweet
Hashtags	#pisco #peru #sismo
Facebook	
Maps and Situational awareness reports	
	Ica (32 mi), Chincha Alta (76 mi) , Puquio (114 mi) and Lima *186 mi) are nearest towns/cities--seeing tweets that buildings in Lima shook
Video link (Embed code please)	
What happened?	6.9 (Preliminary magnitude — update expected within 15 minutes) http://earthquake.usgs.gov/earthquakes/recenteqsww/Quakes/at00ltshuy.html Dat-Time Friday, October 28, 2011 at 18:54:32 UTC Friday, October 28, 2011 at 01:54:32 PM at epicenter Location 14.500°S, 75.800°W Depth 15 km (9.3 miles) set by location program Region NEAR THE COAST OF CENTRAL PERU Distances 52 km (32 miles) S (180°) from Ica, Peru 122 km (76 miles) SSE (163°) from Chincha Alta, Peru 183 km (114 miles) W (278°) from Puquio, Peru 299 km (186 miles) SSE (153°) from LIMA, Peru Location Uncertainty
Survive	
Hospitals	
Sustain	
Shelters	
Reunite	
Red Cross	
Animals in Disaster links and information	Veterinaria Animaniac Jr. casma 295 Urb. Los Pinares - Los Olivos Lima39, Peru Tel: 521-4752 EMail Veterinaria Animaniac Dr. Luis Morales Montejo

	Dr. Luis Morales Montejo Dr. Josue Paz Palacios Information: Animales menores Bertchi Veterinary Clinic Jorge Dintilhac 675 San Miguel L-32, Lima, PERU Tel: +51 (1) 464-3342 Fax: +51 (1) 451-8077 EMail Bertchi Veterinary Clinic Dr. Javier Chávez Dr. Fernando Chávez Dr. Ricardo Matínez
Urgent Needs Information	
General Information	
Transportation, Infrastructure and Road Conditions:	
Picture Links	
Video Link	

Figure 17. Google Document for Event Monitoring – Intermediate State, 3:06 pm PET

Around that same time, the Skype chats go almost completely silent, as volunteers have virtually exited one workroom and moved to another. Though 30 messages were posted in the *HR Urgent Events* window in the ten minutes before Catherine linked to the Google Doc, only six messages are posted there in the ten minutes after the new tool is introduced, and a majority of those are positive comments about how well the new tool is working. Although Alexander continues to post to the window to ask clarification questions and share situational awareness information with the others, the bulk of the information organizing activity and the resource creation work has shifted from the Skype chat environment to the Google Document platform.

[Peru Excerpt Part 14: Skype Textual Chat – *HR Urgent Events* window – 10/28/11]

164 **HR Alexander Breuer** (3:01:32 pm PET): does anyone already know aboout SHOA? must be an agency. Tweeters report that SHOA dismisses the danger of a tsunami

165 **HR Alexander Breuer** (3:01:51 pm PET): Catherine: (thumbs up)(thumbs up)(thumbs up)

166 **HR Alexander Breuer** (3:03:57 pm PET): @ALertaNews24 URGENT: RPP reports some new houses in Pisco have cracked product of 6.9-magnitude earthquake in Peru today.

167	**HR Kaitlin Rice** (3:04:36 pm PET): TY, Cat! Wow!
168	**HR Chris Thompson** (3:05:22 pm PET): Now you can open the google doc and see what we have and have not collected and even put your name in the doc to say "Chris - getting xxx"
169	**HR Kaitlin Rice** (3:05:45 pm PET): this is SOOO cool, IMHO!

Working together, volunteers quickly move the information we previously collected into the new work environment, and then continue to collect information and fill it into other areas of the document. Though we can see the information that others have added to the document, we cannot yet see where other volunteers are working. Adapting our coordination needs to this constraint, HR Chris Thompson proposes a technique (line 168) of writing our names in the left column of the row we are working on, underneath the category title, which allows us to communicate to other volunteers that we have that information category covered in a way that remains visible as activity continues (Figure 18).

Humanity Road Event Monitoring
6.9 Peru Earthquake
Volunteers reporting in:
Alexander, Kate, Sally, Chris, Cat, Kaitlin, Sam

National and Regional Links	@AlertaNews24 @info_emergencia http://www.peru.gob.pe/
Emergency Numbers	Ambulance: 131 Fire: 132 Police: 133. 137 maritime emergencies, Conaf 130, PDI 134.
Nearest Airports (Kate looking)	Pisco Airport (PIO) GPS: 13° 44' 41" S, 76° 13' 13" W
Other Airports	
News and Social Media Twitter links (Sam Getting)	@Info_emergencia @SubTVChile @terratvperu @news_in_peru @AlertaNews24 @perutweet
Hashtags (Sam Getting)	#pisco #peru #sismo #Earthquake #Temblor #Lima

Figure 18. Google Document for Event Monitoring – Intermediate State, 3:08 pm PET

Because we can all see now which information collection tasks are already complete, which are in progress, and which remain, newly arrived volunteers no longer need to be told what to do. Instead, we all can easily identify where the holes are and can self-select our tasks. Additionally, active volunteers do not need to remember what everyone is doing or continuously track back through Skype conversations to figure this out, and so when we finish one task, we can simply select another without asking a series of questions. Along those same lines, leaders do not need to give as many directives. The Skype chat room remains a place for collaborative work and for sharing situational awareness information, but it not longer functions as the sole place for information collection and activity coordination as well.

The Google Document has quickly become the center of gravity for our activity coordination efforts and is quite effectively structuring our current work practice. Directives given from one volunteer to another are no longer as necessary, shifting our organizing from a more hierarchical model to a more lateral one. Volunteers are self-selecting tasks instead of waiting to be told what to do. This new tool is shaping how we communicate, how we divide our labor, and even what our work products *look like*. Towards this latter point, a few minutes later, HR Catherine Graham comments on the success of the document, and reports that she is going to start a “living” Event Diary:

170 **HR Catherine Graham** (3:11:12 pm PET): :) ok.. i see its filling up nicely.. i'm going to start a "living" Event Diary entry

What Catherine is planning, and soon executes, is a copy-and-paste of the Google Document into the Event Diary blog entry on the HR website. From this point forward in this event, and many subsequent events, Catherine will periodically capture the content and structure of the volunteers’ resource creation work exactly as it is produced within the Event Monitoring Google Document, and paste that into the content of our website (Figure 19).

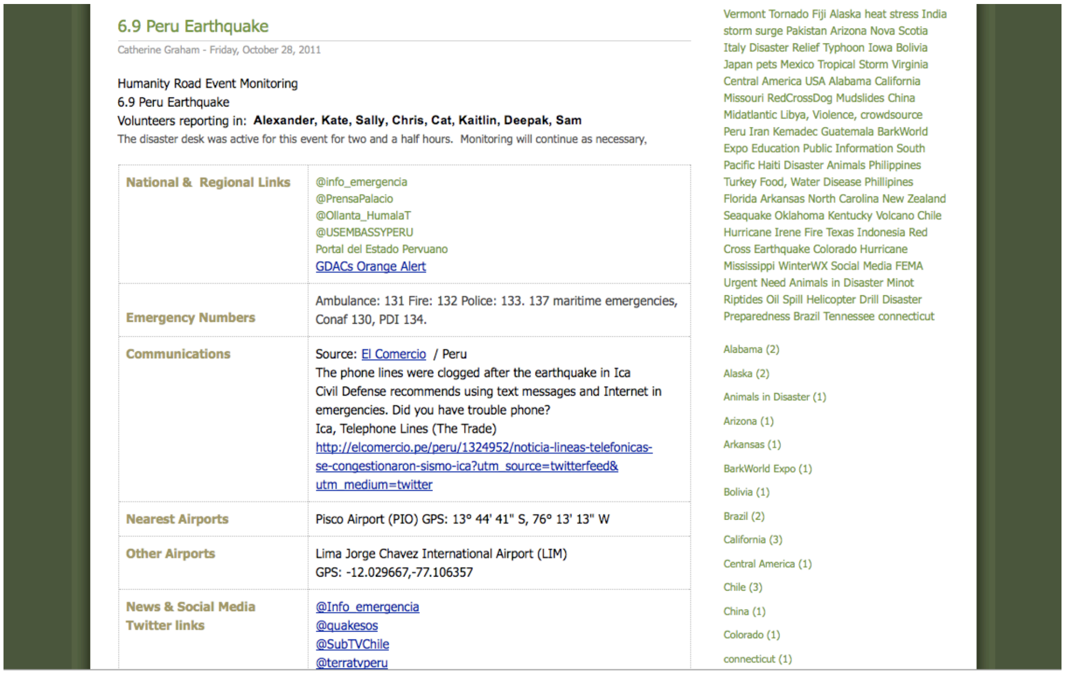


Figure 19. Event Diary for Peru Earthquake Reflecting Google Document Structure

The introduction of the new tool was deemed a "smashing success" by HR co-leader Chris Thompson, who later summed up that day's response efforts in the HR Work Diary window:

[Excerpt: Skype Textual Chat – *HR Work Diary* window – 10/28/11]

> **HR Chris Thompson** (5:18:36 pm PET)
>
> Today the team monitor a 6.9 earthquake just offshore the central coast of Peru. The disaster desk was active for 2.5 hours. Catherine created a new google doc that replicates the grid form we used in the old wordpress design that allowed multiple volunteers to collect information simultaneously. It was a smashing success - allowing volunteers to work side by side and independently while not overlapping on data mining and collection of critical links. The table was used to populate the event diary - with only minor retyping required. The event diary entry was immediately posted after the disaster desk stood down and is here http://ow.ly/7cw60

In the five months following the document's first use, HR has continued to use the Event Monitoring Google Doc, relying on it for activity coordination and resource creation during several event responses, typically responses to large events when multiple volunteers are active. The document plays an ongoing role in organizing their disaster response activities. Like other documents that both reflect and enforce the existing structure of the organization, the Google Doc has evolved—the current digital version shows slight changes to the categories in column one, and contains in the space below the active grid both a list of "tips" for how to use the document and a copy of the Disaster Checklist.

Volunteers now use this document as a "springboard" for launching a response. They have set up a technique where, when they decide to do a full response to an unfolding event, one volunteer creates a copy of the document, carrying over a list of "collaborators"—current HR volunteers—who are automatically contacted through the Google Doc platform when the document becomes live. So, the creation of the Event Monitoring document now literally activates the HR Disaster Desk.

6.5 Discussion: Establishing and Sustaining a Virtual Response Organization

This section shifts to a broader view of the organization, examining how HR continually works to define and redefine its role across time and events and within the larger ecosystem of disaster and humanitarian response.

6.5.1 Sustaining the Virtual Volunteer Organization: Ethics of Responsibility

The above description of HR refers, in several places, to the organization making decisions about whether or not to respond to specific events. This section examines how this issue of when and how to respond is an ongoing sense-making activity within the organization, where the outcomes of previous decisions lay down rationale for future ones. These decisions take into account several issues. A first order concern is whether or not there is a need for the organization's services. As described in the extended excerpt regarding the organization's response to the Peru earthquake, in the early stages of an event HR volunteers assess the impact of the situation to determine if there is a need for them to activate, and at what scale.

The next issue is resources—HR's goal of responding, "24 hours, always on," to events around the world brings up questions of capacity. For each event, they must ask, does the organization have enough volunteers available to respond at the level necessary for this event? Additionally, do volunteers have the right skill sets? Are there volunteers available who are fluent enough in the language of the affected area to translate incoming and outgoing communications? During fatal flooding and mudslides near Rio de Janeiro in Brazil in January 2011, HR recruited three volunteers from Brazil who helped the group understand the geographical and cultural aspects of the disaster and translate incoming and outgoing communications. With those volunteers eventually taking the lead, HR supported that event for several weeks. A comment by HR Chris Thompson, revisited from the transcript of the group's Skype conversation in the moments leading up to the Peru earthquake, indicates that HR has had less success supporting similar flooding events in other parts of the world:

> **HR Chris Thompson** (1:49:18 pm PET): the 5.4 kermadec didnt produce any effects, the Bangkok flooding is still bad - but with language barriers - we have not had much success in that area of the world

Because their work is digital and potentially global, HR must constantly make choices about which events to respond to, and this example shows how human resources guide the small organization in these decisions.

The role of the organization in relation to other groups is another piece of the equation. Are there other organizations responding, physical or digital? For instance, are there organizations responding from the ground of the event, and if so, should HR work with them directly to fulfill a recognized need? In the digital volunteer space, is another group, like the Standby Task Force, setting up a crisis mapping effort? If so, will HR volunteers work within that effort or will they work separately? HR is just one entry in an emerging ecosystem of digital volunteer organizations who simultaneously collaborate and carve out roles for themselves in the digital response space, just as traditional NGOs do in the physical response arena.

Throughout their first two years of existence, the group has continuously negotiated, both with its own members and with members of other organizations, what their role in response efforts is and should be.

A question of both capacity and role surfaced during the group's response to Hurricane Irene in August 2011, a collaborative effort of digital volunteers across organizations during which HR took a lead role. On August 26, the day before projected impact, as Irene approached the U.S. East Coast, threatening a large swath of the coast from North Carolina to Maine, HR and collaborating volunteers from the Ushahidi community and the Standby Task Force formed a media monitoring team for crisis mapping the hurricane. Two major factors shaped the group's early strategy for responding to that event. First, many of HR's veteran volunteers, including leader Chris Thompson, live along the coast or are personally connected to areas predicted to be impacted by the storm. Additionally, before the storm came ashore, HR connected to a media outlet, WRAL in Raleigh, NC, who requesting mapping assistance for an Ushahidi instance they intended to carry on their website. Due in part to these connections, HR leaders initially urged their volunteers and others invited into their collaborative window to focus their efforts on the impact potential in North Carolina:

[Irene Excerpt, Part 1: Skype Textual Chat – *Irene Map: Media Monitoring* window – 08/26/11]

> **HR Chris Thompson** (3:54:23 PM): Focus on North Carolina
>
> **HR Chris Thompson (**3:54:45 PM): Hashtags #ncwx #irene #obx

On the morning of August 27th, as hurricane impact was immanent, HR Catherine Graham summarized the strategy of the collaborative group as negotiated on an AM conference call:

[Irene Excerpt, Part 2: Skype Textual Chat – *Irene Map: Media Monitoring* window – 08/27/11]

> **HR Catherine Graham (**10:47:24 AM): Call Notes: (*) Humanity Road responds on the internet to provide information on how to survive, sustain and reunite. Yesterday we stood up the Skype window, prior to the Ushahidi map (platform) being launched. As we were being engaged we were tapped to assist with a map being launched in NC for WRAL in Raleigh, NC. 1.5 million viewing the Ushahidi map and will be embedding it on their website. We are focused on emergency information,

but the NC map is currently focused for situational awareness only. There is another map for NYC that may be launched. As this event unfolds, please pace yourself, please make a decision where you want to focus and stay with that, it will help you be more focused. Right now our focus is NC, as the day progresses our focus may slide up further North but for today we're going to focus on NC in getting info into the map. Some of you are Standby Task Force members, some are Humanity Road members and some of you are new.

Indeed, North Carolina was the site of initial impact of the storm on August 27, but as that day progressed and the storm swept north up the coast, several volunteers in the collaborative chat room began to question directives to concentrate on North Carolina.

[Irene Excerpt, Part 3: Skype Textual Chat – *Irene Map: Media Monitoring* window – 08/27/11]

Valoria (14:19:43 PM): Are we covering NY already or keep focusing on NC?

HR Chris Thompson (14:21:15 PM): Meg - we are just focusing on NC right now not NYC right?

Melissa Elliot (14:21:41 PM): errrrr....really??

HR Chris Thompson: we are not staged at this point for this map for NY

HR Chris Thompson (14:22:16 PM): we can be - but our sponsor is NC local

Kirk (14:22:51 PM): i think we are looking at the coast as [it moves] NE?

HR Chris Thompson (14:23:39 PM): we can - and probably should - but before we get to NY - we have VA and DC/MD and NJ

Eventually, this conversation moved from the Skype window designated for several media monitoring volunteers to the *Irene Map HQ* window for the Irene map, a Skype chat consisting of veteran volunteers from Ushahidi and Standby Task Force mapping efforts, where strategic and technical issues of the mapping effort were discussed and coordinated. Here, HR Chris Thompson presented her rationale for focusing first on North Carolina and then moving up the coast with the storm.

[Irene Excerpt, Part 4: Skype Textual Chat – *Irene Map HQ* – 08/27/11]

Melissa Elliot (14:45:11 PM): given that the MM team is being told to only submit reports for NC, should the name not reflect that?

Aaron Huslage: that's not what they should have been told

HR Chris Thompson (14:45:34 PM): @Meg - for now - we are focused on NC

HR Chris Thompson (14:45:51 PM): as we move northward - with the storm - we can move northward with the reports

HR Chris Thompson (14:46:55 PM): I don't think we have a large enough team of mm's to handle the entire east coast at this time - for manual entry of shelters as well as after effects of urgent needs for that many states

Aaron Huslage (14:48:08 PM): i think we need to do a reset. can we all gather here at 3pm?

Patrick Meier (14:48:22 PM): Thanks for the explanation, Chris.

…

HR Chris Thompson (14:51:29 PM): sure - we can throw every state in the window and we will lose effectiveness overall - I think if we use an approach of joint planning to all move in unison - targeting highest areas of impact - we may be able to manage the volume. We can use geography to focus - move to VA next? Focus on VA Beach, Norfolk, coastal communities - hardest areas of impact - I am 150 miles inland of VA Beach - I have just rain and some wind

HR Chris Thompson (14:52:02 PM): Open to input and suggestions

Patrick Meier (14:53:39 PM): Makes sense to me given the resource constraints, @Chris

Aaron Huslage (14:53:47 PM): agreed chris

Aaron Huslage (14:54:09 PM): if we get more volunteers who want to target a geography, then we can change strategy

This extended excerpt demonstrates how HR and other digital organizations work out their policies about which events or which parts of events to respond to, often as the events are in motion. In the final section of the chat excerpt, Chris argued that with limited resources (people) and a massive area that could potentially be affected, the group should concentrate on areas of highest impact using a focused approach that moved north as the storm moved north. Eventually, the team of digital volunteers, collaborating across organizations, agreed to go along with this strategy. Though early resources concentrated on North Carolina, the mapping efforts did shift as the event progressed and went on to include event reports from all of the impacted states along the coast, from North Carolina to Maine.

Ethical concerns about the impact of the volunteers' work also factor in to discussions on whether and how to respond, especially during events like political protests, where it may be hard to determine who might be helped by the real-time information-processing efforts of the organization. For HR, a partner in the Standby Task Force's project to map violence in Libya during the political unrest there in the Winter and Spring of 2011, the decision to not take part in similar efforts to assist the reporting of the political violence in Syria, which began not long after and was ongoing in the Spring of 2012, is significant. Interestingly, a "decision" to not respond was never formally made. Instead, HR leaders and veteran volunteers simply chose not to initiate discussion about the event, and did not engage in conversations about Syria begun by others in the chat rooms. The follow excerpt is one of three times that the topic Syria is mentioned in HR chatrooms:

[HR Conference Call Excerpt: Skype Textual Chat – *HR Work Diary* – 04/30/11]

> **HR Catherine Graham** (10:59:12 AM): For Libya, Humanity Road volunteers provided a total 188 man hours of coverage which equated to 40% of the total volunteer hours scheduled for the initial 14 day rotation
>
> **Jimmy** (10:59:27 AM): Is there a Syria map in play yet?
>
> **HR Catherine Graham** (10:59:32 AM): not sure
>
> **Jimmy** (10:59:33 AM): WOW on Libya
>
> **Shiban** (10:59:47 AM): I don't think so. no syria map i guess.

At the time of Jimmy's comment, nine volunteers, including two of HR's leaders and several veteran volunteers, were present and tuned in to the *HR Work Diary* window as the group discussed its recent accomplishments during a conference call. Catherine's reply here to Jimmy's question about digital volunteer efforts to support Syria, "not sure," serves to answer his question, but does so in an interesting way. This interaction provides a contrast to her reaction, during another event, to another volunteer calling the group's attention to an emerging event. In the early moments of HR's response to the Peru earthquake in October 2011, recounted in Part 4 of the Peru Excerpt earlier in this chapter, HR veteran volunteer Sally posted information about the recent earthquake and Catherine reacted by posting "ohh" in the window, a comment that communicated a very different orientation towards a possible response to an emerging event. Both comments are short responses, but while the latter is an exclamation that opens up the floor for more discussion, the post in the Syria-related excerpt confers little interest and does not provoke discussion. A bit later, another conversation participant, but one who was not a veteran HR volunteer, also responded to the Syria question, in a slightly more engaged tone. However, the conversation ended there as no member with the authority to do so initiated conversation on whether the group should do something for Syria.

There is strikingly little conversation in the HR chat rooms about Syria in the months that follow, and it is unclear if this decision not to entertain the possibility of initiating media monitoring or mapping for Syria was related to need, resources, role, or the potential that their work could have negative impact on those opposing the Syrian government—the protestors and other government opposition groups in that event. Discussions about Syria in other crisis mapping communities suggest that ethical concerns probably played a part in HR's choice to avoid engaging in Syria-related activities, but their digital traces record this non-decision as one of omission.

6.5.2 Sustaining the Virtual Volunteer Organization: Motivations for Participation

Related to the capacity issue above is a larger issue of how the virtual volunteer organization can sustain itself over time, especially in a domain with so much (potential) work to do. Responding 24-7 to disaster events all around the world requires a consistently available workforce with global reach and the ability to expand rapidly. Humanity Road meets this capacity challenge in a unique way—for both the domain of disaster response and formal volunteer organizations—by maintaining a strong central core of dedicated volunteers and being flexible in the ways that it incorporates volunteers at its boundaries.

HR emerged from digital, spontaneous volunteerism, and unlike other non-profits in the disaster domain and beyond, it continues to accept spontaneous volunteers into its activities during events, relying on experienced volunteers to quickly train new recruits, giving them clear tasks at first and later, when response efforts calm down, attempting to recruit them into official membership. HR also allows for episodic volunteers that self-activate only occasionally, usually when they recognize a large event has occurred. Part 9 of the Peru Excerpt marks the arrival of a group of episodic volunteers into HR's response efforts during that event. Voida et al. (2012) notes that social computing technology encourages episodic volunteerism, and cites this as a negative phenomenon for volunteer coordinators. But the disaster response domain may be positioned well to accommodate this style, because during event responses there is a need for more work at the same time that these volunteers are most likely to engage.

Also contributing to the stability and the viability of the organization, HR maintains a strong internal core that includes its two leaders and a small group of veteran volunteers who contribute several hundred hours (each) to the organization annually. HR relies on these experienced volunteers to recruit, train and mentor newer members, to work in-between events on other organizational needs, and to monitor for emerging events. This style of volunteerism, also critical for HR, requires the sustained participation over time by increasingly skilled members.

There are different motivations for this kind of sustained participation versus the spontaneous or episodic volunteerism described above. Discussing motivations for participating in collectively intelligent crowd work, Malone et al. (2009) offer love in the form of *contributing to a cause* as one of several possible motivations for participation. This benevolent incentive is perhaps the primary motivation for all volunteers, especially at first, but over time other factors come into play. Ye and Kishida (2003) claim that the learning that occurs through LPP can serve as motivation for participating in open source communities. Similarly, learning new skills may be an incentive for ongoing participation in HR. Malone et al. (2009) also suggest that love in the form of *socializing with others* can motivate virtual crowd workers, a class of motvaitions that can be unpacked and expanded using ideas of capital (Bourdieu, 1986; 1998; Putnam, 2000).

Bourdieu (1998) offered the concept of *symbolic capital* to describe how reputation, honor and status confer benefits to those who possess them. The development of symbolic capital can be an incentive for certain kinds of volunteer work, but only where the activity or recognition of the acitivty is visible to others. Since Twitter activity is essentially public, messages sent by HR members from personal accounts, while serving the primary function of educating the public about how to respond to current or future events, also act as visible markers of their volunteer work within their Twitter friend-follower networks, presenting the opportunity for volunteers to receive public recognition for their work. Developing and maintaining a public identity as a crisis tweeter and thereby receiving recognition for that activity may therefore be a secondary incentive for some HR volunteers.

The benefits of socializing with others can be mapped to ideas of social capital (Bourdieu, 1986; Putnam, 2000). Expanding on earlier conceptions of social capital, Putnam (2000) explained a difference between bridging capital that connects heterogeneous groups, and bonding capital that reinforces connections within tight groups. While digital volunteerism can increase bridging capital, as volunteers connect to new people, for long-term HR volunteers, the development of bonding capital in the form of strengthened relationships with other volunteers, supported by their interactions within the chat rooms, may be a significant motivation for sustained participation in the community. Voida et al. (2012) assert that fostering community and supporting deeper engagement are aims of many volunteer organizations. Explaining the importance of supporting these kinds of interactions in mediated group work, Schmidt and Bannon write:

> "The 'informal' interactions that take place in the office thus not only serve important psychological functions in terms of acting as a human support network for people, for example, providing companionship and emotional support, but are crucial to the actual conduct of the work process itself." (1992, p 23)

This research shows HR using the Skype chat environment as an ongoing, interactive space where volunteers can develop and strengthen ties with each other and with the organization.

Like open source communities (Ye & Kishida, 2003), HR can be seen as a community of practice (Lave & Wenger, 1991), a group of people who come together within the shared domain of digital volunteerism, form a community, and work together towards a common goal of, in this case, assisting in disaster response efforts. Both within HR and outside that group in the larger space of digital volunteerism, there is evidence of legitimate peripheral participation (LPP) whereby new members slowly increase their engagement and their centrality in the community, progressing from more simple tasks to more complex and more risky ones (Lave & Wegner, 1991). During their event response to the Peru Earthquake, while only team leaders were permitted to publish resources to the outside world through the organization's website and Twitter account, inexperienced volunteers were assigned information-gathering tasks that were simple and finite, allowing them to engage in the larger work of the group. Over time, as they learn more about how the organization works, HR volunteers are expected to take on more complex tasks, gradually drifting from peripheral participation to leading the Disaster Desk, tweeting from the @HumanityRoad account, managing Event Diary entries, representing the organization to outside entities, etc. Though some volunteers remain at the outside rungs of the group, dropping in from time to time to take on basic information gathering tasks, others (Sally, Alexander, and Kaitlin in this example) have moved to positions of increased responsibility.

6.5.3 Sustaining the Virtual Volunteer Organization: ICT, Structure and Action

In all of their activities, ICT acts as a structuring force for the organizing. The ICT they appropriate—often widely adopted platforms like Twitter, Skype, and Google Docs—simultaneously host the material, the site, and the products of their work. These tools shape not only what they can do, but also *how* they do it, and their work to mold these tools to the goals and values of their organization is a core element of their constant effort to sustain.

6.5.3.1 The Recursive Relationship between Action and Structure

This description of HR's history and mission, as well as the analysis of an extended excerpt of their activity during a large response effort, support the supposition of a recursive relationship between action and structure within the virtual volunteer organization. Many of HR's current documents reflect the structuring influence of earlier volunteer action, in some cases going back to activity that took place prior to the formalization of the organization itself. For instance, their Mission Statement carries a faint echo of early work by founders Graham and Thompson to use "Internet and mobile communications" tools—Internet Cafés—to help affected people "reunite with loved ones" after Hurricane Katrina. Their core mission also incorporates lessons learned from early "crisis tweeting" efforts by volunteers during the Iran Election protests and other events leading up to the Haiti Earthquake, an event that acted as a catalyst for HR's incorporation.

In the first weeks of the Haiti response, before HR formalized as a 501(c)(3), the small amount of structure they had previously developed acted as an organizing point for the group, a center of gravity that volunteers could organize around, and something that pulled other volunteers in. Describing HR's efforts to support the Haiti earthquake response, Thompson claimed that they "gained additional volunteers because of [their] structured approach," which included the Twitter Commandments published on their website and their "founding principles of locating facts from rumors and tweeting responsibly during disaster" (personal communication). These comments suggest that by offering guidance to others for how to do crisis tweeting, HR's founders were able to recruit others to their cause. This is supported by findings in Study 2 (Chapter 5) that show inexperienced digital volunteers who were newly arrived on the virtual scene of disaster response relished this structure—they wanted to know that they were doing things correctly. This initial structure, a formalization of the right way to do crisis tweeting, became something that early volunteers recruited with and organized around. This structure was both *shaped by* previous action and acted *to shape* future action.

6.5.3.2 ICT as a Structuring Force

This examination of virtual organizing highlights the influence of a third shaping element: ICT. Describing the start-from-scratch conditions of virtual organizing, Finholt et al. write that the "absence of prior structure means group members must develop new structures for sharing information, for example, norms or rules for reporting progress and division of labor" (1990, p. 292). This ethnographic examination of HR reveals how some of these structures come into being under conditions where ICT plays an early and foundational role in shaping the organization.

The leaders of HR, when they incorporated, had some idea of how they wanted to do things. Graham and Thompson both had previously volunteered with the American Red Cross and brought that experience into their early conceptions about what their new organization would do and how it would be structured. However, the ICT that the group used to mediate almost all of their interaction, the tools that they incorporated into their work, and the very domain in which they worked—digital volunteerism—introduced new constraints and new opportunities and those, in turn, begot new structures to deal with the possibilities and limitations of virtual organizing.

In this research, the role of ICT in shaping HR's work practices can be seen in their ICT use and their articulation work. Part 5 of the extended excerpt above shows how the availability of certain tools shapes the kinds of tasks that HR volunteers take on, and how these tools then act as a mechanism for the division of labor. ICT, along with the expertise that certain volunteers have with particular tools, shapes what work can be done as well as how this work is done and by whom. In this virtual organization working in a digital domain, ICT is a central organizing force that leaves its mark on everything from the overriding mission of the organization—helping people during disasters using "Internet and mobile communications technology"—to the minutiae of how they make this happen.

One place where it is easy to witness the shaping force of the ICT is within the articulation work that volunteers do to coordinate their activity. The Skype textual chat platform is an integral component of HR's organizing efforts, as it is for several other virtual volunteer groups. However, the platform has constraints, and in evolving to work within those constraints volunteers have developed a method of communicating that involves explicit articulation of current work, future intentions, and directives to others. Examples of this articulation work can be see throughout the extended excerpt presented in this chapter: from the conversations before the earthquake struck that demonstrate how volunteers signal availability to others; to the early coordination work by seasoned volunteers who clearly announce exactly what they were doing or were going to do; and into the sections where leaders begin to give very specific orders to less experienced volunteers.

Articulation work is a feature of all social work environments, not just virtual ones. In arguing for the development of systems to support this work, Schmidt and Bannon asserted that articulation work is integral to organizing and that it is the work that keeps information flowing (1992). Though Skype is clearly the platform of choice for current virtual volunteer groups, activity coordination within that platform requires considerable articulation work. By directing the group to her new Event Monitoring Google Document, Catherine Graham introduced an innovation to support both resource creation and articulation work. Within the new interaction space of the Google Doc, the status of what needed to be done became consistently visible—unlike in the Skype chats where that information would scroll up and out of sight. Volunteers quickly adapted their use of the new tool to articulate who was currently working on which task, also in a way that was continuously visible to other volunteers. The amount of articulation work necessary to efficiently coordinate their activities was significantly reduced by the adoption of a new tool and a few small adaptations of their work practice to this new environment.

The ICT HR uses also leaves an imprint on the organization's work products—the resources they provide to the outside world. This effect can be seen throughout the long excerpt in this chapter: in the way that the format of earthquake report from the USGS website carries over first to their Skype chats, then to their Google Doc and then into their Event Diary entry on our website; in the configuration of certain types of information as they are transferred from the tools in which volunteers do their monitoring to the resources they create; and now in the structure of Event Diary entries which pull over the spreadsheet format of our Event Monitoring Google Document.

6.5.3.3 The Role of ICT in Passing On Organizational Know-How

Another intersection of ICT, action and structure within the virtual organization occurs as the development of routines. One aspect of this is the intentional use of digital traces for shaping work practices. DeSanctis and Monge suggest the possibility of this device as a stabilizing force within virtual organizations:

> "It may be that electronic communication products, such as conversations and documents stored in knowledge repositories, can provide stability to otherwise tenuous relationships. Perhaps communication histories from one setting can be carried into the communication future of other settings via evolving databases" (1999)

Using communication history to structure future action is an ongoing practice at HR, and may indeed be an adaptation to constraints of the virtual organizing environment that enables them to successfully and efficiently establish procedures and transfer domain knowledge.

In describing the dynamics of ad hoc organizing, Weick (1995) refers to a popular quote attributed to E.M. Forster, “How do I know what I think until I see what I say?” People working together to solve a problem, especially a new problem or an old problem under new conditions, are constantly improvising solutions. For HR, an emerging organization in an emerging domain, much of their work involves improvisation. However, as they enact solutions to a particular problem, they discover how they do something and in many cases they try to draw on this experience to guide them in future action, and in this way standardize some of their procedures. During the Peru Earthquake, HR deployed a new tool and developed new procedures for doing work using this tool. Afterwards, group leaders tried to pass on these discoveries of “this is how we do it,” embedding directions for using the document in the future, derived from improvisation in the moment, into the Google Doc’s content.

Traditionally, storytelling has been a significant conduit for organizational culture and know-how (Kelly, 1985; Orr, 1986; Boje, 1991). Orr (1986) described how a group of Xerox technicians used war stories and other narratives as a method for incorporating experience into the shared expertise of the community, and wondered how a computer system might be used to facilitate this practice of passing on experience through storytelling. The virtual interaction space of HR, which includes Skype chats, Twitter messages and now the Google Document, is not a place where group members often tell war stories or relate long anecdotes of previous work. Occasionally longer stories are told during monthly “All-Hands” conference calls, but most communication in the day-to-day interaction consists of short, rapid messages exchanged via text by people who are often multi-tasking. Instead of storytelling, HR relies heavily on the digital traces of previous activity to transfer what they have learned how to do to other and future members of the organization. During the Peru earthquake response, when less experienced volunteers checked into the active window and asked how they could help, HR Chris Thompson directed them to a resource that volunteers had created in a past event as an example for what they should do during this event (Part 10 of the extended excerpt). This shows how virtual volunteers use the affordances of the digital work environment to efficiently pass on know-how to other members, who, due to dynamics of virtual work, may have differing understandings of what the organization is and does at any given time.

Digital traces can be seen as an external counterpart to Giddens' *memory traces*, which he describes as internal structures that orient the conduct of human agents Along these lines, digital traces in the virtual environment can be intentionally accessed and used as cognitive aids to shape action. The use of external resources in a system where knowledge and action are distributed across people and ICT fits within a distributed cognition perspective, explored in Chapter 8.

There may be positive and negative impacts to shifting from storytelling as a method for transferring organizational knowledge to using transcripts of communication history to teach. In a storytelling model, details are embellished, interpretations are made, and previous sense-making efforts are embedded into future tellings of the shared stories. Stories can be shifted in progress to highlight different aspects and to focus on specific details relevant to the problem at hand (Orr, 1986). Orr's narratives were a "verbal process," and much of this process can be lost in communicating this kind of information by referring back to previously created documents or other digital traces of earlier activity. Within virtual organizing, pointing another group member to a specific piece of the digital record can highlight certain types of organization knowledge, but the process is not as dynamic. Sense-making and interpretation efforts that volunteers took part in after the fact may not be captured in the digital history, though other digital documents, like the Checklist, can be designed to include content from post hoc analysis.

One way in which HR members do capture lessons learned and the interpretation layered on top of that is within the *Work Diary* window in their Skype environment. Volunteers are encouraged to go to the *Work Diary* to report what they have been working on at the end of the day, after participating in a response effort, or after executing another HR-related task. The *Work Diary* keeps this self-reported history of the volunteers' efforts, which contains some of the interpretation and sense-making work missing from the action transcripts in other windows. This window also acts as a reference for HR leaders, helping them maintain a record of who is doing what within the organization.

6.5.3.4 Infrastructure and Change: Assembling and Re-Assembling

In some ways, the HR organization emerged from infrastructure. In Study 2 (Chapter 5), we describe how during the early aftermath of the Haiti Earthquake, technical resources such as the Internet, Twitter, Tweak the Tweet, and other tools, created the initial opportunities for action that voluntweeteers took—action that eventually led to the defining of tasks, a division of labor, and the emergence of their networked organization. HR grew out of that network and still relies on many of these same tools in their work. As they continue to develop, the group assembles an infrastructure from available tools, relying in most cases upon robust, commercial platforms designed for mass adoption—often ones that volunteers are already familiar with before joining the organization (e.g. Twitter, Skype, Google Documents). These are powerful platforms that experience few breakdowns, but their users are powerless to change them. For HR, breakdowns in this infrastructure are opportunities for reassembling and reconfiguring the infrastructure, but within the constraints of volunteers' expertise.

An example of this reconfiguration occurred during the organization's response to the earthquake in Peru, when HR leader Catherine Graham introduced a new tool (a Google Document) for coordinating the real-time information-processing activity of the organization. The introduction of this innovation marked a shift away from the top-down organizing necessitated by the linear nature of the Skype chat environment back to a "self-directed work team approach," the intended operational strategy of the organization (Graham, personal communication).

This example surfaces a relationship between the tools appropriated and the work practices and values of the organization, demonstrating value-sensitive appropriation and reconfiguration of ICT—a correlate perhaps of value sensitive design (Friedman, 1996; Friedman et al., 2006)—by users not empowered to change the design or implementation of those tools. Catherine's effort to re-assemble the organizational infrastructure was work of sustaining the organization through re-aligning its work practices with its efforts to facilitate incorporation of new volunteers and allow existing volunteers to participate through self-directed work.

6.6 Discussion: The Communicative Flow of Information Processing Work

McPhee and Zaug (2009) propose that the communicative processes that constitute an organization can be divided into four flows, and much, though importantly not all, of the communication of HR can be classified as one or more of these types. As this account reveals, *activity coordination* is a huge component of the work that volunteers do, and organizational members have developed methods of clearly articulating their coordination efforts within the constraints of the ICT they use. The Google Doc introduced during the Peru Earthquake response became a catalyst for a drastic re-shaping of how HR coordinates during response efforts when multiple volunteers are contributing. *Membership negotiation*, discussed within the account of my joining the organization (Section 6.3.2), occurs within the group's Skype chats as spontaneous volunteers are activated and then moved towards official membership. HR's website maintains specific information pages and sign-up services designed to recruit and transition members into the organization. Membership negotiation also takes place within HR's monthly, multi-party voice calls where volunteers discuss their current HR activities and during periodic events like a recent volunteer appreciation banquet (which was also mediated by voice chat) where HR leaders discussed the achievements of all the volunteers. The communicative flow of *organizational self-structuring* can be observed in the group's mission statement, in the tweetables they design and distribute to the public, and in their articulated routines, like the Disaster Checklist. Additionally, much of their public messaging, including their website and their outgoing tweets, especially those sent in-between events, functions as *institutional positioning*.

Though all these flows are present in the digital traces observed here, there are other communicative processes that do no fit any of these flows. This research provides evidence of a fifth communicative flow for the virtual organization whose work is embedded in ICT. This fifth flow is literally the work itself—

i.e. the communicative acts of information manipulation that function to do the work of HR. It occurs through the Twitter platform, within their Skype chatrooms, and now through their Event Monitoring Google Doc, and it includes the volunteers' communicative acts of sharing information with each other, the tweets they amplify, route and add hashtags to, and the resources they create and publish. This fifth flow is the part of HR's work that is most closely aligned with the larger theme of this research, the information-processing activity of the organization.

6.7 Conclusion

This study examines the case of a purely virtual organization based on volunteerism in the domain of disaster response, following its growth from an emergent group to a formal organization. It describes how HR emerged from earlier spontaneous digital volunteerism focused on helping to process information and get it into the hands of people who need it during disaster events, and how the group leveraged that previous experience (along with their leaders' and members' volunteer experience in other capacities) as a foundation for their new organization. Exploring the relationship between ICT, structure, and action, the study shows the group assembling an infrastructure using available ICT, and demonstrates how their assembled infrastructure shapes their work practice. It also explores how the group works to sustain—to continue to respond to events all over the world at a rate of more than one per day, to maintain participation of core volunteers, to recruit new members, and to carve out a role for itself in an emerging domain of ICT-enabled, digital volunteerism during disaster event.

Significant to the larger theme of this research effort, HR demonstrates a different kind of crowd work from the distributed human computation activity of TtT Translation (Study 1, Chapter 4) and the self-organizing emergent response efforts of the Voluntweeters (Study 2, Chapter 5). Considering the real-time disaster response functionality of the group, the crowd work of HR involves organizing a small, but persistent group of geographically dispersed members into a remote information-processing workforce. Though the organization leverages the remote crowd, it is not "crowdsourcing" in the microwork sense. Their activities are rich, diverse, interactive and collaborative, and instead resemble more of an open source community, where a small group of leaders and experienced volunteers function as a central core of the organization, and other members are encouraged to participate along the periphery from where they can slowly move towards the center through increased experience and responsibility. This participation model allows HR to effectively incorporate the spontaneous and episodic volunteers that can be a problem for traditional volunteer organizations (Voida et al., 2012)—volunteers that are critical to meeting the larger goals of the organization: to continue to respond, globally, to disaster events.

CHAPTER 7
STUDY 4. (HOW) WILL THE REVOLUTION BE RETWEETED? INFORMATION DIFFUSION & THE 2011 EGYPTIAN UPRISING[90]

Addressing Research Question 4 of this larger research[91], this study examines how all participation in the broad interactive space of social media during mass convergence events, even when not intentionally volunteer-oriented, can be considered a form of crowd work. This research views this crowd activity as productive "work" along multiple dimensions, including its functioning as a massive collaborative filter for information coming from the ground during mass disruption events.

7.1 Introduction

Microblogging tools have been appropriated for a wide range of applications, including, but certainly not limited to, the networking and socializing we have come to associate with social media. Several research studies have documented the use of platforms like Twitter (the most popular microblogging tool, which has global reach) and Sino-Weibo (primarily used in China) during mass emergencies and large-scale crises (e.g. Messina, 2007b; Heverin & Zack, 2010; Starbird et al., 2010; Vieweg et al., 2010; Qu et al., 2011). Messages about the use of microblogging, specifically Twitter, during political protests conflict. Though celebrated by the Western media as a force for good during the Iran Election protests in June 2009 (Grossman, 2009), follow-up research suggests the role of Twitter had been over-stated, noting that the social media revolution failed to affect the change it intended, and that the service may have been used by the government to crush opposition protests and identify protesters, whose lives were put in immediate danger (Burns & Eltham, 2009).

---

[90] This work is an adaptation of an earlier work:

Kate Starbird and Leysia Palen. (2012). (How) Will the Revolution be Retweeted: Information Diffusion and the 2011 Egyptian Uprising (CHI '12). ACM, New York, NY, USA, 7-16. DOI=10.1145/2145204.2145212 http://doi.acm.org/10.1145/2145204.2145212

To cite material from this Chapter, please cite the original paper.

[91] See Chapter 3 for an explication of research questions and their mapping to the separate studies of this work.

During the “Arab Spring”—a string of political uprisings that took place in the winter and spring of 2011 across the Arab world—attention was again paid to the emerging role of social media. Following Tunisia’s successful demonstrations, protesters took to the streets of Egypt to demand reform (New York Times, 2011a). As had happened during the Iranian Election protests, people across the world tuned into the event via social media; there again mainstream media described a significant role being played by platforms like Facebook and Twitter (New York Times, 2011b).

In his attention-grabbing editorial about the difficulties of using social media to affect revolutionary change—change he claimed required high-risk activism—Gladwell wrote that the “revolution will not be tweeted” (Gladwell, 2010). Our empirical consideration of the events in Egypt calls this claim into question.

Here, we report on this research, finding—to foreshadow a small bit of it—that 30% of the 1000 mostly highly retweeted Twitterers who were using popular hashtags related to the protests *were in Cairo* during the event, and that many of these Twitterers were “on the ground” in street protests. Tweets from these users contained information about meeting times, injuries, violence, supplies needed, etc. Revolutionaries were clearly using social media services to coordinate their actions and garner support. While we will not confront head-on Gladwell’s argument that social media fail as organizing tools for high-risk activism, this study will address a piece of that argument, asserting that the low-risk “activism” enabled by social media—what others have termed “slacktivism” (Morozov, 2009)—may indeed have been a productive component of this revolution.

In this study, we examine the use of one Twitter feature: the retweet, a mechanism by which Twitter users pass on or forward information to other users. We use the retweet mechanism as a means to understand broader Twitter behavior around these protests, and to demonstrate how remote individuals participated in Egypt’s 2011 revolution through low-risk, social media-enabled activities. To be clear about aims, we neither present a full description of Twitter activity during the protests, nor an analysis of the role that social media played (or didn’t play) in *orchestrating* this event. Rather we show how consideration of the retweet mechanism reveals a good deal about information contagion across a large number of people and how this behavior figures into social movements.

To that end, we theorize that "the crowd" participated in at least two ways. By uncovering interaction via information propagation between those on the ground and those in the broader Twitterverse, we first demonstrate how the crowd, through a show of interest, expressed solidarity with the cause. We then show how some protesters embraced this show of interest and aligned with it, giving credibility to the idea of solidarity through social media. Second, we show how the crowd did the "work" of information processing (through retweeting), and discuss how we might be able to leverage that recommendation process to increase situational awareness during events. This latter question speaks to the larger theme of this research, framing the collective activity of the participating crowd as a massive information filter during mass disruption and therefore mass participation events.

7.2 Background

7.2.1 Twitter and Mass Disruption

Twitter is a popular microblogging service with global reach, boasting well over 100 million registered users worldwide (Twitter Blog, 2010). Although considerable attention has been paid by the media to the use of Twitter during political protests after the Iran Election and again during the Egyptian protests, few empirical studies have been done on the use of microblogging during political disruption. There is some literature that addresses this issue from a political perspective. Mungiu-Pippidi and Munteanu (2009) give an accounting of the role of microblogging during the failed "Twitter Revolution" in Moldova in 2009. Burns and Eltham (2009) warn that the hype around the role of Twitter during the Iran Election protests has been overblown and blinded to the negative consequences of the service's use.

A growing body of research in the area of crisis informatics, including several studies on the use of microblogging during mass emergencies and disasters, helps elaborate the topic of civil disruption and ICT. Though obvious disparities exist between these types of social convergence events—not least of which are the special brand of concerns about privacy and security that arise during planned internal protests—the behavior of the attendant audience as mediated by social media likely has some similarities by virtue of the disruptive nature of either kind of event.

Microblogging has been used during crisis events all over the world, including the Haiti earthquake as examined in Study 2 (Chapter 5) of this larger work, the 2007 San Diego Wildfires (Messina, 2007b), the Oklahoma wildfires and Red River floods of 2009 (Starbird et al., 2010; Vieweg et al., 2010), the 2010 Yushu earthquake in China (Qu et al., 2011), and other more recent earthquakes in Christchurch New Zealand (Manhire & Tran, 2011) and Japan (Neubig et al., 2011). Research indicates that Twitter has been employed to seek information (Vieweg et al., 2010), to publicize the names of missing persons (Manhire & Tran, 2011), and to broadcast immediate needs and solicit donations (Qu et al., 2011). Studies 2 and 4 (Chapters 5 and 7) show Twitter also being used by remote volunteers to organize and provide informational aid.

7.3 Information Propagation and Twitter

This research study focuses specifically on how information spreads through Twitter during political protests by way of the retweet mechanism. The retweet is a user-driven convention—now formally supported by one-click functionality on the Twitter platform—that acts to forward tweets, giving attribution to the original (or another upstream) author. Boyd et al. (2010) found that retweets are used both for information diffusion and for engaging others. That study also noted the difficulty of identifying retweets and traces due to different syntaxes, user truncation, and added commentary across the tweet propagation. In the data collected for this study, we found four forms of the retweet convention in use: the most popular is `RT @username`. Also used at high volumes are `via @username` placed at the end of the tweet, `R @username` followed by the message, and `"@username: <original tweet text>"`, where the entire tweet is quoted.

Several studies have described dynamics of the retweet across large, random samples of tweets. Examining features that lead to increased retweetability, Suh et al. (2010) found that tweets with URLs and hashtags were more likely to be retweeted. They also noted a strong linear relationship between a user's number of followers and the likelihood that that user's tweet would be retweeted. They suggested that "social context"—information about the Twitter author including followers, following, account inception date, etc.—can help identify the "value" of information. Kwak et al. (2010) observed that retweets spread quickly and broadly, noting that the speed of diffusion could be an indication of strength of influence of the Twitterer, and suggesting that the number of followers a user has and the number of times that user is retweeted are different measures of popularity. Comparing Twitter and Digg, Lerman and Ghosh (2010) found that information diffuses faster on Twitter, and that initial number of followers is not as closely correlated with retweetability as it is on Digg. Van Liere (2010) used retweets to study the distance that a tweet travels and identified a type of Twitterer—the "information broker"—who connects to others according to shared interest and who spreads information across geographic distance.

Others have addressed information diffusion through microblogging sites during crises. In a previous study looking specifically at the retweet mechanism during crises, we found that retweets with topical keywords were more likely to be on-topic related to the disaster than non-retweets (Starbird & Palen, 2010), suggesting a role of information recommendation performed by the crowd. We also noted different retweet patterns between those local to a disaster, who preferred to retweet messages written by people who were also locals, over those who were remote, who retweeted almost exclusively messages that possessed "broad appeal." In Vieweg et al. (2010), we noted that retweeted information is more likely to include information that contributes to situational awareness than non-retweeted information, and Qu et al. (2011) also found that reposts were more likely to contribute to situational awareness as well as to contain "action-related" information.

Complex contagion, a concept from sociology related to information contagion (Centola, 2007), may be an important dynamic of social media use during political protest. Romera et al. (2011) researched how socially sensitive topics, including political ones, propagated through Twitter. They identified two different properties of diffusion: *stickiness*, the likelihood of information being spread after one contact;

and *persistence*, the likelihood of information being spread after repeated contacts. They found that complex contagion to be at work on Twitter, reporting that political hashtags were more persistent than other types of tags, meaning that they were more likely to be spread after multiple exposures. Twitterers who are initially unlikely to join a conversation on a sensitive topic become more likely to join as they see increasingly more people becoming involved.

7.4 Method

7.4.1 Event Description: The 2011 Political Uprising in Egypt

This study investigates Twitter activity during the 2011 political uprising in Egypt, a "mass disruption" event with mass social and informational convergence properties. Mass protests of Egyptian autocratic governance began on January 25, 2011, and continued for eighteen days until Egyptian president Mubarak resigned on February 11 (New York Times, 2011a). Early on, social media appeared to have an active presence. In reaction to the initial, mostly peaceful protests, the Egyptian authorities cut internet access to major providers on January 25 at 12:20am EET[92], with service largely resumed by February 2. The government also moved to disband protesters using a security police force notorious for torture. By January 28, hundreds of thousands of people had gathered in Tahrir Square (New York Times, 2011a). An ad-hoc medical facility was set up near the square to tend those wounded in skirmishes that had broken out between security forces and the protesters (Al-Ghazawy, 2011). February 2 marked a significant shift from relatively peaceful protests to violent clashes between pro-Mubarak groups and anti-Mubarak protesters. Several non-Egyptian reporters on the ground in Tahrir Square, a central location for the protests, and other parts of Cairo were reportedly attacked by pro-Mubarak "thugs," as they came to be called (New York Times, 2011a; Sweney, 2011). Protests continued and tensions increased over the following week, finally culminating on February 11 when Mubarak stepped down.

---

[92] All times in this study in local Cairo time, GMT +02:00

7.4.2 Data Collection

Though political unrest is ongoing in Egypt at the time of this writing, this study focuses on an early window of these protests. It is based on Twitter data collected between February 2, hours before Internet access was restored in Egypt, and February 15, four days after Mubarak stepped down. Using sophisticated search architecture that takes advantage of several Twitter APIs (Anderson & Schram, 2011), we collected both tweet and Twitterer data. For tweet collection, we relied on Twitter's Streaming API to collect tweets in real-time, filtering on the following terms: `egypt`, `#egypt`, and `#jan25`.[93] This captured tweets in English, as well as a range of other languages, including tweets written in Arabic, which constitute 15% of the data. For each unique Twitterer who contributed a tweet to this collection, we also captured all of their Twitter profile information including follower count at the time of the first tweet grabbed by our collection. These data—2,229,129 tweets and 338,895 unique Twitterers—comprise the *Egypt Twitterverse Proxy* set. Due to technical issues with Twitter data collection, we were unable to collect data during the early stages of the event (between January 25 and February 2) and lost data during three short windows within the collection period. However, the type of analysis conducted here does not depend on a full accounting of all possible tweets. Additionally, the delayed capture of profile data, which resulted in elevated initial follower counts for local Twitterers, likely weakened some of the effects we report on later in this chapter.

7.4.3 Highly Retweeted Twitterers

Accepting the retweet feature to be a measure of popularity (Kwak et al., 2010) and possibly a recommendation feature during mass disruption events (Starbird et al., 2010), we then attempted to identify and examine the most highly retweeted Twitterers. Retweets (using `RT @` and `via @` conventions) make up 58.5% of the *Egypt Twitterverse Proxy* set (all languages combined). Furthermore, 56.3% of all tweets with non-English characters (which include Arabic language tweets) in that set are

[93] Collecting Twitter data during mass disruption events requires rapid selection of search terms in an evolving information space (Romero et al., 2011). Arabic speakers in our research lab selected these terms as the most popular during the early days of this event, resulting in a large, low-noise sample.

retweets, which suggests that the retweet conventions have been at least partially adopted by Arabic language tweeters, and indicates that retweeting continues to be an important behavior during mass disruptions.

Searching against the two retweet conventions mentioned above, we calculated how many times each Twitterer was retweeted within the *Egypt Twitterverse Proxy* set, and then identified the top 1000 most-retweeted accounts. We then cross-referenced these usernames against the usernames of the Twitterers whose profile information we had collected, and located the profile and follower data of 956 of the 1000 highly retweeted usernames. The 44 missing usernames are likely absent because users protected (or restricted access to) their accounts or because they untraceably changed their account names within the data collection window. The 956 remaining users and all of their keyword tweets comprise the *Egypt Highly RTed Twitterers* set.

For further analysis, we randomly selected a one-quarter sample from these 956 Twitterers. This resulted in 254 initial Twitterers. We then removed six accounts: three were removed due to a zero following count, which indicates a problem with the account or an account suspension; two were removed because their account names had changed during the event, altering their retweet numbers; and a final account was removed because it was a bot with more than 10,000 automatically sent tweets during the time period. The remaining 248 Twitterers comprise the *Egypt Highly RTed Sample* set.

7.4.4 *Measuring diffusion of individual tweets*

To further understand information diffusion, we traced the propagation of all retweeted tweets that were originally authored or attributed to a Twitterer in the *Egypt Highly RTed Sample*. We were able to trace the history of 313,662 retweets back to 34,605 original tweets authored by a Twitterer in this sample. We calculated both how many times an author was retweeted, as well as how many different tweets of theirs were retweeted. We also captured how many times each tweet was retweeted.

DATA SET	# Twitterers	# Tweets
Egypt Twitterverse Proxy	338,895	2,299,129

Tweets sent during Egyptian protests Identified using keyword search: `#jan25 #egypt` Egypt *Used to identify highly RTed & track tweet/meme diffusion.*		
Egypt Highly RTed Subset of Egypt Twitterverse Proxy Top 1000 most RTed accounts, where profiles could be found.	956	282,010
Egypt Highly RTed Sample 25% sample from Egypt Highly RTed *Coded for Location of Author.*	248	69,461

Table 8. Egyptian Protest Data Set Statistics

7.4.5 Qualitative Coding

To understand the types of accounts that are most retweeted, we categorized the Twitterers in the *Egypt Highly RTed Sample* along several criteria, including their account affiliation and location during the event. Because users occasionally include locations in their profile that are not their actual locations, or because they move around during events, we do not accept the location character string in the user profile as the true location. Instead, we manually "code" users for location by reading through each user's entire collected twitter stream (which ranged from tens to thousands of tweets for each Twitterer) looking for assertions of or references to being in Cairo during the period of the protests. Because this study centers around the activities in Cairo's Tahrir square, we coded each account's location relative to Cairo, using the following categorizations: *In Cairo, In Egypt but not in Cairo* and *Not in Egypt*. Twitterers who were in Cairo at any time during the event were coded as being *In Cairo*.

Many of the top 1000 most retweeted accounts in the Egypt data set had Arabic language tweets in their user streams. To make determinations of location and affiliation for Arabic language Twitterers, we first analyzed the profiles and user streams using an English translation generated by Google Translate software. When coding determinations could not be made in this way, we asked a native Arabic speaker to translate the tweet and profile data. Ultimately, we could not confidently determine the locations for three of 248 Twitterers in the *Egypt Highly RTed Sample*.

7.5 Findings

Using both qualitative and quantitative analyses, we examine how the retweet was used to diffuse information during the Egyptian protests. We begin at the tweet level, investigating the propagation and re-appropriation of a single tweet meme that spread widely in the set. We then use a measure of the number of times a tweet is retweeted to uncover distinctions in retweetability among different types of information. Next, we shift back to look at the retweet as a method of recommending Twitterers as well as tweets, and examine the features of Twitterers who were most highly retweeted in our data.

7.5.1 Propagation of a Metaphor for the Revolution

The most popular tweet in the *Egypt Twitterverse Proxy* set is a proclamation of support for the removal of Mubarak, represented by an ascii-created "progress bar" graphic:

> **@adelshehadeh** (10 Feb, 18:37): Uninstalling dictator ... 99% complete
> ████████████████████████████░ #egypt #jan25 #tahrir #mubarak

Variations of the "Uninstalling dictator" with progress bar tweet appear 19,836 times in the data set. With this finding, we further investigated the morphology of the progress bar text and graphic, discovering that its use was widespread, cutting across Twitterverse, and began much earlier, progressing to this most popular form. It appears more frequently among those who are not local, though those who were on the ground also propagated it—a point to which we return at the conclusion of this section.

Over 1% of our *Egypt Twitterverse Proxy* set (23,012 tweets) reference the progress bar theme, as determined by inclusion of the block characters as well as occurrence of the % character. In total, of 338,895 Twitterers in the entire set, we found 20,727 who tweeted at least one progress bar tweet. We examined these tweets by hand to identify the major forms of the progress bar tweet and how they shifted over time, as well as the smaller changes that Twitterers made as evidence of the "work" that was put to sustaining and adapting the *meme* over time.

Throughout the time period of the protests, variations on the progress bar meme appear, morphing into new forms, but in most cases, keeping to and elaborating on a common "computing" metaphor. The first occurrence happens the morning of 2 February, suggesting that the progress bar was already propagating, though in smaller numbers, prior to data collection start. The tweet is retweeted twice:

@NotNaif (2 Feb, 11:49): Hosni #Moubarak Escape loading... ██████████████ █████] 99% #alarabiya #egypt #cairo #jan25 #FREEEGYPT #Internet #25jan #Arab

The next form extends the computing metaphor that is implied by the progress bar. It is only retweeted once in this form, but introduces both the ideas of "installing" and "freedom" into the information space:

@AntiMubarak (3 Feb, 01:41): RT @thameralzaidy Installing Freedom in Egypt ████████████████████ 99% #Jan25

The first Arabic language tweet with a progress bar appears almost two days later. Adjusting the sentiment of the tweet to reflect a slightly less positive outlook after the violent clashes that took place on 2 February in Tahrir Square, @almuraisy suggests that "some more patience" is needed. This tweet was retweeted 6 times.

@almuraisy (4 Feb, 14:26): ███████████████████░
99% هانت يا جماعة #Egypt #Jan25 #Tahrir

Of note, only 52 instances of the progress bar tweet (much less than 1%) had Arabic characters in the text.

The next form expands upon the "freedom" idea, and makes a strong connection to the progress bar with "loading." This is the first progress bar to diffuse broadly in our set, appearing 5977 times in a similar format:

@fakroona (4 Feb, 14:39): FREEDOM LOADING ███████████████████░ 99% #Egypt #Jan25 #Tahrir #Cairo #mubarak #sidibouzid #fridayofdeparture #yemen #syria #jordan

Interestingly, even though the "FREEDOM LOADING" and progress bar text, as well as the attribution to @fakroona, diffuses widely, the #fridayofdeparture hashtag is dropped in about half of the retweets, like the one below.

> **@StephanJourdan** (4 Feb, 15:00): RT @fakroona: FREEDOM LOADING ████████████████████░ 99% #Tahrir #egypt #jan25

All these major progress bar tweet forms include variations in hashtag listings, as well as attributions to the original and downstream Twitterers. Boyd et al. (2010) report that text and attribution often change over the course of a retweet's life, an effect they attribute partly to character length limits of tweets and to added commentary by downstream Twitterers. We see evidence in this study of people in different locales trying to make the tweet their own by slightly adjusting the meaning to take into account the event progression, or contributing new tags or short comments.

In a move similar to @almuraisy's Arabic tweet above, @RuwaydaMustafah also adjusts the popular meme to acknowledge renewed uncertainly in the success of the protests, introducing an "error message" on 4 February:

> **@RuwaydaMustafah** (4 Feb, 15:25): FREEDOM LOADING ███████████████████░ 99% [Error : Please remove Mubarak and try again!] #Jan25 #Egypt

This tweet results in 324 instances, but is distributed fairly steadily until Mubarak steps down on the 10th. Notice that this tweet drops attribution to the original "freedom loading" author. Other variants of the metaphor appear, but are not as frequently propagated. Their presence nevertheless suggests attention to a "collective riff," and includes "downloading freedom" and the "blue screen of death" which appear 3 and 9 times respectively in our set:

> **@JameedKaraky** (4 Feb, 17:21): #Jan25 Downloading Freedom in #Egypt ███████████████████████░ 99%

> **@BanglarPain** (4 Feb, 14:50): @fakroona #Mubarak Overload ████████████████████████░ 99% Blue screen of death! #Tahrir #jan25 #Egypt

The latter tweet is a direct reply to @fakroona, explicitly playing off of @fakroona's popular progress bar tweet. The investment into the collective riff continues with some attention to incrementing the percentage value associated with the progress bar by 0.1%:

> `@Ahmoooos (10 Feb, 17:57): FREEDOM LOADING ██████████████████░ 99.8% [ removing mubarak in processing right now please wait ! ] #Jan25 #Egypt #Mubarak #tahrir #cairo`

New forms continue with the addition of more, and shortened progress bars, here with attention to leaders. Another variation on this style appears in the form of other nation states in the Arab world where recent protests have taken place or future protests have been proposed:

> `@alihabibi1 (10 Feb, 18:21): Ben Ali ████████ 100% Mubarak ███████▓ 99% Qaddafi ███▓____░ 42% Boutaflika ██▓_______░ 19% #Jan25 #SidiBouzid`

It was @adelshehadeh's tweet (at the beginning of this section), however, that generated the most propagation, and it didn't appear until 10 February, when the resignation of Mubarak was appearing imminent. That tweet accounts for 78% of the "progress bar" tweets in the set, and introduces the morphological change of "uninstalling."

Still more extensions of the computing metaphor continue to appear after the first appearance of the most diffused form of the meme, with references to "Egypt 2.0," "Fatal error," "kernel error," and "install Democracy, Disk #1 into drive E: gypt." Figure 20 illustrates the volume of progress bar meme tweets over time, marking some of the specific variations.

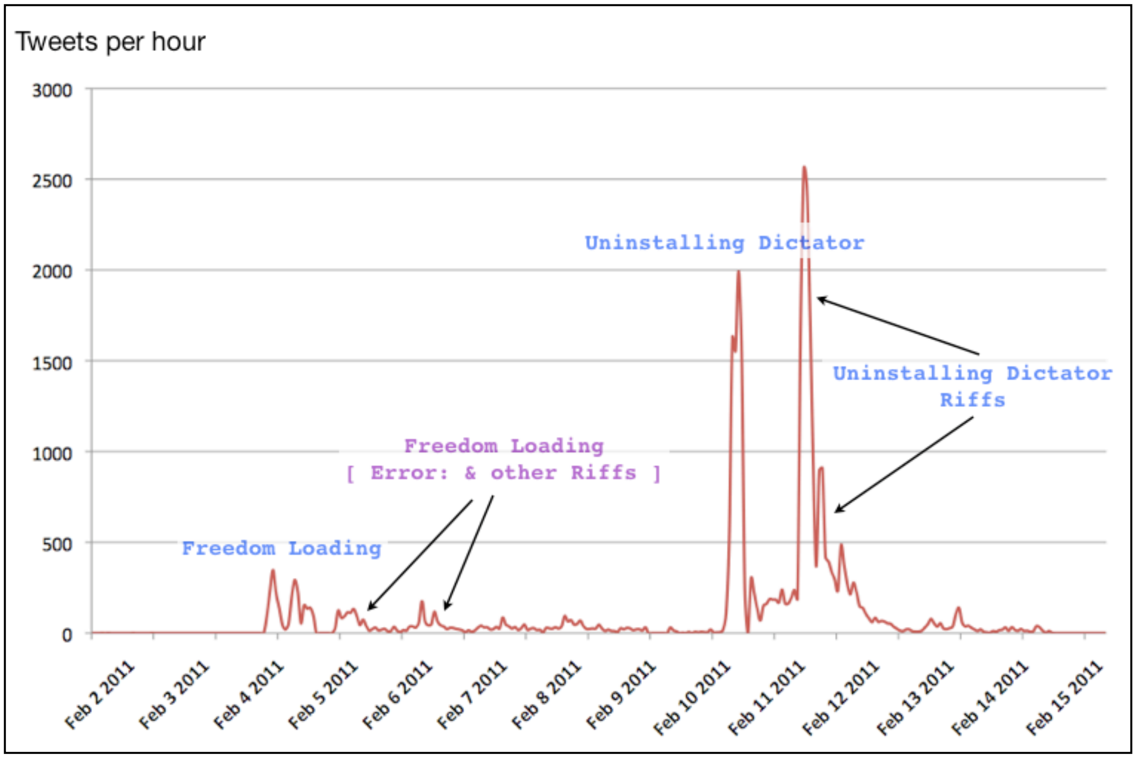


Figure 20. Volume of Progress Bar Meme Variations in Tweets per Hour[94]

In these extractions of the major forms of the meme morphology identified here, we see the evolution of a tweet meme. The progress bar ascii graphic is probably not something that everyone can easily create. Once someone introduced it to the public sphere, it got picked up and modified. Though these different forms comprised a fairly high absolute number, their totals paled in comparison to one particular form that did not appear until 10 February. In the time leading up to that date, however, variations on the first form transformed it into a meme, a widely understood idea. Centola and Macy (2007) call this "complex contagion," where multiple exposures from multiple "long-tie" sources ignite and support diffusion. The remixing of elements, all keeping within a computing metaphor, shows some degree of shared understanding of its purpose, and kept the meme propagating. The variations show collective "work" happening on the meme to keep it relevant and interesting, until it reached the point of widespread coverage and possibly saturation with the 10th of February "uninstalling dictator" tweet.

---

[94] This Figure was first presented in the CSCW talk related to this work, but does not appear in:
Kate Starbird and Leysia Palen. (2012). (How) Will the Revolution be Retweeted: Information Diffusion and the 2011 Egyptian Uprising (CHI '12). ACM, New York, NY, USA, 7-16.

We note that most progress bar tweets appeared in English, though a small number also appeared in Arabic, Japanese, Korean, Dutch, German and Spanish. We suspect that the number of tweets in Arabic was lower than expected, and that the overall distribution tended strongly toward English because of the desire for contributions to the collective meme to be widely understood. To this point, many of the authors who originated new versions of the progress bar were tweeting from other locations in the Middle East.

7.5.1.1 Propagation by Others vs Locals

To the question of whether tweet propagations like this effectively contribute to the basic work of the revolution, we look to how much this meme appeared in the tweet streams of those who were on the ground in Cairo versus those who were not.

In the *Egypt Highly RTed Sample*, where we coded for location, Twitterers who were not in Cairo were considerably more likely (27.6%) to send a progress bar tweet than those who were in Cairo during the protesting (17.6%). This suggests that this "sticky" meme had broad appeal in the wider Twitterverse, offering a short abstract of the event (Starbird & Palen, 2010). Indeed, another interpretation of this finding might be that non-local Twitterers were boosted into highly retweeted status by the popularity of the progress bar meme. We will explore this further in the latter half of the Findings section.

Perhaps more interesting than the discrepancy here is that 13 of our sample's 74 highly retweeted Twitterers who were tweeting from Cairo *were* taking part in propagating the progress bar tweet, including @sandmonkey, one of the ten most retweeted accounts in our set. Tracing the origin of several of the most popular variations, we find little evidence that any began in the streams of Cairo Twitterers. This suggests that Twitterers on the ground were, to some extent, embracing external declarations of solidarity as legitimate messages and voices within their cause.

In sum, the propagation of the progress bar meme by both those on the "inside" as well as the large attendant audience suggests interplay between both realms. Though such a meme could be perceived as "silly" and incidental to the event—and indeed, some people publicly asked for its distribution to stop—

such phenomena need attention to fully understand their impact on newsworthy and risky events like a political revolution. It could be that wide propagation of a tweet-idea helps cast a light on events happening in other parts of the world. Participation by those in the midst of the geographical event suggests active endorsement of the meme. In other words, if those on the ground find value in the propagation of a tweet, then—in contrast to Gladwell's (2011) point—perhaps that is one of the real measures by which one decides how much a role social technology plays in revolutions.

7.5.2 Broad Appeal v Local Utility

By tracing the most popular (highly retweeted) tweets originated by authors in the *Egypt Highly RTed Sample,* we see a similar bias towards broad appeal tweets, messages of solidarity with the Egyptian cause, as well high-level news.

> **@JoeUnfiltered:** The people of #Egypt have shown the world that youth activism can change ANYTHING. 18 days has ended 30 years of oppression ***[RT 3989x[95]]***

> **@DominicKavakeb:** Its so vital that world understands this is not a divided #Egypt. This is the state attacking the people. #jan25 ***[RT 2660x]***

> **@HuffingtonPost:** BREAKING: Mubarak will step down, hand power to military tonight - Reuters http://huff.to/egyptnews #Egypt #jan25 ***[RT 1233x]***

These were the three most highly retweeted tweets originated by any author in our sample. The top tweet was retweeted 3989 times; the second 2660 times and the final tweet 1233 times. As above, none of these tweets were originally authored by individuals who were in Cairo at the time of the protests.

Highly propagated tweets originated by locals often follow the broad appeal pattern as well. Tweets about violence, detained friends or colleagues, requests for solidarity, and humorous tweets are often the most retweeted messages authored by the group. The following tweet, retweeted 998 times, is the most popular tweet authored by a local.

95 This notation indicates the number of retweets.

@sharifkouddous: A couple just got married in Tahrir in front of army tanks. A revolution wedding. #Egypt #Jan25 ***[RT 998x]***

While witty and humorous tweets experienced "sticky" retweetability, serious tweets remarking on violence, especially violence against the media, and asking for support were also among the most popular. These two tweets asking for support for the release of Ayman Mohyeldin, a reporter for Al Jazeera, were retweeted a total of 1386 times.

@nolanjazeera: the hashtag is #freeayman please retweet as much as u can to get this trending. Attacks on media in #Egypt MUST end now ***[RT 985x]***

@AymanM: Ayman has been detained by #Egypt military. Will keep everyone posted on his status as things develop (tweeted by friend) ***[RT 401x]***

Popular tweets from locals also included first hand reports of violence or tactical information. The following tweet, a real-time accounting of violence on the ground, written by an NBC news reporter, was retweeted 591 times. The two subsequent tweets, which received considerably fewer retweets relayed tactical information about the evolving protests.

@richardengelnbc: #egypt.. Just saw a protester hit by a molotov .. Saw him catch on fire.. Other protesters managed to put him out.. ***[RT 519x]***

@alaa: Army gave up and let us control flow of ppl at tv building #Jan25 ***[RT 336x]***

@3arabawy: The army has allowed a group of 100 Mubaraks thugs into Qasr el-Nil bridge now. #Jan25 ***[RT 175x]***

Less "popular" tweets by locals, as measured by the breadth of their diffusion within the larger Twitterverse, often contained more detailed information from the ground. The following tweets were retweeted fewer than 20 times each, but offer very specific details about the situation in Cairo.

@wilyawil: Im hearing from friends inside Tahrir Sq that best entrance is from Falaki st. #jan25 #egypt ***[RT 18x]***

@mosaaberizing: Weve taken over Talaat Harb entrance as well. Just the musuems left now. #Jan25 ***[RT 14x]***

The popularity of individual tweets shows a bias towards English language tweets. Though 34% of tweets in the *Egypt Highly RTed Sample* contain Arabic characters, only 20% of tweets that were retweeted from these Twitterers had Arabic characters. Arabic language tweets that were retweeted at least once received an average of five retweets, while non-Arabic tweets were retweeted an average of eight times. The first Arabic tweet from the Twitterers in our highly retweeted sample, a tweet from the @nytimes requesting an interview with someone on the ground, appears as the 83rd most popular tweet, retweeted 242 times.

@nytimes: في القاهرة او في اسكندرية او
في سويز و تحب ان تكلم عن ال ضع؟ كـ @Nadia8 في انجليزي او عربي.
#Egypt #Jan25 ***[RT 242x]***

Analyzing the diffusion of popular tweets through retweeting suggests that actionable or action-related information (Qu et al., 2011), information coming from the ground and, perhaps, appearing in the information space for the first time, can be found in the streams of local Twitterers, but not necessarily in their most popular tweets. Tracing the most popular retweets, even those sent by locals, may result in broad appeal messages. But tweets like the following may be more valuable to those acting on the ground, and perhaps to others who are remote, trying to better understand and support the situation on the ground:

@occupiedcairo: Medical supplies needed at temp. hospital at tahrir: neck supports and stitching thread #jan25 ***[RT 39x]***

How might the noisy retweeting by the well-intentioned and easily entertained crowd be leveraged to home in on actionable information being tweeted from the ground?

7.5.3 Who Were the Most Retweeted Accounts

Possibly a more effective strategy than investigating the most popular tweets during the Egyptian protests is identifying the most popular Twitterers. The number of times that an account is retweeted can be used as a measure of popularity (Kwak et al., 2010) or, during a specific, mass disruption event, an implicit recommendation by the crowd that an account is a useful source of information (Starbird & Palen, 2010). In this section, we will describe features of the most highly retweeted accounts, and investigate the utility of using the retweet mechanism, in concert with other aspects of an account's "social context" (Suh et al., 2010) to home in on accounts with "local authority" (Starbird & Palen, 2010).

Research has shown that many features of information propagation through social media have power law or other heavy-tailed distributions (Leskovec et al., 2007; Lerman & Ghosh, 2010). For our *Egypt Highly RTed* set, the number of times that an account was retweeted has a heavy-tailed distribution. Figure 21 shows this distribution, the number of accounts with x number of tweets, in linear-linear and log-linear scale. Note that the first bar in this distribution is at 170 tweets, the cut-off point for inclusion in the top-1000 Twitterer sample. The mean number of times retweeted is 1089, and the most highly retweeted Twitterer, @ghonim, was retweeted over 35,000 times.

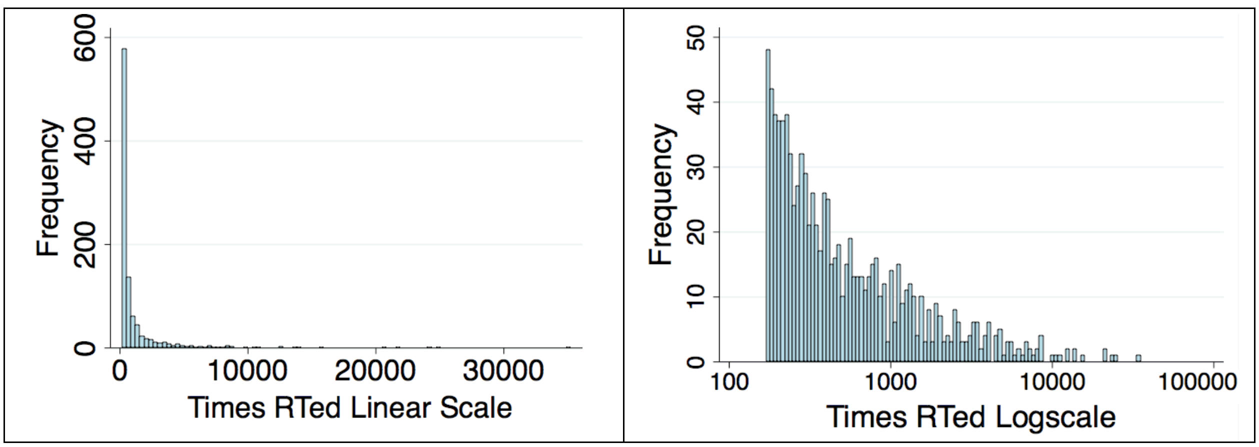


Figure 21. Times Retweeted, Linear-Linear & Log-Linear Scale

7.5.3.1 Top Ten Most Retweeted Accounts

Examining the top ten most retweeted accounts in the *Egypt Twitterverse Proxy* set (Table 9), we see the retweet acting to recommend Twitterers who, for the most part, were actively tweeting from Cairo. Many of these highly retweeted accounts were operated by journalists, or mainstream media with journalists on the ground. Others were popular Egyptian bloggers, and all have above average follower counts (see Figure 23).

Name	Times RTed	Followers	Affiliation	Location
@ghonim	35265	12491	individual	Cairo

@dima_khatib	25062	11320	journalist	Latin America
@bencnn	24066	21147	journalist	Cairo
@3arabawy	21607	9022	journalist	Cairo
@sandmonkey	20714	6943	blogger	Cairo
@alarabiya_ar	15681	14011	mainstream media	Arab World*
@monaeltahawy	14150	16048	journalist	Cairo
@ajenglish	13791	145246	mainstream media	Arab World*
@ajarabic	12687	30244	mainstream media	Arab World*
@monasosh	12609	2535	blogger	Alexandria

Table 9. Top 10 Most Retweeted Accounts (* denotes organizations with reporters on the ground.)

7.5.3.2 Top 1000 Most Retweeted: Location

Examining the location of the Twitterers in the *Egypt Highly RTed Sample* (Table 10), we see a relatively high number of local Twitterers. 89 of 248 Twitterers in the sample (36%) have evidence within their tweet streams of being in Egypt at some point between February 2 to February 15. Of these, 74 (30% of the sample) were in Cairo during that time. 43 others were located elsewhere in the Middle East and the rest were spread across the globe, with the highest concentrations in Europe (27) and North America (54). Among the 248 there were three for whom we could not determine if they were in Egypt or not.

Location	# Twitterers	% of Sample
Egypt – Cairo	74	29.8%
Egypt – Other	15	6.0%
Unknown, possibly in Egypt	3	1.2%
Arab World	43	17.3%
Outside Arab World	113	45.6%

Table 10. Location for the *Egypt Highly RTed Sample*

Though we do not have the locations of the remaining 800,000 plus Twitterers who sent one or more tweets hashtagged with `#egypt` or `#jan25`, it is probable that nowhere near a third of them were on the ground in the protests in Egypt. Highly retweeted accounts appear to be more likely than other Twitterers to be local in this event. This finding suggests that during mass-participation events of global import, the retweet may serve as a mechanism through which the larger Twitterverse identifies and recommends local Twitterers.

7.5.3.3 Top 1000 Most Retweeted: Number of Retweets

As mentioned in the above section on Broad Appeal versus Local Utility, tracing the propagation of an individual tweet offers some insight into the value of that tweet measured by its contribution to situational awareness and newness in the information space. Similarly, the number of unique tweets, sent by the same author and propagated by the crowd, may be a measure of overall value of that user's tweet stream, in relation to a single event. In other words, though their overall number of times retweeted is the same, a Twitterer retweeted 500 times for a single tweet may be more likely to be a broad appeal Twitterer who is not on the ground, than a Twitterer retweeted 50 times for 10 different tweets.

Figure 22 shows the distribution of the numbers of individual popular tweets for each user in the *Egypt Highly RTed Sample* in linear-linear and log-linear scale. The number of popular tweets for highly retweeted Twitterers has a log-normal distribution. The mean is 108 retweeted tweets, with seven Twitterers in our sample having tweeted only one Egyptian protest tweet that was retweeted, albeit very broadly. Twitterers in Cairo were retweeted, on average, for a higher number of tweets (123), than those not in Cairo (102). Though this comparison of the number of retweeted tweets between locals and non-locals shows some difference, things become more interesting when we break down this effect in relation to different variables, below.

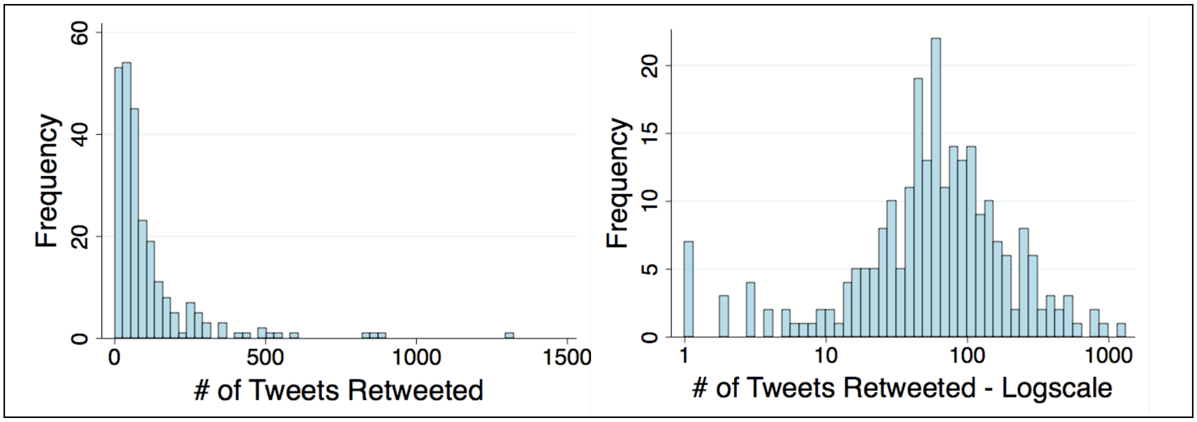


Figure 22. Number of Popular Tweets by Twitterer

7.5.3.4 Top 1000 Most Retweeted: Followers

For Twitterers in the *Egypt Highly RTed Sample*, the number of followers has a log-normal distribution with a long tail. The mean number of followers is 116,318, with a high standard deviation of 500,000. The median is 1730. Comparing the count across Twitterer location shows a stark difference between the median number of followers for a highly retweeted local (790) and a highly retweeted non-local (2609).

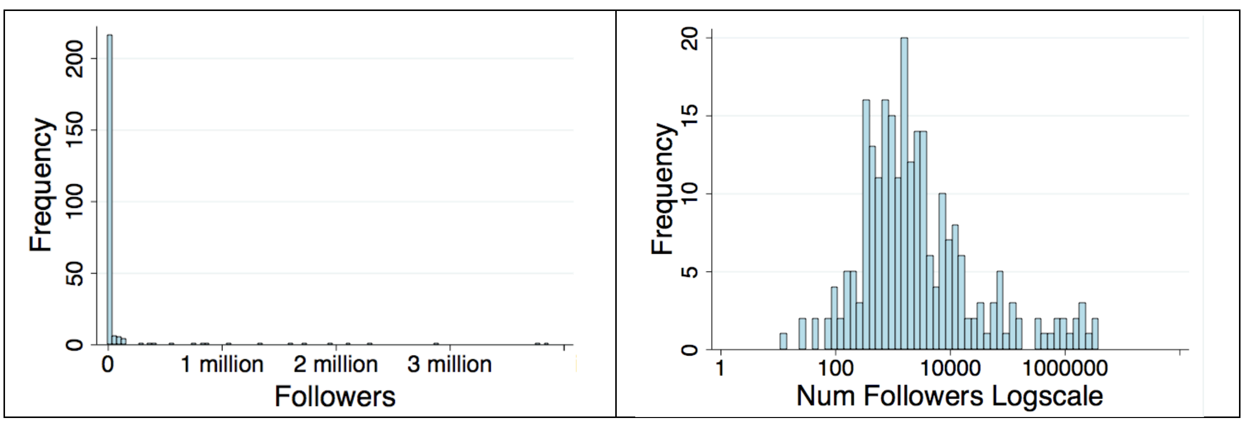


Figure 23. Number of Followers, *Highly RTed Sample*

Accounts with a higher number of followers get more exposure for their tweets, and therefore, each tweet they send has a proportionally higher chance of "infecting" the recipient or being passed on to others (Suh et al., 2010; Qu et al., 2011). In our sample, the number of followers is positively correlated with the number of times a Twitterer is retweeted. However, the Pearson coefficient between the two variables is only 0.1306, indicating that follower number does not fully account for retweet popularity.

While recognizing that retweets are one measure of popularity, Kwak at al. (2010) noted this gap between the popularity measures of follower count and retweet count. The stark difference between average number of followers for highly retweeted locals and highly retweeted non-locals may suggest that during mass disruption events the follower-retweet gap is related to whether or not the Twitterer is in or near the affected area.

7.5.4 Using retweets and "social context" to identify locals

In the final section of this analysis, we will address the possibility of leveraging what we know about the retweet-follower popularity gap, as well as other measures of retweet behavior and social context to identify Twitterers and tweets from "the ground."

The preceding analysis suggests a relationship between different measures of tweet behavior, social context, and location. To evaluate these relationships we did a multinomial logistic regression, using location (*In Cairo, In Egypt, Not in Egypt, Unknown*) as a polytomous (non-ordered) dependent variable. For our independent variables, we used follower count, number of times an individual is retweeted, number of different tweets authored by a Twitterer that are retweeted by the crowd, and the total number of total keyword tweets that each user sent during the collection window. To normalize the heavy-tailed distributions, we did natural log transformations on all of the independent variables.

Variable	Coef.	Std. Err	Z	P>|z|
log_followers	-.691792	.1286891	-5.38	0.000
log_timesRTed	.5020533	.1802369	2.79	0.005
log_numtweets	-1.08295	.3713641	-2.92	0.004
log_uniques	1.18107	.4061834	2.91	0.004
_cons	1.716295	1.44004	1.19	0.233

Table 11. Multinomial Logistic Regression: *In Cairo* case vs *Not in Egypt* base case

Treating *Not in Egypt* as the base case for the analysis, we find statistically significant effects from all four independent variables, but only for those who were in Cairo during the event.[96]

Log number of followers and log number of tweets both have significant negative effects on a Twitterer's odds of being in Cairo during the event. The log number of times retweeted and the log number of different tweets that were retweeted were significantly, positively related to being in Cairo. These descriptive statistics suggest that a *highly retweeted* Twitterer is only more likely to be local when their initial follower count is low. Number of followers and total number of keyword tweets contributed raise a Twitterer's chance of being retweeted, but do not make them more likely to be local. These results also support the idea that a Twitterer with only one or only a few popular tweets related to an event is less likely to be local than a Twitterer with many different popular tweets.

7.6 Conclusion

Twitter content during mass emergencies consists of both "original" information—that coming into the information space for the first time—and "derivative" information, which can be found elsewhere, as our previous research indicates (Starbird et al., 2010). Retweets, which constitute nearly 60% of tweets sent with tags referencing the Egyptian protests, are one form of derivative information. One perspective considers retweets then as merely noisy output and members of the crowd only cheerleaders.

Another view that incorporates the ideas put forth in this larger research effort is that retweets serve as a crowd-powered recommendation system. Plenty of Twitterers put thought into what they retweet (boyd et al., 2010). This constitutes a form of work—and sometimes a form of collective work, as we saw in the propagation and morphing of a metaphor-driven tweet across the duration of the Egyptian protests.

We also demonstrate how considering the crowd from this angle of a working collective can help identify, among a vast number the tweeters, locals who may be introducing "original" information into the space. However, this process is not so straightforward such that finding the most popular tweets or

---

[96] One other statistically significant effect was found for log number of followers on the Unknown case, which had a slightly negative effect, with P = 0.025 significance.

most popular Twitterers will solve the "noisy crowd" problem. Instead, it is important to understand how and why people retweet what they do, including the distinction between broad appeal and local utility tweets, and leverage this understanding to home in on new, locally relevant information. This can be useful as state of the art moves forward toward more automatic forms of curation and collation of vast computer-mediated communication.

When we consider these tweet activities of the locals—who, in this case, were closest to the activist cause and performed some of the work of activism as represented by their tweets—in relation to the activities of the on-lookers, the interaction between them suggests a more integrated relationship than the "slacktavism/activism" (Gladwell, 2010; Morozov, 2009) construct suggests. Here we see that parts of the crowd activity do more than create noise, and that the crowd's zealous but safe vicarious participation has perhaps a more meaningful and new connection to the hard work of the social movement than is readily visible.

This research on the use of the retweet mechanism shows the crowd doing work in two ways that make it a *functional*, if not necessarily central (and that debate needs to continue), aspect of social movements. First, expression of social solidarity is found through the collective, observable creation of a tweet-based, visual and metaphorical meme. This is a deliberate form of collective work. It is critical to recognize that the meme creation was also visibly acknowledged by local activists. These expressions of social solidarity, we propose, are the kinds of activities that draw and sustain attention on a cause, which in turn may sustain the cause itself. The second form of work happens individually but collectively has an effect—retweeting as information filtering and recommendation.

The behaviors of people both on the perceived "inside" and "outside" of the social movement show that responsibilities and weights of responsibility differ across the collective. However, the behaviors are interrelated, and as such, have elements of mutual benefit. Social movements indeed happened prior to social media, but we would be remiss to imagine social movements of the future not leveraging and adapting these tools to gather more to a cause.

CHAPTER 8
CROWDWORK: A NEW PERSPECTIVE FOR MASSIVELY-CONNECTED, COLLABORATIVE WORK[97]

In recent years, efforts by the crowd to contribute to disaster response, self-organized and otherwise, have often been referred to as *crowdsourcing*. During the early aftermath of 2010 Haiti earthquake, popular media and the digerati praised the role of "crowdsourcing" in the response efforts (e.g. Biewald & Leila, 2010; Mullins, 2010). The Disaster Relief 2.0 Report also used that term—28 times—to characterize the activities of digital volunteer groups during that event, and crowdsourcing efforts have been credited as serving important roles in response efforts for numerous crisis events since. However, the widespread use of this popular term, employed as a blanket descriptor for a variety of activities, may be obscuring the complexity of human behaviors and computational systems that support them.

The four studies presented in this dissertation offer several examples of how the crowd works to organize information during mass disruption events, revealing this work to be extremely diverse. Study 1 (Chapter 4) shows how Twitter users appropriated the Tweak the Tweet innovation to self-organize into a remote workforce for filtering and structuring social media updates. Study 2 (Chapter 5) reveals TtT use by remote volunteers to be one small component of much richer digital volunteer activity occurring on and through the Twitter platform—carried out by *voluntweeters*. An extension of that work is explored again in Study 3 (Chapter 6), which examines the Humanity Road organization, a group of online volunteers that performs as a remote information-processing workforce during disaster events. Finally, Study 4 (Chapter 7) looks at crowd work on a massive scale, examining how the actions of every member of the social media crowd can act as a collaborative filter for information coming from the ground of disaster events. Collectively, these studies describe crowd work as rich in content, interactive and often collaborative in nature, and almost always, in at least some ways, emergent.

---

[97] Some portions of this chapter have been adapted from:
Starbird, K. (2012). What "Crowdsourcing" Obscures: Exposing the Dynamics of Connected Crowd Work During Disaster. Presented at *Collective Intelligence 2012*, Cambridge, MA.

This research on crowd work in the domain of mass disruption serves as an inroad for revealing several inconsistencies between the definition of crowdsourcing and how the term is employed to characterize productive crowd activity. This chapter first explores some of the possible reasons for, as well as the impact of, this divergence in meaning. Next, it looks towards other approaches towards characterizing crowd work, Quinn and Bederson's taxonomy of human computation (2011) and Malone et al.'s genome of collective intelligence (2009). Concluding that current terminology and frameworks for characterizing crowd-powered systems and communities are limited by their focus on the systems themselves instead of the behaviors that those systems enable and constrain, I offer a new term, *crowdwork*, through which to examine applied collective intelligence of the connected crowd through the activities of its participants rather than the features of its systems.

I then present two approaches for understanding crowdwork during mass disruption events. The first is a descriptive model of crowdwork during mass disruption events as a layered filtration system that integrates human crowdwork with computational systems that enable this work. This model points towards future work, highlighting opportunities for the design of technological systems to support and leverage digital volunteerism, specifically, and crowdwork, generally. The second perspective considers this connected crowdwork to be more than mere filtration, offering a framework for examining applied collective intelligence as distributed cognition.

8.1 Crowdsourcing = Microwork? Addressing a Problematic Drift

Howe's initial description of crowdsourcing (2006; 2008) included a wide variety of crowd activities, including crowd brainstorming, crowd voting, collaborative creation, the division of complex tasks into micro-tasks for geographically dispersed workers, citizen science, collaborative filters and more. In recent years, however, the crowdsourcing term has become increasingly synonymous a single one of those forms: *microtasking*.

Microtasking involves breaking complex problems down into finite tasks, typically ones that require human cognitive abilities, and distributing those across a large number of remote workers through

Internet or mobile connections. Amazon's Mechanical Turk (mTurk) is perhaps the quintessential microtasking platform, maintaining a market for project owners to distribute "human intelligence tasks" (HITs) to geographically dispersed workers. These HITs can typically be completed in a small amount of time (seconds to minutes), and workers earn a few cents to a few dollars for each, depending upon the type and duration of the work (Ross et al., 2010). Mechanical Turk and other microtasking systems utilize a top-down task-assignment strategy. The initiator of the project either designs a system specifically to collect the human intelligence actions desired or defines a series of HITs for completion within an existing platform, like mTurk. In both cases, there is a leader who generates the project, defines the goal, and owns the resulting product or products. The workers merely complete the tasks.

Microtasking is also referred to as *microwork*, but this research suggests a distinction between the two terms where microtasking is a sub-section of microwork that suggests a top-down assignment strategy. As described in Section 8.1.2.4, microwork can manifest in more organic ways as well—ways that support collective ownership of the activity.

Though microtasking represents one of many forms of crowdsourcing according to its original definition (Howe 2006; 2008), within the Human-Computer Interaction (HCI) research community, the *crowdsourcing* term has experienced significant compression, moving closer and closer to this particular type of crowd work. At the CHI 2011 workshop on Crowdsourcing[98], 21 of 43 papers focused exclusively on the Mechanical Turk platform or projects that used Mechanical Turk either as a functional component of the system or for gathering research data. Seven other papers talked about different systems that supported some other form of microtasking. Researchers may be drifting into this narrow view (microtasking) and single instantiation (mTurk) of crowdsourcing because Mechanical Turk is easy and cheap to study—Ross et al. (2010) found that workers make about $2.00 an hour—and because Mechanical Turk can serve multiple purposes in research: i.e. a research site or object (e.g. Callison-Burch, 2009; Ross et al., 2010; Hullman, 2011); as a tool for coding research data (Paul et al., 2011); and

[98] http://crowdresearch.org/chi2011-workshop/

as a resource for generating classification data for machine learning algorithms (Callison-Burch, 2010; Carlson et al., 2010). However, collectively intelligent crowd work is far more varied than microtasking, and the crowdsourcing term itself, as outlined by Howe (2006, 2008), was originally offered as an explanation for a much wider variety of collective behavior.

8.1.1 Competing Definitions for Crowdsourcing: Open Source or Outsourcing?

Just as the roots of crowdsourcing lie in both *outsourcing* and *open source*, tensions in how the term is understood and employed may stem from the differences between those two terms. On his website[99], Jeff Howe provides two definitions—his White Paper Version and his "Soundbyte" Version—that illustrate, perhaps unintentionally, this discrepancy:

> **The White Paper Version**: Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call
>
> **The Soundbyte Version**: The application of Open Source principles to fields outside of software

In the "White Paper Version," Howe compares crowdsourcing to the outsourcing business model, while in the "Soundbyte Version," he uses open source principles as the comparison point. By offering these two definitions as functionally equivalent, Howe appears to suggest that these two things are essentially the same. However, open source organizing principles and outsourcing business models are inherently different things.

Outsourcing is a strategy used by companies to utilize workers outside their employee base. Harland et al. (2005), define outsourcing as a company going outside its internal workforce to procure something, a service or product, that the company either previously performed or created in-house or could have performed or created in-house. Considering this, the *source* in out*sourcing* refers to the workforce, or how the work is distributed. Conversely, the source in open source refers not to the workers but to the activity, or how the work is done. An open source community is a group of computer programmers who come

[99] http://www.crowdsourcing.com/

together to a create computer program, or source code, that is open—i.e. freely available for others to use. The *open* of open source is therefore the distribution license, and the *source* is the code.

These terms, both used as reference points for crowdsourcing, imply very different kinds of things. Open source has been referred to as the "bazaar" model of software development (Raymond, 2000), where workers self-organize from the ground-up to create systems where the end product belongs to the public domain. As Howe's definition above suggests, crowdsourcing involves extending open source organizing techniques to other kinds of problems—not just computer programming. In the studies presented in this work, the efforts of the *voluntweeters* can be seen as crowdsourcing of the open source variety. In a powerful example from Study 2 (Chapter 5), Twitterer @MelyMello initiated an effort to fill up the cell phones of people affected by the earthquake. At first, she worked alone, but as she collected more and more numbers, she recruited help from other Twitterers to identify numbers that needed minutes, solicit funds, and add time to the phones. Though @MelyMello can be seen as the leader of this effort—much as Linus Torvald was the leader of the Linux open source effort—the group came together from the ground-up, self-organizing into a functional "organization" that worked to assist Haitians in reconnecting after the earthquake. The Humanity Road organization, the focus of Study 3 (Chapter 6), can also be seen as an open source community, using internet and mobile technologies to come together and organize around a common goal—educating the public before, during and after crisis events. Humanity Road workers complete a range of tasks, and though top-down communication and organizing have become increasingly prevalent in their interactions, leaders continue to hold up their "self-directed work team approach" as an ideal for their work practice (personal communication, Cat Graham).

Outsourcing, the other proposed analogue for crowdsourcing, carries very different connotations. The focus there is not on how the work is done, but on how it is distributed, and that is almost always from the top-down. For an outsourcing effort, the goal is typically defined not by the group who comes together to address it, but by the project originator who then pushes it out as work for someone else. Crowdsourcing of the outsourcing variety implies connecting a pre-defined problem to the remote crowd.

8.1.1.1 Virtual Teamsourcing: Crowdsourcing as Outsourcing to a Team of Remote, Connected, Crowd Workers

Additionally, there are two different types of outsourcing strategies for crowdsourcing projects. The first, which follows the original outsourcing business model, occurs when an organization distributes a problem to another, outside organization, who then determines how to solve that problem using its own resources. In these cases, the organization completing the work has structure that both precedes and extends beyond any particular project or task, and its workers are typically connected to each other at one or multiple levels. In this arrangement, tasks may be broken down and work organized by the teams themselves, in a lateral or top-down fashion, or both, depending upon the structure of the outsourced-to organization.

This is the type of crowdsourcing that most closely aligns to outsourcing in the offline world, with a significant difference in that the organization to which the work is outsourced is a virtual one. For the purposes of differentiating between types of crowd work, this research offers the term, *teamsourcing*, to characterize crowd work that outsources a problem to a virtual team. Though they have not established relationships to do this, Humanity Road has the potential to become a teamsource organization, where organizations in need of their services could outsource specific information-processing tasks to the group. The concept of a Virtual Operations Support Team (VOST) (St Denis et al., 2012) is an example of teamsourcing already taking place in the crisis response domain.

8.1.1.2 Microtasking: Crowdsourcing as Outsourcing to Remote, Disconnected Crowd Workers

However, crowdsourcing as outsourcing rarely looks like a VOST, where work is distributed as a large task to a team of connected individuals. Instead, it typically follows the other variety of outsourcing-crowdsourcing: microtasking. Microtasking, a form of crowd work newly-enabled by Internet connectivity, involves breaking down a problem into finite, pre-defined tasks and distributing those tasks to remote workers who are typically not connected to each other. Though not equivalent to human computation, microtasking is a type of human computation that takes place within platforms that both coordinate these tasks and encapsulate the resulting work products.

Though Section 8.1.2.4 discusses how other forms of human computation can manifest outside of technological platforms specifically designed to exploit then, the majority of human computation systems featured in HCI research—e.g. reCAPTCHA (von Ahn et al., 2003), the ESP game and Peekaboom (von Ahn & Dabbish, 2008), and Soylent (Bernstein et al., 2010)—are examples of microtasking, where a technological platform coordinates and contains the activities, and where workers may connect to each other as necessitated by the task—e.g. the ESP game (von Ahn & Dabbish, 2008)—but are not otherwise connected within an organization or social network. Microtasking is a therefore a variant of the outsourcing model of crowdsourcing, as well as its most popular form. As opposed to Raymond's "bazaar" model of open source development (2000), microtasking represents a return to a "cathedral" participation model where work is organized entirely from the top-down. It is the assembly line of crowd work. Though microtasking does not always imply paid work, it does imply that tasks are defined by someone other than the workers who complete them. Work within these systems is also rigidly schematized, leaving little room for emergence. It is therefore a substantially different crowd work model from open source, and represents a single variant of the much larger crowdsourcing phenomenon.

8.1.2 The Crowd Work Drift: From Emergence to Open Source to Teamsourcing to Microtasking

It is possible to view crowd work along a continuum where open source, virtual teamsourcing, and microtasking are salient positions, marking both types of work and, for some systems and organizations, points in time for an evolving project (Figure 24). While microwork is at one end of this spectrum, open source and teamsourcing have more central positions. At the opposite end lies emergent collective behavior. In a comprehensive literature review, Salminen (2012) determined emergence to be one of three "levels of abstraction" from which to consider collective intelligence, but interestingly, emergence is not specifically addressed in Howe's crowdsourcing descriptions (2006; 2008) or in Malone et al.'s (2009) framework for characterizing collective intelligence—described in detail later in this chapter.

During crisis events, the emergence of new social and organizational configurations arises as a result of disruptive conditions that cause breakdowns to normal routines and open up opportunities for improvisation (Langewiesche, 2002; Kendra & Wachtendorf, 2003b), related to the convergence by large amounts of people, including volunteers, onto the scene (Dynes, 1970; Kendra & Wachtendorf, 2003a). The conditions of mass disruption, therefore, set the stage for emergent collective behavior, a significant feature of many of the examples of crowd work explored in this research.

Aligning with this view, findings for the interviews conducted for Study 2 (Chapter 5) indicate that digital convergers arrive on "the scene" looking for things to do and from there may develop into an emergent response organization. For the voluntweeters, resources were appropriated and activities begun before tasks were defined, and the concept of an organization with a common goal developed after that. There was no specific "call" from one person to a larger community. There were role models, but no single leader or organized leadership preceded the activities that came to define the group. The organization emerged through the interactions of multiple people who converged and were able to connect to each other through social media.

The crowd work defined in Study 4 (Chapter 7), where Twitterers worked to show solidarity with Egyptian protesters and to spread the word of what was going on there, was also emergent. Hundreds of thousands of Twitterers, without any pre-existing organization outside of the affordances and norms of the Twitter platform, functioned as a collaborative filter to identify accounts communicating from "the ground." Some 20,000 Twitterers participated by showing solidarity with the cause through the propagation of a collaborative meme about the revolution. Though not explicitly coordinated, the work was both collective and emergent, as the activity of some influenced the behavior of others, establishing new norms.

These findings show emergence to be an important component of crowd work during disaster. For some crowd work communities, it is also a first phase in their development into other types of organizations. Some organizations and crowd work systems demonstrate a temporal progression along this spectrum (Figure 24) from emergence to open source or teamsourcing and even to microtasking. The reasons behind this progression may echo some of the rationale behind the movement in the HCI research community to conflate crowdsourcing with microtasking.

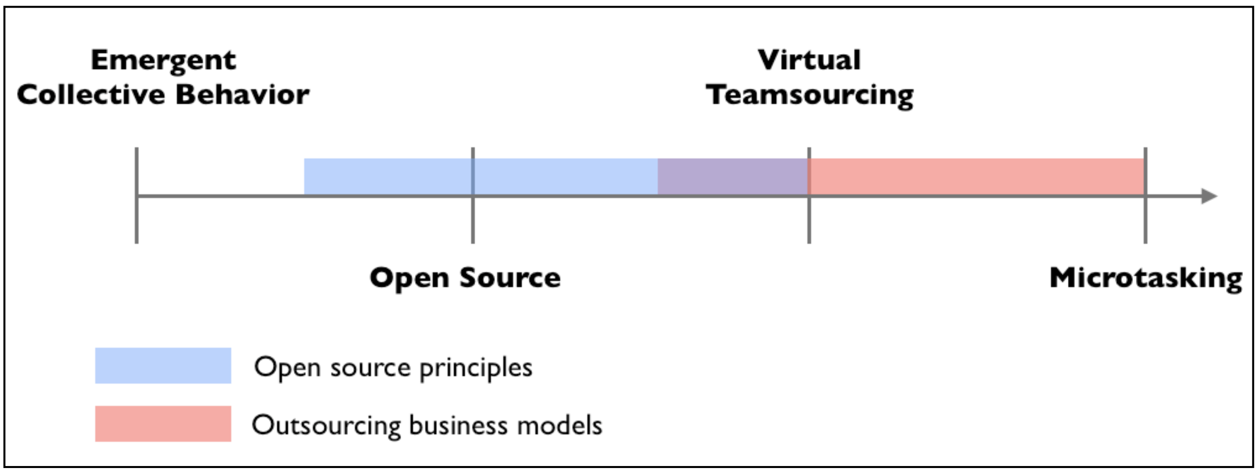


Figure 24. Crowd Work Drifts towards Microtasking

8.1.2.1 Voluntweeters to Humanity Road: Temporal Progression from Emergent Collective Behavior to Open Source

Considering their origins in crisis tweeting during the Iran Election protests and the Haiti earthquake, Humanity Road has progressed from emergent collective behavior to an open source community. As described in Study 2 (Chapter 5) and Study 3 (Chapter 6), the early organizational structure of Humanity Road emerged with the voluntweeters. Many of the original Humanity Road volunteers were voluntweeters, and their organization grew out of those efforts, incorporating many of the routines and norms they had established through collective activity as digital volunteers. Study 2 (Chapter 5) frames the voluntweeters as an emergent organization who progressed, temporally, from resources and activities to tasks and a division of labor. Eventually, some subsections of that group achieved the stage of a "domain"—a collective representation and its reasons for being (Kreps & Bosworth, 1994). As described in Starbird and Palen (2011), Humanity Road represented the culmination—the domain—of the *R→A→T→D* temporal progression of organizing by the voluntweeters. Considered within Dynes (1970) classification scheme for organization types that manifest post-disaster, the voluntweeters were an *emergent* organization and Humanity Road is now an *expanding* organization. Though Humanity Road leaders and volunteers are continuously negotiating who they are and how they do their work, they have—and have defined for themselves—their goals and reasons for being. They maintain an ongoing presence and exhibit increasingly stable rules and routines in their work practice, though there is still opportunity for new structures to emerge, as detailed in Study 3 (Chapter 6). Considered along the continuum for crowd work (Figure 24), the Humanity Road organization now operates as a community of digital volunteers within the open source definition of crowdsourcing, with a strong internal core as well as participation by those at the periphery—spontaneous and episodic volunteers. The organization also has potential application within virtual teamsourcing as a remote organization to whom others "outsource" work.

8.1.2.2 Ushahidi: Temporal Progression from Open Source to Microwork

Another example from the crisis response domain of temporal progression across the crowd work continuum—from emergent collective behavior to microwork—involves the development and use of the Ushahidi crisismapping platform. Ushahidi is an open source platform, originally developed to support citizen reporting of ethnic violence after a contested election in Kenya in 2007 (Okolloh, 2009). Okolloh, herself a Kenyan blogger, explains that as that violence unfolded, she proposed an idea to her "techie" countrymen and others, calling for the creation of a website that would enable the collection of citizen reports of ethnic violence through the web or through SMS messages, and then allow volunteers to filter and map those reports. The original website was built in less than a week by 15-20 developers, the majority of whom were Kenyans. In its first deployment, Ushahidi represented an extremely rapid self-organizing effort by volunteers to connect via social media, create a tool to support their work, and develop complex work processes to maintain their ad hoc group (Okolloh, 2009).

Originally, Ushahidi was very much an open source effort. Following the precise definition of that term, the initial volunteers were software developers who came together to create a publicly available software system. After the first version of the system was deployed, volunteers worked to operate the platform as a crowdsource/open source community whose work was no longer necessarily technical. Their activity was, as Okolloh describes, initially very ad hoc (2009). As is typical of open source communities, they self-organized around a common goal that they defined and re-defined themselves.

Though originally created for a very particular purpose, the technology underlying the Ushahidi platform continued to undergo development after the violence in Kenya eased, and has since been deployed for mapping efforts around other events. Ushahidi was deployed for the first time in a natural disaster event in the early aftermath of the Haiti earthquake, and as that event progressed, the volunteer work organized around and through that platform began to drift from its open source origins towards microwork. Initially launched in conjunction with an SMS shortcode to collect reports from affected people, Ushahidi leveraged various groups of volunteers to identify actionable information, translate it into English and French, geolocate it, and structure it into reports stored within the system. As Patrick Meier, then Director of Crisis Mapping and Partnerships within an established (by that time) Ushahidi organization, explained in his blog (Meier, 2010a), the effort to support the Haiti earthquake with crisismapping using the Ushahidi platform began by leveraging a co-located volunteer workforce of Tufts University students in conjunction with a group of trusted volunteers within Meier's personal social network. Meier termed this effort, *netsourcing*. As the project went on and began to receive media attention, remote volunteers, connected via the Internet, began to participate in the crisis mapping effort, and at that point the collective work became a *crowdsourcing* project (Meier, 2010a).

The work of the crowdsourced volunteers was coordinated through the Ushahidi platform as well as within Skype chatroom and via Twitter messages. The process of collecting and mapping reports evolved to include eleven different teams—e.g. technology, media monitoring, translation, geo-location, reports, etc. Tasks and work practices varied across the groups and many volunteers participated in multiple groups at once. Within the collective work of a single group, activity was often coordinated through an interactive Skype chat where volunteers worked together and supported each other's work. This virtual organization that supported the Ushahidi instance developed routines for organizing their work practices over time and eventually some aspects of work became more regimented. The development of several different groups supported and contributed to the larger work being broken down into smaller tasks. Though the Ushahidi support team was largely self-organizing during its initial deployment for Kenya, organizing became more hierarchical or top-down during the Haiti efforts. However, at the stage described here, it was still an open source/crowdsourcing project. Many volunteers self-deployed, and they often worked on self-selected and/or self-directed tasks. The goal of the organization was self-defined, and the possibility for emergence remained.

Then an interesting thing happened. On March 16, about two months after the Haiti earthquake, Meier proposed moving a large section of the work to a "Turksourcing" model, referencing Amazon's Mechanical Turk platform (Meier, 2010b). To increase the scale of what the crowd could accomplish, Meier suggested disaggregating the media monitoring and geolocation tasks of Ushahidi into human-intelligence tasks (HITs) that could be handled by workers in systems like Mechanical Turk. Later, the rationale changed, and this shift to microwork was advocated with the goal of incorporating Haitians into the Ushahidi crisis mapping efforts, and eventually some portion of the responsibility of processing reports was transferred to paid Haitian workers coordinated through Crowdflower and Samasource (Munro, 2010).

The trajectory described here exemplifies the drift from open source crowd work to microwork. Ushahidi began as a self-organized effort by a group of volunteers, who defined their own goals, then set about working towards them, developing their own platform and work practices along the way. Meier, a leader of Ushahidi during the Haiti response period, attempted to push the work configuration towards a highly regimented (and paid) work environment—for at least some subsection of the work. Possible motivations for this kind of change abound, including scalability, the reason cited by Meier (2010b), as well as efficiency, repeatability, dependability, etc. However, the shift is also problematic for several reasons. First, it leaves behind a great deal of functionality—e.g. collaborative verification behavior that was supported by the interactive work environment. Along those same lines, the potential for emergence in terms of new work practices and new goals is diminished. Additionally, the loss of interactivity and the inclusion of work-for-pay change the motivations for participation, and this may impact work quality.

It is important to point out that the shift proposed by Meier was never fully consummated. The model for using *turksourcing* to support an Ushahidi instance has not been repeated. There have been dozens, perhaps hundreds of deployments of that platform for different events in the last two years and almost all of them have been supported by small communities of volunteers using Skype windows to coordinate.

8.1.2.3 Standby Task Force: Temporal Progression from Open Source to Teamsourcing

The Standby Task Force (SBTF), an outgrowth of volunteer groups that formed to support Ushahidi instances during the response efforts for Haiti and other crisis events that followed, demonstrates a shift from open source crowd work to outsourcing in the form of virtual teamsourcing. The SBTF is a network of volunteers who collaborate to provide live mapping support to organizations involved in humanitarian response, human rights and election monitoring, and media coverage of crisis events[100]. The group, which has an ongoing relationship with the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA), can be "activated" by organizations needing support during mass disruption and/or violent events. During 2011, events spawning activations included natural disasters like the earthquake in Christchurch, NZ and political disruptions like the world witnessed in Libya.

As with the above case of Ushahidi's drift towards microwork, the movement towards a virtual teamsourcing model was also proposed by Ushahidi leader Patrick Meier, this time at the (October) 2010

[100] http://blog.standbytaskforce.com/about/introducing-the-standby-task-force/

International Conference on Crisis Mapping[101]. The rationale for creating the group involved connecting the work of the skilled, virtual volunteers to formal organizations in disaster and humanitarian response, as well as media and election monitoring (Meier, 2010c).

In its creation, the SBTF appropriated many of the existing resources of the volunteer community that had developed to support crisismapping efforts for the 2010 Haiti earthquake and the Pakistan floods in the summer of that same year. The new organization relied on the same tools, work practices, and many of the same volunteers. Many of the organizing structures developed through volunteers' actions during those previous events were carried over into the new instantiation of the volunteer community, including the division of labor into specialized workgroups. Like the Ushahidi crowdsourced efforts that came before it, the SBTF is still largely self-organizing with some hierarchical structure. How the group does the work, who does what, and how tasks are broken down are still determined within the virtual organization, by its volunteers and leaders. The one significant difference between the old form and the new form is who determines what problems volunteers work on and when. For the SBTF, an event sponsor defines the problem for the group to address—i.e. an outside entity requests that the group respond to a particular event. That project initiator can be, for instance, UNOCHA, a media outlet near an affected area, or established leaders of the SBTF. For that latter case, the SBTF is still an open source community of crowd workers albeit with a more hierarchical structure than before, but for the other cases, it has developed a closer resemblance to traditional outsourcing.

8.1.2.4 Tweet the Tweet: Insight from Emergent Microwork

Though the crowd work continuum (Figure 24) provides a framework for understanding much of the crowd work covered in this dissertation research, as well as how certain forms of crowd work evolve into other forms, this model for differentiating between types of crowd work is challenged by the description of Tweak the Tweet translation in Study 1 (Chapter 4). TtT translation can be seen as existing on two

[101] Video and blog available: http://blog.standbytaskforce.com/about/introducing-the-standby-task-force/

sides on the spectrum. In the findings of that study, TtT translation is described as an emergent form of distributed human computation, a type of un-encapsulated microwork. Study 1 (Chapter 4) reports that though Project EPIC researchers distributed thousands of tweets instructing others how to use the convention, and provided many other resources to teach the crisis reporting syntax, the production of TtT translated tweets drove the adoption of the syntax—people learned how to use the syntax by seeing it in use in others' tweets. As users appropriated the TtT microsyntax, they also began to introduce alterations to its use—e.g. introducing new keywords and reporting new categories, and employing the syntax for a different purpose than originally intended. Though the "task" of creating a TtT tweet had prescribed rules, users appropriated the innovation for alternative uses and in many cases adjusted the rules of the syntax to their own work practice. They self-selected their own tasks and many were able to employ these tasks in a way that aligned with their own goals, and not necessarily the goals of the overriding system. For these TtT users, the innovation was a tool that enabled their own emergent work practice, at the far left of the spectrum in Figure 25. However, for a small portion of TtT translators, their use of the microsyntax resembles microtasking, and belongs at the far right of that schematic. Figure 13, an image of the network of TtT translators, shows four Twitterers who translated tweets into the microsyntax but who did not participate in the broader activities of the voluntweeters. These Twitterers can be viewed as accepting TtT translation work as microtasks, following the defined goal and work process laid out by EPIC researchers in some of publicized resources on using the syntax.

The adoption of hashtags had a similar trajectory. Like Tweak the Tweet had with Project EPIC, hashtag use had an early advocate in Chris Messina, who wrote blogs encouraging others to use the technique and offering them rationale for why they should (Messina, 2007a, 2007b). However, the development of norms around hashtag use was largely emergent collective behavior. One could argue that few hashtag users ever saw Messina's blog. Instead, they learned how to use hashtags by seeing others use them, and they did not always use them exactly as Messina had specified or intended.

These examples suggest that emergence is a potential property of all types of crowd work, even microwork, and that though emergence and rigidly structured tasks are somewhat (inversely) correlated, they are not perfectly so. Indeed, the limiting factor for emergence may not be the amount of specificity in the task or how far it has been broken down, but the amount of connectivity between participating individuals and the empowerment of individuals to affect change in their work practices. The most significant difference between self-organizing human computation like Tweak the Tweet and the microtasking of turksourcing may be that the work within the latter platform is rigidly structured and *encapsulated* in a technological platform.

8.1.3 Why Crowd Work Drifts towards Outsourcing and Microwork

The Tweak the Tweet project was intentionally designed to exist within a social network—to diffuse and be used "in the wild." Through dozens of deployments, that innovation has not moved far along the crowd work continuum, and was consciously altered by researchers at Project EPIC during its first deployment to reduce the number and rigidity of rules to support emergent work practices that we were already noticing. However, as the other examples above demonstrate, many crowd work systems and communities do drift towards outsourcing and encapsulated microwork. Part of this may be a natural progression from highly emergent groups to organizations with increasingly rigid rules and routines, an effect of having certain ways of doing things repeatedly enacted, an outcome of structuration (Giddens, 1986).

8.1.3.1 Intentionality and Strategy in Shaping Crowd Work as Outsourcing or Microwork

The above examples also show that shifts along the spectrum can be intentional and strategic, as has twice occurred with work centered on the Ushahidi platform. In the first case, attempting to address the scalability of crisismapping, Meier proposed moving some Ushahidi tasks into a microtasking environment. Later, this shift was also framed as addressing the ethics of humanitarian response work, as proponents claimed it empowered people who were affected by the disaster to participate in their own recovery/relief and to receive financial compensation for this work instead of being preemptively replaced by remote volunteers (Munro, 2010). In the second example, attempting to connect volunteer's crisismapping efforts to formal organizations, Meier and other Ushahidi representatives created the Standby Task Force, essentially moving the open source community of Ushahidi volunteers more towards an outsourcing model. This can be viewed as a strategic and functional decision, as leaders of the Standby Task Force may have hoped to legitimize the work of Ushahidi crisismappers and to make it tangibly and traceably useful. A shift towards an organizational configuration with more rigidly defined work practices and products may make an organization's role more explicit and externally knowable. The rationale provided by Meier for these shifts indicates a belief that this change would help other response organizations tap into the crisismapping community as a resource.

8.1.3.2 What Role does ICT Play?

Other contributing factors in the shift towards encapsulated microwork may be the ICTs appropriated by crowd workers or created to support their work. Software systems can provide a particularly rigid structure for how work is done that is unchangeable except by their developers. Technological systems designed to enable certain kinds of crowd work may codify previous action as strict procedures and push work from ad hoc and improvisational to inflexible configurations where the workers themselves no longer organize their own work. It is hard to know what work practices and functionality of the original work practice Ushahidi volunteers lost when they sanctioned the creation of a platform to help them with their work of collecting and mapping reports of violence in Kenya; the accounts of the development of the platform (e.g. Okolloh, 2009) do not describe in detail how the work was done before they began to use the software. Certainly, the software system enabled them to do new kinds of things. As Giddens reminds us, structures both enable and constrain action (1986). It is easier to contemplate what would have been lost if work to support the Ushahidi platform had moved (in all or part) permanently to the "turksourced" model proposed by Meier (2010c). The highly improvised work practice where the volunteers collaborate within an interactive virtual environment and continuously negotiate what they do and how they do it would have been reduced—in places, at least—to a system where pre-specified tasks were pushed out to unconnected crowd workers through mTurk or another platform. Gains in scalability produced by using a platform to assign microtasks out to the crowd may have changed the kinds of agency that volunteers possessed.

8.2 Existing Frameworks for Characterizing Crowd Work

Figure 24 diagrams a perceived drift by some crowd work communities towards the outsourcing side of the crowd work spectrum, but does not provide a method for classifying all varieties of crowd work. Relying on examples from the studies presented in this research, this section explores other frameworks that have been proposed for characterizing crowd work—Malone et al. (2009) and Quinn and Bederson (2011). Both of these frameworks offer insight into how to think about and differentiate between different configurations of crowd leveraging systems. However, both fall short in significant ways when considered within the frame of this research. For instance, neither framework addresses the role of emergence in crowd work. Drawing on these frameworks, as well as Howe's original definition of crowdsourcing, this analysis serves the purpose of expanding the explanation of crowd work begun in the above section, making explicit some aspects of that model (like motivations) that were previously overlooked and revealing features of crowd work (like emergence, visibility and connectivity) that current frameworks obscure.

8.2.1 Quinn & Bederson's Delineation of the Boundaries of Human Computation

In presenting a taxonomy for classifying systems within the field of human computation, Quinn and Bederson (2011) carve out an understanding of what human computation is and how it differs from other forms of crowd work. They provide a Venn diagram (Quinn & Bederson, 2011a, Figure 1) that delineates between related terms in the crowd work domain—collective intelligence, crowdsourcing, human computation, social computing, and data mining—along with an explanation of how and why they separate the concepts in this way. They appropriately position collective intelligence as an overarching category, in this case encompassing all of crowdsourcing and social computing, most of human computation, and the portion of what they call "data mining" that is collaborative filtering. However, they appear to overlook or under-represent substantial overlap between some of the other terms. For example, in differentiating between human computation and other forms of collective intelligence, they admit a small overlap between crowdsourcing and human computation, but assert that an important difference exists: "Whereas human computation replaces computers with humans, crowdsourcing replaces traditional human workers with members of the public" (p. 3). That demonstrates a far more limited view of crowdsourcing than Howe's (2008), pushing the definition of crowdsourcing towards outsourcing (again). They also locate crowd-driven collaborative creation projects, like Wikipedia and open source, outside the realm of crowdsourcing, and claim that crowdsourcing is separate from collaborative filtering. Both of these categorizations conflict with Howe's (2008) early definition of crowdsourcing.

Evidence within the studies of crowd work presented here demonstrates some of the shortcomings in dividing these concepts in this way. Study 4 (Chapter 7) describes how the collective Twitter activity of the remote crowd worked as a collaborative filter to identify individuals tweeting from the ground during the 2011 Egyptian uprising. It can be argued that the work of these individuals was an emergent form of human computation, in that many people in the crowd set about identifying information coming from the ground of the disaster and then retweeted this information. This is a small task that computers cannot do, but humans can. Quinn and Bederson might term this social computing (from one perspective) or data-mining (from another), depending upon whether the "system" being assessed was the platform that enabled the behavior (i.e. Twitter) or the technology for finding patterns in it (e.g. a collaborative filtering algorithm), a perspective that is dictated by those researchers' focus on the systems and not the behavior. Alternating to a behavior-focused analysis, separations between the related categories of crowd work grow more ambiguous: the remote crowd work of Twitterers during the Egyptian protests functioned as an emergent, collaborative filter through thousands of related, though not explicitly coordinated, acts of human computation.

8.2.2 Malone et al.'s Map of the Genome of Collective Intelligence

Aligning with Quinn and Bederson's positioning of collective intelligence as an umbrella category, Malone et al. (2009) characterize almost all of the examples of crowdsourcing offered by Howe (2006; 2008) as collective intelligence. In classifying these systems further, they identify four building blocks of systems that leverage collective intelligence, asking: 1) *Who* does the work? Is it done within an organizational hierarchy or by the crowd? 2) *Why* are they doing it? What motivates them: money, love, or glory? 3) *What* are they accomplishing? Are they creating something or deciding something? 4) *How* is it being done, independently or dependently?

Their genome model provides a useful mechanism for classifying different systems and for characterizing different sub-processes within complex crowd work systems at a micro level of analysis. As they explain, many crowd work systems have different components that manifest as different "genes"—i.e. different answers to the above four questions. For instance, they describe the Linux open source development community as involving two different gene combinations: in the first, the crowd works to create new software modules through collaboration for motivations of money, love and glory; and in the second, Linus Torvalds and his lieutenants decide which modules to include, using a hierarchical decision-making process, for motivations of love and glory. Applying the genome model to the work of Humanity Road during a disaster event nets three combinations: the leaders and veteran volunteers decide, usually collaboratively, if the group should respond to an event; the crowd (all active members) decides, individually, which pieces of information to bring into the shared window; and, using the information they collect as raw material, the crowd then works through collaboration to create informational resources for affected people. These are not stages, but different crowd work types—according to Malone et al.'s model (2009)—that can all occur at the same time. For all of these, motivations for most volunteers are love, though glory factors in for leaders of the group.

Though being able to classify different components of crowd work in this way is helpful for understanding some aspects of a system or sub-component of a system, this model has some shortcomings. Perhaps the most significant drawback is that the language of genes, while permitting a micro-level description of different components of crowd work, is not easily communicated. Speaking in combinations of genes, especially when describing complex systems or groups that manifest many different gene combinations in their crowd work simultaneously, may not be accessible to possible clients of crowdsourcing work in crisis context. It also may obscure differences between systems or communities. Expanding on these points, the classification for Humanity Road using this scheme required three different sets of genes, and these would almost exactly match those of another virtual volunteer organization, the Standby Task Force, without revealing what is different about those two groups—the former is a more open source model, while the latter has officially moved to an outsource model.

Taking into consideration many of the examples explored in this research of crowd work during mass disruption events, Malone et al.'s genome (2009) also needs more types of genes, as several salient aspects of crowd work in this context are not present in their model. For instance, Malone et al. discuss social networks as collective intelligence, assigning them the genes of independent decisions by crowd members, but they do not address properties or conditions of emergent collective behavior. Additionally, some gene types are inherently related or dependent on others, and these dependencies are overlooked.

Finally, some of Malone et al.'s categories are not fully explicated in a way that allows us to understand how certain crowd work configurations function. For example, their *why* category (motivations) is divided into three different genes: money, love and glory. This represents a good foundation, but there are two problems: first, the glory category is not fully unpacked (which I address later in this chapter); and second, their "love" gene compresses several different types of motivations into a single subcategory, obscuring subtle differences in incentive structures. Malone et al. themselves suggest that there are several different aspects to "love," offering three specifically: entertainment, socializing with others, and contributing to a cause (2009). These three are starkly different things, though, and map to different kinds of crowd work and different systems to support that work.

Several examples in this research demonstrate the implications of these differences. Almost all of the crowd work examined in this research is motivated in all or part by the desire to contribute to a cause. The voluntweeters and members of the Humanity Road organization did and do their work to help people affected by disasters. Remote Twitterers during the 2011 Egypt protests demonstrated solidarity with protesters by identifying information from locals and spreading messages about what was occurring there. For that group, enjoyment or entertainment can also be seen as incentive to participate. Love in the form of socializing with others is important to several of the groups examined in this research, including the Humanity Road volunteers, who can be seen between events in their chatrooms (Study 3, Chapter 6) hanging out, sharing information about themselves, telling jokes, etc. This subtype of the love gene requires that crowd workers can interact with each other and is supported, in the Humanity Road environment, by the interactive environment and the small group size. These examples from the domain of mass disruption demonstrate that the subcomponents of Malone et al.'s love gene, conflated in their model, manifest differently across different crowd work configurations. They also suggest that the type of "love" motivation supported by a platform or community has a substantial effect on who does what—i.e. which crowd workers participate in what kinds of activities.

8.3 Characterizing Crowdwork: Shifting Perspective from Systems to Work Practices

This inquiry reveals shortcomings of existing frameworks for classifying crowd work by demonstrating how they overlook salient features of the productive crowd activity that occurs during and immediately after mass disruption events, including properties of emergence and connectivity, subtleties in motivations, and dependences across features. These noted weaknesses may result from how those studies approached the subject. All three of the frameworks for understanding crowd work examined here—Howe's definition of crowdsourcing (2008), Malone et al.'s genome for collective intelligence (2009) and Quinn & Bederson's schematic for differentiating between human computer and related terms (2011a)—focus on the technological systems and platforms that enable crowd work. For instance when

Malone et al, (2009) classify online social networks—as the crowd making independent decisions—they do so from a perspective of the platform's affordances. Quinn & Bederson's model does a similar thing, classifying collective behavior on social media either as social computing, looking at the platform that enables the behavior, or data mining, looking at the tool that perceives patterns from that behavior. In contrast, this research looks at the communities and the work practice of members of those communities, for which systems and tools are contributing factors, but not the primary foci of analysis. This shift in perspective is what brings emergence to the forefront as a salient feature of crowd work. Emergence is a property of behavior that can be enabled or constrained by features of a system, but it is not a property of the system itself. These efforts to characterize technological systems designed to support and leverage crowd work were therefore methodologically prevented from perceiving emergence as a feature.

Shifting from a systems perspective to one that looks at work practices and socio-technical arrangements sets the stage for a new approach to examining productive work by the connected crowd. I offer a new term, *crowdwork,* to characterize applied collective intelligence by a crowd or group of distributed workers connected to each other directly or indirectly through ICT. Assuming a crowdwork-focused perspective unveils several significant features of collective intelligence in action during mass disruption events.

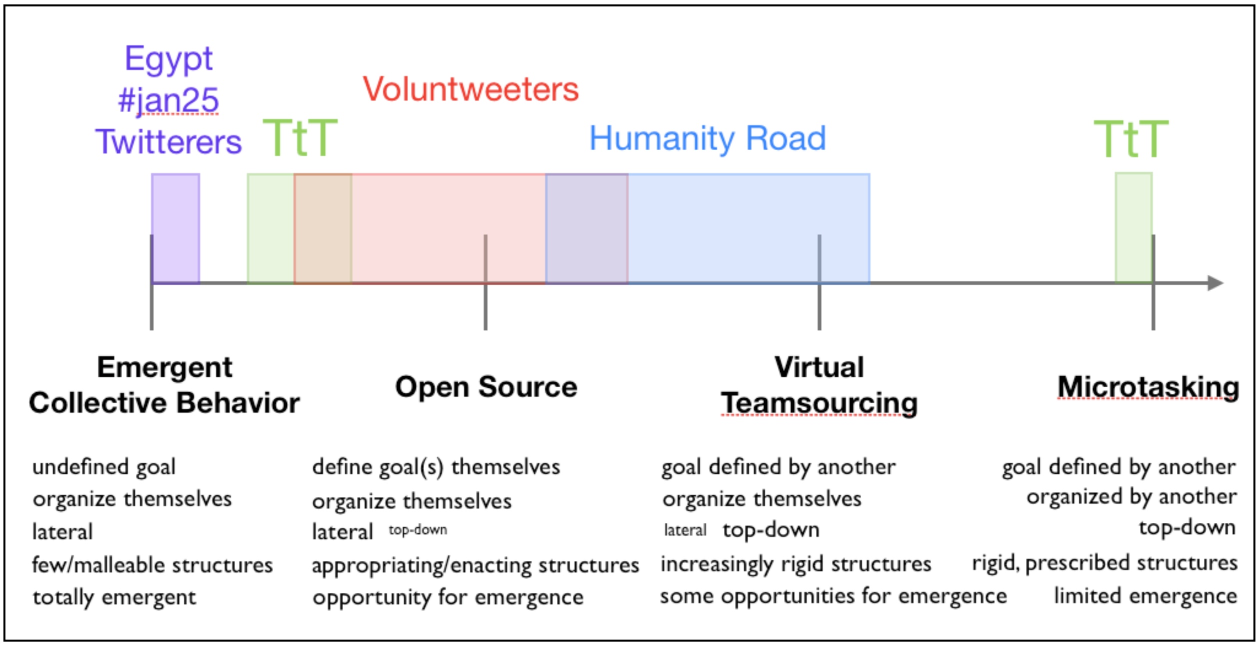


Figure 25. Characterizing Crowdwork during Mass Disruption Events

8.3.1 Reworking Motivations

Though not unique to this domain, benevolence is a major factor in motivating crowdwork during mass disruption events. For the voluntweeters interviewed in Study 2 (Chapter 5), benevolence was cited as an entry point for digital volunteer activities—several reported being motivated by a need to help the people of Haiti after the catastrophic earthquake there. This drive to *contribute to a cause* (Malone et al., 2009) is a catalyst for volunteer action and may be an incentive for continued participation in volunteer activities, but other motivations manifest over time and become increasingly important for sustained participation in volunteer efforts.

Many of these other motivations for crowdwork participation can be mapped to notions of capital: including financial capital, social capital (Bourdieu, 1986; Putnam, 2000; Ellison et al., 2007), cultural capital (Bourdieu, 1986), and symbolic capital (Bourdieu, 1988). Table 12 offers seven categories of motivations for crowdwork along with a related, though not in every case equivalent, category from

Malone et al.'s framework (2009) and a description. In addition to five categories that map to concepts of capital are two others: benevolence and entertainment.

Motivation Category	Malone et al.'s (2009) Category	Definition
Economic capital	Money	For personal financial gain
Social capital - bridging	Love – socializing with others	Increased connections to new people, weak ties Putnam (2000); Ellison et al. (2007)
Social capital - bonding	Love – socializing with others	Increased strength of existing connections Putnam (2000); Ellison et al. (2007)
Benevolence	Love – contribute to a cause	Wholly for the benefit of others
Entertainment	Love – enjoyment	For personal enjoyment
Symbolic capital	Glory	Increased honor, status or reputation (Bourdieu, 1988)
Self-improvement		Embodied cultural capital, knowledge gained through self-improvement (Bourdieu, 1986)

Table 12. Motivations for Participating in Crowdwork

In the crowd work analyzed for this research, benevolence is assumed as at least a partial motivation for most crowdworkers, motivations related to economic capital are minimal, and entertainment, though perhaps a factor for many, is hard to assess. However, several of these other incentive categories can be seen to be important motivators to crowdworkers in the domain of mass disruption.

8.3.1.1 Motivations of Social Capital: Making New Friends and Strengthening Relationships

Love in the form of *socializing with others* can be related to ideas of social capital (Bourdieu, 1986; Putnam, 2000; Ellison et al., 2007). Bourdieu defines social capital as "the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance and recognition – or in other words, to membership in a group" (1986, p. 248). These relationships can be considered within a social network model, with social capital being the amount of resources that can be accessed through connections within the network.

Putnam (2000) divides the idea of social capital into two components: bridging and bonding capital. Bridging capital occurs through building new connections between heterogeneous groups (Putnam, 2000).

Ellison et al. (2007) relate this to online social networks and Granovetter's (1973) concept of "weak ties." In the crowdwork described here, increased bridging capital can be observed in the social network connections generated by voluntweeting activity (Study 2, Chapter 5). The Skype work environments of Humanity Road and the SBTF also allow for the development of bridging capital, as they bring together volunteers from all over the world. Social network technologies that allow users to create new ties are important for building bridging capital.

Bonding capital refers to reinforcing existing connections in close communities (Putnam, 2000; Ellison et al., 2007). Veteran Humanity Road volunteers experience increases in bonding capital during their repeated conversations within the various Skype windows of their virtual environment. The between-event interactions described in Study 3 (Chapter 6) show volunteers engaging in casual chatting, telling jokes and sharing information about themselves. Peer-to-peer communication is important for bonding capital, and is supported within digital volunteer communities by social media tools, including Skype and Twitter.

Distinctions between these two subtypes of social capital are tied to motivations for crowdwork participation. Different configurations of crowdwork enable the development of bridging and bonding capital to differing degrees, and features of the technological platforms developed or appropriated to support these crowdwork activities play an important role in these distinctions.

8.3.1.2 Motivations of Symbolic Capital: Developing a Reputation as a Volunteer

An important aspect of crowdwork incentives, especially those where benevolence plays a role, is what Bourdieu (1998) calls *symbolic capital*:

> "Symbolic capital is any property (any form of capital whether physical, economic, cultural or social) when it is perceived by social agents endowed with categories of perception, which cause them to know it and to recognize it, to give it value. For example, the concept of honor in Mediterranean societies is a typical form of symbolic capital which exists only through repute, that is, through the representation that others have of it to the extent that they share a set of beliefs

liable to cause them to perceive and appreciate certain patterns of conduct as honourable and dishonourable" (1998: 47).

In the framework for motivations being developed here, symbolic capital is assigned to benefits for the participating crowdworker of increased reputation. For instance, developing a reputation as a crisis tweeter may increase the amount of symbolic capital that volunteer possesses within the community or communities that can perceive that reputation. Along these lines, symbolic capital in a digitally-connected social world can be measured in two parts. The first of those is strength of reputation—i.e. how good of a volunteer is she? How skilled? How active? How many events has she responded to? How many communities is she involved in? The second component of symbolic capital is breadth—i.e. how many different people can perceive this reputation? The distinction between the strength and breadth of symbolic capital aligns well with, but is not exactly the same as Putnam's separation of bridging and bonding capital (2000). A critical factor in the development of symbolic capital through crowdwork is the visibility of activity. If other people cannot see the activity, or if it is done anonymously, then symbolic capital does not increase. Online social networks are highly conductive of symbolic capital. Study 1 (Chapter 4) offers the visibility of Tweak the Tweet within the Twitter social network platform as a motivating factor for TtT use. The public tweets of the Haiti voluntweeters and Humanity Road volunteers all serve a primary purpose of sharing crisis response information with the general public, but simultaneously function to show others in their Twitter friend-follower networks that crisis tweeters are actively volunteering, presenting the opportunity for them to receive public recognition for their work. For work performed wholly within the smaller and insular crowdwork communities, like the activity in Humanity Road's Skype chatrooms, the breadth of symbolic capital related to that digital volunteer activity is significantly less, but the strength of their symbolic capital may be greater for those within that smaller group. Symbolic capital may be an unintentional benefit of crowdwork motivated primarily by benevolence; even when not felt or acknowledged as a motivation, gains in symbolic capital accompany acts motivated by benevolence where those acts can be publicly witnessed.

8.3.1.3 Motivations of Self-Improvement: Developing New Skills

A subcomponent of Bourdieu's concept of cultural capital (1986), that of embodied capital in the form of self-improvement or knowledge gained, may also be relevant to crowdwork across domains. Ye and Kishida (2003) claim that learning new skills is a motivation for participating in open source communities. Dow et al. (2012) position learning as a desirable outcome of microtasking work as well.

Task granularity impacts the potential for cultural capital gains in a crowdwork environment. Complex tasks and work practices require multiple skills and can lead workers to develop those skills through their crowdwork activities. Through the course of their volunteer work, Humanity Road members are exposed to new technological tools and platforms—e.g. Skype, Google Documents, Twitter search tools, etc.—and they must become proficient at using these tools to succeed in their volunteer roles. Study 3 (Chapter 6) demonstrates how veteran Humanity Road members help less experienced volunteers by advising them on what tasks need to be done and pointing them to the right tools to complete these tasks. This shows that cultural capital is also tied to connectivity, as social interaction can provide support for learning in crowdwork communities (Dow et al., 2012). Learning opportunities beyond technical skills present as well, e.g. volunteers working to help geolocate reports during event response must understand geographical and cultural features of the affected area, and often this necessitates rapid learning about a part of the world with which the individual was previously unfamiliar. For some kinds of crowdwork, especially crowdwork that is both complex and interactive, self-improvement can be a significant motivating factor.

8.3.1.4 Motivations Vary across Individuals and Time

Motivations for crowdwork vary across platforms, communities and participating individuals. Towards that latter point, the findings of this research show that motivations for a single individual can vary over time. Several of the Humanity Road volunteers began their digital volunteerism as crisis tweeters with the primary goal of helping out during disaster events, motivated by what Malone et al. (2009) describe as *love* in the form of *contributing to a cause*. Over time, volunteers, e.g. members of

Humanity Road's interactive community, may begin to reap other rewards as well, such as building friendships and maintaining public identities as crisis tweeters. These secondary motivations may be vital for sustaining crowdwork participation over time.

8.3.1.5 Motivations Align with Other Crowdwork Features

As revealed through the examples above, different crowdwork configurations rely on different kinds of incentives. These incentives often align with other features of the platform or community, including work ownership and agency, connectivity, and visibility.

8.3.2 Emergence

Emergence is a property of some collectively intelligent crowdwork. The work of the voluntweeters, described in Study 2 (Chapter 5), was nearly totally emergent. That network of volunteers came together after the Haiti earthquake and, leveraging some previous experiences of network members who had participated in crisis tweeting before, developed the beginnings of a commonly understood work practice as they attempted to help identify, verify and route important information. The collective behavior of the remote Twitterers who participated in the Egyptian protests by tweeting and retweeting what they deemed to be valuable information was also highly emergent, and again, like that of the voluntweeters, manifested through interactions across a large, loosely connected network during an international response to a single event. Organized crowdwork that persists over time can also have emergent properties, as shown in Study 3 within the description of the Humanity Road community adopting a new work practice during an event response. Emergence is a property that manifests, not something that a community or system *has*, but it can be supported—or not supported—by the design of a crowdwork system or organization. Like motivations, emergence is also intricately tied to other features of a crowdwork environment, including connectivity and visibility.

8.3.3 Progression of Participation

Examples from the studies on digital volunteerism, Studies 1-3 (Chapters 4-6) within this dissertation, show digital volunteers progressing from initially passive roles, to simple tasks like retweeting important information, and on to more complex and collaborative activities, including taking positions of increased responsibility. Study 3 (Chapter 6) describes the integration of new volunteers into Humanity Road's work processes as legitimate peripheral participation (Lave & Wegner, 1991), and compares this organization to an open source community. Extending the view of this digital volunteer activity back to its roots in voluntweeting, this movement of convergers from passive observers, to simple amplifiers to voluntweeters and then members and leaders of virtual volunteer organizations also follows the *reader-to-leader* framework offered by Preece and Shneiderman (2009) to characterize individual progressions within technology-mediated social participation. Activities in the crisis context, because they are tied to the roles of participants as affected people, formal volunteers, spontaneous volunteers, or responders, may not map precisely to this framework—e.g. affected people may be more likely to be readers and contributors than members of the remote crowd, who will often be readers, collaborators and/or leaders. However, among digital volunteers and remote crowdworkers, we do see this kind of progression taking place, and, as Preece and Shneiderman explain, it is also closely tied to motivations (2009).

8.3.4 Visibility

Visibility is an important feature of crowdwork configurations that relates to both motivations and properties of emergence. Visibility implies that other members in a crowdwork environment and/or the surrounding public can see the work activities of the crowdworkers. Visibility enables new norms to develop and spread. Collective behavior cannot emerge without some form visibility—i.e. participants have to be able to see each other's actions in some way in order to influence the behavior of others, a prerequisite for emergence.

Visibility in crowdwork can be examined as a combination of three factors. The first of these asks if the activity is public or otherwise visible to others. If crowdwork is invisible, symbolic capital will not be an incentive to work. If visible, a second classification assesses to whom the work is visible: the larger public, other workers within the crowdwork environment, or some subgroup of fellow crowdworkers? For instance, the work of TtT translators and Egypt retweeters was visible to anyone monitoring the Twitter platform during those events, while the work that occurs within the Skype chatrooms of Humanity Road is visible only to other members of that community. These visibility categorizations affect the types of motivations that come into play, the potential for emergent work practices and other structures, and how work is organized.

A third factor looks at whether the crowdwork is done anonymously, pseudonymously (Pfitzman & Köhntopp, 2001) or in a way in which the worker's true identity is known. The connection between the work provided and an ongoing reputation of the crowdworker, whether pseudonymous or known, can have a positive effect on the quality of work in communities where that is a value, and can also be used to assess the quality of work (Kane, 2011). At the same time, anonymity obstructs the movement of symbolic capital and will therefore have an impact on motivations.

There is a downside to visibility and that is that it violates one of Surowiecki's (2005) three principles of Wisdom of Crowds—independence. Crowd work configurations that are designed to take an average across a population or rely on independence to encourage creativity rather than convergence, may not want the activities of some workers to be visible to others, at least not in real-time.

8.3.5 Connectivity

Connectivity can be thought of as a combination of visibility plus feedback. Though crowdwork necessarily implies that the work is connected either through or to a system or within a community, the connectivity referred to here is that between individuals—i.e. does the system enable crowdworkers to interact with each other? For all of the crowdwork configurations analyzed within this research, participants could interact with each other through the systems appropriated for their work. Massive amounts of people interacting together can give rise to patterns of collective behavior, like those observed in the remote Twitter crowd during the Egypt protests (Study 4, Chapter 7). In smaller groups, like Humanity Road in Study 3 (Chapter 6), connective technology can allow crowdworkers to coordinate and negotiate their activity. The ad hoc improvisation of the voluntweeters (Study 2, Chapter 5) was supported by technological tools that allowed them to connect, form a group, and develop and pass along norms (e.g. verifying information). Convergence, a property of social behavior during mass disruption events, is related to emergence, a property of crowdwork systems and communities, and this relationship is mediated by connectivity.

Connectivity, as a feature of crowdwork configurations, is also related to motivations. Connectivity can support the development of social capital, of both the bridging and bonding varieties, symbolic capital, and embodied cultural capital in the form of increased knowledge or new skills. Crowdwork that takes place within massively interactive spaces like Twitter tends to best support bridging and symbolic capital, while crowdwork within smaller groups like Humanity Road encourages the development of social bonding capital, though all four motivations are to some extent present in both types of connective crowdwork environments. For instance, for the voluntweeters in Study 2 (Chapter 5), the Twitter platform supported the creation and strengthening of relationships (bonding capital) and allowed experienced volunteers to transfer knowledge to new volunteers (self-improvement). Connectivity encourages learning, as volunteers can support others in gaining new knowledge or new skills. Microwork environments that support little or no connectivity between participants must rely on motivations other than social capital and embodied cultural capital development, though they may be able to confer symbolic capital through another system feature: visibility.

8.3.6 Ownership of the Work Products and Agency within the Work Practices

Two other important and interrelated features of crowdwork systems and communities relate to who owns the work products and who organizes the work. These resemble Malone et al.'s (2009) categories of *Who does the work? (Crowd or hierarchy)* and *How is it organized? (Laterally or top-down)*, but this is not a perfect mapping. The distinctions drawn here are: 1) *Who determines what work is done and what the products of that work are?* and 2) *Who determines how that work is done?* In answering this latter question, the "determination" of how something is done is not necessarily made before action is begun. In fact, the ways in which a crowdwork community works can be changed continuously, as Study 3 (Chapter 6) shows. The question here addresses who has the power to enact those changes, whether from the outside the community, inside the community from its leaders (top-down), or inside the community from the members themselves (laterally).

Table 13 demonstrates how different combinations of answers to these two questions can be used to differentiate between the four separate crowdwork types on the continuum in Figure 24.

Crowdwork Type	**Who determines *what* is done?**	**Who determines *how* it is done?**
Emergent Collective Behavior	Crowdworkers	Crowdworkers - laterally
Open Source	Crowdworkers	Crowdworkers – laterally and top-down
Teamsource	Project Owner	Crowdworkers – laterally and top-down
Microtasking	Project Owner	Project Owner

Table 13. Classifying Crowdwork by Ownership and Agency

8.4 A Holistic View of Crowdwork during Mass Disruption Events: A Layered System Integrating Human and Machine Computation

Moving beyond the characterization of different crowdwork configurations, the final half of this chapter takes a step back to look holistically at the multiple, interacting forms of crowdwork taking place during mass disruption events and conceiving of this as a single, collaborative, integrated system that works to process information during these events.

The convergence behavior of the now highly-connected public during crises and other mass disruption events provides both a challenge and an opportunity. Citizen reporters from affected areas along with digital convergers from remote locales meet in shared media spaces to seek information, to share information, to collectively make sense of the event (Palen & Liu, 2007, Hughes et al., 2008; Shklovski et a., 2008), and to attempt to provide assistance to responders and affected individuals (Qu et al., 2011). This activity generates and circulates huge amounts of information during and after large-scale events. As the studies presented here show, the collective activity that produces this information also has the power to help organize this information through a variety of different mechanisms I characterize as *crowdwork*. While no single crowdworker, virtual organization, or response organization can be expected to make sense of the entirety of public, online communications during a large-impact event, the collective work of multiple crowdwork communities can perform as a massive filter and information-organizer. It is possible to consider (much of) this activity as a large, layered system, with different types of crowdwork plugging in at different stages of an information-processing problem. In presenting a system designed to allow forum members to collectively organize and synthesize information within their discussions, Nam and Ackerman (2007) introduced the concepts of *incremental diagenesis* and *incremental summarization*, describing how participants could gradually transform their space to increased levels of order. Both of these processes can be seen occurring occuring organically within crowdwork in the crisis context. Though, as this research describes, much of this work is emergent, as we learn more about how people come together to act on and organize this information, we may be able to devise strategies for leveraging and supporting this collective crowdwork with computational tools.

Figure 26 presents a diagram, based on this dissertation research, of this layered model of how crowdworkers and crowdwork communities, supported by computational platforms and tools, help to organize information during mass disruption events. This model also includes rationale for how human crowdworkers and machine computation tools could work together to improve this filtration process.

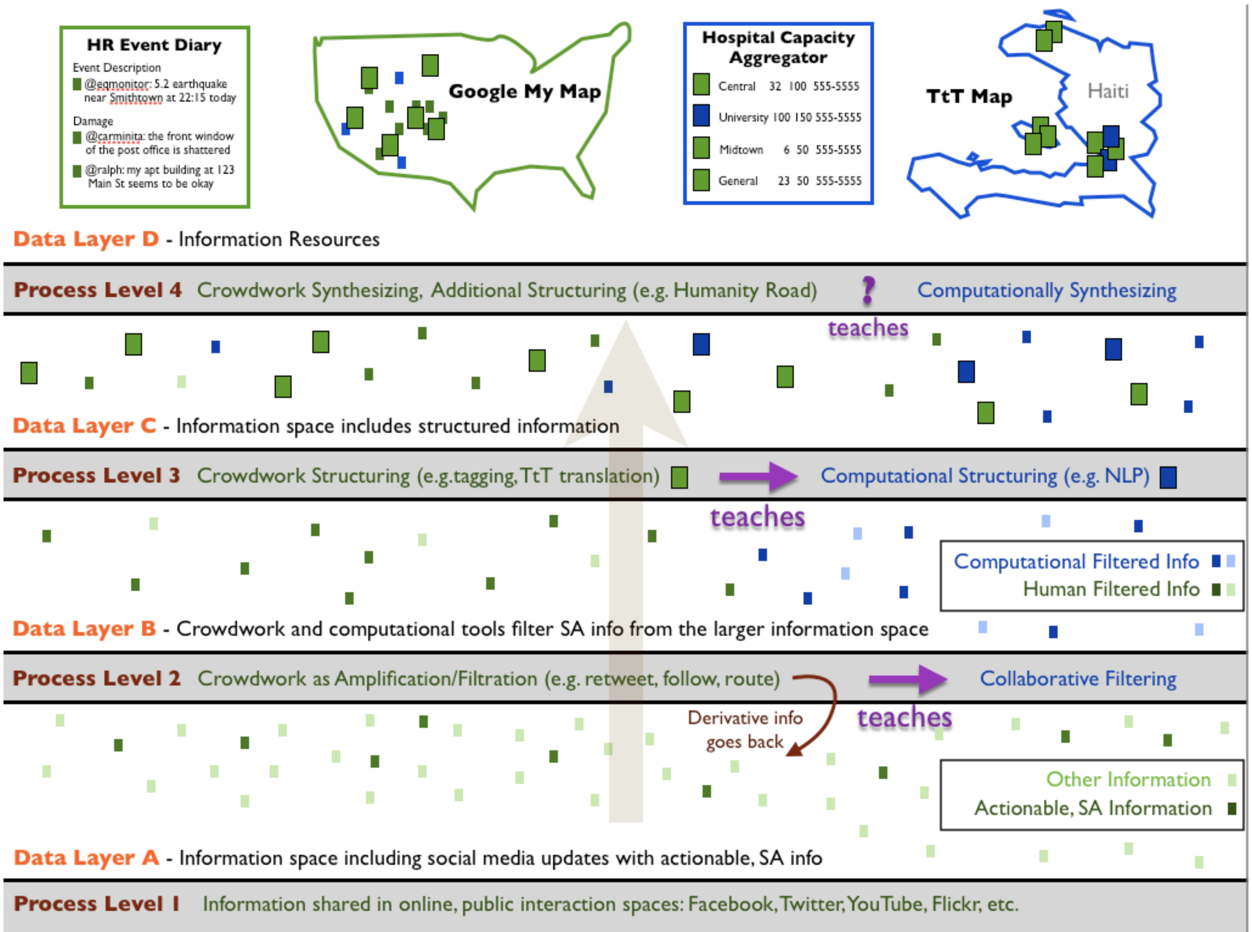


Figure 26. Information Organizing during Mass Disruption Events by Crowdworkers & Computational Systems

This schematic divides the work of information processing into *Processing Levels* (1-4) and *Data Layers* (A-D). The Processing Levels represent different types of processing work (e.g. filtering, structuring, synthesizing) with some separation between crowd-powered work (green, left side) and purely machine-computational work (blue, right side). For this model, the machine computation designation (blue) is assigned to work that relies wholly on artificially intelligent processes, like machine learning algorithms, though these can and do learn indirectly from human behavior. The crowdwork processes here (green) can be supported by computational tools and systems (e.g. Twitter search tools), but require directly tapping the cognitive work of human beings. Different types of crowdwork plug into different levels of this filtration model.

Though the products of all this processing activity in fact exist together within the shared media spaces of social media and the surrounding Internet, the Data Layers in this model are conceptual states of the information space. Each layer in this model contains the products of earlier processing activities, which act as material for subsequent Processing Levels. This is the data that feeds into the next level of the filtration model, though some processes can work on data from any layer below—e.g. Humanity Road volunteers working to synthesize information into resources at Process Level 4 can manually comb through social media data at Layer A. In Figure 26, the products of purely machine computational processing appear in blue in the Data Layer directly above them, and the products of crowdwork activity appear in green in the next level.

8.4.1 Process Level 1: Information Sharing via Social Media

Here is an explanation of how this model works. At the bottom (Process Level 1), people put information into the media space through media articles, blogs, tweets, Facebook posts, etc. This information-sharing includes the citizen reporting variety of crowdwork, though a large portion of the information is "derivative"—new information generated by acting on existing information in the surrounding information space, which includes social media platforms and the Internet around them (Starbird et al., 2010). From previous research (Starbird et al., 2010; Vieweg et al., 2010) as well as the studies described in this dissertation, we know that only a small amount of this information is both new to the space, on-topic, and either actionable or information that would otherwise contribute to situational awareness. The activities of digital volunteers are often focused around finding this information that could be *useful* or potentially useful to affected people, responders, or other volunteers. Data Layer A in Figure 26 illustrates the presence of useful information within a noisy interaction space by marking it as small, dark green/dark blue squares and displaying other information as light green/light blue. In this model, information takes on the blue hue after a computational algorithm processes it.

8.4.2 Process Level 2: Creating Derivative Information through Amplifying, Filtering, and Routing

People act on the information available in Data Layer A, essentially the social computing space, by reposting or retweeting information, routing information, posting links to information and creating following connections within social networking sites. At this processing level (2) on the left side, crowdworkers are, in many cases, acting in and on social media. At Process Level 2, the remote Twitterers "working" during the Egyptian protests were following good sources and retweeting information from on-the-ground Twitterers—as well as others. Many voluntweeters began their digital volunteerism with Process Level 2 activity, by retweeting and otherwise amplifying information. Workers at this level can be intentionally or unintentionally trying to help. Motivations of entertainment, benevolence, social bridging capital and symbolic capital are all at play within this level of crowdwork.

This work can function directly to filter information on to other levels of crowdwork (dark green squares at Data Layer B). Study 1 (Chapter 4) explains how manually filtered tweets passed on as retweets from recognized digital volunteers served as the material for TtT translators to structure (a processing activity that occurs at Level 3). Level 2 processing activity can also function indirectly as human computation data to feed into machine learning algorithms (suggested by the purple arrow in the schematic). Even as it acts to amplify and filter information, this "work" also creates more information—derivative information (Starbird et al., 2010)—and the brown arrow at Process Level 2 demonstrates how activity here, while performing as a filter in some cases for upper Data Layers, also feeds back into Data Layer A. Collaborative filtering techniques, the computational processing side of Processing Level 2, can infer patterns from the social media activity at Data Layer A, essentially a blend of citizen reporting and noise (which includes the derivative information), and can use those patterns to filter information computationally. Figure 26 shows the products of computational processing at Level 2 as blue squares in Data Layer B. Dark blue is desirable information, and light blue represents information that is not useful—allowing for some error in algorithmic processing. Within Data Layer B there is higher noise (more light colored squares) among the products of computational processing techniques than of crowdwork processing techniques. Manually processing by a human filterer may be more accurate than machine filters. In another study related to this research (Starbird, Muzny & Palen, 2012), we report that computational filtering techniques may only be partially successful and that the products of these tools should be distributed to human decision makers to further filter or classify.

8.4.3 Process Level 3: Structuring Information into Standard Formats

The next processing level (3) is structuring, though it can occur simultaneously with other processing activity, and can pull information either from Data Layer A, by taking data directly from social media or other information space, or Data Layer B, by monitoring products of other crowdwork systems or communities. Structuring by crowdworkers is an intentional activity of re-contextualization, whereby individuals categorize data, change its format, geo-locate it, etc. TtT translators operated on Process Level 3 by "translating" information (from Data Layers A or B) into a prescribed microsyntax. This added structure allowed computational tools to automatically process that information into records that could be searched and sorted across several different categories. Simply adding metadata to information, as voluntweeters from Study 2 (Chapter 5) did when they added the `#haiti` or the `#rescuemehaiti` tag to tweets before retweeting them, is structuring activity that would take place at Process Level 3 in this model as well. This type of processing also includes the work of Ushahidi volunteers to create "reports" from social media updates and other media activity, structuring them by entering them into a web form, accompanied by GPS coordinates for the report when available.

This research indicates that crowdwork at this processing level varies across almost all of the salient features of crowd work outlined in Section 8.3, including motivations, visibility, connectivity, features of emergence, work product ownership, and agency. TtT translation was an emergent activity that took place publicly within a social media environment. Formatting information into Ushahidi reports is a private activity that occurs through the Ushahidi crowdsourcing platform, though communities have arisen to support this activity—e.g. some of the voluntweeters from Study 2 (Chapter 5) worked together to process Ushahidi reports, using both Twitter and Skype.

Currently, crowdworkers use their own social networks and simple search tools with simple filters (geographic, keyword terms, limiting to RTs or non-RTs, etc.) to find information to work on for structuring activities. It is possible that they could be fed filtered information by collaborative filtering tools designed to support their activities (right side, Process Level 2, results in blue). Computational tools also could be plugged in to process information at this level, taking as their raw material information filtered at Process Level 2 by computational tools and/or crowdworkers, using machine learning techniques, like NLP (Verma et al., 2011) to categorize information and structure it into databases or redistribute it in a structured format or with added metadata.

For the processing work on both Level 2 and Level 3, this model envisions a blend of human and computational processes, tightly integrated, where the crowdwork (on the left side of Figure 26) feeds the machine computation (on the right). The purple arrows in Figure 26 demonstrate opportunities for feeding the results of crowdwork activity as classification data into machine learning algorithms. In this way, machine-learning algorithms can learn from how crowdworkers "work": learning who might be on the ground from who is being retweeted (Study 4, Chapter 7), learning how to categorize/structure information by inferring how volunteers are manually categorizing/structuring that information, etc.

In terms of how the products from work at Level 3 are made available to the public at the next level, some structuring activity, like the creation of Ushahidi reports, is essentially private and goes directly into a platform designed to facilitate the work. However, this information can be made publicly available—e.g. Ushahidi provides public access to crisismapped reports in spreadsheets (CSV or XLS) formats[102]. The structuring activity of hashtagging and TtT translation is done publicly, with the products of that work remaining in the larger social media space (the Twitter platform) where other people can access them manually and other tools can collect them using the Twitter Search API. Data Layer 4 therefore includes a collaboratively filtered and structured social media space as well as structured data feeds—e.g. RSS and GeoRSS feeds, CSV files, public databases, etc.

[102] http://community.ushahidi.com/research/datasets/

8.4.4 Process Level 4: Synthesizing Information into Resources

The final process level (4) features work to take material from previous data levels and synthesize it into resources for affected people. At this level of the schematic, resources outlined in green are synthesized by human workers and resources outlined in blue are generated by computational tools that pull in structured information—information structured either by humans or computational tools at previous process levels or within a platform that combines structuring and synthesis in a single step.

On the human-powered side of this equation lies the work of Humanity Road volunteers to build "Event Diary" pages by synthesizing information from multiple sources, as well as the work of crisismappers who use platforms like Google Maps to manually create maps of evolving events. Though supported by the platforms that enable them to assemble and publish the information and tools like Twitter Search, Tweet Tracker, etc. that help them to collect it, the synthesis activity for HR workers and Google crisismappers is currently a primarily manual activity. However, it could be possible to use machine learning tools at Process Levels 1, 2 and 3 to feed data into this collective synthesis activity.

On the computational side, there is opportunity for developing applications that pull structured information into synthesized resources. Currently, many of the computationally-synthesized resources used by digital volunteers and others in the crisis domain rely on human-structured data. TtT, which collects and maps human-structured tweets (from earlier process levels), is an example of that strategy. Ushahidi also relies on human-structured information, but combines the work of structuring and synthesizing (Process Levels 3 and 4) into a single step within its platform. Though that strategy has been effective for organizing the work of digital volunteers in a single site, structured information made public separate from the platform that collects it opens up possibilities for more creative delivery of the aggregated information—essentially allowing the development of resources to be "crowdsourced" as well.

8.4.5 Where the Filtration Model Falls Short

This model offers an explanation for how the connected crowd functions as a massive information filtration and processing system during mass disruption events, relying on multiple types of crowdwork and computational tools to build resources like Humanity Road's Event Diaries, Ushahidi's crowdmaps, Tweak the Tweet maps, etc. It also highlights opportunities for the development of tools and systems to support and leverage crowdwork and to better meet the goal of creating resources for responders and affected people during these events.

However, the model has significant drawbacks. As revealed through the findings in the studies presented here, crowdwork during mass disruption events is more than filtration and resource building. The model glosses over a major contribution of the crowdwork of the voluntweeters, many of who perform in real-time to move information to people who need it. What happens on the other side of this model? This system, consisting of crowdwork and tools, works to create a resource, perhaps a map or a single report of a hospital that has available beds. But how does it connect this information to the people who can use it? These questions lead to a final perspective on crowdwork, conceiving of the combined work of connected people and their tools as collective intelligence understood through the lens of distributed cognition.

8.5 A Distributed Cognition View of Collective Intelligence

Existing frameworks for characterizing crowd work, including Malone et al.'s (2009) and Quinn & Bederson's (2011), approach crowd work from a perspective that concentrates on the configurations of the systems that shape the activity. A large portion of the analysis presented in this work takes an alternative, yet related tack, positioning human behavior in the form of organizing as the focal point of the analysis. For instance, Study 2 (Chapter 5) characterizes the work of voluntweeters as an emergent response organization, using the literature on the sociology of disaster to frame that discussion, and Study 3 (Chapter 6) uses structuration theory as a lens for understanding the activity of Humanity Road's digital volunteers. In this section, I shift to a perspective that focuses instead on information, looking at how it moves and changes, and how it is moved and is changed, and seeing collective intelligence in those transformations. This section examines crowdwork as the movement of information and frames collective intelligence as distributed cognition.

8.5.1 An Information-Centric Perspective

The various activities that constitute crowd work during mass disruption events as described in the studies presented here—i.e. citizen reporting, recommending, filtering, verifying, structuring, mapping, integrating, synthesizing, routing—can all be viewed as serving one particular type of task, organizing information. It is possible to approach the analysis of this information-organizing activity as a large information processing system, where the "organization" is the crowd itself, and the glue that holds it together is information. In Studies 1 and 2 (Chapters 4 and 5), I describe the crowdwork of TtT translators and voluntweeters as purposeful activity by digital volunteers that contributed to increased information organization. Studies 2 and 3 (Chapters 5 and 6) demonstrate how groups of crowdworkers, like the voluntweeters and Humanity Road, come together and self-organize, using various tasks of information processing to shape their work processes. Study 4 (Chapter 7) demonstrates how even "passive" or natural activity within the social media space helps to organize the information. In all cases, human beings are converging via ICTs and performing tasks of information processing—tasks whose products could be useful to people affected by and responding to crises. For crowdwork during mass disruption events, both the source and the site of the action is information, which serves to organize the activities around it.

> "The historian Brian Stock has shown how, from the eleventh century on, the spread of the written word and literacy together allowed "textual communities," particularly religious communities, to organize themselves. The most distinctive of these were groups of heretics or reformers who organized around new interpretations of sacred text, developing new ideas of how to live. These new groups, Stock argues, were important 'laboratories of social organization.'" (Brown & Duguid, 2002, p. 192)

Brown and Duguid recount Stock's analysis of how information and shared discourse helped to evolve new idea-based communities in the eleventh century, and relate that to the role of information in our connected society. This dissertation's description of the activities of crowdworkers during mass disruption events demonstrates how social media act in this same way, as conduits for information, and as sites of new opportunities for organizing around that information. Rather than placing the technological platforms (or the human behavior) at center stage, the alternative perspective explored next considers information and information organizing as the central components of crowdwork.

8.5.2 Framing Collective Intelligence as Distributed Cognition

The theory of distributed cognition (Rogers & Ellis, 1994; Hutchins, 1995; Hollan et al, 2000) provides a methodological framework for examining connected crowd work from an information-centered perspective. Distributed cognition posits cognition not solely within individual minds, but distributed through social systems and technological artifacts. Cognition can be distributed physically (and digitally) across artifacts and tools. A simple example of this is the intentional off-loading of cognition into the physical environment by taking notes or visually embedding a memorized speech into physical space by mapping sections of the content onto objects. More complex examples of distributed cognition demonstrate how cognition happens through interactions between people and resources—e.g. one's understanding of how a tool works develops through manipulating it and interpreting the results of those manipulations. Physical and digital tools and artifacts are all components of distributed cognition. Cognition can also be distributed socially, across different people in a group. Just as cognition within the brain exists as movements of signals across synapses between neurons, socially distributed cognition exists as the movement of information between nodes across network connections.

Though presented as a theory of human cognition, several research studies (e.g. Flor & Hutchins, 1991; Rogers & Ellis, 1994; Hutchins & Klausen, 1996; Hutchins & Palen, 1998) have employed distributed cognition as a methodological framework, within which "the unit of analysis is a culturally constituted functional group rather than an individual mind … (permitting researchers) to describe cognitive processes by tracing the movement of information through a system and characterize the mechanisms of the system which carry out the performance, both on the individual and the group level"

(Hutchins & Klausen, 1996, p. 15) In this way, the cognition of the system occurs as transformations of representations. This model provides an obvious approach for analysis at the level of these transformations: using Hutchins' strategy[103] to ask, "What information goes where, when and in what form?"

8.5.3 Tweak the Tweet: Crowd Filtering and Structuring as Distributed Cognition

Analyzing the collective activity of Tweak the Tweet translators using the framework of distributed cognition provides an interesting analogy between how the emergent social network processes information and how the structures and processes of a single brain produce "thought." This perspective uses as the unit of analysis *the crowd*—i.e. the entire network of event-related Twitterers in combination with their tools, resources, and artifacts—and addresses two parts of Hutchins' question: What information? And in what form?

Focusing on individual transformations, each act of translation into TtT syntax takes place as a combination of first, the identification of a specific kind of information—actionable or otherwise relevant information; and second, the performance of one or more transformations that reconfigure the information into a state of increased structure. The following example demonstrates:

> **@NolaBird** (2011-05-23 00:21:26): RT @TVJoe: Animal shelter established in #Joplin, Missouri. Call (620) 674-3634 if you need animal shelter services.

> **@TheFireTracker2** (2011-05-23 00:21:54): MT @TVJoe: #Offer Animal shelter #Loc Joplin, Missouri. #contact Call (620) 674-3634 if you need animal shelter services. #Joplin

In the early morning hours after the Joplin tornado, a handful of TtT translators, all with previous experience using the syntax, were working to identify tweets with actionable information and translate them into TtT syntax. In the above example, a singular act of TtT translation similar to thousands of

[103] This is an oft-repeated line in Hutchin's classroom lectures on distributed cognition. It also appears: http://hci.ucsd.edu/hutchins/datalink.html

others witnessed across the many events for which TtT was deployed, @TheFireTracker2, a veteran digital volunteer, identifies information that might be useful to affected people and creates a TtT tweet by adding three TtT tags (`#offer`, `#loc`, and `#contact`) and reorganizing the tweet text to match the prescribed format and conform to the 140-character limit. The consistent word order between the original and the translation suggests a copy and paste strategy. @TheFireTracker2 uses “MT” instead of “RT” at the beginning of the tweet to indicate a “modified tweet,” a strategy adopted by TtT translators to both give attribution to an upstream author and acknowledge that the tweet text has been changed. Often, TtT translators create modified tweets like these by initially generating a retweet using Twitter functionality and then editing the text accordingly before sending it out. For practiced TtT translators, using a series of small micro-structuring actions like adding a few tags and copying-pasting pieces of text, it can take several seconds to few minutes to generate a tweet like this. The results of these transformations are new information representations that are later circulated through the digital communication system.

Few cases of TtT translation are as simple as the base translation case: TtT translators recognizing information entering the system for the first time and rapidly translating it into the format. Information most often takes a much less direct route from original source to structured state. The following example demonstrates this effect, revealing an important role that the larger network plays in relaying the signal through the information space until it reaches a node with the potential to process it.

> **@ally123** (2011-05-23 01:15:33): #Mo #Joplin The emergency animal hospital on 7th and Illinois behind Sonic in the old Sears Plaza is taking an… (cont) http://deck.ly/~abc12

> **@medmrsmith** (2011-05-23 01:22:01): #joplin Emergency Animal Hosp on 7th and Illinois behind Sonic is OPEN and taking found animals

> **@ktkeyboard** (2011-05-23 01:25:00): RT @medmrsmith: #joplin Emergency Animal Hosp on 7th and Illinois behind Sonic is OPEN and taking found animals

@Jeannie_Hartley (2011-05-23 01:37:12): RT @ktkeyboard: #joplin Emergency Animal Hosp on 7th and Illinois behind Sonic is OPEN and taking found animals @franyafranya

@maryslate (2011-05-23 01:42:57): RT @Jeannie_Hartley: RT @ktkeyboard: #joplin Emergency Animal Hosp on 7th and Illinois behind Sonic is OPEN and taking found animals ...

@SoutholdVOICE[104](2011-05-23 01:47:54): RT @maryslate: RT @ktkeyboard: #joplin Emergency Animal Hosp on 7th and Illinois behind Sonic is OPEN and taking found animals

@CrisisMappers (2011-05-23 02:57:44): #tornado #offer Emergency Animal Hosp behind Sonic is OPEN & taking found animals #loc 7th & Illinois #Joplin #src @medmrsmith @SoutholdVOICE+

In this example, the report of an emergency animal hospital announcing that it will accept animals originally enters the system in @ally123's tweet, though the content exceeds 140-characters and is truncated by a client application that supports tweet-shortening, leaving the remainder of the text accessible, but only through the link. A few minutes later, @medmrsmith tweets out this same report in a much shorter format. A strong overlap with the word order from @ally123's tweet suggests that @medmrsmith's tweet is a shortening of that original report or was drawn from the same source. After this, the information bounces quickly off three more Twitterers who retweet the text and give attribution in each case to an upstream author, though the original author begins to be excluded by the second retweet. Eventually the tweet is translated into TtT syntax by @CrisisMappers, who manages to track down the original author of this variation (@medmrsmith) and cites that account along with the author immediately upstream (@SoutholdVOICE).

Here the "cognition" is two-fold, measurable in both the form of the information and the movement of information across different nodes in the network. Distributed cognition as transformations of information format occurs through @CrisisMappers' TtT translation at the end of this tweet excerpt.

[104] This account is now @SoutholdSMEM, operated by @joannalane. Cited with permission for direct attribution.

Cognition was also happening in the actions above, through the many retweets whereby nodes in the network acted to pass on the information. This activity is an important feature of a collaborative information-processing system where different nodes have different information-processing capabilities. Tweets have a very short half-life (Kwak et al., 2010) and are essentially ephemeral, but retweets and other derivative tweets serve to keep information alive by repeatedly posting it to the public stream (Starbird et al., 2010). Though the intermediary accounts might not have had the expertise or awareness of TtT translation, their work to rebroadcast information can be seen to increase the odds that it will reach the account of someone who can use it. In this case, eventually the report came to the attention of @CrisisMappers, a user who had previous experience translating tweets, and it was subsequently structured into TtT syntax. The tweet may have also been noticed by Twitterers who could act on the information in other ways, perhaps even people who were seeking a place to take found animals, though that use is not clear from the data we have.

At times, TtT translation also occurs in stages, as structure is added incrementally and digital volunteers pull multiple pieces of information into a tweet:

@SteveHaiti (2010-01-13 15:46:17): Hopital Sacre Coeur in Milot, #Haiti is open and available to provide medical service to the injured. http://bit.ly/6cnaDg #haitiquake

@Haiti123 (2010-01-15 <16:00:00): contact Angela Jones Crudem Support - Hopital Sacre Coeur in Milot, Haiti (904) 555-5555 ajones@anon.net

@rqskye (2010-01-16 ~10:15:00): SacredHeart Hosp, Milot, has landing pad, staff, supplies for 200 patients NOW. Call C Angela Jones Crudem @904-555-5555 #haiti

@epiccolorado (2010-01-16 12:45:31): #haiti #offering #med #loc SacredHeart Hosp, Milot #num 200 #contact Angela Crudem #contact 904 555 5555 #info has landing pad, staff, suppl

@karlapeter (2010-01-17 18:39:44): #haiti #offering medical teams waiting, 2 ORs, full staff supplied #loc HospSacreCoeur,Milot #contact Angela Jones 904.555.5555 #rescuemehaiti

In the first tweet from this excerpt, @SteveHaiti shares a link to the website of a medical facility in Haiti that is advertising availability for treating injured people. His tweet, which uses one of the less popular hashtags for the event (`#haitiquake`), is not widely retweeted. Two days later, this same information begins to circulate more broadly on Twitter and elsewhere, partially driven by efforts of volunteers at CRUDEM, the organization that runs the hospital, to communicate their availability through their website, press releases, blogs, emails and media interviews. In the context of a series of tweets about the hospital, @Haiti2015 sends out a tweet with the contact information for Angela Jones[105], a staff member of the organization who has been desperately trying to get the word out, through multiple channels, about the hospital having capacity to take patients. A day later, @rqskye broadcasts a tweet that synthesizes the other information in the report with the contact information, adding `#haiti` to the end of the tweet to push it into searches for that popular hashtag. Later that day, @epiccolorado takes the information in @rqskye's tweet and produces the first TtT tweet about this report. A day later, @karlapeter follows up—also using TtT syntax—to relay new information that medical teams are on site and have set up two operating rooms. This final tweet from the excerpt also contains the `#rescuemehaiti` hashtag, a specialized hashtag discussed below.

This example reveals structuring as collective, collaborative work, with each Twitterer building upon the work of others. In some cases, structure is added incrementally, one transformation on top of another. Again, distributed cognition can be seen in the chain of retweets and textual transformations that move information from unstructured content on Twitter and elsewhere to tweets structured in TtT syntax. Angela Jones, the woman listed as the contact for Sacre Coeur Hospital, worked tirelessly to communicate the availability of their facility to other responders and volunteers, but she did not use Twitter. Digital volunteers, combing through information available in many forms online, located her report on the hospital's website and in blogs on other websites—many tweets with information about the hospital link to these sources—and then moved that information, manually, onto the Twitter platform.

[105] Name anonymized to protect her identity.

This activity indicates that the boundaries of the cognitive system of TtT translation and digital volunteerism in general extend beyond Twitter into other resources and tools that digital volunteers use in their work.

The `#rescuemehaiti` hashtag, added to the last tweet in the above excerpt, is another way that digital volunteers, though not necessarily TtT translators, worked to structure information during the Haiti response efforts. An interview with a digital volunteer from this event revealed a subgroup of volunteers who created a system where they would tag information with `#rescuemehaiti` to call it to the attention of other digital volunteers and response agencies. Initially, they directed others to use it just for marking tweets with information about people who were trapped in the rubble and needed rescue. Later, they extended the recommended use of the tag to tweets about people needing other kinds of immediate assistance. Members of that sub-community of digital volunteers both tuned in to the stream of `#rescuemehaiti` and monitored the larger `#haiti` tweet stream, adding their specialized tag to tweets containing certain kinds of information. Although `#rescuemehaiti` promoters directed the crowd to use the tag in this way, it was often used by group outsiders to tag tweets with other types of information, including information that was not actionable. This example of user-driven structuring shows that information processing activity similar to, yet distinct from, TtT translation work is occurring "naturally" within digital volunteer communities, introduced by the volunteers themselves.

In fact, user-driven structuring is a common occurrence in mass disruption events. Twitterers introduced and adopted special hashtags for categorizing certain types of response-related information during the Alabama tornados in April 2011 (`#ALneeds` and `#ALhaves`), and passed on variations of those tags and that process to digital volunteers responding to the Joplin tornado a month later (`#MOneeds` and `#MOhaves`).

> **@jsandford** (2011-05-03 19:27:35): Please use #ALHaves or #ALNeeds if you have or need supplies in Alabama storm recovery to help better org tweets. #WeAreAlabama @spann

@two_slice (2011-05-06 20:55:30): Theres a need for doctors to help National Guard soldiers with meds and such, in Greensboro. #WeAreAlabama #ALneeds

…

@jsandford (before collection began at 2011-05-23 00:10:55): Encouraging #joplin to use #MOHaves and #MONeeds to classify where people have or need supplies. #mowx #WeAreAlabama

@cattycate (2011-05-24 13:14:18): #Joplin College Heights Church Distribut Center needs FOOD, Paper Plates, Canned Goods & duct tape! located at 4311 E. Newman #moneeds

User-driven tweet structuring can be seen as a self-organizing human computation system providing real-time information processing. During mass disruption events, groups of self-organizing volunteers self-deploy and work to monitor the incoming tweet stream and mark up information according to a set of rules, often suggested and taught by other volunteers. During the aftermath of the Joplin tornado, this behavior began soon after the event. @jsandford sent his tweet recommending the use of `#MOhaves` and `#MOneeds` (the third tweet in the excerpt above) a few hours after the initial impact of the tornado. We deployed Tweak the Tweet for that event a little while later, and TtT coverage for Joplin stretched out over several weeks, with dozens of volunteers working together to identify important information and tag it with the `#MOneeds` convention or translate it into the TtT syntax.

A similar network developed during the relief period of the Haiti earthquake a year earlier. Volunteers even developed an informal system where those who did not have time to translate information into TtT syntax would rely on other Twitterers who were more skilled at translation to format their tweets.

> Meg: There were some ppl that were REALLY good at it and they would catch many of our forgotten tweets and redo them. Sometimes we were going so fast that it was hard to remember to use it.

This suggests a multi-tiered system where some volunteers monitor incoming streams to identify and amplify actionable information and other volunteers tune in to those amplification efforts and add structure to the information, sometimes through multiple transformations. With TtT use, this activity results in machine-readable information. Other structuring activity, e.g. adding a specific tag like `#MOneeds` or `#rescuemehaiti`, makes the information more useful to other digital volunteers and also, potentially, to responders who may be monitoring the stream. In all cases, at the level of the individual, this cognitive system enables every node within a social media platform to be an emergency operator of sorts who can provide real-time information processing assistance. Collectively, this activity is socially distributed across growing networks of volunteers where veteran volunteers bring experience and teach new volunteers during their efforts, often simply through their own actions. The cognitive system works to process information in real-time during mass disruption events through the collective action of individual nodes that transform information by moving it from one place to another, rebroadcasting it to keep it alive in a temporal context, or altering its form by adding micro-structure to its textual content.

8.5.4 Voluntweeters: Amplifying, Routing & Verifying as Distributed Cognition

The section above explores how the collective work of TtT translation fits a model of socially distributed cognition. The Tweak the Tweet project provided a window through which to observe digital volunteer activity, but later studies in this work show crowdwork to be much richer than the tasks performed for TtT translation. Study 2 (Chapter 5) gives a detailed description of a digital volunteer community that emerged after the Haiti earthquake, and though TtT translators made up a subset of it, the voluntweeters' network was a much larger group who participated in a range of activities. The diverse activities of digital volunteers during that event and other mass disruption events that followed also fit within a perspective that views crowdwork as distributed cognition. An examination of three voluntweeting behaviors that can be witnessed across events—amplifying, verifying, and routing—reveals how different kinds of transformations taking place across a large, distributed network of people and resources function to process information on a large and potentially massive scale.

8.5.4.1 Amplifying

Perhaps the most simple transformation visible in our analysis of digital volunteerism on Twitter is the act of amplification. The section above illustrates how amplification behavior fed into the TtT translation activity examined in Study 1 (Chapter 4). Study 4 (Chapter 7) focused heavily on amplification behavior, using the retweet as a window to examine some of the broader dynamics of crowdwork, and for the digital volunteers whose efforts are described in Study 2 (Chapter 5), amplifying information constituted a significant part of their voluntweeting work. For many Twitter users who become digital volunteers, amplification serves as the entry-point to their volunteerism. Interviews with TtT translators, as well as tweet histories from the larger group of voluntweeters during the Haiti response efforts, suggest that many began to help or to attempt to help by retweeting information they thought was important. Others first went on Twitter and saw other users retweeting information and followed their lead by doing it themselves. Casey, a voluntweeter who did a lot of retweeting during that event, explained this behavior, stating, "[I wanted to] pass around information I thought was relevant."

Amplifying occurs as the combination of two separate actions: first, identifying actionable information or information deemed relevant to affected people or responders; and second, rebroadcasting that information to increase its range of exposure. The example below demonstrates how the crowd does this:

> **@sergegilles** (2010-01-26 ~15:03): I Heard that there is a Human Traffic of childen in Hopital Espoir, Delmas 75. Can @UNICEF Check this out?

This tweet, a report of human trafficking by a Haitian Twitterer who was there during the event and the response and relief periods, was sent two weeks after the earthquake, at a time when the issue of child trafficking was a growing concern (Gupta & Agrawal, 2010) that had been brought to international attention through several mainstream media reports (e.g. Evans, 2010). The Twitter crowd quickly amplified this tweet, retweeting it in a variety of forms. Within the contextual streams of a large subset of the Haiti voluntweeters, this tweet was retweeted 16 times: eleven times in the first hour after the original tweet and four times in the second hour. 118 other tweets were sent referencing the same report about trafficking at Hopital Espoir, but with slightly different content from @sergegilles's original. 42 of 339 Twitterers in the set of voluntweeters we identified—or 12% of the network—sent at least one tweet about this report of child trafficking. This collective activity by separate nodes within a network of digital volunteers functioned to amplify relevant, actionable information—in this case information from a trusted source who was on the ground at the time.

The work of the remote crowd during the Egypt protests in 2011, described at length in Study 4 (Chapter 7), performed a similar function, to amplify the voices of those on the ground. Amplifying therefore works in two slightly different ways: 1) to identify good information, i.e. first-hand information that is actionable or helps in other ways to improve situational awareness; and 2) to identify good information sources, i.e. people who are on the ground, have special insight on conditions in the affected area, or represent an official voice during the event.

> Amanda: "I quickly identified the sources of good information ... the people who meant well but got tricked by hoaxers and tricksters ... and the people actually IN Haiti, both locals, journos and aid workers."

The interview response above reveals how some voluntweeters approached this aspect of their work intentionally, but Study 4 (Chapter 7) demonstrates that Twitterers can also perform this identification function less purposefully. That study shows how seemingly uncoordinated activity of amplification can have an aggregate effect of adding organization to the information space. These related forms of crowdwork, though perhaps motivated by different factors, can perform as a sensor network for relevant, actionable information during mass disruption events. These activities can also be seen to feed other parts of the cognitive system, for instance by providing source material for translating as outlined in the previous section, as well as for the routing and verifying behavior described next.

8.5.4.2 Routing

Evidence within both the tweet streams of and interviews with voluntweeters in Study 2 (Chapter 5) suggests that many saw routing information as an important task within their digital volunteer work. Routing behavior falls within the first part of Hutchins' leading question for assessing distributed cognition: "What information goes where?" Voluntweeters during the Haiti earthquake response wanted to connect certain information with people who needed or could use it. Though other channels, including private direct messages (DMs), emails and phone calls, were used as well, on the Twitter platform routing was often done using @mentions. The following example, linked to the above discussion of amplification, highlights routing behavior that occurred in the response to the report of human trafficking on January 26.

> **@MarkJones** (2010-01-26 15:28:21): @navynews My friend @Sergegilles is getting reports of kids being trafficked @ Hopital Espoir, Delmas 75. Please contact authorities NOW!

> **@MLBaxter** (2010-01-26 15:35:41): @andersoncooper RT @MarkJones 100% sure of source. It may be too late by the time we get there. Kids being sold @ Hopital Espoir, Delmas 75

> **@mommajulie** (2010-01-26 15:39:25): @Internethaiti RT @MarkJones 100% sure of source. It may be too late by the time we get there. Kids being sold @ Hopital Espoir, Delmas 75

> **@skinterfy** (2010-01-26 16:51:40): @redcross RT@MarkJones 100% sure of source. It may be too late by the time we get there. Kids being sold @Hopital Espoir, Delmas 75

These tweets are selections from 64 different tweets sent in the wake of the report and the resulting amplification efforts, directing that information to specific accounts. The accounts on the receiving end of these directed tweets were typically accounts that voluntweeters thought could act on the information. In the cases of @navynews and @redcross, these accounts were connected to organizations that had people on the ground in Haiti and voluntweeters might have thought that they could pass this information on to their operatives there. In the interview excerpt below, Meg, a high-volume and highly-connected voluntweeter, describes how she attempted to connect people on the ground with actionable information, by combining strategies for finding on-the-ground Twitters with directing behavior:

> Meg: By searching Twitter and finding these ppl, we could send them details on where to go and who needed what... If we saw that they were headed to a particular area...

Directed-to accounts also included those of media and celebrities whom voluntweeters felt could further amplify the report—e.g. the second tweet in the above excerpt was directed to @andersoncooper, a CNN reporter who covered the Haiti earthquake response and relief efforts extensively. Sometimes, Twitterers would direct the tweet on to accounts of other voluntweeters, hoping to continue to move the information if not directly to, then at least closer to someone who could use it. @mommajulie directed her tweet to @InternetHaiti, an account whose owner was not on the ground in Haiti, but who was performing as a digital volunteer at the time.

Through their routing actions, voluntweeters intentionally move information across nodes in the network, attempting to push it directly to a node that can use it, or move it closer to such a node. Like TtT translation, the collaborative activity of routing overlays well on the model of internal cognition, as signals moving across synapses from one node to the next. While amplification activity is essentially just a mass broadcast out to every node in the network that an account can reach through following connections or through possible searches on terms in their tweet, the processing work required for routing involves both identifying where to send the packet of information and directing it there.

8.5.4.3 Verifying

Another way that voluntweeters attempt to help shape the information flowing through Twitter and other channels is through verifying information. Many of the voluntweeters interviewed for Study 2 (Chapter 5) reported that they realized through early experiences in crisis tweeting, some during the Haiti earthquake response and some in previous events, that verifying information was an important aspect of their work. As Amanda related in the quote presented in section 8.5.4.1 (above) on amplifying, voluntweeters wanted to prevent misinformation and disinformation from entering the system and diverting vital resources. One particularly vexing part of this problem within the Twitter space was keeping track of the original time of specific reports. Reports would sometimes get reposted hours or even days after initially entering the system, and were often no longer valid at that time. Those mistakes might then be amplified by well-intentioned volunteers trying to promote actionable information. The re-introduction of out-of-date information became a focus of volunteer efforts to remove bad information in the system. Maria describes her own orientation towards verifying, touching on this issue of timeliness, in her interview response:

> Maria: “Crisis tweeting" is an art really. Accuracy can mean the difference between life and death for the people directly affected. That is why we ask that ppl not RT info that they havent confirmed.... Sometimes, during Haiti, ppl would tweet info or urgent needs (like someone trapped) and it was several days old. Ppl want to help but dont appreciate that they may cause harm.

The report of child trafficking, used as an example in several places within this section, eventually became a site of verification work by the digital volunteer network. Several hours after the initial tweet and after most, but not all, of the amplification efforts had faded, a few volunteers began to question the validity of the report and attempted to have someone confirm the information.

> **@JaneSM** (2010-01-26 ~20:31): @sergegilles my friend can you please elaborate on the source that came to you with the trafficking report. We thank you ...

> **@radioto** (2010-01-27 00:15:01): Is this story is true or not ? "...Heard that there is a Human Traffic of childen in Hopital Espoir, Delmas 75..." PLEASE OFFICIAL SOURCE

As occurred in this example, verification work often took place across a section of the digital volunteer network, with several volunteers joining in on a collaborative effort. At times, these interactions could be contentious, because reputations of individuals Twitterers were on the line. Digital volunteers may lose the trust of others in the network if they are thought to have put bad information out into the space. When a Twitterer would openly question information, others sometimes stepped in publicly to support the questioner, as did the two in this example. Later in the digital history of this report of child trafficking, @sergegilles attempted to confirm the report by contacting his source, but determined it to be a rumor and shared that with the network, which then worked to spread the retraction—though significantly less broadly than the original report.

Public, collective questioning can work to identify potentially false or temporally invalid information at both a small and a large level. Verifying behavior builds off and feeds into amplifying, which provides its source material and often spreads its findings. Occasionally, verifying is paired with structuring. Voluntweeters during several events have introduced conventions where they add "verified" or "confirmed" (with or without the hashtag symbol) to tweets to indicate that verification work has been done on that information—though certainly not all "confirmed" tweets have timely, accurate information. One interviewee from the Haiti voluntweeters recounted using the TtT syntax as a marker for signaling expertise and letting others know that the information within her tweets had been verified.

Verifying again demonstrates distributed cognition work by the socially-networked crowd, speaking to the issue of "what information." Through their collective work to confirm information, the crowd works as a collaborative sensor network for good and bad information as well as relevant and actionable information. In a large, quantitative study, Mendoza et al. (2010) demonstrate this effect, suggesting that large-scale patterns of social activity can be used to identify rumors in the social media space after disaster. At the micro-level, this kind of processing is done by individuals in interaction with others within the network. Good information is passed along, while rumors and other bad information are, at times, openly questioned. The public nature of the Twitter platform and other sites that connect information to the online reputations of its providers generates social pressure to avoid distributing false information. It also enables individual actions of information questioning to draw others' attention to an issue and rally increasingly larger groups to address problems. Within verification work, there is often some back and forth between nodes as those questioned rise up to defend their information, concede to the accusations, or—as happened in several cases during the Haiti response period—simply disappear after a large enough group begins to confront them.

Though the examples presented here all revolve around the Twitter platform, verifying work moves to other parts of the information system as well, as volunteers look to triangulate reports and find outside sources to confirm. Some voluntweeters in Study 2 (Chapter 5) reported calling phone numbers listed as contact information in tweets to verify that the information was true and still valid.

8.5.5 Humanity Road: Virtual Organizing as Distributed Cognition

The work of Humanity Road, a virtual organization of digital volunteers, offers additional insight into the dynamics of crowd work and crowd cognition during disaster. Information processing within Humanity Road is done through continuous interactions between people, tools and resources, and the theory of distributed cognition insists that cognition exists within these interactions.

Humanity Road, along with other similar organizations like the Standby Task Force, Crisis Commons and emerging VOST teams, are relatively small groups compared to the networks of voluntweeters examined above (a whole crowd view). In these smaller digital volunteer communities with established boundaries and developing norms, social organization becomes more salient. Study 3 (Chapter 6) describes how social and technical structure shapes the work processes of Humanity Road volunteers. Within a distributed cognition perspective, these features of the physical-social-digital environment provide a *cognitive architecture* that shapes the cognitive processes of the group (Hutchins, 1995; Hollan, Hutchins & Kirsh, 2000). This cognitive architecture and its role in structuring the information processing activity of Humanity Road are revealed within the description of the organization's work offered in Study 3 (Chapter 6).

8.5.5.1 Synthesizing: How Humanity Road Works as an Integrated Subsection of a Larger Cognitive System

Synthesizing information is a primary activity of the Humanity Road organization, something that they devote considerable resources to, especially during active crisis events. An analysis of how they do this work reveals a collaborative process that is both contained within the group and integrated with the cognitive architecture of the massive connected crowd.

Synthesizing is essentially the work of pulling information together from disparate sources to create more complete representations of an unfolding event. At the micro level, synthesis work is again enacted as transformations of representations, and like many of the activities described above—e.g. amplifying, structuring, verifying, and routing—it relies first on the work of identifying actionable or otherwise useful information during an event. Humanity Road volunteers bring together multiple pieces of information, including damage reports, nearest airports, location and capacity information for hospitals and shelters, lists of local government and response media accounts, etc. The team works together to filter information, identify important pieces, synthesize them together into a resource, and publicize that product of their work in various ways to make it available to affected people, responders, and other digital volunteers. They complete this work in interaction with each other, within conversations in Skype windows as well as through tweets, emails, and other channels.

Significantly, the synthesizing work of Humanity Road volunteers is also embedded within and connected to other digital volunteer communities and the massive connected crowd surrounding them. Volunteers pull information from other Twitter accounts, media reports, official websites, and blogs. Many are connected through Twitter to other digital volunteers and in many cases, to other communities to which those volunteers belong. Information filtered by one network or community becomes input for another. Humanity Road volunteers will often recognize that an event has recently occurred by reading a tweet from a crisis tweeter outside their organization. They will also incorporate information into their resources that they learn from other voluntweeters in their personal networks. At the same time, other organizations may get information for their resources from Humanity Road. The cognitive system of

Humanity Road is plugged into a much larger cognitive architecture that works to process information during mass disruption events.

8.5.5.2 Distributed Expertise: Transactive Memory as Distributed Cognition

Attempting to expand the understanding of distributed cognition in organizations, Theiner (2010) proposed a model of group cognition based on transactive memory, a theoretical view of group memory as a combination of the cognitive capacities of individual members with individual knowledge about the other members' areas of expertise (Wegner, 1986). Section 6.4.2.7 of this work demonstrates transactive memory in action during Humanity Road's response to an October 2011 earthquake in Peru. In an except from that section (lines 79-87), a Humanity Road volunteer, Sally, wished to access a tool that could execute a keyword filter for Twitter messages that would translate them automatically from one designated language to another. She had used this tool during a previous event, but could not remember what it was called or where to find it, so she sent out two different requests to other Humanity Road members looking for assistance. The first was a general question to the group through the Skype window, pushing her information request out to any and all other members who were tuned in to the conversation. A short time later, she used a different approach, directing a tweet to another volunteer (me) who Sally knew had expertise with using that online tool, and therefore might be able to guide her to it. Both requests netted Sally prompt replies and she was quickly able to locate the desired tool.

This account shows that organizations know how to get things done often by knowing which group members know how to do which tasks. Knowledge and expertise are distributed through the social system. Collaboration allows the system as a whole to function by accessing the expertise of different nodes. The group works through interaction by directing tasks and questions towards members who can address them. This is similar in many ways to how less integrated groups like the voluntweeters use strategies of retweeting and directing tweets to relay and route information and information requests, pushing tasks towards people or tools that can handle them. This knowledge of who and where to move information is an integral part of distributed cognition.

8.5.5.3 The Role of Shared Digital Artifacts in Distributed Cognition

Though interactions between people are an important part of distributed cognition, many studies of distributed cognition place considerable focus on the interactions between people and other tools and resources (e.g. Flor & Hutchins, 1991; Hutchins, 1995; Hutchins & Klausen, 1996; Hutchins & Palen, 1998). Tools and other resources play essential roles in both "individual" cognition and cognition distributed across a group.

Study 3 (Chapter 6) focuses heavily on the information-processing work done by Humanity Road volunteers during their response to an earthquake in Peru. Section 6.4.2.9 highlights a critical period of group interaction during that event, showing how the adoption of a new tool into volunteers' work practice radically re-shaped their information collation activities. As that event unfolded, volunteers began to recognize a problem in their current work practice. Initially, several volunteers were working at once to both coordinate their activity and place the products of their activity—pieces of information they were collecting—in the same Skype window. This was causing considerable confusion at the time, as volunteers were having trouble remembering who was doing what and keeping track of what had been done and what needed to be done. There was also an extra burden, which fell onto one of Humanity Road's leaders, of re-collecting the disparate pieces of information dropped into the window, then reorganizing them and formatting them into a separate document. In the midst of this action, Humanity Road leader Catherine introduced a new tool for the group to use to collect and collate information, a shared Google spreadsheet document (Figures 16-18). Almost immediately, the group reorganized their collation work around the Google document, successfully eliminating much of the confusion and improving their efforts to filter and synthesize information about the unfolding event.

The Google document changed how information about the group's current work practice was represented. It embedded, within the visible structure of the document itself, information about what needed to be done, what had already been done, and—after volunteers adopted a new procedure for claiming work areas—who was doing what. The document became not just a place for collating but also a resource for organizing the work of collation. By preserving and displaying this coordination work and making it visible in a shared place, the document lowered the cognitive load of the volunteers, who no longer had to mentally keep track of what everyone else was doing—or, alternatively, repeatedly track back up through a Skype conversation to recreate awareness of what everyone was doing. In addition to demonstrating how cognition is distributed across group members and resources, this excerpt reveals the benefits of facilitating effective cognitive off-loading into the physical or digital environment. Just as virtual team members can offload knowledge and expertise by relying on other members of the team to fill their gaps, these things can also be distributed among digital artifacts and other tools in the collaborative, virtual workspace.

8.5.6 Crowdwork as Distributed Cognition

As Bush's memex machine attempted to create an information retrieval system based on the way the human brain works using what he referred to as associative "trails" (1945), this work sees online, collective intelligence manifesting in a similar manner to individual cognition. Where internal cognition exists as transformations and movements of signals between neurons in a massive and massively connected system, the distributed cognition of crowdwork occurs through transformations and movements of representations between nodes of the expanding and increasingly connected network. Digital traces of many of these transformations preserve representational states and provide an important data source for cognitive ethnography (Hollan, Hutchins & Kirsh, 2000), allowing researchers to see many of the movements and transformations that constitute socially distributed cognition.

Many forms of connected crowdwork have a clear analogue in the traditional model of internal cognition. The Tweak the Tweet translation activity described in Study 1 (Chapter 4) and the various forms of voluntweeting activity examined in Studies 2 and 4 (Chapters 5 and 7) offer evidence of socially distributed cognition within a network of individuals that can be mapped onto internal cognition within a network of neurons. Nodes in the connected crowd, like neurons in the brain, process and transmit information to other nodes in the system. Nodes in the form of people and/or computational tools can change the form of representations—as physical and chemical structures in the brain determine the content of the signals that are distributed. Nodes in the crowd also act as relays, transmitters that send out information, often specifically directing it towards other nodes.

But mapping crowdwork to internal cognition falls short of fully explaining how virtual crowdworkers help to process information during crisis events. Distributed cognition offers a framework that explains both how cognition can be distributed across a social group and how cognition occurs through interactions with tools and other resources. Hollan et al. (2000) argue that distributed cognition both reveals how social groups think and can be reversed to demonstrate how individual cognition is distributed as well.

> "Cognitive processes involve trajectories of information (transmission and transformation), so the patterns of these information trajectories, if stable, reflect some underlying cognitive architecture. Since social organization—plus the structure added by the context of activity—largely determines the way information flows through a group, social organization may itself be viewed as a form of cognitive architecture" (Hollan, Hutchins & Kirsh, 2000, p 177)

Those authors explain that a consequence of this view is that we can use models of social cognition to explain individual cognition. Cognition occurs in interactions between a person—consisting of a brain in concert with a physical body—and her environment, which includes other people, resources, tools, artifacts, etc. Taking an example from the description of Humanity Road volunteers in action in Study 3 (Chapter 6), a Humanity Road volunteer's understanding of the event unfolding in Peru emerges from her interaction with information flowing through multiple sources, her interactions with other group members, her use of tools and the outputs they produce, the kinds of messages she puts out into the space and the answers she receives in return, and so on.

In this view, cognition is always emergent. Hollan et al. explain, "From the perspective of distributed cognition, the organization of mind—both in development and in operation—is an emergent property of interactions among internal and external resources" (2000, p 177). Just as the cognition of the overall system is distributed and emergent, so is the cognition of each individual. It is shared, it is distributed, and it exists only in interaction with other parts of the system.

In this examination, the theory of distributed cognition performs as a lens for studying how the connected crowd works to process information during mass disruption events. At the level of massive interaction, the crowd works as a filter for actionable and otherwise significant information, relying on collections of a large number of individual actions to relay and amplify certain kinds of information. Intentional activity of crowd workers who take on roles as digital volunteers functions to process information further. Networks of connected individuals direct information through the system, add structure to it, and synthesize it into resources. When uniform, these micro-actions of transforming information can be collected and processed using simple computational algorithms—the idea behind Tweak the Tweet. Less standardized actions can perform information processing "manually" as layers of transformations across networks of volunteers move information into states of increased organization. The end results of these efforts, in some cases, are digital artifacts like crisis maps and spreadsheets—collaboratively built representations of information that can act as resources for responders and affected people during crises and other mass disruption events.

8.6 Conclusion

In the context of crises and other mass disruption events, the connected crowd works in a variety of different ways to help filter, process and move information. Using the framework of distributed cognition to examine this activity—looking at what information moves where, when, and in what form—reveals dynamics of collective intelligence in action, applied to a complex and crucial challenge in a time- and safety-critical domain. Engelbart (2004) has argued that this type of problem is an ideal target for what he refers to a "collective IQ":

> "Consider a community's "Collective IQ" to represent its capability for dealing with complex, urgent problems—i.e., to understand them adequately, to unearth the best candidate solutions, to assess resources and operational capabilities and select appropriate solution commitments, to be

> effective in organizing and executing the selected approach, to monitor the progress and be able to adjust rapidly and appropriately to unforeseen complications, etc." (Engelbart, 2004, p. 1).

Drawing from the four studies presented in this dissertation, this chapter has explored several perspectives for examining applied collective intelligence during mass disruption events, including: identifying salient features for crowd activity in this context; noting a drift in the organizational configurations that support this work; and conceptualizing this whole of this work as a massive, multi-layered information filtration and processing system. Informed by this exercise, I offer a new term, *crowdwork*, to characterize this activity from a human- rather than a system-centered viewpoint, and explain how the framework of distributed cognition can be used as a lens to view this activity from an information-centered one.

CHAPTER 9
CONCLUSIONS

This dissertation on crowdwork shows social media being utilized by citizens all over the world during mass disruption events. These platforms are known to be sites of citizen reporting, information seeking, and information sharing, but increasingly are also sites where people go to act on this information, to process and organize it for others. Aligned with traditional, offline dynamics of human behavior during disaster (e.g. Fritz & Mathewson, 1957; Dynes, 1970; Kendra & Wachtendorf, 2003a, 2003b), social media users are currently appropriating the tools that connect them to improvise creative solutions to the demands of these events. In this emerging space of digital volunteerism, we see remote operators trying to get information to those who need it, emergent response groups forming to fill cell phones of affected people, and virtual organizations working to quickly assemble information resources for affected people and responders.

9.1 Emerging Role of Digital Volunteers and other Crowdworkers

The four studies in this dissertation reveal *crowdwork* in this context to be rich, diverse, interactive, and collaborative, consisting of both digital volunteerism—the individuals, emergent groups and now ongoing organizations who purposefully work to attempt to assist in response efforts—and the activities of remote crowd members who may not identify as volunteers, but whose mere participation in the space contributes in productive ways, e.g. by demonstrating solidarity with protesters or helping to identify information coming from the ground.

Observations from multiple studies, including one that included more than 14 months as a participant observer in a virtual volunteer community, suggest that the connected crowd will continue to create and evolve work practices to help affected people and responders during these events. New volunteer groups and organizations continue to arrive on the virtual scene of disaster response. Many of the groups discussed in this research, i.e. Humanity Road, SBTF, CrisisCommons and CrisisMappers, along with others such as the Humanitarian OpenStreetMap Team (HOT) and Sahana, were early arrivals in the space. But the ecosystem continues to expand and diversify. For instance, the American Red Cross recently launched a Digital Operations Center for Humanitarian Relief,[106] and the Digital Humanitarian Network[107] recently formed with the goal of bringing together various groups in the digital volunteer space and connecting with traditional response groups. Smaller groups are developing in the space as well, e.g. info4disasters, doyourpart.org, and an Oklahoma crisismapping group now advertising itself as a VOST. The VOST concept (St. Denis et al., 2012) is becoming an increasingly hot topic in the Social Media for Emergency Management (SMEM) scene—a collection of emergency managers, social media enthusiasts and other interested individuals who meet online to discuss the challenges and opportunities related to social media use for emergency responders. Seeded by the VOST concept, several support teams have formed and a community of VOSTs[108] has developed which now maintains a website with contact information for more than 20 VOST groups, along with blogs from community members containing explanations for how the concept works and input from formal emergency responders who are early proponents of the idea.

Meanwhile, new events continue to act as catalysts for new groups to emerge. Indeed, a recent article describes how during the 2011 Joplin tornado, social media users developed their own solutions to what they perceived to be inadequacies of response efforts (Mazmanian, 2012), and currently, in ongoing fire events in the state of Colorado, local citizens and remote volunteers are working to identify the needs of firefighters and evacuees, using social media to find out what can be donated and where they can take it. These observations, in combination with what we already know about human behavior during disaster, suggests that there will always be self-organizing groups of spontaneous volunteers cobbling together solutions to perceived problems. During disaster events, individuals are going to continue to converge

---

106 http://www.redcross.org/portal/site/en/menuitem.94aae335470e233f6cf911df43181aa0/?vgnextoid=1cc17852264e 5310VgnVCM10000089f0870aRCRD

107 http://digitalhumanitarians.com/

108 http://vosg.us/active-vosts/

digitally, appropriating available resources including social media, connecting to others, and self-organizing into emergent response groups. And in every event, new “crisis tweeters” will be born, arriving on the scene intent on helping but not knowing how, then seeing others in the space, connecting with them, learning from them, and joining them in a developing cause.

Digital participation is a new reality of emergency and humanitarian response. Many of these early groups were volunteer-based and not connected to traditional response organizations, but as more formal organizations enter the space, debate may arise—and indeed is already arising—over the alternative strategies of using “trusted” versus spontaneous volunteers. The VOST concept is an example of the former strategy, one that is popular among the early-adopter emergency managers who are cautiously moving into the space. However, Humanity Road and the SBTF are developing and demonstrating effective models for incorporating spontaneous and episodic volunteers into consistent crisis response activities as well.

9.2 Contributions

The broad contribution of this work is the revelation and exploration of this new role in disaster and humanitarian response for digital volunteers and other remote crowdworkers. This dissertation has sought to describe the resources, capacities, work practices and work products of these new responders, with goals of both helping formal responders understand and therefore be able to utilize these efforts and informing new design opportunities for supporting and leveraging this work.

Though not all future activities of these crowdworkers will be information-based, many of them will be, as volunteers and affected communities try to both take advantage of social media platforms and address some of the new challenges they create, like dealing with the problems of noise, unstructured messages, lost context, and misinformation. This research has focused both on the organizing of people and on this very micro-level, information-organizing activity—the filtering, amplifying, re-contextualizing, mapping, sorting, synthesizing, tagging, and verifying work of the remote, connected crowd. Shifting between the analytical foci of people-organizing and information-organizing, this dissertation offers several other contributions:

9.2.1 Unpacked the Popular "Crowdsourcing" Term

One goal of this research has been to unpack the popular "crowdsourcing" term—a word that has received a good deal of attention related to online crisis response efforts, especially during the aftermath of the 2010 Haiti earthquake—and to reveal the diverse and complex work practices that constitute that phenomenon in this context. This work exposes the roots of crowdsourcing in open source development communities and outsourcing business models, and ties the term to related concepts of collective intelligence and human computation. Identifying the places where some of these delineations fall short of helping to explain the dynamics of online participation during mass disruption events, this research offers a new term, *crowdwork*, to characterize the diverse information-processing activities occurring in this context.

9.2.2 Identified a Drift in Crowdwork Configurations from Open Source to Outsource

This research also captures evidence of a drift in crowdwork configurations from emergence and open source towards outsourcing and microtasking. As virtual organizations develop the intent to continue to respond to an ongoing event or subsequent events, their processes shift, becoming increasingly formalized. This shift can be seen to relate to new goals that include plugging into formal response and sustaining activity over time, and these in turn drive groups to deal with new issues like funding, recruiting, and maintaining volunteers. The drift also may be an effect of affordances of ICTs built to support the work, or result from natural organizing processes of structuration whereby repeatedly enacting work practices functions to reinforce them.

9.2.3 Identified Salient Features of Crowdwork in the Context of Mass Disruption

These four studies of the diverse crowd activities taking place during mass disruption events reveal the salient of features of crowdwork in this context to include properties of emergence, motivations for participation, connectivity between crowdworkers, visibility of work, progression of participation, and work ownership and control. Many of these features are shown to be interdependent—e.g. visibility of crowdwork affects motivations to participate in that type of activity. Building off the motivation "genes" offered by Malone et al. (2009) in their framework for classifying collective intelligence, and incorporating ideas of social capital (Bourdieu, 1986; Putnam, 2000; Ellison et al., 2007) and symbolic capital (Bourdieu, 1984), this work offers an expanded scheme for understanding and characterizing motivations for crowdwork.

9.2.4 Examined ICT as a Structuring Mechanism of Crowdwork

This work maintains, throughout, a partial focus on the role of ICT as a structuring feature of crowdwork. Connective technology both enables and constrains collective action, as Orlikowski's (1992) concept of *duality of technology*—itself an extension of Gidden's idea of *duality of structure*—insists it will. While non-technical crowdworkers appropriate available ICTs for their collective work, digital volunteers in technical, open source communities build, adapt, and hack tools to help them help others during crisis events. These appropriated and developed technical resources both enable and constrain the work of digital volunteers, shaping what kinds of tasks they take on and how they do them. For example, available crisis mapping platforms, e.g. Google MyMaps and Ushahidi, have become organizing points for volunteer efforts, spontaneous and otherwise. Communities of digital volunteers are becoming skilled at crisis mapping. These tools give them something tangible to do with tangible outcomes—i.e. they can see the maps they create, as can others. Is this the best application of what has become a large proportion of digital volunteer resources? That question is not yet answered. However, because these tools are available, volunteers have organized around them and incorporated them into their crowdwork practice. This research draws heavily on tweets for the primary reason that Twitter is a platform that hosts considerable volunteer activity during mass disruption events. Certainly, the affordances of that platform are intricately connected to what digital volunteer work is, currently. Drawing from this understanding of the mutual shaping of ICT, structure, and action in crowdwork, one lesson going forward is that the broad community of crowdworkers and response agencies may need to work together to identify what types of tasks responders and affected people need assistance with, and help to design and deploy tools that help volunteers do those kinds of work. However, as the voluntweeters so aptly demonstrate, people will consistently appropriate the ICTs they have on hand, the ones they use in their daily lives, to cobble together their own infrastructures for spontaneous, digital response activity in the wake of disaster events—and in the midst of other kinds of disruptions such as political protests as well.

9.2.5 Described Crowdwork as a Massive Information Filtration-Processing System

After determining that current frameworks for characterizing collective crowd activity fall short of illuminating what is salient in this context, this dissertation offers two new perspectives for considering crowdwork. In the first, the activity of the entire ICT-enabled crowd is viewed as a massive, layered system of information processing. In this model, different types of crowd activity work to process information at different levels of the system. Social media tools are both sites of information sharing and places where volunteers go to work on this information, helping to process and organize it. The collective activity of filtering, routing, structuring, geolocating, synthesizing, etc. moves information into increasingly organized states. This view brings to light several design opportunities for supporting and leveraging this activity, including opportunities for solutions that integrate machine learning with human computation, with machines both feeding and learning from the activity of humans.

9.2.6 Framed Crowdwork as Distributed Cognition

Shifting to an information-centered viewpoint, the final contribution of this work is a description of connected crowdwork as applied collective intelligence, observed through the lens of distributed cognition. By examining the micro-actions of information transformations—identifying what information moves where and in what form and mapping to these activities of filtering, routing, structuring, etc.—this analysis reveals a shared, distributed cognitive system, spanning the breadth of the connected crowd with its tools and resources, where its collective activity works to shape and re-shape the information space. This perspective incorporates both purposeful volunteer activity and the "natural" activity of social media users converging onto the virtual disaster scene to participate in the information seeking, sharing and sense-making processes. Though concentrating on connected activity applied towards a certain kind of problem and within a specific domain—i.e. information processing during mass disruption events—this distributed cognition view of crowdwork may provide a method for other examinations of social media interaction where analysis focuses on information transformations.

9.3 Final Thoughts

Interest in possible novel human capacities afford by our newfound ability to connect has been exploding since the emergence of Web 2.0 platforms that afforded mass participation through many-to-many interaction. Considering the range of problems that we face as a society, technology mediated social (and civic) participation (Preece & Shneiderman, 2009) is a vital area for research and development efforts, and digital volunteerism and other forms of crowdwork during mass disruption events are fascinating new phenomena within this growing area. The human instinct to converge and offer help after disaster events, combined with communication tools of new media that enable people to tune in to events from all over the world, and, most importantly, to connect to others, are already demonstrating tangible impact on real lives during crisis events—e.g. see Study 2 (Chapter 5) in this work for a description of remote volunteers filling cell phones for affected Haitians after the 2010 earthquake there. This research also shows that the newly enabled ability to share information from the ground, along with the opportunity for mass and in many cases global interaction, participation and collaboration, have potential that is not yet fully realized for assisting in gathering, processing and distributing information to people who need it during mass disruption events. The workforce is mobilizing, but there is work to do in better connecting the products of this work to the needs of affected people and disaster.

Open questions remain. Emergency responders are currently grappling with the issues of whether and how to incorporate, leverage or manage this activity, and they will continue to have questions about and concerns with how to interact with crowdworkers and virtual volunteer organizations. At the same time, there are a multitude of design opportunities for building tools to support, leverage, and learn from this activity.

Looking beyond the context of mass disruption events, another question arises: Can these dynamics of mass participation be applied to other kinds of problems that we face as a society? Howe's *Crowdsourcing* volume focused on new economic opportunity embedded in the connected crowd (2008), but it was in addressing the vexing problems of our time—the important and complex issues like human rights (already a focus of crisismapping communities), climate change, poverty, energy and economic issues—that Engelbart proposed tapping into the collective IQ (2004) of the crowd. Considering the capacities of mass participation and the complexities of organizing and sustaining participation over time, it is important to recognize that the disasters and political uprisings examined in this work were events that sparked mass, and in some cases global, participation. But it is possible that these dynamics are unique to this crisis domain where spontaneous volunteerism has traditionally been an active feature. Will problems in other domains manage to motivate and assemble similar crowds to address them on the scale of mass participation?

Clearly, crowdwork will be a fixture of the informational landscape around disasters and political protests going forward. Hopefully, this research will help researchers, affected people, responders, and even the volunteers themselves understand better the dynamics of their diverse and ever-changing work practices, and through that understanding be better able to connect the products of crowdwork to the intense needs of affected populations during and after mass disruption events.

BIBLIOGRAPHY

Ahuju, M.K., & Carley, K. (1998). Network Structure in Virtual Organizations, *Journal of Computer-Mediated Communication, 3*(4*)*.

Al-Ghazawy, O. (2011). Tahrir Field Hospital Saves Hundreds of Lives, *Nature Middle East*. Retrieved March 15, 2011. http://www.nature.com/nmiddleeast/2011/110214/ full/ nmiddleeast.2011.18.html

Al-Ani, B., Mark, G. & Semaan, B (2010). Blogging in a Region of Conflict: Supporting Transition to Recovery. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems* (CHI 2010). New York: ACM, 1069-1078.

Allan, S. & Thorsen, E. (eds) (2009). *Citizen Journalism: Global Perspectives.* Peter Lang: New York.

American Red Cross (2010a). Web Users Increasingly Rely on Social Media to Seek Help in Disaste*r. American Red Cross.* Available at: http://www.redcross.org/portal/site/en/menuitem.94aae335470e23 3f6cf911df43181aa0 /?vgnextoid=6bb5a96d0a94a210VgnVCM10000089f0870aRCRD

American Red Cross (2010b*). The Case for Integrating Response with Social Media. Washington DC: American Red Cross. Available at:* www.slideshare.net/americanredcross/the-case-for-integrating-crisis-response-with-social-media

Anderson, K. & Schram, A. (2011). Design and Implementation of a Data Analytics Infrastructure In Support of Crisis Informatics Research. In the *33rd International Conference on Software Engineering*, 21-28. (Honolulu, Hawaii, 2011).

Association of the United States Army (2004). ES2: Every Soldier is a Sensor. Available at: www.ausa.org.

Ben-Hur, A. & Weston, J. (2010). A User's Guide to Support Vector Machines Methods. *Molecular Biology* 609, 223-239.

Bennett, W.L. & Segerberg, A. (2012). The Logic of Connective Action. *Information, Communication & Society, 15*(5), 739-768.

Bernstein, M.S., Little, G. Miller, R.C. Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D. & Panovich, K. (2010). Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (UIST 2010). New York: ACM, 313-322.

Biewald, L. & Leila, J. (2010). Crowdsourcing Disaster Relief. *TechCrunch*. (August 12, 2010). Retrieved July 24, 2011. Available: http://techcrunch.com/2010/08/21/crowdsourcing-disaster-relief/

Boje, D.M. (1991). The Storytelling Organization: A Study of Story Performance in an Office-Supply Firm. *Administrative Science Quarterly, 36*(1), 106-126.

Bourdieu, P. (1984). *Distinction, a Social Critique of the Judgement of Taste*. Harvard: Routledge.

Bourdieu, P. (1986). The Forms of Capital. In J. Richardson (Ed.) *Handbook of Theory and Research for the Sociology of Education*, New York: Greenwood, 241-258.

Bourdieu, P. (1998). *Practical Reason: On the Theory of Action*. Cambridge: Polity Press.

boyd, d., Golder, S. & Lotan, G. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter, In *Proceedings of HCSS 2010*, (Koloa, Hawaii), 1-10.

Boyd, E.B. (2011). How Social Media Accelerated the Uprising in Egypt. *Fast Company*. (January 31, 2011). Retrieved October 21, 2011. Available at: http://www.fastcompany.com/1722492/how-social-media-accelerated-the-uprising-in-egypt

Brabham, D. C. (2010). Crowdsourcing as a Model for Problem Solving: Leveraging the Collective Intelligence of Online Communities for Public Good. PhD Dissertation, University of Utah.

Brown, J.S. & Duguid, P. (2000). *The Social Life of Information*. Boston: Harvard Business School Press.

Brownstein, J. (2011). Using Social Media for Disease Surveillance. *CNN World: Fareed Zakaria GPS*. Posted August 18, 2011. Available at: http://globalpublicsquare.blogs.cnn.com/2011/08/18/using-social-media-for-disease-surveillance/

Burns, A. & Eltham, B. (2009). Twitter Free Iran: An Evaluation of Twitter's Role in Public Diplomacy and Information Operations in Iran's 2009 Election Crisis. In *Communications Policy & Research Forum*.

Bush, V. (1945). As We May Think, *Atlantic Monthly*, July 1945, 101-108.

Callison-Burch, C. (2009). Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1* (EMNLP 2009), Association for Computational Linguistics, Stroudsburg, PA, USA, 286-295.

Callison-Burch, C. & Dredze, M. (2010). Creating Speech and Language Data with Amazon's Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk* (CSLDAMT 2010). Association for Computational Linguistics, Stroudsburg, PA, USA, 1-12.

Carlson, A., Betteridge, J., Wang, R.C., Hruschka, H.R., & Mitchell, T.M. (2010). Coupled Semi-Supervised Learning for Information Extraction. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining* (WSDM 2010). New York: ACM, 101-110.

Casner, S.M. (2010) Why Don't Pilots Submit More Pilot Weather Reports (PIREPs)? *National Journal of Aviation Psychology 20*(4), 347-374.

Cavallo, E, Powell, A, & Becerra, O. (2010). *Estimating the Direct Economic Damage of the Earthquake in Haiti*. February 2010. IDB Inter-American Development Bank.

Centers for Disease Control and Prevention. (2011). Cholera in Haiti: One Year Later. CDC Website. Available at: http://www.cdc.gov/haiticholera/haiti_cholera.htm

Centola, D. & Macy, M. (2007). Complex Contagion and the Weakness of Long Ties. *American Journal of Sociology*, 113, 702-734.

Chen, K. (2011). Human Computation: Experience and Thoughts. Position Paper for the *CHI 2011 Workshop on Crowdsourcing and Human Computation*. (CHI 2011).

Cohn, J.P. (2008) Citizen Science: Can Volunteers Do Real Research? *BioScience 58*(3), 192-197.

Corvey, W.J., Vieweg, S., Rood, T. & Palmer, M. (2010). Twitter in Mass Emergency: What NLP Techniques can Contribute. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media* (Los Angeles, California, June 2010), 23–24.

Corvey, W. J., Vieweg, S., Verma, S., Palmer, M. & Martin, J.H. (2012). Foundations of a Multilayer Annotation Framework for Twitter Communications During Crisis Events**,** In *Proceedings of the Eighth International Conference on Language Resources and Evaluation* (LREC 2012), May 21-27, 2012, Istanbul, Turkey.

"crowd, n., 1b" (1989) *The Oxford English Dictionary*. 2nd ed. OED Online. Oxford University Press. Available at: http://dictionary.oed.com.

DeSanctis, G., P. Monge. (1999). Communication Processes for Virtual Organizations. *Organization Science, 10*(6), 693–703.

DeSanctis, G. & Poole, M. S. (1994). Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory. *Organization Science, 5*(2), 121-147.

Diakopoulos, N.A. & Shamma, D.A. (2010). Characterizing Debate Performance via Aggregated Twitter Sentiment. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems* (CHI 2010). New York: ACM, 1195-1198.

Dow, S., Kulkarni, A., Klemmer, S., & Hartmann, B. (2012). Shepherding the Crowd Yields Better Work. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (CSCW 2012). New York: ACM, 1013-1022.

Drabek, T. E. (1986). *Human Responses to Disaster: An Inventory of Sociological Findings.* New York: Springer-Verlag.

Dynes, RR. (1970). *Organized Behavior in Disaster*. Heath.

Eler, Alicia. (2012). Why Instagram is the New Camera. Blog in *ReadWriteWeb.* (April 3, 2012). Available at: http://www.readwriteweb.com/archives/how_instagram_became_the_new_camera.php

Eliot, D. & Banda, P.S. (2010). Colorado Wildfire Evacuees Cautiously Allowed Back Home. *USA Today*, September 10, 2010. Retrieved from: http://www.usatoday.com/weather/ wildfires/2010-09-10-colorado-wildfire_N.htm. Accessed on December 12, 2010.

Ellison, N. B., Steinfeld, C., & Lampe, C. (2007). The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication, 12* (4), article 1.

Engelbart, D. (1962). *Augmenting Human Intellect: A Conceptual Framework*. AFOSR-3233, Summary Report. Menlo Park, CA. Stanford Research Institute. Retrieved from: http://www.invisiblerevolution.net/engelbart/full_62_paper_augm_hum_int.html

Engelbart, D. (1968). The oNLine System (NLS) Demo. *Fall Joint Computer Conference*. Dec 8, 1968. San Francisco, CA. http://video.google.com/videoplay?docid=8734787622017763097

Engelbart, D. (2004). Augmenting Society's Collective IQs. In *Proceedings of the 15th ACM Conference on Hypertext and Hypermedia* (HYPERTEXT 2004). New York: ACM, 1-1.

Evans, T. (2010). Traffickers Targeting Haiti's Children, Human Organs, PM Says. *CNN World.* (January 27, 2010). Available at: http://articles.cnn.com/2010-01-27/world/haiti.earthquake.orphans_1_haitian-children-human-organs-port-au-prince-airport?_s=PM:WORLD

Facebook Newsroom. (2012). Key Facts. Accessed June 3, 2012. Available at: http://newsroom.fb.com/content/default.aspx?NewsAreaId=22

Finholt, T. & Sproull, L.S. (1990). Electronic Groups at Work. *Organization Science* , *1*(1), 41-64.

Finholt, T., Sproull, L., & Kiesler, S. (1990). Communication and Performance in Ad Hoc Groups. In J.Galegher, R.Kraut & C.Egido, (Eds.), *Intellectual teamwork* (pp. 291–325). NJ: Lawrence Erlbaum Associates.

Fischer III, HW. (1998). *Response to Disaster: Fact Versus Fiction & Perpetuation*, 2nd ed. New York: University Press of America.

Flickr. (2011). Photos and Video from Everyone in Flickr. Accessed: June 22, 2011 Available at: http://www.flickr.com/photos

Fiol, M. & O'Connor, E. (2005). Identification in Face-to-Face, Hybrid, and Pure Virtual Teams: Untangling the Contradictions. *Organization Science, 16*(1), 19-32.

Flor, N. & Hutchins, E. (1991). Analyzing Distributed Cognition in Software Teams: A Case Study of Team Programming During Perfective Software Maintenance, in *Empirical Studies of Programmers: Fourth Workshop*, J. Koenemann-Belliveau, T.G. Moher, & S.P. Robertson, Editors. Ablex, 36-59.

Fort, K., Adda, G., & Cohen, K.B. (2011). Amazon Mechanical Turk: Gold Mine or Coal Mine? *Computational Linguistics*, *37*(2), 413-420.

Friedman, B. (1996). Value-Sensitive Design. *Interactions* 3, 6 (December 1996), 16-23.

Friedman B., Kahn P. H., Jr., & Borning, A. (2006). Value Sensitive Design and Information Systems. In *Human-computer Interaction in Management Information Systems: Foundations*, M.E. Sharpe (Ed.), NY: Armonk, 348-372.

Fritz, C. E., & Mathewson, J. H. (1957). *Convergence Behavior in Disasters: A Problem in Social Control*. Washington, DC: National Academy of Sciences.

Giddens, A. (1979). *Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis*. London: The Macmillan Press.

Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California Press.

Giglio, M. (2011). Is Egypt Next? *Newsweek*, January 23, 2011. Retrieved July 3, 2011. http://www.newsweek.com/2011/01/22/the-revolution-comes-to-egypt.html

Gilley, K.M. & Rasheed, A. (2000). Making More by Doing Less: An Analysis of Outsourcing and its Effects on Firm Performance, *Journal of Management*, *26*(4), 763-90.

Gillmor, D. (2004). We the Media: The Rise of Citizen Journalists. *National Civic Review*, Fall 2004, 58-63.

Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J. & Smith, N.A. (2011). Part-of-speech Tagging for Twitter: Annotation, Features, and Experiments. In *ACL* (Short Papers), 42-47.

Gladwell, M. (2010). Small Change - Why the Revolution Will Not Be Tweeted, *New Yorker*, October 4, 2010.

Goldberg, D., Nichols, D., Oki, B.M., & Terry, D. (1992). Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM 35*, 12 (December 1992), 61-70.

Goodchild, M.F. (2007). Citizens as Sensors: The World of Volunteered Geography. *GeoJournal 69*(4), 211–221.

Goodchild, M.F. & Glennon, J. A. (2010). Crowdsourcing Geographic Information for Disaster Response: A Research Frontier. *International Journal of Digital Earth 3*(3), 231-241.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1360-1380.

Greenberg, P. (2011). Engaging Citizens the Right Way: Government Uses Twitter During Hurricane Irene. *Blog in ZDNet,* September 12, 2011. Retrieved September 12, 2011. Available at: http://www.zdnet.com/blog/crm/engaging-citizens-the-right-way-government-uses-twitter-during-hurricane-irene/3478

Grier, D.A. (2005). *When Computers Were Human*. Princeton, NJ: Princeton University Press,.

Grier, D.A. (2011). Foundational Issues in Human Computing and Crowdsourcing. Position Paper for the *CHI 2011 Workshop on Crowdsourcing and Human Computation*. CHI 2011.

Grossman, L. (2009). Iran's protests: Why Twitter is the medium of the movement. *Time*, 2009. Retrieved May 22, 2011. http:// www.time.com/time/world/article/ 0,8599,1905125,00.html

Gupta, J., & Agrawal, A. (2010). Chronic Aftershocks of an Earthquake on the Well-Being of Children in Haiti: Violence, Psychosocial Health and Slavery. *Canadian Medical Association Journal*, *182*(18), 1997-1999.

Guy, M, Earle, P, Ostrum, C, Gruchalla, K & Horvath, S. (2010). Integration and Dissemination of Citizen Reported and Seismically Derived Earthquake Information via Social Network Technologies. *Advances in Intelligent Data Analysis IX, Lecture Notes in Computer Science,. 6065*(2010), 42-53.

Hagar, C. & Haythornthwaite C. (2005). Crisis, Farming & Community. *The Journal of Community Informatics, 1*(3), 41-52.

Handy, C. (1995). Trust and Virtual Organization, *Harvard Business Review, 73*, 40-50.

Harland, C., Knight, L., Lamming, R., & Walker, H. (2005). Outsourcing: Assessing the Risks and Benefits for Organisations, Sectors and Nations. *International Journal of Operations & Production Management*, *25*(9/10), 831-850.

Harvard Humanitarian Initiative. (2011). *Disaster Relief 2.0: The Future of Information Sharing in Humanitarian Emergencies*. Washington, D.C. and Berkshire, UK: UN Foundation & Vodafone Foundation Technology Partnership.

Hecht, B., Hong, L., Suh, B. & Chi, E.H. (2011). Tweets from Justin Bieber's Heart: The Dynamics of the Location Field in User Profiles. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (CHI 2011). New York: ACM, 237-246.

Hersman, E. (2009). Crisis Info: Crowdsourcing the Filter. *Ushahidi Blog*. Retrieved Sept 19, 2011. Available at: http://blog.ushahidi.com/index.php/2009/02/04/crisis-info-crowdsourcing-the-filter/

Heverin, T., & Zach, L. (2010). Microblogging for Crisis Communication: Examination of Twitter Use in Response to a 2009 Violent Crisis in Seattle-Tacoma, Washington Area. Paper presented at the *2010 Information Systems for Crisis Response and Management Conference, ISCRAM 2010*, Seattle, WA.

Hiltz, S.R. & M. Turoff (1978). *The Network Nation: Human Communication via Computer*, Addison-Wesley.

Hollan, J., Hutchins, E. & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction. 7*(2) (June 2000), 174-196.

Honeycutt, C. & S.C. Herring. (2009). Beyond Microblogging: Conversation & Collaboration via Twitter. HICSS 2009, 1-10.

Howe, J. (2006). The Rise of Crowdsourcing. *Wired*. Issue 14.06 (June 2006). Retrieved June 2, 2011. Available at: http://www.wired.com/wired/archive/14.06/crowds.html

Howe, J. (2008). *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business.* New York: Crown Business.

Hughes, A., L. Palen, J. Sutton, S. Liu, & S. Vieweg. (2008). “Site-Seeing” in Disaster: An Examination of On-Line Social Convergence. In *Proceedings of Information Systems for Crisis Response and Management Conference* (ISCRAM 2008).

Hughes, A. & Palen, L. (2009). Twitter Adoption and Use in Mass Convergence and Emergency Events. In *Proceedings of the 2009 Information Systems for Crisis Response and Management Conference,* (ISCRAM 2009).

Hullman, J. (2011). Not All Hits Are Created Equal: Controlling for Reasoning and Learning Processes in MTurk. Position Paper for the *CHI 2011 Workshop on Crowdsourcing and Human Computation*. (CHI 2011).

Humanity Road (2011). *Humanity Road, Inc. 2010 Annual Report*. Boydton, VA: Humanity Road. Available at: www.humanityroad.org

Hutchins, E. (1995). *Cognition in the Wild.* Cambridge, Mass: MIT Press.

Hutchins E. & Klausen, T. (1996). Distributed Cognition in an Airline Cockpit. In Middleton, D. & Engeström, Y. (Eds.) *Communication and Cognition at Wor*k. (pp. 15-34) Cambridge: Cambridge University Press.

Hutchins, E. & Palen, L. (1998). Constructing Meaning from Space, Gesture and Speech. In L. B. Resnick, R. Saljo, C. Pontecorvo, & B. Burge (Eds.) *Discourse, Tools, and Reasoning: Situated Cognition and Technologically Supported Environments*, Springer-Verlag, Germany.

Kane, G.C. (2011). A Multimethod Study of Information Quality in Wiki Collaboration. *ACM Transactions on Management of Information Systems, 2*(1), Article 4 (March 2011), 16 pages.

Kaplan, A.M. & Haenlein, M. (2010). Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons, 53*(1), 59-68.

Kelly, J. (1985). Storytelling in High Tech Organization: A Medium for Sharing Culture. *Journal of Applied Communication Research, 13*, 1.

Kendra, J. M. & Wachtendorf, T. (2003a). Creativity in Emergency Response After The World Trade Center Attack. In *Beyond September 11th: An Account of Post-Disaster Research*, (pp. 121-146). Institute of Behavioral Science, Natural Hazards Center, University of Colorado.

Kendra, J. M. & Wachtendorf, T. (2003b). Reconsidering Convergence and Converger: Legitimacy in Response to the World Trade Center Disaster. *Terrorism and Disaster: New Threats, New Ideas: Research in Social Problems and Public Policy, 11*, 97-122.

Kittur, A. Chi, E.H. and Suh, B. (2008). Crowdsourcing User Studies with Mechanical Turk. In *Proceeding of the 26th annual SIGCHI conference on Human factors in computing systems* (CHI 2008). New York: ACM, 453-456.

Kreps, G. A. & Bosworth, S.L. (1994). *Organizing, Role Enactment, and Disaster: A Structural Theory*. Cranbury, NJ: Associated University Presses.

Kruk, S.R., & Decker, S. (2005). Semantic Social Collaborative Filtering with FOAFRealm. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) *ISWC 2005. LNCS, vol. 3729*. Springer, Heidelberg.

Kwak, H., Lee, C., Park, H. & Moon, S. (2010). What is Twitter, a social network or a news media? *Intl. WWW Conference,* (Raleigh, NC, 2010), New York: ACM, 591-600.

Kurczy, S. (2010). Haiti's Tent Cities to Bear Worst of Potential Hurricane Tomas. *Christian Science Monitor* (November 4, 2010). Available at: http://www.csmonitor.com/World/Americas/2010/1104/Haiti-s-tent-cities-to-bear-worst-of-potential-hurricane-Tomas

Langewiesche, W. (2002). American Ground: Unbuilding the World Trade Center. *The Atlantic Monthly, 290*(2), 44-79.

Large, T. (2010). Cell Phones and Radios Help Save Lives After Haiti Earthquake. *Reuters.* (January 25, 2010). Retrieved August 1, 2011. Available at: http://www.reuters.com/article/2010/01/25/us-haiti-telecoms-idUSTRE60O07M20100125

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Lerman, K. & Ghosh, R. (2010). Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks, in *4th International AAAI Conference on Weblogs and Social Media*, (AAAI 2010), 90-97.

Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., & Hurst, M. (2007). Cascading Behavior in Large Blog Graphs, in *SIAM International Conference on Data Mining*.

Leuf, B. & Cunningham, W. (2001). *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley Professional.

Lévy, P. as translated by Bononno, R. (1997). *Collective Intelligence*. Cambridge, Mass: Perseus Books.

Liu, S. Iacucci, A.A., & Meier, P. (2010). Ushahidi Haiti & Chile: Next Generation Crisis Mapping. *American Congress on Surveying and Mapping (ACSM) Bulletin*, August 2010.

Liu, S., Palen, L., Sutton, J., Hughes, A., Vieweg, S. (2008). In Search of the Bigger Picture: The Emergent Role of On-Line Photo-Sharing in Times of Disaster. In *Proceedings of the International Conference on Information Systems for Crisis Response & Management Conference,* (ISCRAM 2008).

Lohr, S. (2011). Online Mapping Shows Potential to Transform Relief. *New York Times: Business Day.* (March 28, 2011). Retrieved August 1, 2011. Available at: http://www.nytimes.com/2011/03/28/business/28map.html

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I. & boyd, danah. (2011). The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions. *International Journal of Communications* 5, 1375-1405.

Madrigal, A. (2011). The Tech Used to Help in Haiti's Earthquake Recovery. *The Atlantic.* Posted January 11, 2011. Retrieved Oct 20, 2011. Available at: http://www.theatlantic.com/technology/archive/2011/01/the-tech-used-to-help-in-haitis-earthquake-recovery/69327/

Malone, T. W., Laubacher, R. & Dellarocas, C. N., Harnessing Crowds: Mapping the Genome of Collective Intelligence (February 3, 2009). MIT Sloan Research Paper No. 4732-09.

Manhire, T. & Tran, M. (2011). Christchurch Earthquake: Hunt for Survivors Begins. *The Guardian*. Retrieved May 31, 2011. http://www.guardian.co.uk/world/2011/ feb/22/christchurch-earthquake-hunt-for-survivors

Mathes, A. (2004). Folksonomies – Cooperative Classification and Communication Through Shared Metadata. UIC Technical Report, 2004.

Mazmanian, A. (2012). Of Hurricanes and Hashtags: Disaster Relief in the Social-Media Age. *NationalJournal*. (June 3, 2012). Available at: http://www.nationaljournal.com/tech/of-hurricanes-and-hashtags-disaster-relief-in-the-social-media-age-20120603

McAdam, D. (1986). Recruitment to High-Risk Activism: The Case of the Freedom Summer. *American Journal of Sociology*, *92*(1), 64-90.

McPhee, R.D. & Zaug, P. (2009) The Communicative Constitution of Organizations. In Putnam, L.L. & Nicotera, A.M. (Eds.) *Building Theories of Organization: The Constitutive Role of Communication*, (pp. 21-47). New York: Routledge

Meier, P. (2010a). Ushahidi and the Unprecedented Role of SMS in Disaster Response. Blog in *Haiti Rewired*. Posted February 23, 2010. Available at: http://haitirewired.wired.com/profiles/blogs/ushahidi-amp-the-unprecedented

Meier, P. (2010b). From Netwsourcing to Crowdsourcing to Turksourcing Crisis Information. Blog in *iRevolution: From innovation to Revolution.* Posted March 16, 2010. Available at: http://irevolution.net/2010/03/16/netsourcing-to-turksourcing/

Meier, P. (2010c). Introducing the Standby Task Force. Blog in *The Standby Task Force: Online Volunteer Community for Live Mapping.* Posted October, 2010. Available at: http://blog.standbytaskforce.com/about/introducing-the-standby-task-force/

Meier, P. (2011). Want to Help Humanitarian Operations in Libya Using Your Laptop? Join UN Volunteers Online Now. *Standby Task Force blog*, March 10, 2011. Available: http://blog.standbytaskforce.com/want-to-help-humanitarian-operations-in-libya-using-your-laptop-join-un-volunteers-online-now/

Meier, P. & Munro, R. (2010). The Unprecedented Role of SMS in Disaster Response: Learning from Haiti. *SAIS Review*, *30*(2), 91-103.

Meier, P., Valuch, J., Munro, R., Iacucci, A., & Larrauiri, H.P. (2011). Why We Need a Disaster 2.1 Report. *Standby Task Force blog*. Retrieved June 20, 2011. Available at: http://blog.standbytaskforce.com/?p=456

Mendoza, M. Poblete, B. & Castillo, C. (2010). Twitter Under Crisis: Can We Trust What We RT? In *Proceedings of the First Workshop on Social Media Analytics* (SOMA 2010). New York: ACM, 71-79.

Messina, C. (2007a) Groups for Twitter; or A Proposal for Twitter Tag Channels. Aug 25, 2007. Blog in: *FactoryCity*. URL: http://factoryjoe.com/blog/2007/10/22/twitter-hashtags-for-emergency-coordination-and-disaster-relief/

Messina, C. (2007b) Twitter Hashtags for Emergency Coordination and Disaster Relief. Oct 22, 2007. Blog in: *FactoryCity*. http://factoryjoe.com/blog/2007/10/22/twitter-hashtags-for- emergency-coordination-and-disaster-relief/

Messina, C. (2009). Stowe Boyd Launches Microsyntax.org. May 16, 2009. Blog in: *FactoryCity*. URL: http://factoryjoe.com/blog/ 2009/05/26/stowe-boyd-launches-microsyntax-org/

Morozov, E. (2009). The Brave New World of Slacktivism. May 19, 2009. Blog in *Net.Effect Foreign Policy*. Retrieved: May 31, 2011. Available at: http://neteffect.foreignpolicy.com/posts/2009/05/19/the_brave_new_world_of_slacktivism

Mullins, J. (2010). How Crowdsourcing is Helping in Haiti. *New Scientist*. (January 27, 2010). Retrieved August 1, 2011. Available at: http://www.newscientist.com/article/mg20527453.600-how-crowdsourcing-is-helping-in-haiti.html

Mungiu-Pippidi, A., & Munteanu, I. (2009). Moldova's "Twitter Revolution" *Journal of Democracy 20*(3), 136-142.

Munro, R. (2010). Crowdsourced Translation for Emergency Response in Haiti: The Global Collaboration of Local Knowledge. In *Proceedings of the AMTA Workshop on Collaborative Crowdsourcing for Translation*.

Nam, K. & Ackerman, M. (2007). Arkose: Reusing Informal Information from Online Discussions. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (GROUP 2007). New York: ACM, 137-146.

Neubig, G., Matsubayashi, Y., Hagiwara, M., & Murakami, K. (2011). Safety Information Mining – What Can NLP Do in a Disaster. In *Proceedings of 5th Intl. Joint Conf. on Natural Language Processing*, (Chiang Mai, Thailand), 965-973.

New York Times (2010a). Digital Help for Haiti. *New York Times: Personal Tech*. January 27, 2010. Retrieved October 20, 2011. Available at: http://gadgetwise.blogs.nytimes.com/2010/01/27/digital-help-for-haiti/

New York Times. (2010b). Haiti Earthquake of 2010. Sept 23, 2010. *The New York Times*. http://www.nytimes.com/info/haiti-earthquake-2010/

New York Times. (2011a). Egypt News—Revolution and Aftermath. *New York Times*, Retrieved May 31, 2011. http://topics.nytimes.com/top/news/international/ countriesandterritories/egypt/index.html

New York Times. (2011b). Times Topics – Wael Ghonim. *New York Times*. Retrieved 15 March 2011. URL: http://topics.nytimes.com/top/reference/timestopics/ people/g/wael_ghonim/index.html

New York Times. (2011c). World—Tunisia. *New York Times*. Retrieved July 3, 2011. http://topics.nytimes.com/top/news/international/countriesandterritories/tunisia/index.html

Okolloh, O. (2009). Ushahidi, or 'Testimony': Web 2.0 Tools for Crowdsourcing Crisis Information. *Participatory Learning and Action 59*, 65–70.

Orlikowski, W.J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations, *Organization Science, 3*(3), 398-427.

Orlikowski, W.J. (2000). Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations, *Organization Science*, *11*(4), 404-428.

Orr, J.E. (1986). Narratives at Work: Story Telling as Cooperative Diagnostic Activity. In *Proceedings of the 1986 ACM Conference on Computer-Supported Cooperative Work* (CSCW 1986). New York: ACM, 62-72.

Palen, L., & Liu, S. B. (2007). Citizen Communications in Crisis: Anticipating a Future of ICT-Supported Participation. *Proceedings of the 2007 ACM Conference on Human Factors in Computing System* (CHI 2007). New York: ACM, 727-736.

Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., & Grunwald, D. (2010). A vision for technology-mediated support for public participation and assistance in mass emergencies and disasters. In *Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference* (Edinburgh, United Kingdom, April 14 – 16, 2010). Swinton, UK: ACM-BCS Visions of Computer Science, British Computer Society, 1-12.

Palen, L., Vieweg, S., and Anderson, K. (2011). Supporting "Everyday Analysts" in Time- and Safety-Critical Situations. *The Information Society Journal, 27*(1), 52-62.

Palen, L., Vieweg, S., Liu, S., Hughes, A. (2009). Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007 Virginia Tech Event. *Social Science Computing Review,* Sage, 467-480.

Paul, S.A., Hong, L. & Chi, E. (2011). What is a Question? Crowdsourcing Tweet Categorization. Position Paper for the *CHI 2011 Workshop on Crowdsourcing and Human Computation* (CHI 2011).

Pfitzman, A. & Köhntopp, M. (2001). Anonymity, Unobservability, and Pseudonymity – A Proposal for Terminology. In Hannes, F. (Ed) *Designing Privacy Enhancing Technologies*. Berlin/Heidelberg: Springer. pp. 1-9.

Phillips, J. (2011). VOSG/T Concept Report. Google Document. Available at: https://docs.google.com/document/d/1cj5Ta_Ar%20D4ji7ZngZiirhuw_q46J3_vsNnOY_k0gg_c/edit?hl=en_US

Preece, J., & Shneiderman, B. (2009). The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. *AIS Transactions on Human-Computer Interaction, 1*(1), 13-32.

Puertas, L. & Elsen, J. (2007). Earthquake in Peru Kills Hundreds. *The New York Times*. (August 16, 2007). Available at: http://www.nytimes.com/2007/08/16/world/americas/16cnd-peru.html

Putnam, R. D. (2000). *Bowling Alone*. New York: Simon & Schuster.

Putnam, L.L., Nicotera, A.M., & McPhee, R.D. (2009) *Building Theories of Organization: The Constitutive Role of Communication*, Putnam, L.L. & Nicotera, A.M. (eds.), Routledge: New York.

Quinn, A. & Bederson, B. (2011a). Human Computation: A Survey and Taxonomy of a Growing Field. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (CHI 2011). New York: ACM, 1403-1412.

Quinn, A. & Bederson, B. (2011b) Hybrid Human-Machine Hybrid Computation. Position Paper for the CHI 2011 Workshop on Crowdsourcing and Human Computation. (*CHI 2011)*.

Qu, Y., Wu, P.F., Wang, X. (2009). Online Community Response to Major Disaster: A Study of Tianya Forum in the 2008 Sichuan Earthquake, *42nd Hawaii International Conference on System Sciences* (*HICSS '09),* 1-11.

Qu, Y., Huang, C., Zhang, P. & Zhang, J. (2011). Microblogging After a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake, In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*. (CSCW 2011). New York: ACM, 25-34.

Queensland Police Service (2011). Disaster Management and Social Media – a Case Study. *Media and Public Affairs Branch, Queensland Police Service.*

Raymond, E.S. (2000). *The Cathedral and the Bazaar*. *Version 3.0*. Retrieved June 7, 2011. http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/

Reinikainen, E. (2009). #iranelection Cyberwar Guide for Beginners. Blog in *Networked Culture.* Retrieved Sept 21, 2011. Available at: http://reinikainen.co.uk/2009/06/iranelection-cyberwar-guide-for-beginners

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of NetNews. In *Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work*. (CSCW 1994) Chapel Hill, NC: ACM, 175-186.

Reuters (2011a). Death Toll From Joplin Rises to 142. *Reuters.com*. May 28, 2011. Available at: http://www.reuters.com/article/2011/05/28/usa-tornadoes-joplin-idUSN2814172620110528

Reuters (2011b). Joplin Tornado Death Toll Revised Down to 161. *Reuters.com*. Nov 12, 2011. Available at: http://www.reuters.com/article/2011/11/12/us-tornado-joplin-idUSTRE7AB0J820111112

Rogers, Y. & Ellis, J. (1994). Distributed Cognition: An Alternative Framework for Analyzing and Explaining Collaborative Working. *Journal of Information Technology, 9*(2), 119-128.

Romero, D., Meeder, B. & Kleinberg, J. (2011). Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter, in *International WWW Conference*, (Hyderabad, India), New York: ACM, 695-704.

Rosenberg, T. (2011). Crowdsourcing a Better World. *New York Times: Opinion Pages.* (March 28, 2011). Retrieved August 1, 2011. Available at: http://opinionator.blogs.nytimes.com/2011/03/28/crowdsourcing-a-better-world/

Ross, J., Irani, L., Silberman, M.S., Zaldivar, A. & Tomlinson, B. (2010). Who are the Crowdworkers?: Shifting Demographics in Mechanical Turk. In *Proceedings of the 2010 Annual Conference on Human Factors in Computing Systems* (CHI 2010). New York: ACM, 2863-2872.

Ross, N. & Potts, M. (2011). ABC's Crowdsourced Flood-Mapping Initiative. *ABC: Technology+Games*, (January 13, 2011). Retrieved August 1, 2011. Available at: http://www.abc.net.au/technology/articles/2011/01/13/3112261.htm

Russell, P. (1983). *The Global Brain.* Los Angeles: J.P. Tarcher.

Salminen, J. (2012). Collective Intelligence in Humans: A Literature Review. In *Proceedings of Collective Intelligence 2012* (Cambridge, Mass).

Sarcevic, A., Marsic, I., Lesk, M.E. Burd, R.S. (2008). Transactive Memory in Trauma Resuscitation. In *Proceedings of the 2008 ACM conference on Computer Supported Cooperative Work* (CSCW 2008). New York: ACM, 215-224.

Sarcevic, A. Palen, L., White, J., Starbird, K., Bagdouri, M. & Anderson. K. (2012). "Beacons of hope" in Decentralized Coordination: Learning from On-the-Ground Medical Twitterers during the 2010 Haiti Earthquake. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (CSCW 2012). New York: ACM, 47-56.

Schmidt & Bannon (1992). Taking CSCW Seriously: Supporting Articulation Work. *Computer Supported Cooperative Work 1* (1), 7-40.

Shamma, D.A., Kennedy, L. & Churchill, E.F. (2009). Tweet the Debates: Understanding Community Annotation of Uncollected Sources. In *Proceedings of the 1st SIGMM Workshop on Social Media* (WSM 2009). New York: ACM, 3-10.

Shepard, A. (2011). NPR's Giffords Mistake: Re-learning the Lesson of Checking Sources. *NPR blogs*. Posted January 11, 2011. Retrieved September 9, 2011. Available at: http://www.npr.org/blogs/ombudsman/2011/01/11/132812196/nprs-giffords-mistake-re-learning-the-lesson-of-checking-sources

Shirky, C. (2008a). *Here Comes Everybody*. New York: The Penguin Press.

Shirky, C. (2008b). It's Not Information Overload: It's Filter Failure. *Web 2.0 Expo, New York.* Accessed June 27, 2011. Available at: http://www.web2expo.com/webexny2008/public/schedule/detail/4817

Shklovski, I., Palen, L. & Sutton J. (2008). Finding Community Through Information and Communication Technology in Disaster Events. In *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work* (CSCW 2008), New York: ACM, 127-136.

Singh, P. Lin, T., Mueller, E.T., Lim, G., Perkins, T. & Zhu, W.L. (2002). Open Mind Common Sense: Knowledge acquisition from the general public. *Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*. Irvine, CA.

Singel, R. (2011). Burning Questions: Why Am I Being Followed by Twitter Robots? *Wired*. (May 31 2011). Retrieved June 9, 2012. Available at: http://www.wired.com/magazine/2011/05/pr_burning_twitter/

Smith, J. (2010). Tufts Map Steered Action Amid Chaos. *Boston Globe: boston.com*. April 5, 2010. Retrieved October 21, 2010. Available at: http://www.boston.com/news/education/higher/articles/2010/04/05/tufts_project_delivered_aid_to_quake_victims/

Sorokin, A. & Forsyth, D. (2008). Utility Data Annotation with Amazon Mechanical Turk. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08*. (Anchorage, Alaska).

Standby Task Force (2011). Libya Crisis Map Deployment 2011 Report. Blog. Retrieved Sept 21, 2011. Available at: http://blog.standbytaskforce.com/libya-crisis-map-report/

St. Denis, L., Hughes, A., & Palen, L. (2012). Trial by Fire: The Deployment of Trusted Digital Volunteers in the 2011 Shadow Lake Fire. In *Proceedings of the Information Systems for Crisis Response and Management Conference* (ISCRAM 2010). Vancouver, B.C., Canada.

Star, S.L. & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research 7*(1), 111-134.

Starbird, K. (2008). E-Data Visualization for Qualitative Research. Interactive Poster presented at *ACM 2008 Conference on Computer Supported Cooperative Work* (CSCW '08), San Diego, CA.

Starbird, K., Palen, L., Hughes, A. & Vieweg, S. (2010). Chatter on The Red: What Hazards Threat Reveals about the Social Life of Microblogged Information. In *Proceedings of the ACM 2010 Conference on Computer Supported Cooperative Work* (CSCW 2010). New York: ACM, 241-250.

Starbird, K., & Palen, L. (2010). Pass It On?: Retweeting in Mass Emergencies. In *Proceedings of the International Conference on Information Systems for Crisis Response & Management Conference,* (ISCRAM 2010).

Starbird, K., & Palen, L. (2011). 'Voluntweeters': Self-organizing by Digital Volunteers in Times of Crisis, In *Proceedings of the 2011 ACM Conference on Human Factors in Computing Systems* (CHI 2011). New York: ACM, 1071-1080.

Starbird, K. Palen, L., & others. (2012). Promoting Structured Data in Citizen Communications During Disaster Response: An Account of Strategies for Diffusion of the 'Tweak the Tweet' Syntax. In Christine Hagar (Ed.), *Crisis Information Management: Communication and Technologies,* (pp. 43-63), Cambridge, UK: Woodhead Publishing Limited.

Starbird, K. & Palen, L. (2102). (How) Will the Revolution be Retweeted?: Information Propagation in the 2011 Egyptian Uprising. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (CSCW 2012). New York: ACM, 6-17.

Starbird, K., & Stamberger, J. (2010). Tweak the Tweet: Leveraging Microblogging Proliferation with a Prescriptive Syntax to Support Citizen Reporting, *Proceedings of the International Conference on Information Systems for Crisis Response & Management Conference* (ISCRAM 2010).

Stork, D. G. (1999). The Open Mind Initiative. *IEEE Intelligent Systems & Their Applications*, *14*(3), 19-20.

Suh, B., Hong, L., Pirolli, P. & Chi, E.H. (2010). Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network, in *IEEE Intl Conference on Social Computing*, IEEE, 177-184.

Surowiecki, James. (2004). *The Wisdom of Crowds.* New York: Doubleday.

Surowiecki, James. (2005). *The Wisdom of Crowds.* First Anchor Books Edition. New York: Anchor Books.

Sutton, J., Palen, L. & Shklovski, I. (2008). Backchannels on the Front Lines: Emergent Use of Social Media in the 2007 Southern California Wildfires. In the *2008 Proceedings of Conference on Information Systems on Crisis Response and Management* (ISCRAM 2008).

Sweney, M. (2011). Al-Jazeera Office Attacked in Egypt protests. *The Guardian.* (February 4, 2011). Retrieved March 15, 2011. http://www.guardian.co.uk/media/2011/feb/04/al-jazeera-office-attacked-egypt-protests

Theiner, G. (2010). Making Sense of Group Cognition: The Curious Case of Transactive Memory Systems. In W. Christensen, E. Schier, & J. Sutton (Eds.), ASCS09: *Proceedings of the 9th conference of the Australasian Society for Cognitive Science*. Sydney: Macquarie Centre for Cognitive Science, 334-342.

Tierney, K, Lindell, M, & Perry, R.W. (2001). *Facing the Unexpected: Disaster Preparedness and Response in the United States*. Washington, DC: John Henry Press.

Toffler, A. (1971). *Future Shock.* New York: Random House.

Tsukayama, H. (2011). YouTube Users Upload Two Days of Video Per Minute. *The Washington Post, Blogs*: *Faster Forward*. (Posted May 25, 2011). Retrieved June 22, 2011. Available at: www.washingtonpost.com/blogs/faster-forward/post/youtube-users-upload-two-days-of-video-per-minute/2011/05/25/AG9Z9DBH_blog.html

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind*, *59*, 236, Oxford University Press, 433-460.

Twitter Blog. (2010). The Evolving Ecosystem. Retrieved May 30, 2011. http://blog.twitter.com/2010/09/evolving-ecosystem.html

Twitter Blog. (2012). Posted March 21, 2012. Accessed June 3, 2012. Available at: http://blog.twitter.com/2012/03/twitter-turns-six.html

van Liere, D. (2010). How Far Does a Tweet Travel? Information Brokers in the Twitterverse, *Proceedings of the International Workshop on Modeling Social Media* (MSM 2010).

Verma, S., Vieweg, S., Corvey, W., Palen, L., Martin, J., Palmer, M., Schram, A. & Anderson, K. (2011). NLP to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency. In the *Fifth International AAAI Conference on Weblogs and Social Media*, 17-21.

Vieweg, S., Palen, L., Liu, S., Hughes, A., & Sutton, J. (2008). Collective Intelligence in Disaster: An Examination of the Phenomenon in the Aftermath of the 2007 Virginia Tech Shootings. *Proceedings of the Information Systems for Crisis Response and Management Conference* (ISCRAM 2008).

Vieweg, S., Hughes, A., Starbird, K., & Palen, L. (2010). Micro-blogging during two natural hazards events: What Twitter may contribute to situational awareness, *Proceedings of the 2010 ACM Conference on Human Factors in Computing System* (CHI 2012). New York: ACM, 1079-88.

Voida, A. Harmon, E. & Al-Ani, B. (2012) Bridging between organizations and the public: volunteer coordinators' uneasy relationship with social computing. In *Proceedings of the 2012 ACM Conference on Human Factors in Computing System* (CHI 2012). New York: ACM, 1967-1976.

von Ahn, L. Blum, M., Hopper, N.J. & Langford, J. (2003). CAPTCHA: Using Hard AI Problems for Security. In *Proceedings of the 22nd International Conference on Theory and Applications of Cryptographic Techniques* (EUROCRYPT 2003), Eli Biham (Ed.). Berlin, Heidelberg: Springer-Verlag, 294-311.

von Ahn, L. (2005). Human Computation. Doctoral Thesis. UMI Order Number: AAI3205378, CMU, Pittsburg, PA,

von Ahn, L. Games with a Purpose. (2006). *IEEE Computer Magazine*, *39*(6), 92–94.

von Ahn, L., Maurer, B., McMillen, C., Abraham, D. & Blum, M. (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science*, 321 (12 September 2008), 1465-1468.

von Ahn, L. & Dabbish. L. (2008). Designing Games with a Purpose. *Communications of the ACM 51*(8) (August 2008), 58-67.

de Vries, Imar. (2011). #Egypt, Social Media, and Determinism. *New Media Studies.* (July 2, 2011). Retrieved October 21, 2011. Available at: http://www.newmediastudies.nl/magazine/egypt-social-media-and-determinism

Wachtendorf, T. (2004). *Improvising 9/11: Organizational Improvisation Following the World Trade Center Disaster*. Dissertation, Univ of Delaware Disaster Research Center.

Wallop, H. (2011). Japan Earthquake: How Twitter and Facebook Helped. *The Telegraph*. (March 13, 2011). Retrieved August 1, 2011. Available at:

http://www.telegraph.co.uk/technology/twitter/8379101/Japan-earthquake-how-Twitter-and-Facebook-helped.html

Wegner, D.M. (1986). Transactive Memory: A Contemporary Analysis of the Group Mind. In B. Mullen, G.R. Goethals (Eds.), *Theories of Group Behavior*. New York: Springer.

Weick, K.E. (1976). *The Social Psychology of Organizing*, 2nd Ed. Reading, MA: Addison-Wesley.

Weick. K.E. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: Sage Publications.

Wikipedia Charts (2011). Wikipedia Charts, All Languages. Retrieved June 22, 2011. Available at: stats.wikimedia.org/EN/ChartsWikipediaZZ.htm

Ye, Y., & Kishida, K. (2003). Toward an Understanding of the Motivation of Open Source Software Developers. In *Proceedings of the 25th International Conference on Software Engineering.* New York: ACM, 419-429.

Zhang, D., Zhou, L., & Nunamaker Jr., J.F. (2002) A Knowledge Management Framework for the Support of Decision Making in Humanitarian Assistance/Disaster Relief. *Knowledge and Information Systems 4*, 370-385.