Verifying Crowdsourced Social Media Reports for Live Crisis Mapping: An Introduction to Information Forensics

Patrick Philippe Meier
patrick@iRevolution.net
www.iRevolution.net/bio

Feedback welcome

Verifying Crowdsourced Social Media Reports for Live Crisis Mapping: An Introduction to Information Forensics

Abstract

False information can cost lives. But no information can also cost lives, especially in a crisis zone. Indeed, information is perishable so the potential value of information must be weighed against the urgency of the situation. Correct information that arrives too late is useless. Crowdsourced information can provide rapid situational awareness, especially when added to a live crisis map. But information in the social media may not be reliable or immediately verifiable. This may explain why humanitarian (and news) organizations are often reluctant to leverage crowdsourced crisis maps. Many believe that verifying crowdsourced information is either too challenging or impossible. The purpose of this paper is to demonstrate that concrete strategies do exist for the verification of geo-referenced crowdsourced social media information. The study first provides a brief introduction to crisis mapping and argues that crowdsourcing is simply non-probability sampling. Next, five case studies are analyzed to demonstrate how different verification strategies work. In conclusion, the paper argues that some information is ultimately better than no information since some information can at least be verified.

Key words: crowdsourcing, crisis mapping, verification, social, media

1: Introduction

The most significant revolution in cartography over the past ten years has perhaps less to do with geography than with time. The radical shift from static, "dead" maps to live, dynamic maps requires that we re-conceptualize how we think about maps and the way we use them.[1] The growing volume of real-time geo-referenced data is largely responsible for this shift towards real-time. Physical sensors around the world generate a considerable amount of this data but so do humans. Indeed, human beings equipped with mobile phones make for formidable mass multi-media sensors. This evolving network of human sensors has even been described as a new "nervous system" for our planet—a crowdsourced nervous system that generates a significant amount of real-time data via SMS and social media platforms like Twitter, Facebook, YouTube and Flickr—particularly in times of crisis. The live crowdsourced maps deployed in response to the disasters and crises that have occurred in Haiti and Libya are possible thanks to this new nervous system. In Haiti, micro-needs assessments were crowdsourced via SMS and then mapped in near real time. In Libya, the UN used a crowdsourced social media map of the country to provide great contextual information on the evolving situation.

The rise of live crisis maps while exciting does present a myriad of new challenges. Perhaps chief among them is the verification of crowdsourced data. Humanitarian organizations cannot run the risk of making decisions based on unreliable data. Indeed, they cite this specific concern to explain their hesitation in considering the use of crowdsourced information more seriously.[2] This concern is certainly understandable. Can reports from the crowd be trusted? How does one verify crowdsourced information in near real-time? Is verification possible under such strict time constraints? Many humanitarian professionals are at best skeptical (Tapia *et al.* 2011). In contrast, respected journalist Mark Colvin seems to think the verification of crowdsourced information is possible based on his experience in covering the 2009 Iran elections:

[1] What is Crisis Mapping? An Update on the Field and Looking Ahead: http://iRevolution.net/2011/01/20/what-is-crisis-mapping

[2] Seeking the Trustworthy Tweet: Can "Tweetsourcing" Ever Fit the Needs of Humanitarian Organizations? http://iRevolution.net/2011/06/05/tweetsourcing

> "There was what the intelligence people call a lot of static or chatter - a vast amount of miscellaneous material coming out of Iran on Twitter, with no way of verifying it absolutely. But if you had a sharp eye for detail, you could pick up indications of whether someone was reliable or not. If it referred to protest action in a particular square, were other people saying the same thing? Quite often I would find three twitter sources, all apparently in different parts of the crowd, reporting on the same event, though from a different perspective. Time came into the judgment as well. Was this someone whose tweets yesterday and the day before had been proved to be true by other reports later? One twitter user, Change_for_Iran, reported in a series of messages being under siege in a dormitory. Another Twitterer, from another part of the campus, reported seeing the same thing from another angle. A day later, pictures started coming through of the damage done in the attack. It starts to add up to something like credibility" (Colvin 2010).[3]

The purpose of this study is to delve deeper into these types of strategies and create a more comprehensive how-to guide on verifying crowdsourced information for live crisis mapping projects.[4] The point is to demonstrate that concrete, replicable steps can be taken to verify crowdsourced data. While some uncertainty may very well remain, this uncertain can be minimized by following systematic strategies.

The study is structured as follows. The first section introduces the new field of live crisis mapping. The second outlines the opportunities and challenges that crowdsourcing crisis information presents. Section three analyzes 5 real world case studies that comprise various efforts to verify social media. The five case studies are: Andy Carvin and Twitter; Kyrgyzstan and Skype; BBC's User-Generated Content Hub; the Standby Volunteer Task Force; and U-Shahid in Egypt. The final section concludes the study with specific recommendations.

[3] See also Wag the Dog, or How Falsifying Crowdsourced Data Can Be a Pain: http://iRevolution.net/2010/04/08/wag-the-dog

[4] See also How to Verify Social Media Content: Some Tips and Tricks on Information Forensics: http://iRevolution.net/2011/06/21/information-forensics

2: An Introduction to live Crisis Mapping

The proliferation of live maps is driven by the increasing availability of real-time geo-referenced data and new mapping technologies that are often free, open source and easier to use than earlier, proprietary systems. This new frontier in the field of geography is commonly referred to as "neogeography" and consists of

> "techniques and tools that fall outside the realm of traditional GIS, Geographic Information Systems. Where historically a professional cartographer might use ArcGIS, talk of Mercator versus Mollweide projections, and resolve land area disputes, a neogeographer uses a mapping API like Google Maps, talks about GPX versus KML, and geotags his photos to make a map of his summer vacation. Essentially, Neogeography is about people using and creating their own maps, on their own terms and by combining elements of an existing toolset. Neogeography is about sharing location information with friends and visitors, helping shape context, and conveying understanding through knowledge of place" (Turner 2006).[5]

The birth of neogeography is often traced back to Google's acquisition of Keyhole Inc. in 2004, which lead to the launch of Google Earth that same year. Google Maps went live shortly thereafter. Together, these mapping platforms went a long way to democratize interactive mapping and broaden public access to satellite imagery. In 2007, the Harvard Humanitarian Initiative (HHI) at Harvard University launched a 2-year Program on Crisis Mapping and Early Warning to study the potential use of live mapping technologies in humanitarian response.[6]

The focus on crisis meant that collecting and displaying information in real-time was imperative. The program thus catalyzed conversations between a wide number of technology professionals, geographers and seasoned humanitarian practitioners. Recognizing the tremendous potential that

[5] See also Neogeography and Crisis Mapping Analytics: http://iRevolution.net/2009/02/24/neogeography-and-crisis-mapping-analytics

[6] The author, Patrick Meier, co-founded and co-directed this program with Dr. Jennifer Leaning. For more information on the initiative, please see: http://hhi.harvard.edu/programs-and-research/crisis-mapping-and-early-warning

existed, HHI launched the International Network of Crisis Mappers, a global network of some 2,500 members in over 120 countries actively interested in the application of live mapping technologies to crisis situations. Established in 2009, the Crisis Mappers Network has since become an important part of the neogeography story.

Another milestone was the launch of the first Ushahidi map in 2008.[7] This simple web-based platform allowed Kenyans to report human rights violations during the post-election unrest. They submitted these reports via web-form, email and SMS. Reports from the mainstream media were also mapped. Some 20,000 Ushahidi maps have been launched in more than 130 countries since 2008. The launch of a hosted version of the Ushahidi platform—Crowdmap—in 2010 accounts for the majority of these 20,000 maps. What is perhaps novel about the Ushahidi mapping technology is that it is free, open source and easier to use than proprietary tools. In addition, the information mapped on the platform is often crowdsourced live rather than collected months later. This is particularly true for crisis mapping applications of the Ushahidi platform. Of note are the crisis maps launched in Haiti, Chile, Pakistan, Russia, Tunisia, Egypt, New Zealand, Sudan, Libya and most recently Somalia.[8]

The use of Ushahidi technology by the Standby Volunteer Task Force (SBTF) for live mapping has spurred much of the recent conversations about the verification of crowdsource data for crisis mapping—specifically SMS and social media content.[9] The SBTF is a global network of some 700 volunteers from more than 60 countries who are trained in live crisis mapping operations. The group was instrumental in Libya when the UN Office for the Coordination of Humanitarian Affairs (OCHA) activated the SBTF to provide a live, crowdsourced social media of the

[7] http://www.Ushahidi.com

[8] See for example, Haiti and the Power of Crowdsourcing: http://iRevolution.net/2010/01/26/haiti-power-of-crowdsourcing; Information and Communication Technology in Areas of Limited Statehood: Russian Fires: http://iRevolution.net/2011/04/03/icts-limited-statehood; Crisis Mapping Egypt: Collection of Protest Maps: http://iRevolution.net/2011/01/29/crisis-mapping-egypt; How Egyptian Activists Kept Their Ushahidi Project Alive Under Mubarak: http://iRevolution.net/2011/05/25/u-shahid-interviews; Crisis Mapping Sudan: Protest Map of Khartoum: http://iRevolution.net/2011/01/30/protest-map-of-khartoum; Crisis Mapping Libya: This is No Haiti: http://iRevolution.net/2011/03/04/crisis-mapping-libya; Crowdsourcing Satellite Imagery Analysis for UNHCR-Somalia: http://iRevolution.net/2011/11/09/crowdsourcing-unhcr-somalia-latest-results

[9] http://blog.standbytaskforce.com

unfolding situation (LibyaCrisisMap.net). The SBTF's verification strategies forms one of the five case studies analyzed in section 3.

The next section considers the advantages and disadvantages of using crowdsourcing as a methodology for collecting information in near real-time.

3: Crowdsourcing: A form of non-probability sampling

The concept of crowdsourcing may be relatively new to many in the humanitarian sector, but when it comes to statistics, crowdsourcing is a well-known and established sampling method.[10] The crowdsourcing of crisis information is simply an application of non-probability sampling. In the field of statistics, this sampling technique describes an approach in which some units of the population have no chance of being selected or where the probability of selection cannot be accurately determined. An example of non-probability sampling is convenience sampling. The main drawback of such techniques is that "information about the relationship between sample and population is limited, making it difficult to extrapolate from the sample to the population." Contrast this approach to probability sampling in which every unit in the population being sampled has a known probability (greater than zero) of being selected. This approach makes it possible to "produce unbiased estimates of population totals, by weighting sampled units according to their probability selection" (IAE 2011).

The point, however, is not to justify the use of crowdsourcing for statistical analysis but to demonstrate that this sampling technique is a well-known methodology for the collection of information. Using non-random sampling to collect information has obvious disadvantages when compared to random sampling. But several compelling reasons exist for using this technique in spite of those disadvantages. For example, non-probability sampling is a quick way to collect and analyze data in a range of settings with diverse populations. The approach is also a cost-efficient means of greatly increasing the sample, thus enabling more frequent measurement.

[10] See Demystifying Crowdsourcing: An Introduction to Non-Probability Sampling: http://iRevolution.net/2010/06/28/demystifying-crowdsourcing

Crowdsourcing can also be a form of exhaustive sampling, such as the crowdsourced damage assessment of aerial imagery following the Haiti earthquake.[11] In some cases, the non-probability sampling may actually be the only approach available—a common constraint in a lot of research, including many medical studies, not to mention sudden-onset humanitarian crises.

The method is also used in exploratory research, e.g., for hypothesis generation, especially when attempting to determine whether a problem exists or not. Non-probability sampling can save lives, many lives. Much of the data used for medical research is the product of convenience sampling. When a patient walks into a doctor's office or becomes hospitalized, that is not a representative sample. In addition, emergency phone calls to 911 in the US and 999 in the UK are also examples of crowdsourced crisis information.[12] It would be absurd to abolish this crowdsourcing mechanism on the basis that the information collected is non-representative of the overall population because it is a form of non-probability sampling.

Perhaps the main disadvantage with the use of non-random sampling is that the sample is not necessarily representative of the overall population. This limitation is further compounded by the fact that the identity of those individuals who comprise the sample is not readily known when crisis information is crowdsourced remotely, e.g., via the Web or SMS. In other words, information collected using non-random sampling methods carried out at a distance may yield a slice of information that says nothing about the overall population and which may not even be trustworthy. This perhaps explains why the notion of "bounded crowdsourcing" has entered the discourse and debate on the verification of crowdsourced data. The point of bounded crowdsourcing is to begin with a trusted network of participants and to expand this network over time. Participants can be selected in a way that they represent the broader population but more importantly they are selected based on trust. This overcomes the immediate need to verify the data they generate.

While "bounded crowdsourcing" may be a foreign term to those in the humanitarian community, the technique is a well-known sampling method in statistics: purposive sampling. This approach

[11] http://gfdrr.org/gfdrr/labs
[12] See also Calling 911: What Humanitarians Can Learn from 50 Years of Crowdsourcing: http://iRevolution.net/2010/09/22/911-system

involves targeting experts or key informants. The number of informants can then be increased over time. Snowball sampling is one way to do this. In snowball sampling, a number of individuals are identified who meet certain criteria but unlike purposive sampling they are asked to recommend others who also meet this same criteria—thus expanding the network of participants. Although these "bounded" methods are unlikely to produce representative samples, they are more likely to produce trustworthy information.

In addition, there are times when it may be the best—or only—method available. To be sure, a recent study that analyzed various field research methodologies for conflict environments concluded that snowball sampling was the most effective method (Cohen and Arieli 2011). Indeed, snowball sampling is especially useful when trying to reach populations that are inaccessible or hard to find. In contrast, probability sampling often requires considerable time and extensive resources. Furthermore, non-response effects can easily turn any probability design into non-probability sampling if the "characteristics of non-response are not well understood" since these modify each unit's probability of being sampled. This is not to suggest that one approach is better than the other since this depends entirely on the context and research question. Just like any other type of data, crowdsourced data is confronted with issues of reliability and validity.

The next section analyzes several case studies to compare different approaches to verifying crowdsourced data.

4: Challenges in the verification of crowdsourced data

Verifying crowdsourced data is certainly a challenge and in some situations not possible. But the five short case studies below describe various approaches that demonstrate the challenge is not always insurmountable.[13] The first case focuses on Andy Carvin's novel verification efforts using Twitter during the escalating crisis in Libya. This example was selected to demonstrate that investigative journalism is possible via micro-blogging. The second focuses on an

[13] See original blog post on How to Verify Social Media Content: Some Tips and Tricks on Information Forensics: http://iRevolution.net/2011/06/21/information-forensics

innovative strategy that leveraged Skype to verify information during widespread violence in Osh, Kyrgyzstan in 2010.[14] This example demonstrates the value of a snowball sampling approach. The third example examines the strategies that the BBC's User-Generated Content (UGC) Hub employs to verify information in the social media space. This case highlights the importance of working across different media. The fourth case study describes the strategies that the Standby Volunteer Task Force (SBTF) uses to verify social media data. This approach was used to verify crowdsourced social media for the UN in Libya. The fifth and final example relates to the verification strategies developed by the U-Shahid team for the Egyptian Parliamentary elections of 2010.[15] This last case was selected because it demonstrates a field-based approach to information verification in an insecure environment.

4.1 Andy Carvin and the Arab Spring

Andy Carvin is a Senior Strategist at the National Public Radio (NPR) in Washington DC. An avid Twitter user, Carvin played an active curative role during the Arab Spring by verifying social media content using his Twitter feed and followers. He began tweeting about the Arab Spring in December 2010, as events in Tunisia were just beginning to unravel. Carvin interrogates his follower base and places the burden on them to prove that events being reported in the social media space have actually taken place. Indeed, "Carvin's followers are the engine that drives his reporting. They help him translate, triangulate, and track down key information" (Silverman 2011).

For example, Carvin will retweet news he receives about an incident and will add "Anyone else reporting on this yet?" or "How unusual is this?" or just: "Source?" He also asks for pictures or videos to confirm or dispel a rumor. When asked how he judges the accuracy of a tweet, Carvin notes that, "a red flag for him is when non-journalists adopt the language of breaking news." He elaborates further: "Some of the rumors I see floating around seem to be accompanied by the words 'breaking' or 'confirmed' or 'urgent' all in capital letters," he said. "I think it's partially

[14] See original blog post on How to Use Technology to Counter Rumors During a Crisis: http://iRevolution.net/2011/03/26/technology-to-counter-rumors
[15] See also http://iRevolution.net/dissertation

because you've got people on the ground in the Middle East hearing information and they're very excited about getting it, or feel like it needs to be out there as quickly as possible. They start using phrases that reporters use but they are using them in a very different way" (Silverman 2011).

One of Carvin's main verification success stories relates to rumors that Muammar Gaddafi had attacked rebels using mortars made in Israel. A photograph that accompanied these rumors purported to show a Star of David with an odd multi-crescent shape above it. Carvin got on twitter and asked his followers for help in identifying exactly what kind of mortar the photo displayed and whether it was indeed Israeli. This spurred a flurry of activity, which rapidly helped to debunk the story "even as other news outlets, including Al Jazeera's Arabic TV channel, continued to report the bogus link to Israel" (Farhi 2011).

Interrogating sources and triangulating content is in many ways traditional, investigative journalism. As Carvin rightly notes, "the notion of journalists gathering, analyzing and disseminating relevant information isn't new at all" (Stelter 2011). The difference is that Carvin has "turned the newsgathering process inside out and made it public. He's reporting in real time and you can see him do it. You can watch him work his sources and tell people what he's following up on" (Farhi 2011). Moreover, Carvin has not met the vast majority of twitter users he uses for tips and verification. Still, Carvin says he "relies on sources who've proven to be reliable and drops those with dubious track records" (Farhi 2011). Carvin also challenges the validity of certain reports on a regular basis.

"The vast majority of folks that are posting information, their hearts are in the right place but sometimes the fog of war affects them just as it would any other journalist" (Lefkow 2011). Carvin doesn't always get it right, however. His followers often correct him when his tweets are wrong. The biggest lesson that he has learned from this experience in real-time curation and verification is that "most of the people who try to reach out to you are not trying to mislead you. It doesn't necessarily mean that everything they send you is true but there's generally a grain of truth in most of what you see" (Lefkow 2011). The key, according to Carvin, "is disclosing what he doesn't know and asking others to fill in the blanks." He considers this a "self-correcting

mechanism." Incidentally, this perceived dynamic is supported by a recent empirical study that assessed the veracity of tweets following the earthquake in Chile (Mendoza *et al*. 2010).

It's important to note that a lot goes on behinds the scenes with respect to Carvin's verification strategies. He has extensive "backchannel conversations on Facebook, on Skype, on email, and occasionally on the phone […] Facebook and YouTube and other content-sharing sites have been a goldmine of new content" (Lefkow 2011). The fact that he had a network of blogger contacts in the region to begin with, is also key (Katz 2011). There are of course a myriad of challenges with the approach that Carvin takes. "For one thing, Twitter can echo in the sense that it's loud at first then reverberates for a while. So something one person might've posted 12 hours ago gets retweeted by someone who's just checking twitter for the first time, causing it to propagate further" (Zuckerman 2011). Ultimately, Carvin refers to what he does as more art than science. But as this study seeks to demonstrate, the field of information forensics can be systematized and codified.

4.2 Using Skype to counter rumors in Kyrgyzstan

The regions of Osh and Jalal-Abad in Southern Kyrgyzstan experienced widespread violence during May and June of 2010. The escalating violence resulted in the country's interim government declaring a state of emergency on June 12, 2010. Reports on how many people were killed are disputed, with figures ranging between 200 and 2,000. Displacement figures range from 100,000 to 400,000. Misinformation and disinformation was widespread throughout this period, both in the form of SMS and YouTube videos.

For example, one challenge that local groups faced during periods of ethnic tension and violent conflict last year was the spread of rumors, particularly via SMS. These deliberate rumors ranged from humanitarian aid being poisoned to cross border attacks carried out by a particular ethnic group. But many civil society groups were able to verify these rumors in near real-time using Skype. When word of the conflict spread, the director of one such group got online and invited her friends and colleagues to a dedicated Skype chat group. Within two hours, some 2,000

people across the country had joined the chat group, with more knocking, but the group had reached the maximum capacity allowed by Skype. (They subsequently migrated to a web-based platform to continue the real-time filtering of information from around the country).

The Skype chat was abuzz with people sharing and validating information in near real-time. When someone got wind of a rumor, they'd simply jump on Skype and ask if anyone could verify. This method proved incredibly effective. Why? Because members of this Skype group constituted a relevant, trusted and geographically distributed network. A person would only add a colleague or two to the chat if they knew who this individual was, could vouch for them and believed that they had—or could have—important information to contribute given their location and/or contacts.

The degrees of separation needed to verify a rumor was close to one. In the case of the supposed border attack, one member of the chat group had a contact with the army unit guarding the border crossing in question. They called them on their cell phone and confirmed within minutes that no attack was taking place. As for the rumor about the poisoned humanitarian aid, another member of the chat found the original phone numbers from which these false SMS's were being sent. They called a personal contact at one of the telecommunication companies and asked whether the owners of these phones were in fact texting from the place where the aid was reportedly poisoned; they weren't. Meanwhile, another member of the chat group had them self investigated the rumor in person and confirmed that the text messages were false.

This Skype detective network proved an effective method for the early detection and response to rumors. Once a rumor was identified as such, 2,000 people could share that information with their own networks within minutes. In addition, members of this Skype group were able to ping their media contacts and have the word spread even further. In at least two cases and in two different cities, telecommunication companies also collaborated by sending out broadcast SMS to notify subscribers about the false rumors.

4.3 The BBC's approach to verification

The BBC's User-Generated Content (UGC) Hub has been operational since 2009. The team is largely responsible for newsgathering via social media. "Hub journalists scour the Internet for pictures, videos, and other content that might contribute to a story, which they then verify and clear for use. But they also find people, sources who can be contacted by reporters in other departments within the BBC" (Stray 2011). According to UGC journalist Silvia Costeloe, the most interesting stories they crowdsource comes from the comments forms displayed at the end articles on the BBC website.

Just like Andy Carvin's approach described above, the UGC's approach is not strictly passive but rather active newsgathering. The comments generated through these comments sections are usually very targeted. Worth noting is that the forms don't require users to sign in. In addition, the use of comments forms allows the BBC's UGC Hub to identify eyewitnesses and thus create a network of contacts on the ground.

According to UGC journalist Silvia Costeloe, the verification process is "mostly a matter of persistence and organization. Still, she offered a few practical hints, such as searching for people with a specific location listed in their Twitter profile, or putting 'pix' or 'vid' in your search to find multimedia content, or watching who local news organizations are watching" (Stray 2011). During the election protests in Iran, the team worked with the Persian Service to verify the authenticity of videos being shared in the social media space—and in particular the dates of those videos. One way they do this is by looking for shadows to determine the possible time of day that a video was shot. In addition, they examine weather reports to "confirm that the conditions shown fit with the claimed date and time" (Murray 2011). These same strategies can also be applied to photographs.

Videos and pictures can also be partly authenticated by investigating whether they were actually taken in the location being claimed by users of social media. For example, the UGC team analyzes visual content for any clues vis-à-vis the location it was taken based on buildings, signs, cars, etc. If video footage is being analyzed has audio, then the vocabulary, slang, accents can be

analyzed to determine whether they are correct for the location that a source might claim to be reporting from. In Syria, a video showing protesters being brutally beaten was denounced by the regime as being fake. Several days later, however, one of the protesters who was captured in the original video footage created a video response: "Holding up an I.D. card to identify himself as Ahmad Bayasi–a 22 year old Syrian national–he stood in front of Bayda's town square, where he confirmed the events took place. Several others anonymously stepped forward to reveal signs of injuries as further evidence of what had happened in Bayda" (O'Carroll 2011).

4.4 The Standby Volunteer Task Force

The Verification Team of the Standby Volunteer Task Force (SBTF) is responsible for verifying all types of reports including social media data identified by the Media Monitoring Team.[16] The SBTF is developing a two-pronged strategy to verify crowdsourced content. The first aims to assess the trustworthiness of a source while the second seeks to use triangulation techniques to verify the veracity of social media reports. Triangulation techniques seek to identify multiple reports on a given event for cross-verification purposes.

If every source monitored in the social media space was known and trusted, then the need for verification would not be as pronounced. In other words, it is in part the plethora and virtual anonymity of sources that creates the need for verification in the first place. The process of verifying social media data thus requires a two-step process: the authentication of the source as reliable and the triangulation of the content as valid. If the source can be authenticated and tagged as trustworthy, this may be sufficient to trust the source's content and be marked as verified depending on context. If source authentication is difficult to ascertain, then the content itself needs to be triangulated. As a rule of thumb, however, both steps should be carried out.

The SBTF's first step is to try and determine whether the source is trustworthy. This first involves reviewing the tweeter's bio on Twitter. Does the source provide a name, picture, bio and any links to their own blog, identity, professional occupation, etc., on their page? If there's a

[16] See http://blog.standbytaskforce.com

name, does searching for this name on Google provide any further clues to the person's identity? Perhaps a Facebook page, a professional email address, a LinkedIn profile? Next, the number of tweets by that tweeter is considered. Is this a new Twitter handle with only a few tweets? If so, this makes authentication more difficult. "The more recent, the less reliable and the more likely it is to be an account intended to spread disinformation" (Arasmus 2011). In general, the longer the Twitter handle has been around and the more Tweets linked to this handle, the better. This gives a digital trace, a history of prior evidence that can be scrutinized for evidence of political bias, misinformation, etc. "What are the tweets like? Does the person qualify his/her reports? Are they intelligible? Is the person given to exaggeration and inconsistencies?" (Arasmus 2011).

The tweeter's followers are then reviewed. Does the source have a large following? If there are only a few, are any of the followers know and credible sources? Also, how many lists has this Twitter handle been added to? Also, How many Twitter users does the Twitter handle follow? Are these known and credible sources? In terms of retweets, what type of content does the Twitter handle retweet? Does the Twitter handle in question get retweeted by known and credible sources? Location is also an important parameter in the verification process. Can the source's geographic location be ascertained? If so, are they nearby the unfolding events? One way to try and find out by proxy is to examine during which periods of the day/night the source tweets the most. This may provide an indication as to the person's time zone. The timing of tweets can also be a telling factor. Does the source appear to be tweeting in near real-time? Or are there considerable delays? Does anything appear unusual about the timing of the person's tweets?

If the Verification Team is still unsure about the source's reliability, they use their own social network via Twitter, Facebook and LinkedIn to find out if anyone in their network knows about the source's reliability—much like Andy Carvin's approach. Finally, the team will also seek to triangulate the identity of the tweeter via the mainstream media. For example, is the source quoted in the mainstream media or is the story reported by the source also being reported by the mainstream media?

The above strategies are obviously not full proof. This explains why the SBTF's second step is to try and triangulate the content. For example, the Verification Team will seek to establish whether other sources on Twitter or elsewhere are reporting on the same event. They "remain skeptical about the reports that [they] receive" and "look for multiple reports from different unconnected sources" (Arasmus 2011). The team seeks to identify as many independent "witnesses" as possible in order to render the need for individual identity authentication less critical. In addition, if the user reporting an event is not necessarily the original source, the original source is sought for authentication. In particular, if the original source is found, the time/date of the original report is checked to determine whether it makes make sense for given the situation. Finally, if a twitter user shares "visual evidence", the SBTF analyzes the photograph or video for any clues about the location it was taken based on buildings, signs, cars, etc., in the background? This strategy draws directly on the BBC UGC Hub's approach.

Some of the above strategies are still being mainstreamed into the SBTF's protocols. As such, they should be considered a work in progress.

4.5 U-Shahid's field-based verification efforts

The U-Shahid team in Egypt developed specific verification strategies to verify crowdsourced content related to the 2010 Parliamentary Elections in 2010.[17] The team's first step is to define concrete criteria for what types of reports require verification. In other words, they don't attempt to verify all the content that is available. For example, one type of report that U-Shahid requires verification for is one relating to an immediate threat or act of violence. Another is "grave electoral fraud" the gravity of which is calculated by "the importance of the parties involved (celebrities, known PMs, government officials) or the importance of the event itself or it's location" (U-Shahid 2010).

If a report meets U-Shahid's criteria, the team tags the report as verified if one or more of the following requirements are met: "It is supplemented by video or pictures that clearly confirms

[17] See also http://iRevolution.net/dissertation

what has been reported; It has been reported by two or more independent sources; Messages coming from social media (Twitter and Facebook) needs to be confirmed by an SMS, a media report or a direct witness before being flagged ad verified; At least one of the sources of the information must be clear and known (i.e., 2 SMSs from unknown sources cannot verify each other)" (U-Shahid 2010).

U-Shahid uses four core strategies to try and meet the above requirements. The first involves calling or tweeting the person who sent the report that is being investigated. If the report was sent by SMS, that number is called to verify the person's identity. The witness is asked if they observed the event themselves or if they simply learned about the event from someone else. More specifically, details on who did what, to whom, how and where are asked. If the event being reported is still unfolding, the witness is asked if anyone else nearby is able confirm the information. They are also asked to provide a video or picture of the event but only if it is safe to do so. If the report came from Twitter itself, the account of the tweeter is reviewed. Simple content analysis of previous tweets and the account holder's bio is carried out. In addition, the Tweeter's followers are also reviewed. To acquire additional information, U-Shahid will tweet back to the original Twitter user asking for more information—again using the "who did what, to whom, how and where" format. Like Andy Carvin, U-Shahid also uses Twitter to ask followers to confirm if the information is indeed correct.

The second core strategy involves in-person verification via a trusted source. The U-Shahid team determines whether one of their election monitors was close to the area referenced in a report that required verification. If a monitor is indeed on site, that person is asked to verify the report. If U-Shahid does not have any monitors in the area, the check whether any of their NGO partners may have monitors in the area. If so, those individuals are asked to confirm the validity of the report being investigated. U-Shahid's third core strategy leverages the mainstream media for confirmation. They use web-based research to look for any evidence that is specific to the event that was report as well as that location. They look for article, blogs, video or pictures that can confirm the information reported. Fourth and finally, the U-Shahid team seeks to triangulate the report being investigated with the reports they have already received.

5. Conclusion

There is no doubt that verifying crowdsourced information is a major challenge. The humanitarian community's reluctance to draw on crowdsourced information is perfectly understandable given the consequences of making decisions using wrong information. At the same time, the case studies presented in this paper clearly demonstrate that leveraging a mix of strategies can at times provide a powerful way to verify crowdsourced information in near real-time.

Take the Amina Abdallah Araf "abduction" story, for example. The authenticity of that story could have perhaps been questioned earlier had some of the verification strategies in this study been used. Amina, a gay activist whose blog and Facebook page had become a prominent example of activism in Syria over the years, was reportedly kidnapped according to a guest blog post on her blog, "A Gay Girl in Damascus." It turns out that "Amina" was actually a 40-year old American who had assumed a fake identify. As Andy Carvin noted at the time, "none of the reports of the arrest of Amina Abdallah Arraf appeared to have been written by journalists who had previously met or interviewed her" (Mackey and Stack 2011). Carvin could also have Tweeted: "Has anyone actually met Amina Abdallah Arraf or her family in person? Prove it." In any case, the very public unravelling of the Amina riddle does mean that journalists and others will be more cautious in the future. This story also emphasizes the value of taking a more bounded approach to crowdsourcing.

To be sure, live crisis mapping projects are increasingly using Storyful to access and map verified content.[18] One of Storyful's comparative strengths when it comes to real-time news curation is the growing list of authenticated users it follows. This represents more of a bounded (but certainly not static) approach. As noted in this paper, following a bounded model presents some obvious advantages. This explains by the BBC recommends "maintaining lists of previously verified material [and sources] to act as a reference for colleagues covering the stories." This strategy is also employed by the SBTF's Verification Team. In addition, the ICT

[18] http://www.storyful.com

for Peace Foundation (ICT4Peace) has been working to provide an intuitive way to at least quantify the perceived reliability of reports being added to live crisis maps (ICT4Peace 2011).

Even if every strategy documented in this study is used to verify crowdsourced information, this still does not guarantee one hundred percent veracity. There are plenty of ways to falsify information in ways that are un-detectable by the strategies listed above. But many public health experts in the field of emergency medicine nevertheless state that they would rather have some data that is not immediately verifiable than no data at all. Indeed, in some ways all data begins life this way.[19] These experts would rather know about a potential rumor regarding a health outbreak so they can follow up and verify than have no "early plausible warning" until it's too late if some rumor turns out to be true.[20] Finally, while wrong data can cost lives, this doesn't mean that no-data doesn't cost lives, especially in a crisis zone. Information is perishable so the potential value of information must be weighed against the urgency of the situation. Perhaps the question is ultimately about tolerance for uncertainty–different organizations will have varying levels of tolerance depending on their mandate, the situation, time and place.[21]

In a "Theory of Justice," the philosopher John Rawls introduces the "veil of ignorance", a thought-experiment designed to determine the morality of a certain issue.[22] The idea goes something like this: imagine that you have to decide on the morality of an issue before you are born, i.e., you stand behind a veil of ignorance as you don't know where you will be born, what race, with what kind of family, etc. As put by John Rawls himself … "no one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like."

---

[19] See also Wag the Dog, or How Falsifying Crowdsourced Data Can Be a Pain: http://iRevolution.net/2010/04/08/wag-the-dog

[20] See also No Data is Better than Bad Data… Really? http://iRevolution.net/2011/06/22/no-data-bad-data

[21] For example, deployments of the Ushahidi platform are increasingly carrying disclaimers vis-à-vis the reliability of the information displayed on the public maps. Take the recent uses of the platform in the US, Liberia and Turkey; each displayed a message asking viewers to make their own decision vis-à-vis the possible veracity of the information displayed on the public map. See also Can Ushahidi Rely on Crowdsourced Verifications? http://www.pbs.org/idealab/2011/11/can-ushahidi-rely-on-crowdsourced-verifications325.html

[22] See Crowdsourcing and the Veil of Ignorance: A Question of Morality? http://iRevolution.net/2010/04/25/veil-ignorance

For example, in the imaginary society, you might or might not be intelligent, rich, or born into a preferred class. Since you may occupy any position in the society once the veil is lifted, this theory encourages thinking about society from the perspective of all members. The veil of ignorance is part of the long tradition of thinking in terms of a social contract.

If you were standing behind this metaphorical veil of ignorance, would you discount crowdsourced information on a live crisis map on the basis that the data may not be accurate or representative? Or would you prefer to have that information on hand just in case it can provide vital contextual information on an unfolding situation?

Acknowledgements

I would like to recognize and thank the following colleagues for their feedback on a preliminary working draft: Christina Corbane, Sanjana Hattotuwa, Jessica Heinzelman and Jonathan Stray.

References

Arasmus 2011. *Mapping violence against pro-democracy protests in Libya*. Arasmus, March 1, 2011. Available from: http://www.arasmus.com/2011/03/01/mapping-violence-against-pro-democracy-protests-in-libya [Accessed August 1, 2011].

Cohen, N and T. Arieli. 2011. Field research in conflict environments: Methodological challenges and snowball sampling. *Journal of Peace Research*, 48(4): 423-435. Available from: http://jpr.sagepub.com/content/48/4/423.abstract?rss=1 [Accessed August 15, 2011].

Colvin, M. 2010. *Taking the twit out of Twitter and finding value*. The Punch.com.au, September 30, 2010. Available from: http://www.thepunch.com.au/articles/taking-the-twit-out-of-twitter-and-finding-value [Accessed August 1, 2011].

Farhi, P. 2011. NPR's Andy Carvin, tweeting the Middle East. *Washington Post*, April 12, 2011. Available from: http://www.washingtonpost.com/lifestyle/style/npr-andy-carvin-tweeting-the-middle-east/2011/04/06/AFcSdhSD_story.html [Accessed August 1, 2011].

IAE. 2011. *Collaborative Research and Development*. Information Age Education, IAE-Pedia.org. Available from: http://iae-pedia.org/Collaborative_Research_and_Development [Accessed August 1, 2011].

ICT4Peace. 2011. *The Matrix plugin for Ushahidi platform.* ICT for Peace Foundation. Available from: http://ict4peace.org/publications/the-matrix-plugin-for-ushahidi-platform [Accessed August 1, 2011].

Katz, Ian. 2011. SXSW 2011: Andy Carvin – The man who tweeted the revolution. *UK Guardian*, March 14, 2011. Available from: http://www.guardian.co.uk/technology/2011/mar/14/andy-carvin-tunisia-libya-egypt-sxsw-2011 [Accessed August 1, 2011].

Lefkow, C. 2011. Tweeting the turmoil in the Middle East. *Agence France Presse*, March 26, 2011. Available from: http://www.google.com/hostednews/afp/article/ALeqM5jueNS4XIU9BguHiMY7Vmq49h4RqQ?docId=CNG.6f7d1adfb97a7cc8154c2507c0688933.1a1 [Accessed August 1, 2011].

Mackey, R. and L. Stack. 2011. After report of disappearance, questions about Syrian-American blogger. *New York Times*, June 7, 2011. Available from: http://thelede.blogs.nytimes.com/2011/06/07/syrian-american-blogger-detained [Accessed August 1, 2011].

Mendoza, M., B. Poblete and C. Castillo. 2010. Twitter Under Crisis: Can we trust what we RT? *1st Workshop on Social Media Analytics (SOMA '10),* July 25, 2010, Washington, DC, USA.

Murray, A. 2011. BBC processes for verifying social media content. *BBC News*, May 18, 2011. Available from: http://www.bbc.co.uk/journalism/blog/2011/05/bbcsms-bbc-procedures-for-veri.shtml [Accessed August 1, 2011].

O'Carroll, T. 2011. Ahmad Bayasi's Story: Citizen Video Authentication in Syria and Beyond. *Witness.org*, July 25, 2011. Available from: http://blog.witness.org/2011/07/ahmed-bayasi%E2%80%99s-story-citizen-video-authentication-in-syria-and-beyond [Accessed August 1, 2011].

Silverman, C. (2011). Is this the world's best Twitter account? *Columbia Journalism Review*, April 8, 2011. Available from: http://www.cjr.org/behind_the_news/is_this_the_worlds_best_twitter_account.php [Accessed August 1, 2011].

Stelter, B. 2011. Twitter feed evolves into a news wire about Egypt. *New York Times,* February 13, 2011. Available from: http://mediadecoder.blogs.nytimes.com/2011/02/13/twitter-feed-evolves-into-a-news-wire-about-egypt [Accessed August 1, 2011]

Stray, J. 2011. Drawing out the audience: Inside BBC's User-Generated Content Hub. *Nieman Journalism Lab*, May 5, 2011. Available from: http://www.niemanlab.org/2010/05/drawing-out-the-audience-inside-bbc%E2%80%99s-user-generated-content-hub [Accessed on August 1, 2011]

Tapia, A, K. Bajpai, J. Jansen, J. Yen and L. Giles. 2011. Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations. *Proceedings of the 8th International ISCRAM Conference*, Lisbon, Portugal, May 2011.

Turner, A. 2006. Introduction to Neogeography. *O'Reilly Media*. Available from: http://oreilly.com/catalog/9780596529956 [Accessed August 1, 2011].

U-Shahid. 2010. U-Shahid Guide for Verifiers. *Unpublished document*, September 2010.

Zuckerman, I. 2011. Interview with Andy Carvin on curating Twitter to watch Tunisia, Egypt. *My Heart's in Accra*. Available from: http://www.ethanzuckerman.com/blog/2011/02/04/interview-with-andy-carvin-on-curating-twitter-to-watch-tunisia-egypt [Accessed August 1, 2011]