Supporting Disaster Reconnaissance with Social Media Data: A Design-Oriented Case Study of the 2013 Colorado Floods

Shideh Dashti
University of Colorado
Boulder
shideh.dashti@colorado.edu

Leysia Palen
University of Colorado
Boulder
leysia.palen@colorado.edu

Mehdi P. Heris
University of Colorado
Boulder
mehdi.heris@colorado.edu

Kenneth M. Anderson
University of Colorado
Boulder
kenneth.anderson@colorado.edu

T. Jennings Anderson
University of Colorado
Boulder
jennings.anderson@colorado.edu

Scott Anderson
Federal Highway
Administration
scott.anderson@dot.gov

ABSTRACT

Engineering reconnaissance following an extreme event is critical in identifying the causes of infrastructure failure and minimizing such consequences in similar future events. Typically, however, much of the data about infrastructure performance and the progression of geological phenomena are lost during the event or soon after as efforts move to the recovery phase. A better methodology for reliable and rapid collection of perishable hazards data will enhance scientific inquiry and accelerate the building of disaster-resilient cities. In this paper, we explore ways to support post-event reconnaissance through the strategic collection and reuse of social media data and other remote sources of information, in response to the September 2013 flooding in Colorado. We show how tweets, particularly with postings of visual data and references to location, may be used to directly support geotechnical experts by helping to digitally survey the affected region and to navigate optimal paths through the physical space in preparation for direct observation.

Keywords

Colorado Floods, Crisis Informatics, Engineering Reconnaissance, Extreme Events, Infrastructure Performance, Situational Awareness, Social Media

INTRODUCTION

Engineering fields that examine the structural consequences of extreme events—earthquake, tsunami, landslide, flood, fire, and so on—conduct research that is usually experience-driven. Geotechnical engineers, and researchers in particular, examine the effects of hazards on the built environment either in experimental settings or by documenting detailed observations of damage in the real world to validate engineering design procedures. In the latter case, "perishable data" must be quickly but systematically collected to advance understanding of the underlying damage mechanisms and minimize future damage in similar scenarios. Careful mapping and surveying of damage can yield well-documented case histories that drive the development of empirical engineering procedures used in design. For instance, following the 2011 New Zealand and Japan earthquakes, an international group of geotechnical engineers, structural engineers, and engineering seismologists worked to quickly assess the effects of soil liquefaction and ground shaking on the performance of buildings, slopes, and pipelines as well as factors related to fault rupture, construction practices, and building code limitations (e.g., Bray et al., 2013). More recently, during major flood events in Colorado in September 2013, the influence of flood on mountain slopes, buildings, foundations, and "lifelines" (e.g., utility and road infrastructure) was a subject of great interest to geotechnical engineers and geologists. Such post-disaster engineering field studies are known as *reconnaissance* missions.

However, despite significant reconnaissance efforts by the engineering community, much data about the performance of infrastructure and the timing of various phenomena (e.g., liquefaction or landslide) are lost. This is because there is a time gap between the occurrence of a disaster and the arrival of experts at the site (Bland and Frost, 2012), in addition to the physical limitations in identifying and documenting extensive areas of damage. The difficulty of collecting *in situ* data often necessitates physical model experiments for investigating the mechanisms of damage under similar loading conditions. This paper is written from the point of view that *a better methodology for reliable and rapid collection of perishable hazards data would enhance scientific inquiry, accelerate the building of disaster-resilient cities, and reduce the need for costly experimental studies.*

We propose that the social media data collected from an affected region—that is, self-reports from primarily members of the public about their local environment—in combination with other remote sources of data collection may be used to support geotechnical reconnaissance work. By making use of real-time reporting by those affected in a region, including their posting of visual data, we aim to directly support the reconnaissance work of geotechnical professionals by helping to digitally survey a region under geophysical change and to navigate optimal paths through the physical space for direct observation of sites.

The needs of and opportunities to support reconnaissance work are new in the crisis informatics space (Earle, 2010; Bland and Frost, 2012). Much of the research and development in information systems for crisis response focuses on supporting "situational awareness" for emergency responders and for residents, usually during the height of an emergency period (Terpstra et al., 2012; Imran et al., 2013). However, little attention has been paid to the roles that non-emergency professionals play in recovery and future mitigation phases, both on behalf of the afflicted region and on behalf of other regions in the world that are at risk for similar hazards. Important issues persist when a response officially graduates to the recovery phase and beyond: an affected society can still face second-order effects of disasters that put its welfare at risk for the longer duration. These include, for example, structural safety, pipeline performance, slope safety, or fire following an earthquake. Some of these engineering problems could be mitigated if reliable and comprehensive geotechnical surveys of the situation—of the type engineers rely on—were made available to those who are able to assist in the aftermath of disasters and in anticipation of others elsewhere. Post-event reconnaissance work and how we might support it through the strategic reuse of social media data and other remote sources of information is the focus of this paper, as examined and articulated in response to the real case of the September 2013 flooding in Colorado.

History of Engineering Reconnaissance

The practice of geotechnical engineering relies on observation. Equations based on the laws of physics are used to approximate how different types of loading would impart stresses and strains on geotechnical structures such as building and bridge foundations, retaining walls, and slopes. Observations are then used to calibrate the models with reality. Observations are made through experimental research, and also throughout the construction, operation, and maintenance of civil infrastructure. Because of uncertainties related to loads and the nonlinear response of geotechnical structures to these loads, structures are designed to have a certain margin of safety: a reserve of extra strength and stability. That acceptable margin is typically larger with respect to routine, everyday loads than it is with loads from rare extreme events. Hence, the structure response to routine loading can be difficult to discern. With the significant loads associated with extreme events, much more is revealed about performance; this is why observation during extreme event reconnaissance is so important to the future mitigation of disasters.

Observations are fed back into engineering models. In this way, we can establish new designs that more accurately target the desired level of structural resilience. Safety and economy can be balanced with respect to a clear understanding of tradeoffs. To reap such benefits, a formal program has been established to maximize *in-situ* scientific explorations: the Geotechnical Extreme Events Reconnaissance (GEER) Association, which is a US National Science Foundation (NSF) funded program chartered in 2006. GEER mobilizes teams from an international network of geotechnical professionals to conduct quick response reconnaissance in the field. Such teams face many issues that can challenge systematic investigation: they need to know where to deploy within an affected region to optimize the remaining time before cleanup ensues. They also must pinpoint precise places of geotechnical damage, navigate a damaged road infrastructure, sample across a wide variety of conditions, and be able to ascertain how representative any one place of interest is relative to a larger region.

Following the rare and extreme storm and the subsequent flooding along the Colorado Front Range in September, 2013, the GEER team consulted researchers at the University of Colorado Boulder—a collaboration that was already in place between the authors but awaiting a disaster event, though not necessarily one in their hometown—to determine how the social media activity they monitor could support reconnaissance work, and mitigate the challenges that reconnaissance teams face following a disaster.

Computer-Mediated Communication in Crises and Emergency Response

When disasters strike, remarkable forms of social collaboration emerge to mitigate loss (Tierney and Quarantelli 1989; Fischer 1998; Palen and Liu 2007)—forms that often surprise us by the magnitude of their ingenuity and generosity. Immediately following a disaster, people often turn to on-line sources to find and offer help. As early as the 2004 Indian Ocean Tsunami, people posted photos of the devastation on Facebook (Liu et al, 2008). During the April 16 Virginia Tech (VT) Shooting event, VT students converged onto Facebook to declare themselves safe and well. These destinations rapidly transformed into places where helpers and families of potential victims tried to deduce from an absence of "I'm Okay" messages who might be hurt. Cumulatively, and without the initial intention of doing so, these multiple sites of "distributed cognition" (Hutchins, 1995) or "collective intelligence" (Hiltz and Turoff, 1978) discovered the entire set of fatalities before VT could officially release the names (Vieweg et al. 2008; Palen et al. 2009). Later, in the aftermath of the January 12, 2010 Haiti Earthquake, people acting on behalf of medical groups issued information via Twitter about their activities on the ground (Sarcevic et al 2012).

Paper Objective

Today's technology creates a digital milieu in which these information-producing activities can support intelligent reuse perhaps beyond what the authors envisioned. It is this information during the 2013 Colorado Floods—textual, photographic, and videographic data—that we sought to isolate to realistically consider how it could support a GEER reconnaissance deployment in a case study examination. What follows is a reporting of how this intelligent reuse of social media played out for the reconnaissance effort and to offer guidelines and research directions for using data in this way for professional but non-emergency disaster work.

THE 2013 COLORADO FLOODS: EXAMINING SOCIAL MEDIA SUPPORT FOR RECONNAISSANCE

The Disaster Events

The Front Range of Colorado experienced flooding in July, August, and September of 2013. The July and August storms were localized convective storms in basins west of Manitou Springs. These storms produced approximately 1 inch of rain in 30 minutes over basins that had recently burned in the Waldo Canyon fire of 2012. As a result of the burned watersheds and the subsequent increased runoff and erosion, flooding was larger than the storm totals would otherwise suggest, and debris flows resulted. GEER mobilized a team after the August storm, which was the larger of the first two (Keaton et al., 2013).

The GEER team was on-site seven days after the flood began in August. This was after emergency response efforts: roads had been cleared and many businesses and residences had begun their own clean up. While there were flood and debris deposits that had not been modified—and thus still reflected what had happened—much of the geotechnical evidence had already been lost. Because of this, the team looked to online footage posted on YouTube to get a better understanding of how structures had been affected and found a tremendous amount of information published by the public, either from their properties or cars, mostly sent from smart phones.

Approximately one month later, when much more widespread flooding occurred from what can be described as a 1000-year rain storm (based on the NOAA Atlas 14 frequency analysis) in the northern part of the Colorado Front Range, members of the same GEER team sought public sources of data as a way to support a second round of reconnaissance planning. The rain and floods in September lasted several days, from September 9 to 14, killing four people, isolating townships, and causing an estimated $430 million in state-owned road damage alone. Because the flooding was widespread, it impacted many canyons and closed off access to communities for a long duration. The continued storms also prevented airborne reconnaissance. During this event, social media and other remote sources of information were sought to obtain reconnaissance information, as detailed in the following sections. This latter event is the focus of this paper, though the work of the responding GEER team (Keaton et al., 2013) was informed by the earlier flooding event.

Social Media Data Collection and Analysis

University of Colorado's Project EPIC (Empowering the Public with Information in Crisis) used its Twitter data collection infrastructure (Anderson and Schram, 2011) to collect data during the Colorado event. This infrastructure makes use of a four-node Cassandra cluster to store tweets from Twitter's Streaming API in a reliable and scalable manner (Anderson and Schram, 2012). For this event, collection began at 10 PM MDT on Wednesday, September 11, 2013. For the purposes of this paper, data collected up to and including Friday, September 20, 2013 (the first nine full days of the event) were analyzed in a "quick response fashion," although EPIC continued to collect data past this date in order to also be able to study the long-term recovery phase.

Twitter's Streaming API can accept a list of case-insensitive keywords, user IDs, and geographic bounding boxes as a query; it then returns all tweets that match that query in a continuous stream. For the purposes of this event, we focused exclusively on keyword search terms. These terms were determined by EPIC analysts via monitoring of the public Twitter stream and reading news articles, looking for a set of important terms (e.g., locations, depiction of hazards, etc.) that described the event in as representative a fashion as possible. As new terms emerged we added those as well. The keywords used to collect data during our data collection window are summarized in Table 1, as they were added to the collection. In total, we collected 212,672 unique tweets generated by 57,049 unique Twitter users during the collection window. 2,658 of these tweets were geo-tagged.

Date	Keyword Set Terms
09/11/2013	boulderflood, cowx, nwsboulder
09/12/2013	coflood, cofloods, coflooding, cuboulder flood, #boulder, #cccf, jeffcoflood, waldoflood
09/15/2013	Boulderfloods
09/19/2013	flood gas, flood infrastructure, #boco_trails, cdot, #cofloodrelief
09/20/2013	#coloradostrong

Table 1. The keywords used to collect Twitter data during the 2013 Colorado flooding event.

The social media dataset gathered from September 11 through 20, 2013 was subject to a rapid analysis to evaluate its potential for aiding reconnaissance efforts. The analysis began while data continued to be collected, so that the teams could strategically deploy. Since in this first case we were *developing* the techniques, the analysis was *designed to optimize a reconnaissance deployment*, rather than to comprehensively study all features of the set over a longer duration.

We began by querying the dataset for the September floods along four general attributes and combinations of these attributes: 1) tweets that were encoded with precise latitude and longitude data ("geo-tagged"), 2) tweets that contained obvious URLs to photos and videos; 3) tweets that contained place names; and 4) tweets that contained structural terms as determined by the engineering team. We knew the latter selection of terms by professionals might neglect terms used more likely by the lay population (Sarcevic et al., 2012), but we executed this as a supporting technique when acting quickly in the absence of a rapid solution, to derive terms as they naturally occur in the tweets themselves. Future solutions are expected to be data-driven.

In the spirit of quick response, we determined that the tweets with all attributes would be most valuable because they would provide redundant information on location (geo-tag and place name) and observations (photo or video and description using keywords) that could be used for verification. Such tweets could be used immediately and with more confidence by a reconnaissance team to understand the type and location of impacts and to plan other remote or onsite activities. Another valuable attribute we discovered after the initial derivation was the mention of date and time within the tweet body—this provided additional data about when a photo for example was taken, versus when the tweet was posted. Though this reconnaissance was being conducted within the first days following the worst of the flooding, those precise indications of time enhanced the value. Most tweets have only some of these desirable attributes, however, and they were made valuable by combining them with other data. For example, the geo-tagged tweets were combined with existing hazard maps and satellite or UAV imagery, as discussed in the following sections.

Distribution of Geo-Tagged Tweets in Boulder County during the Flooding

About 1.2% of the tweets contained geographic coordinates (2,658 of 212,672). We imported each geo-tagged tweet into ArcGIS. Figure 1 shows the distribution of all geo-tagged tweets in the Boulder County from September 11th through 23rd, regardless of their contents, to evaluate their density and progression over an extended time period. This figure generally shows a higher density of tweets in the city of Boulder, particularly in the Boulder Creek area, during the first two days of flooding, after which the number of tweets continued to drop. To explore a possible correlation between tweet density (regardless of content) and the expected distribution of flood-related damage, the geo-tagged tweets were overlaid on flood hazard maps provided by the City of Boulder (Figure 2). Further, Table 2 summarizes the number of tweets falling within the high flood hazard zones, within 600ft of high hazard zones, or within 30ft of a major road or highway. The intensity of tweets during the flood period appeared to be primarily concentrated in the higher flood hazard zones. In line with these observations, Earle (2010) found a high density of tweets containing the keyword "earthquake" near the epicenter and in the first 120 sec of the March 2009 earthquake in Morgan Hill, California (magnitude 4.3).

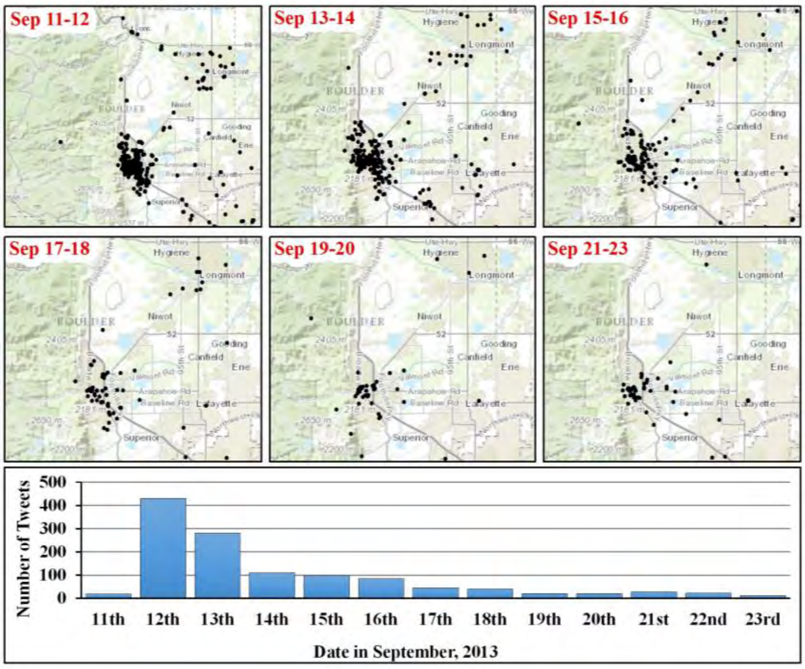
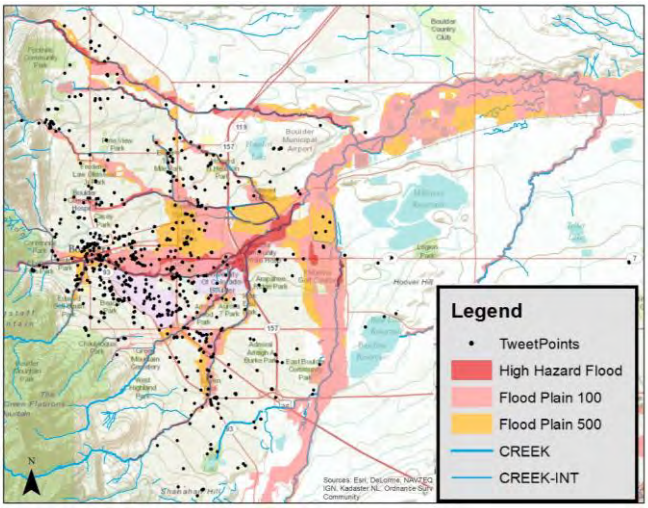


Figure 1. Progression of geo-tagged tweets over time in the Boulder County from September 11 through 23, 2013.

Geo-Tagged Tweets Containing Keywords and Links to Photographs

The geo-tagged tweets were manually filtered to separate those with content in the 140 character-length post that appeared to be directly relevant to reconnaissance activities, specifically the reporting of damage to lifelines (e.g., roads, bridges, walls, water, or sewage lines) and buildings. Forty tweets were identified with the most relevant content, the distribution of which is shown in Figure 3, and the summarized content of which is provided in Table 3. We found the photographs provided by the twitterers (e.g., Figure 3) insightful in identifying the type and extent of damage encountered in different areas of Boulder County and, hence, useful for planning the reconnaissance activities when safe for a site visit.



Sources: Esri, DeLorme, NAVTEQ, IGN, Kadaster NL, Ordnance Survey, Community

Figure 2. Distribution of geo-tagged tweets in the flood plains (September 11 through 23, 2013).

Total No. of tweets in Boulder-Longmont	1145
No. of tweets in high hazard areas	267
No. of tweets within 600 ft of high hazard areas	682
No. of tweets within 30 ft of the main roads	445

Table 2. Tweet distribution in Boulder County (September 11 through 23, 2013).

Damage Category	Tweet Count	Example Content
bridge damage	4	*A bridge collapse means 63rd St in Hygiene is now impassable*
broken sewage	2	*BREAKING Raw sewage from north of Space Sciences (EastCampus) near footpaths. Avoid area.*
High risk bridge	5	*monitoring Fountain Creek at Hwy 24 & Colorado - extremely dangerous water flows*
high risk road	1	*Peoria St in the area of Jewell & Florida is collapsing due to flooding. Asking the public to stay off of Peoria*
path flooded	5	*Report of entire trailer park washed away in Lyons, CO*
property damaged	2	*This cottage washed downhill and is now blocking access to the Flagstaff Viewpoint trail*
road closure	8	*Dispatch: N. 49th Street at Left Hand Creek is washed out*
road flooded	12	*Hwy 36 is completely flooded right after Baseline Road*
sewage back up	1	*Crews working late, pumping out #Boulder sewer lines as people deal with raw sewage backups in their homes.*

Table 3. Content of most relevant geo-tagged tweets reporting damage to infrastructure.

Placename Search

Our content analysis (i.e., Table 3) showed that many of the useful tweets contained locations or names of a road, bridge, or building. To test how well placename keywords could be used to find relevant tweets without limiting the search to the small percentage of geo-tagged tweets, we examined the keyword "36" to represent U.S. Highway 36, which was one of the major highways impacted during the flood and anticipated for field reconnaissance. The database of all tweets (not just geo-tagged) contained 1,911 tweets containing the keyword "36", and most but not all were referring to U.S. Highway 36. No other keyword (e.g., bridge, failure, flood, etc.) was included to constrain this search. Of these 1,911 tweets, 1117 contained a link and 22 were geo-tagged, though most were not specifically geo-tagged to a location on Highway 36 itself. A total of 10 tweets were geo-tagged to the highway itself *and* linked to a photo. This is a relatively small number, but the images provided a clear idea of flooding along a very long reach of the highway that our reconnaissance team could act on, a few examples of which are discussed below.

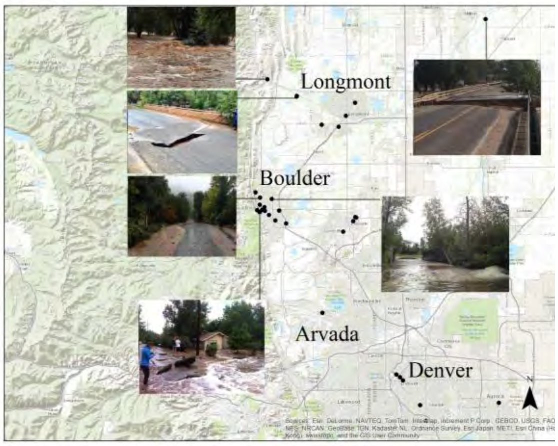


Figure 3. Most relevant geo-tagged tweets reporting damage with example attached photos.

Three geo-tagged and two non-geo-tagged tweets with the "36" keyword indicated failure of an embankment near the Lyon's Bridge on Highway 36. These tweets also included links to photographs (examples shown in Figure 4A and B). The photos were taken on a bridge that was damaged but was still passable and accessible to the public during the flood. The number of similar tweets pointed to the significance of this damage but also, perhaps, the relative ease of public access to this point. One of the tweet photos of the embankment failure (Figure 4B) showed dangerously high water levels, an impassable road, and low cloud cover. The Colorado Department of Transportation (CDOT) attempted to start ground or air reconnaissance during flooding on several occasions, but they were not able to do so for the same reasons indicated in this photo. Further, we took a photo of this same location during field reconnaissance a few days later, which indicated a different water level as expected, although still in the flood stage. The difference in water levels between these photos at different times may be used to learn about basin hydrology and drainage over time, and to evaluate the flood stage in the case of instrumentation damage. Figure 4C shows another tweet photo of extensive damage to County Road 47 along the West Fork Little Thompson River, near its junction with Highway 36. This site was similarly visited and photographed during our field reconnaissance to provide a record of change over time.


A B C

Figure 4. Examples of Tweet Photographs on Highway 36 used during field reconnaissance on September 24, 2013.

Combining Geo-tagged Tweets with other Information Sources to Support Remote Reconnaissance

The content of geo-tagged tweets and their distribution were further investigated in combination with two other remote sources of information to evaluate their combined potential for *remote reconnaissance*. Remote reconnaissance becomes critical when teams cannot get to the field, or cannot survey all areas of an affected area (which is usually the case). Remote reconnaissance may also be done by a virtual support team to further assist those acting on the ground. We examined two remote sources of information: 1) Landsat 8 satellite images captured on September 17 (five days after the start of flooding), produced by NASA and published by USGS; 2) a photo mosaic captured on September 14 by a high resolution DSLR camera mounted on an Unmanned Aerial Vehicle (UAV), provided by Falcon UAV on their website. Because of severe weather conditions, it was not possible to send a reconnaissance team to many of the damaged areas immediately following the flood.

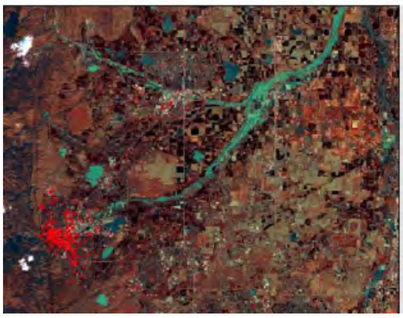


Figure 5. Landsat 8 false color composite bands 7, 4, 1 captured on September 17 with overlaid geo-tagged tweets from September 11-20, 2013 (source: http://landsat.usgs.gov).

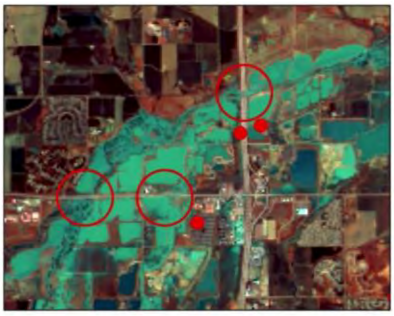


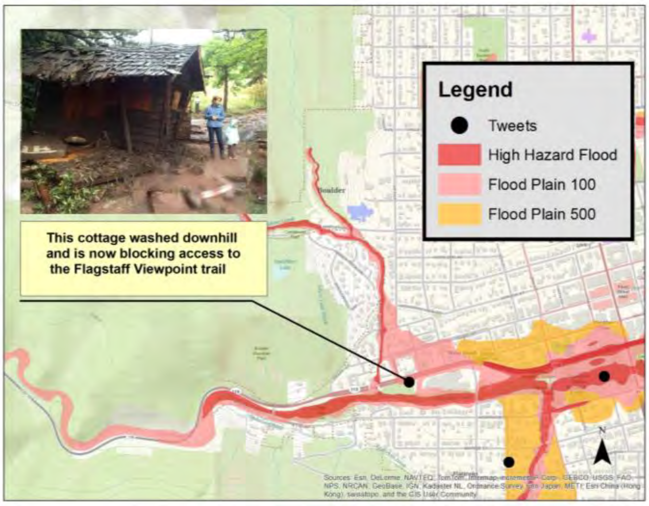
Figure 6. Flooded bridges over I-25 Highway shown on the Landsat 8 image as large circles. Overlaid geo-tagged tweets shown as small filled circles.

Figure 7. Flight path of the UAV and flood damage overlaid on Google map in the City of Longmont on September 14, 2013 (Source: http://sigacts.com/falcon-uav/map.html; their permit status is not known to the authors).

Landsat 8 provides a multi-spectral image with 11 bands that have wavelengths ranging from 0.4 µm to 12 µm, which identify land cover properties including vegetation and soil moisture. Varying the wavelength of the image creates true and false color composites to better distinguish between different surface cover properties. False color composite images are often used to highlight the desired properties of the ground surface (e.g., in this case flooded roads without vegetation). The Landsat 8 image was first captured with its true color composition (bands: Red-4, Green-3, Blue-2) and a spatial resolution of 30 x 30 meters. The image was then sharpened by the panchromatic band (8) with a higher resolution of 15 x 15m. False color composites were then implemented to highlight the flooded areas (bands: Red-7, Green-4, Blue-1), as shown in Figure 5. By using false color composites, specific features or properties may become more visible. Our goal here was specifically to distinguish between vegetation and water. In Figure 5, the turquoise color represents muddy flood water, and the blue color represents deeper and cleaner water. Although the image was captured five days after the flood began, the affected areas are still visible on the map. Figure 6 zooms in on the location of some of the flooded bridges in Longmont. The obtained geo-tagged tweets are also overlaid on the satellite images as red dots.

The high-resolution, UAV images captured on September 14 are overlaid on a Google map in Figure 7, showing the flight path of the UAV. Figure 7 also shows a few UAV photos of flooded roads in the City of Longmont and their locations. Our GEER teams often examine such imagery from sophisticated sources, but we have considered in this case how the human-eye perspective from the ground could be further enhanced by such "sky-views" for remote reconnaissance. We then overlaid the filtered geo-tagged tweets containing the most relevant information (Figure 3) on the hazard map, satellite, or aerial image maps. Most tweets appeared to fall primarily within the high flood hazard zones. Most bridges and roads that were located in the flood plains were expected to experience a high risk of damage, and the tweets and remote data confirmed this pattern. Figure 8 through Figure 10 illustrate how reconnaissance information may be combined from multiple scientific and citizen sources.

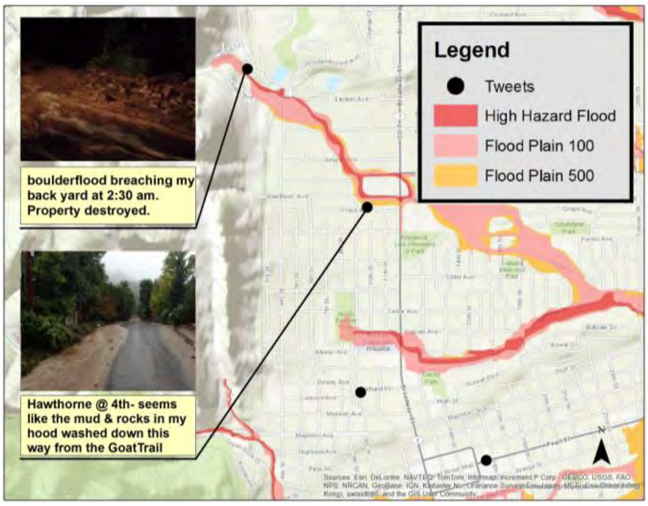
Case 1: A geo-tagged tweet that reported a flooded cottage in the Flag Staff View Point Trail is superimposed on a flood hazard map (Figure 8). The lat-long information falls within the 100-year flood plain zone. The tweet text also indicates written location of the reported damage, while also providing a clear photo of damage *with humans included as scale*. Although the type of damage and structure reported in the tweet content are not detailed enough for systematic reconnaissance measurement, enough information is provided in total to enable a general evaluation of the damage mechanisms and, if necessary and possible, to send a team of experts for further study.



This cottage washed downhill and is now blocking access to the Flagstaff Viewpoint trail

Figure 8. A tweet with reconnaissance relevant information overlaid on the Boulder County flood hazard map.

Case 2: Two tweets reporting flooded streets in North Boulder (Figure 9) lie within the high hazard flood zone and link to photos depicting flood and damage. Importantly, one of the tweets indicates in its text the time of flood breaching the backyard, which is different than the metadata timestamp, showing the potential for more accurately tracking the time progression of events. For example, during an earthquake, following the time progression of damage to geographically distributed lifelines due to soil liquefaction is of particular importance in reconnaissance, which may be obtained from such tweets.



boulderflood breaching my back yard at 2:30 am. Property destroyed.

Hawthorne @ 4th- seems like the mud & rocks in my hood washed down this way from the GoatTrail

Figure 9. Two relevant tweets overlaid on the Boulder County flood hazard map.

Case 3: One tweet reports a flooded pedestrian bridge and a flooded car carried by the stream in the City of Lyons. Downstream, two tweets report a damaged bridge and road closure along Highway 66. These tweets are superimposed on a Landsat 8 image (Figure 10). The tweets are geo-tagged to the location from which the tweets were *posted*. But these coordinates do not *necessarily* coincide with the location of the reported damage. In this case, the Landsat 8 image indicates the geographic coordinates of the tweets follow the expected locations of damage. Like the earlier Highway 36 analysis, the human reference to location in addition to a link to photographs of damage were valuable to reconnaissance planning in addition to the geographic coordinates.

Figure 10. Relevant tweets overlaid on the Boulder County Landsat 8 image.

LESSONS LEARNED AND FUTURE RESEARCH NEEDS

In this paper, we compare the viability of tweet data relative to other scientific sources, and consider how each informs the other for reconnaissance purposes. By making use of real-time reporting by those affected in a region, including their posting of visual data, we show that tweets may be used to directly support engineering reconnaissance by helping to digitally survey a region and navigate optimal paths for direct observation.

In future research, it will be important to examine the *resolution* of data coming from tweets that report damage in *time and space* through other case analyses. This would allow us to further progress the work to a new degree of validation and certainty before the conditions change. It would be advantageous to have a pre-analyzed, data-driven list of terms that tend to be used by the lay public to describe the damage of interest to engineers and scientists. These terms may be used immediately to reduce the large number of collected tweets to a small group that are both geo-tagged and contain relevant reconnaissance information. Future solutions for identifying disaster- and location-specific keywords could be data-driven. Performing word counts and using NLP analysis techniques (Imran et al., 2013; Verma, et al., 2011) may expose hazard-specific and site-specific terms and phrases that the layperson uses to report damage in situ. In addition, a more elaborate campaign that instructs people how to report such damage via tweets, using for example something akin to the hashtag-based syntax of *Tweak the Tweet* (Starbird and Stamberger, 2010; Starbird et al., 2011), may help get better reporting of damage across a region. However, one must be mindful that requesting data from the public can put them in harm's way: in response to requests about particular areas, some may pursue that information with the intention to help the reconnaissance cause, putting themselves at a level of risk to which they may not otherwise be exposed.

SUMMARY

High fidelity, post-event reconnaissance through expert field studies of the effects of disasters on the built environment is critical for improving long-term infrastructural integrity and the resilience of our cities. This paper considers how social media reports about a disaster event could be used to support post-impact engineering reconnaissance work. By combining citizen reports of damage with other remote sources of information (e.g., satellite and high-resolution UAV imagery), and by validating these observations through the reconnaissance work that happened in the wake of the 2013 Colorado Floods, we offer a set of considerations for using such data in support of a growing global, scientific effort to deploy international reconnaissance teams to disaster sites around the world.

ACKNOWLEDGMENTS

We thank GEER for their involvement and support. Portions of this material are based upon work sponsored by the NSF under Grant IIS-0910586.

REFERENCES

1. Anderson, K. & Schram, A. (2011). Design and Implementation of a Data Analytics Infrastructure in Support of Crisis Informatics Research: NIER track. In *Proc. of 33rd International Conference on Software Engineering*, pp. 844-847, Waikiki, Honolulu , HI, USA, May 2011.
2. Anderson, K. & Schram, A. (2012). MySQL to NoSQL: Data Modeling Challenges in Supporting Scalability. *Proc. of the ACM Conference on Systems, Programming, Languages and Applications: Software for Humanity*, pp. 191-202. Tucson, Arizona, USA, October 2012.
3. Bland, H. M., and Frost, J. D. (2012). Opportunities and Considerations for Smart Phone Applications and Mobile Social Media in Post Extreme Event Reconnaissance Data Collection. *Proc. Of 6th Congress on Forensic Engineering*, San Francisco, CA, USA, November 2012.
4. Bray, J. D., O'Rourke, T. D., Cubrinovski, M., Zupan, J. D., Jeon, S. S., Taylor, M., Toprak, S., Hughes, M., Ballegooy, S. V., Bouziou, D. (2013). Liquefaction Impact on Critical Infrastructure in Christchurch. *USGS Technical Report*, March 22, 2013.
5. Earle, P. (2010). Earthquake Twitter. *Nature Geoscience*, 2, pp. 22-222.
6. Fischer III, H. (1998). Response to Disaster: Fact versus Fiction & Perpetuation, 2nd Ed., University Press of America.
7. Hiltz, R. & Turoff, M. (1978). The Network Nation: Human Communication via Computer. MIT Press.
8. Hutchins, E. (1995). Cognition in the Wild. MIT Press.
9. Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013). Practical extraction of disaster-relevant information from social media, pp. 1021–1024.
10. Keaton, J., Anderson, S., Santi, P., Dashti, S. (2013). Geotechnical Effects of Intense Precipitation on August 9, 2013, on Slopes above Manitou Springs, Colorado, that were Burned in the 2012 Waldo Canyon Fire, *Geotechnical Extreme Event Reconnaissance (GEER) Report*, December 2013.
11. Palen, L. & Liu, S. (2007). Citizen Communications in Crisis: Anticipating a Future of ICT-supported Public Participation. *Proceedings of the ACM 2007 Conference on Human Factors in Computing Systems* (CHI 2007), San Jose, CA., pp. 727-736.
12. Palen, L., Vieweg, S., Liu, A. Hughes (2009). Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007 Virginia Tech Event. Social Science Computing Review, Sage, pp. 467-480.
13. Sarcevic, A., L. Palen, J. White, M. Bagdouri, K. Starbird, K. M. Anderson, (2012). "Beacons of Hope" in Decentralized Coordination: Learning from On-the-Ground Medical Twitterers During the 2010 Haiti Earthquake, *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work*, Bellevue, WA.
14. Starbird, K. and L. Palen (2011). "Voluntweeters:" Self-Organizing by Digital Volunteers in Times of Crisis. The *ACM Conference on Computer Human Interaction*, Vancouver, BC, Canada, pp. 1071-1080.
15. Starbird, K., L. Palen, S. B. Liu, S. Vieweg, A. Hughes, A. Schram, K. M. Anderson, M. Bagdouri, J. White, C. McTaggart, and C. Schenk. (2011). 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.
16. Starbird, K. and Stamberger, J. (2010). Tweak the Tweet: Leveraging Microblogging Proliferation with a Prescriptive Grammar to Support Citizen Reporting. *Proceedings of 7th International Information Systems for Crisis Response and Management Conference*, Seattle, Washington, USA, May 2010.
17. Terpstra, T., de Vries, A., Stronkman, R., & Paradies, G. L. (2012). Towards a Realtime Twitter Analysis during Crises for Operational Crisis Management. The 9th Int. ISCRAM Conference. Vancouver, Canada.
18. Tierney, K. & Quarantelli, EL. 1989. Needed Innovation in the Delivery of Emergency Medical Services in Disasters: Present and Future. Disaster Management, 2(2): pp. 70-76.
19. Verma, S., S. Vieweg, W. Corvey, L. Palen, J. Martin, M. Palmer, A. Schram and K.M. Anderson (2011). NLP to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency. To appear in the *Fifth International AAAI Conference on Weblogs and Social Media*, July 2011, Barcelona, Spain.
20. 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 *Conference on Information Systems on Crisis Response and Management* (ISCRAM 2008).