Statistical Journal of the IAOS 31 (2015) 227–236
DOI 10.3233/SJI-150893
IOS Press

A guide to social media emergency management analytics: Understanding its place through Typhoon Haiyan tweets

Cat Graham[a], Chris Thompson[a], Michiko Wolcott[b,*], Joseph Pollack[c] and Minh Tran[b]
[a]*Humanity Road, Boydton, VA, USA*
[b]*Statistics Without Borders, Atlanta, GA, USA*
[c]*Statistics Without Borders, Le Palais, Belle-Île-en-Mer, France*

Abstract. Social media can play a critical role in the dissemination of the information as well as collection of relevant data during natural disasters. The idea of leveraging social media data such as Twitter is intuitively attractive, given their natural ties to mobile devices with obvious disaster response implications.
However, as with any data analysis, the design of the analysis, from the objective definition, the data collection specifics, data management, valuation of the data contents, and analysis methodologies, can determine the ultimate effectiveness and validity of the analysis activities. This is exasperated in the context of disaster response given its time-sensitive nature. The disaster response resources and the analytical resources must work closely together to ensure that the analysis is fit for purpose and meets the ultimate objective. This study discusses the key considerations for such collaboration through an analysis of Twitter data surrounding the 2013 landfall of Typhoon Haiyan in the Philippines.

Keywords: Social media, emergency management, Twitter, Haiyan

1. Introduction: Social media informatics for emergency management

The growth and popularity of social media have provided a rapidly changing field of study for emergency management. Several research studies have investigated the use of social media platforms in crisis situations, such as Starbird's study of social behavior in the context of mass disruptions [1] and Berger's study of influence of social networking communities on disaster response management operations [2].

The application of social media emergency management (SMEM) analysis will not only help identify and reduce risk through collaborative opportunities but will also contribute to long-term disaster risk reduction. This guide was written jointly by Statistics without Borders and Humanity Road to provide insights for emergency managers and analysts leveraging social media data in response to disasters. It addresses research, education and training, implementation and practice for emergency management professionals, while helping the data analyst understand the application of analytics to humanitarian needs. Empowering disaster management with approach and methods for timely decision making can not only help identify current risks for the immediate situation, but also can improve long-term recovery as well as contribute to future disaster risk reduction.

Each disaster and situation will be dynamically different. For this reason, it is difficult to place the exact value on the success of social media emergency management in disaster response, although there is general consensus that the value exists. In practice, navigating the questions between the analyst and the emergency

*Corresponding author: Michiko Wolcott, Msight Analytics, LLC, 2472 Jett Ferry Road, Suite 400-189, Atlanta, GA 30338, USA. Tel.: +1 404 348 4314; Fax: +1 404 348 4315; E-mail: michiko.wolcott@msightanalytics.com.

management professional can be a challenge due to diverse backgrounds of the personnel involved and the dynamic nature of the situation.

We outline two types of application of social media analytics: one as post-disaster assessment and research which aggregates and analyzes data for statistical trending and strategic planning purposes, and the other conducted at the onset, during disaster response, and during recovery phases for rapid assessment and response focused on tactical execution. In general, this guide is meant for the latter, although the principles apply to both.

2. Social media and Twitter

Social media are generally understood to be internet-based applications or platforms that permit users to contribute their own content while accessing the same by others, creating a highly interactive communication environment. There are several commonly used social media platforms, such as Twitter, Facebook, Instagram, Pinterest, and Foursquare among others, each with its own unique features and characteristics.

Twitter allows users to contribute and read short messages of up to 140 characters ("tweets"). A user may also track tweets over mentions of specific topics through the use of the hashtags, subscribe to or unsubscribe from other user's tweets ("follow" or "unfollow"), or forward another user's message ("retweet"). Twitter also keeps limited information about the user's profile which is generally self-reported. The limitation of 140 characters in the tweets have resulted in the development of stylistic particularities unique to the platform with heavy use of abbreviations and other shorthand-like notations.

3. Social media emergency management: Analysis considerations

While the idea of leveraging social media is intuitively attractive, there are a number of key considerations to ensure that the analysis is designed to meet the objective. The opportunity for data analysis must be properly and promptly identified, and the disaster response resources and the analytical resources must work together to determine how to best house, extract, and analyze the data.

3.1. Data management

Data management is a critical area of concern in any analysis of social media data. It requires a good understanding and awareness of issues such as ethical, legal, and societal implications, and its logistics require serious attention and planning.

- Data privacy. Each social media post has been recorded by an individual for uses that may or may not be tied to the research under way, making it highly sensitive. While privacy settings may remove a legal challenge to the use of public data, important ethical considerations remain. Whether the data can be traced directly to an individual or analyzed in such a way as to identify the author, and whether the contents of the message endanger an individual and/or a group, are some of the important issues in data privacy. The need to report intelligence gathered and the protection of the privacy of source of the intelligence must be carefully balanced, and protection of restricted data such as personally identifiable information (PII) and personal health information (PHI) must be strictly observed according to industry standards.
- Ethical uses of data, beyond the legal and contractual requirements, must be considered.
- Widely accepted industry standards in data security must be followed, for data transfer, data storage and retention, etc.

3.2. Data extraction

The parameters and the specifics of the social media data to be extracted for analysis must be carefully defined and understood based on the objectives of the analysis. Considerations include:

- Social media engine. While the population of some countries and geographies are well-versed with the use of Twitter, there are other areas that almost exclusively use other engines such as Facebook with practically no penetration by Twitter. Furthermore, depending on the objective of the analysis, one social media engine may generate more relevant activities than another. It is important to select the engine most appropriate for the geography, the culture, and the analysis objective.
- Availability of a baseline against which the impact of the disaster can be compared. The analysis must be able to isolate the impact of the disaster through comparison against some norm. Since

disaster events often affect communication volume, capacity, and content, having the data from the equivalent non-disaster situation is critical in understanding the true impact of the disaster. This is important for maximizing the value of the data and avoiding biases and misunderstandings which often lead to suboptimal or even incorrect conclusions.

- Speed of the event. Whether the disaster event is a fast-moving one or a slow-moving one has an impact on what data is collected and how.
- Communication trends. Since disaster events are dynamic by nature, it is important to understand the trends in communication.
- Time stamp to be used for data extraction. Whether the time stamp is defined in the local time zone or is simply a time stamp of the system from which the data is extracted is critical not only in the interpretation of the timeline but also in the specification of the timeframe of the data to be included in the extract.
- Relevance of the data to be extracted. Filtering criteria such as country, keywords, hashtags, geolocation, language, type of posts (e.g. organic vs. retweets), and type of poster (e.g., individuals, relief organizations, news organizations, celebrity, etc.) must be carefully considered based on the analysis objective. Different criteria are pertinent for different objectives: some studies examine mostly aggregate trends and may use one set of criteria, while other studies that examine individual tweets for information, such as those pertaining to situation awareness or the identification of tipping points and cascading effects, may use different criteria. Furthermore, the decision to include or exclude retweets is not always straightforward; they may be interpreted as a measure of dissemination, critical in some studies, or they may indicate a substantive similarity to the original message with limited relevancy. These criteria must be determined quickly and appropriately for the intended analysis to fully take place.

3.3. *Implications of social media data on the analysis*

While best analytics practices must be followed – replicability and methodological soundness among others – as with any analysis social media data must be understood in the context specific to disaster response scenarios, with specific implications on the analysis approach and feasibility.

- Geolocation methods. There are a number of ways in which social media data may be geolocated, such as through device GPS, through location associated with IP addresses, and the location information in the user profiles. Each of these has a different level of precision, quality, speed of availability, and interpretability, among other differences. It is important to be sensitive to the requirements of the analysis objective to determine the best approach.
- Type of device. Social media is attractive in the context of disaster response due to its portability and mobility. The device information is critical in understanding the logistics of the communication, among other critical topics.
- Day and time of social media activity. This may impact interpretations around contact logistics of the user and level of difficulty of the contact.
- The source and the basis of the time stamp (e.g. local or system). This has obvious impact on the interpretation of the timing of the posts, such as the actual time of the day relevant to the local community and definitions of the calendar day.
- Geolocation of the specific social media posts. Response operations often participate in social media activities; however, these activities may occur in remote areas which pose challenges in identifying geolocation attributes. For example, tweets by relief organizations while in international waters or airspace present challenges in identifying the country of tweet.
- Impact of relevance. The analyst should continually consider the impact of relevance of the posts on the analysis.
- Language. Beyond its use as a potential filtering criteria, the nature of the social media posts presents unique challenges in identifying the correct language. For example, the limited length of tweet texts leads to heavy use of abbreviations, incomplete sentences, and hashtags (including hashtag-only tweets), and other stylistic deviations from standard written communication are often observed. While standard language identification tools are available, the particularity of the tweets presents a challenge.
- Type of the post and of the poster. Depending on the objective of the analysis, the interpretation of the news posts, retweets, emergency instructions to the public, call for support, etc., and their relative importance are vastly different, making correct categorization of the posts and the posters important.

3.4. SMEM dialogue for the emergency management professionals

For the emergency management professionals, the following are some questions to help the discussion with the data analyst.

Handling and storage of data:

1. What tools will be used to collect data?
2. Who will own the data?
3. How will the data be protected?
4. Where will it be stored?
5. How long will it be stored?
6. How will the results be shared?
7. Who will have access to the data?
8. Will it be anonymized? Who will have access to the raw information?
9. Is there a need for an archive of your social media activity and responses?
10. Which social media accounts need to be monitored and archived?

Creating a baseline; identifying the type of content and trends:

11. Which social media platforms are best sources of data for my geography?
12. Can we create a baseline of data to study now? What questions do you have so that we can establish baseline criteria?
13. What type of information do we want to see in the report (i.e. impacted locations, damage reports, communications challenges, shortages, security issues, types of devices used, best time for marketing messages)?
14. How should we define the time range of our baseline report? Are we studying a location other than our time zone (i.e. local time stamp vs. system time stamp)?
15. What trend information is possible with the data? How and when will it be shared?

Planning the reporting time window, location and language:

16. When will I need the reports, and for how long?
17. What defines actionable for me?
18. Do I want trends? If so, which trends? (I.e. volume is increasing/decreasing, type of device used is changing, number of geolocatable tweets is changing, etc.)
19. Am I interested in information that are only geolocated by GPS, or am I also interested in content geolocated by other means?
20. For which language should I be extracting data?

4. Case study: Typhoon Haiyan

4.1. Background

Typhoon Haiyan (known in the Philippines as Typhoon Yolanda) was one of the deadliest typhoons ever recorded, making landfall on November 7, 2013. The islands of Samar, Leyte, Cebu, Visayas and Siquijor were the most affected. Haiyan caused catastrophic damages to cities and infrastructure, with over two billion dollars of damages, 16 million affected in the Philippines, 4.1 million people displaced, and 6,300 deaths [3]. The damage was mostly caused not by the wind but by the storm surges, knocking out communications, power and transport infrastructure.

As Typhoon Haiyan approached Palau and the Philippines, Humanity Road activated on behalf of the public to monitor for urgent needs and amplify official instructions, publishing its first situation report on November 6, 2013. On November 13, Humanity Road received a phone call in response to our social media situation report. The activation call was for an aid provider that requested situational awareness on urgent needs in their area of operation, specifically to assist in identifying areas for medevac or airdrop of critical supplies. The data set used in this analysis was retrieved based on that specific activation request. In 2014, Statistics without Borders activated to support Humanity Road for the data analysis of Typhoon Haiyan, following a previous engagement in 2013 for the analysis of Hurricane Sandy data.

4.2. Objective of the study

The overall objective of this evaluation was to analyze the tweets from the days surrounding the landfall of the typhoon to 1) identify best practices for data handling, 2) identify analysis approaches for emergency response, and 3) recommend data management approaches. Through this exercise, important considerations and challenges were identified surrounding the use and analysis of Twitter-based data sets for disaster response.

4.3. General data set statistics

The data set for this analysis included tweets from the days following the landfall of the typhoon, from November 7, 2013 through November 11, 2013. As in any other analysis the data set were cleaned before analysis. While over 820,000 tweets were extracted

Table 1
General data set statistics

	Count
Total Valid Tweets Initially Pulled	826,168
Tweets after Additional Filter for Relevance	111,598
Geolocated (Device GPS or Internet Country)	1,680

Table 2
General tweet statistics for the analysis data set

	All tweets	Philippines
Total Relevant Tweets	111,598	620
#Unique Users	66,980	478
#Tweets/User	1.7	1.3
Average Tweet Length	115.1	106.4
#Unique Hashtags	5,826	189
#Hashtags/User	0.087	0.395

originally, additional filters were applied based on further relevance criteria with the resulting data set containing approximately 111,600 tweets. Table 1 shows the basic data set statistics.

The tweets were collected globally, although only 1,680 tweets, or 1.5% of all valid tweets, were geolocated through the device GPS or the IP address. Of those, 620 tweets, or 37%, originated in the Philippines. The location information in the user profile was not used in this analysis. Table 2 shows the basic tweet statistics for the analysis data set after additional relevance filters have been applied. We observe some basic characteristics:

– Users whose tweets are geolocated to the Philippines have a slightly lower number of tweets per user than the users at large for the period.
– Relevant tweets known to have originated from the Philippines are slightly shorter on average than the relevant tweets at large for the period.
– Relevant tweets known to have originated from the Philippines contain more hashtags on average than relevant tweets at large for the period.

4.4. *Filtering for relevance*

While filters and keywords were already applied to the original data set, it remained important to further filter the entries for this analysis. The difference between the original filters and the additional filters is apparent in the resulting top-10 hashtags as can be seen in Table 3. By coincidence, the Miss Universe pageant was held around the time Typhoon Yolanda made landfall, in which Miss Philippines was a finalist. Having a sound understanding of the relevance is critical, and these results reinforce this need prior to analysis.

Table 3
Top-ten hashtags with and without additional filter

Rank	Original filter only	With additional filter
1	#yolandaph	#haiyan
2	#haiyan	#yolandaph
3	#philippines	#philippines
4	#yolanda	#typhoonhaiyan
5	#missuniverse2013	#yolanda
6	#dt	#rescueph
7	#typhoon	#typhoon
8	#rescueph	#filipinas
9	#reliefph	#news
10	#ph	#vietnam

Table 4
Language of tweets

Language	#Tweets	%Tweets
English	94,294	84.5%
Spanish	3,703	3.3%
French	3,535	3.2%
Tagalog	2,567	2.3%
Malay	442	0.4%
Indonesian	250	0.2%
Other/Undetermined	6,807	6.1%
Total	111,598	100.0%

4.5. *Language*

The importance of language when analyzing social media data is intuitively understood, since the post themselves are of interest in addition to the geospatial and other quantitative information. However, the particularities of Twitter makes language identification challenging; the length of messages, heavy use of hashtags and abbreviations, and variations in users' communication styles require considerations beyond straightforward use of standard language identification tools. Furthermore, analyzing social media data in countries such as the Philippines presents additional challenges; the country has several languages in use, of which English is only one, and issues such as variations in transliteration can add to the challenges. For this analysis, the R package textcat [4] was used to classify tweets into languages. Some adjustments were made after the initial classification due to the particularity of Twitter texts. The results, shown in Table 4, indicate that globally, the tweets in this data set were composed most frequently in English (85%) with Spanish and French used by only 3.2% of tweets each.

4.6. *Daily tweet volumes*

Examining tweet volumes over time may highlight important trends. Figure 1 compares the volume trend

Table 5
Tweet device/platform, all countries

(a) Including non-geolocatable tweets

Platform	% of Tweets
Twitter.com	16.5%
Web	14.8%
Twitter for iPhone	11.4%
Twitterfeed	9.9%
Twitter for Android	9.0%
Tweetdeck	5.2%
Facebook	4.9%
Hootsuite.com	3.1%
iPad	3.0%
dlvr.it	2.9%
ifttt.com	2.3%
Twitter for Blackber	2.1%
Twitter.com (Mobile)	1.1%
Instagram	0.7%
Tweetbot for iPhone	0.6%
Twitter for Mac	0.4%
Tweetbot for Mac	0.1%
Foursquare	0.0%
WinPhone	0.0%
Windows	0.0%
Other	11.5%
Unknown	0.3%
Total	100.0%

(b) Geolocatable tweets only

Platform	% of Tweets
Twitter for iPhone	38.2%
Twitter for Android	17.5%
Web	12.7%
Other	8.3%
Instagram	6.4%
iPad	4.3%
Twitter for Blackberry	2.7%
dlvr.it	2.7%
Tweetbot for iPhone	2.4%
Foursquare	1.7%
Twitterfeed	1.6%
Tweetbot for Mac	1.0%
Twitter.com	0.4%
Twitter.com (Mobile)	0.1%
Hootsuite.com	0.1%
Total	100.0%

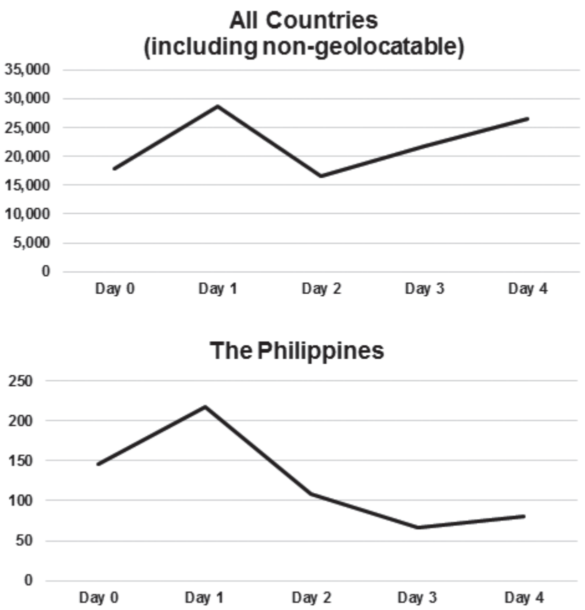

Fig. 1. Daily tweet volumes.

of relevant tweets from the Philippines to that of relevant tweets globally.

Speculations for these phenomena may include the following:

– Interruptions in electricity and communication infrastructure in the Philippines may have contributed to the decrease in volume while there was an increase outside of the Philippines. Emergency management professionals and researchers must be sensitive to such interruptions which may impact the data.
– As Day 0 is the day of the event, the increase in tweet volume on Day 1 globally may reflect initial reaction by the global public. Then as the general awareness increases, the tweet volume picks up in Day 3 and Day 4. It would have been interesting to see when the volume tapered as the news articles, blogs, and calls for support started to subside in the rest of the world.

4.7. *Device/platform*

A platform can be either hardware (tablets, desktops, or mobile phones), or a linked third-party application such as Instagram or Facebook. Such information may, for example, help identify communication patterns or target the scope or channel of potential outreach campaigns. For specialized application development, understanding such patterns and behaviors guides the choices made to encourage usage and adoption.

Table 5 shows the device or platform used for the tweets globally. Considering all relevant tweets globally, we find a highly fragmented range of platforms.

Table 6
Tweet device/platform, the Philippines

Platform	Total
Twitter for iPhone	34.8%
Twitter for Android	25.0%
Instagram	14.2%
Foursquare	4.4%
Twitterfeed	4.0%
iPad	3.5%
Twitter for Blackber	3.5%
Web	0.8%
Twitter.com	0.6%
Tweetbot for iPhone	0.6%
Other	8.4%
Total	100.0%

Table 7
Countries of origin, geolocated tweets

Country	% of Tweets
Philippines	36.9%
US	25.4%
Great Britain	5.6%
Canada	4.5%
Vietnam	2.1%
Other	25.4%
Total	100.0%

Fig. 2. Geolocated tweets in the Philippines.

When limited to only geolocated tweets globally, the mobile phone applications are the most prevalent, with 56% of all tweets coming from mobile applications. Web-based applications account for 13% of the total. Depending on the objective of the study, the platforms with smaller shares may be of high importance.

The geolocated tweets from the Philippines show a pattern different from the geolocated tweets as a whole, as seen in Table 6. This may be because the general awareness about the use of social media during emergencies is relatively well-developed and the use well-supported by the government in the Philippines. In addition, the use of Instagram and other third-party applications such as Foursquare is much higher in the Philippines, with potential implications on the choice of social media applications for data gathering, since some applications may have richer content.

4.8. *Geolocation*

The basis of insights as the one above is the geolocatability of the tweets. Table 7 shows the breakdown of the countries of the relevant geolocated tweets for this analysis. Of the relevant geolocated tweets, 37% come from the Philippines; another 25% from the U.S., followed by tweets from the Great Britain, Canada, and Vietnam. While this shows a global interest and awareness in the event, capturing the specific motivation for the tweets will depend on the objective of the analysis, with important implications in the design of data collection to which emergency management professionals and analysts must be sensitive.

4.9. *Geo-temporal analysis*

Examining the use of social media across space and time may signal a change or development in the behavior of people and their use of social media. For example, the results of this type of analysis may be useful for the design and prioritization of outreach programs as well as top-down communication from the government towards affected populations.

Using a subset of data collected, analysts can provide a geographic display of tweets over time for a specific location. This might be tweets about urgent needs, damage or some other topic. The maps in Fig. 2 are for a subset of geo-tagged tweets from the Philippines based on a list of keywords indicating urgent needs. While the majority of tweets come from Manila, a change over time is indeed observed; there is a decrease in the tweet volume over time including in the region of Cebu. Also observed is the virtual absence of tweets from areas that were heavily affected by the typhoon. This may imply that social media data sets have the intrinsic bias that they depend on access to technology and/or the prevalence of its use.

4.10. Data security and privacy

In order to ensure security and privacy protection of the data, precautions were taken according to generally accepted and reasonable standards. The access to the data were limited to the named persons, and a data security and protection agreement was executed between Humanity Road and the named analysts with access to data, with the following terms:

- Sensitive data must be stored on an encrypted disk and/or in a secure environment that prevents access from persons beyond those authorized.
- Copies and duplication of data must be kept to minimum necessary.
- All data with sensitive information must be removed from equipment not owned by or under the control of Humanity Road within 120 days of project completion.
- Any transfer of sensitive data (email attachments, email body, FTP, removable media, etc.) must be encrypted with at least the minimum level of encryption.
- Encrypted data file and password/key for unencrypting it should not be transmitted via the same channel.
- Sensitive data must be encrypted with PGP, AES-128, or equivalent or stronger.
- All back-ups and copies due to operational process must adhere to the same standards.
- Sensitive information was defined as any identifiers that could be tied back to an individual person or household. This included names, date of birth, address, national ID, account number of any type, Tweet IDs, Twitter handles, and Tweet URLs, among others, as well as any content that could be used to determine the identity of the individuals through triangulation and deduction.

4.11. Comparison to another study: Hurricane Sandy

One of the purposes of statistical analysis is to provide interpretable and comparable results. Since a similar study was conducted by Humanity Road following the landfall of Hurricane Sandy in 2012, an attempt was made to draw comparisons between the two projects.

Several key differences were observed. For example, the data collected during Hurricane Sandy had a higher percentage of geolocated tweets than that of Haiyan (36–67% depending on factors for Sandy vs. 1.5% for Haiyan). However, it is difficult to compare the two studies, because the relevance filters were substantially different. In fact, due to the dynamic nature of emergency management, it is generally difficult to expect one study to be directly comparable to another, further highlighting the importance of the filtering criteria and therefore the need for careful planning and design of the data collection.

4.12. Challenges in this analysis

One of the limitations in this analysis was the lack of baseline against which to compare the results. There was no data collected to establish the baseline of volume and location apart from the storm's landfall, making it difficult to separate the impact of the landfall from the norm. It would have been desirable to determine whether disaster and post-disaster social media behavior was in fact much different than that found under normal conditions.

4.13. Summary and recommendations

In summary, Humanity Road and Statistics without Borders jointly propose the following minimum set of items to be communicated in any analysis of social media for emergency management:

- Relevance
- Geolocatability
- Geography
- Language
- Objective vs. content
- Device vs. impact of the disaster to the infrastructure
- Speed required for obtaining insights (real-time vs. research)

5. Epilogue: Collaborations and relationship building

Innovation in social networking and digital communications and the birth of digital and technology charities have made a diverse set of services, tools and experts available to emergency management and aid responders to augment their response. Corporate responsibility programs as well as digital humanitarian service organizations now offer hardware such as computers, laptops, cell phones, satellite receiver, repeaters and wifi; software programs such as ESRI's ARCGIS, cloud services such as those offered by Microsoft Disaster Response and NCOIC.org and digital services such as social media monitoring, statistical analytics, translations and even GIS Mapping. In many cases these services are at no cost to local communities impacted by disaster. Transformation through innovation is possible through discovering, testing and leveraging these digital humanitarian and technology partners.

By including digital humanitarian response organizations in emergency response planning and through guided dialogues between social media analysts and emergency management professionals, the disaster response cycle can be greatly improved. Participatory recovery planning using social networks as well as social media and technology partners such as Humanity Road and Statistics without Borders should be included in that dialogue. These types of surge support partnerships can assist with analytics, augment emergency management and community responses, can assist to rapidly identify populations at risk, with urgent needs, as well as help to identify how to reach that population for messaging. Taking initial steps toward planning social media tactical and strategic goals is critical to monitoring and understanding the impact to the community, infrastructure and communications.

As part of its mission to test emerging technology to improve disaster response, the tweets collected in this Haiyan data study were retrieved using Scanigo, a social media analytics tool from Progeny Systems which is in TRL7 stage of readiness at the time of this writing. Humanity Road provided Progeny Systems with subject matter expertise and testing of this tool under a small business innovation research grant. Incubator projects funded by research grants, and other local, national or global funding arms are helping to accelerate knowledge management and fuel technology for good. It is also helping to accelerate the testing and adoption of practices in the field of emergency management response in a rapidly changing technology environment.

In response to Hurricane Sandy as well as Typhoon Haiyan, Humanity Road volunteers performed such tasks as data mining, translations (critical support provided by Translators without Borders), triage of emerging data, requesting clarifications, documenting local contact information, routing urgent needs and resolving geolocation discrepancies, using emerging tools to aggregate, filter and triage urgent needs. The operational model used by Humanity Road, "AFTER" (Aggregate, Filter, Triage, Export, Report) became a repeatable process. The information collected and shared by Humanity Road was received and used by organizations such as United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA), Multi-National Force response teams including the U.S. Navy, Americares and other traditional relief organizations via ReliefWeb, All-Partners Network (APAN) and HumanitarianResponse.info.

In 2013, Secretary-General Ban Ki-Moon announced the World Humanitarian Summit (WHS), an initiative by the United Nations (UN) to improve humanitarian action. Through the two-year consultation process, the aim is to build a more inclusive and diverse humanitarian system by bringing all key stakeholders together to share best practices and find innovative ways to make humanitarian action more effective. The process is being managed by the UN Office for the Coordination of Humanitarian Affairs (OCHA) and will culminate in a global summit in 2016. The summit focuses upon four thematic areas, 1) humanitarian effectiveness, 2) reducing vulnerability and managing risk, 3) transformation through innovation and 4) serving the needs of people in conflict.

Transformation through innovation is made possible in today's technology enriched world. We not only have the opportunity to improve the disaster response through technology-enabled tools but through collaborative relationships with organizations such as the Digital Humanitarian Network, and their partners. Digital solution teams like ESRI, Translators without Borders, Statistics without Borders and Humanity Road work together in building rapid responses to quickly identify urgent needs, as well as opportunities for improvement in information management early in the response in a period of information overload.

The report "Humanitarianism in the Network Age" [5] examines the implications for how a world of increasingly informed, connected and self-reliant communities will affect the delivery of humanitarian aid. It lays out some of the most pertinent features of these new technologies, such as SMS, social media and others, and identifies the opportunities and difficulties in applying them. Through their interactions, these groups can work toward rapidly improving disaster response as well as communications on the ground, and through relationship building and collaboration in joint projects with these industry partners, disaster recovery time and cost can be reduced.

Acknowledgements

This paper is closely based on a joint work between Humanity Road and Statistics without Borders. The original guidebook was previously released on November 6, 2014.

References

[1] K. Starbird, "Crowdwork, Crisis and Convergence: How the Connected Crowd Organizes Information During Mass Disruption Events." Ph.D. dissertation, University of Colorado, 2012.
[2] N. Berger, "Digital Humanitarianism: How Do Social Networking Communities Contribute to Humanitarian Response after Disaster?" M.A. thesis, Oxford Brookes University, 2013.
[3] The Republic of the Philippines National Disaster Risk Reduction and Management Council (2014). "NDRRMC Update: Updates re the Effects of Typhoon 'Yolanda' (Haiyan)." Accessed March 23, 2015. https://web.archive.org/web/20141006091212/http://www.ndrrmc.gov.ph/attachments/article/1177/Update%20Effects%20TY%20YOLANDA%2017%20April%202014.pdf.
[4] K. Hornik, P. Mair, J. Rauch, W. Geiger, C. Buchta and I. Feinerer, The textcat Package for n-Gram Based Text Categorization in R, *Journal of Statistical Software* **52** (2013), 1–17.
[5] United Nations Office for the Coordination of Humanitarian Affairs. "Humanitarianism in the Network Age." OCHA Policy and Studies Series. New York: United Nations, 2013.