Countering False Information on Social Media in Disasters and Emergencies

Social Media Working Group for Emergency Services and Disaster Management

March 2018

Homeland Security

Science and Technology

Contents

Executive Summary

Rumors, misinformation and false information on social media proliferate before, during and after disasters and emergencies. While this information cannot be completely eliminated, first responder agencies can use various tactics and strategies to offset bad information. This white paper examines motivations people may have for sharing false information, discusses underlying issues that cause false information and offers case studies from recent disasters to illustrate the problem. Multiple motives lead people to post false information on social media: some posters seek a particular result, such as closing schools for the day; some desire to get attention with a dramatic post; some are pushing a money-making scam or political agenda; and some innocently repeat bad or outdated information.

Best practices for agencies to counter misinformation, rumors and false information are detailed and categorized in this white paper, and challenges and additional considerations are presented for review. This report illustrates methods of countering false information on social media with case studies:

- The 2014 South Napa earthquake: Tweets were filtered by geolocation to eliminate posts from trolls.
- The 2016 Louisiana floods: The Red Cross published and shared a blog to counter rumors and misinformation about food distribution and shelter policies.
- The 2017 Oroville Dam evacuation: An accidentally misleading tweet suggested the evacuation area included all of Sacramento County. Local agencies used traditional and social media to provide correct information.

Examples of best practices include:

- Establishing partnerships with local traditional media outlets before disasters, so means exist to disseminate accurate information;
- Using the Joint Information System to coordinate public information efforts of multiple jurisdictions and agencies; and
- Setting up a central website to debunk bad information.

Introduction

Social media and collaborative technologies have become critical components of emergency preparedness, response and recovery.[1] From international response efforts after large-scale disasters to domestic response and recovery after events affecting the United States, many government officials now turn to social media technologies to share information and connect with citizens during all phases

[1] Social media includes any online or digital medium provided and/or collected through a channel that enables the two-way sharing of information, involving multiple parties. This includes social networking sites, texting, blogs, etc.

of a crisis. Implementing these new technologies, however, requires responding agencies adopt new communication strategies, policies and engagement methods.

Recognizing the need to address these challenges, the U.S. Department of Homeland Security's (DHS) Science and Technology Directorate (S&T) established a Virtual Social Media Working Group (VSMWG) in 2010. After Public Law 114-80 was passed, the VSMWG was re-named as the Social Media Working Group for Emergency Services and Disaster Management (SMWGESDM). The mission of the SMWGESDM is to provide recommendations to the emergency preparedness and response community on the safe and sustainable use of social media technologies before, during and after emergencies. The SMWGESDM is a subcommittee of the Homeland Security Science and Technology Advisory Committee (HSSTAC). The HSSTAC approved the recommendations contained in this white paper by consensus vote on February 22, 2018.

Drawn from a cross-section of subject matter experts from federal, tribal, territorial, state and local responders from across the United States, SMWGESDM members establish and collect best practices and solutions that can be leveraged by public safety officials and responders throughout the nation's emergency response community. Below is a list of agencies and organizations to which the SMWGESDM members belong.

SMWGESDM Member Agencies and Organizations as of March 2018

- American Red Cross
- Argonne National Labs, Public Affairs Science and Technology Fusion Center
- Arlington County [VA] Fire Department
- California Governor's Office of Emergency Services
- Colorado Division of Homeland Security and Emergency Management
- City of Evanston [IL]
- City of Nashua [NH] Office of Emergency Management
- Evacuteer
- Federal Emergency Management Agency (FEMA)
- George Mason University
- Humanity Road
- Indianapolis [IN] Fire Department
- National Institutes of Health
- New York City [NY] Department of Health and Mental Hygiene
- New York City [NY] Emergency Management Department
- Sacramento County [CA] Office of Emergency Services
- Sacramento [CA] Fire Department
- City of St. Louis [MO] Emergency Management Agency
- United States Geological Survey
- University of Washington Emergency Management
- Virginia Department of Emergency Management
- Washington County [AR] Regional Ambulance Authority

Motivations

Social media platforms have allowed individuals and organizations to share information with their peers and specific audiences for more than twenty years.[2] Information typically is shared with good intent; however, some people post on social media to further an ulterior agenda. Their posts may include rumors, false information and misinformation (e.g., deception, propaganda and malicious spamming).

Researchers have identified different characteristics of social media posts that lead consumers of the posts to believe in an alternative, fake reality and suspicious behavior.[3,4] Characteristics of false information may include uncertainty in the "facts," emotional exploitation of a situation, trending topic discussions for hijacking conversations and financial scams, among others.[5,6,7]

An example of false information with these characteristics is deceptive content with a malicious agenda, such as diverting a user towards purchasing a particular product.[8] Such campaigns are also used to lead a user to believe in a fake negative opinion to damage an object's reputation; for example, fake reviews on online e-commerce websites, such as Amazon or Yelp.[9] Likewise, deceptive false information has been posted in large-scale disasters for financial gain.[10] False information with a malicious agenda has long existed in the form of propaganda, which has been used by terror and other extremist/criminal organizations as a tactic to recruit.[11]

---

[2] Weblogs, or blogs, have existed since 1997, and an early example of social media being used to share information is the website Friendster.com, which was launched in 2002. <https://en.wikipedia.org/wiki/Blog> and <https://en.wikipedia.org/wiki/Friendster>.

[3] Pendleton, Susan Coppess. "Rumor Research Revisited and Expanded." *Language & Communication*. 1998. 18,1: 69-86.

[4] Jiang, M., Cui, P., & Faloutsos, C. "Suspicious Behavior Detection: Current Trends and Future Directions." *IEEE Intelligent Systems*. 2016. 31(1), 31-39.

[5] Starbird, K., Spiro, E., Edwards, I., Zhou, K., Maddock, J., & Narasimhan, S. "Could This Be True?: I Think So! Expressed Uncertainty in Online Rumoring." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. May 2016. (pp. 360-371). ACM.

[6] Bessi, A., & Ferrara, E. "Social Bots Distort the 2016 U.S. Presidential Election Online Discussion." *First Monday*. 2016. 21(11).

[7] Huang, Y. L., Starbird, K., Orand, M., Stanek, S. A., & Pedersen, H. T. "Connected Through Crisis: Emotional Proximity and the Spread of Misinformation Online." In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. February 2015. (pp. 969-980). ACM.

[8] Gao, H., Hu, J., Wilson, C., Li, Z., Chen, Y., & Zhao, B. Y. "Detecting and Characterizing Social Spam Campaigns." In *Proceedings of the 10th ACM SIGCOMM conference on Internet Measurement*. November 2010. (pp. 35-47). ACM.

[9] Mukherjee, A., Liu, B., & Glance, N. "Spotting Fake Reviewer Groups in Consumer Reviews." In *Proceedings of the 21st International Conference on World Wide Web*. April 2012. (pp. 191-200). ACM.

[10] Gupta, A., Lamba, H., & Kumaraguru, P. "$1.00 per rt# bostonmarathon# prayforboston: Analyzing Fake Content on Twitter." In *eCrime Researchers Summit (eCRS)*, September 2013. (pp. 1-12). IEEE.

[11] Allendorfer, W. H., & Herring, S. C. "ISIS vs. the U.S. Government: A War of Online Video Propaganda." *First Monday*. 2015. 20(12).

When discussing the online context of false information in today's information age, the concept of false information driven by a motive of a deceptive agenda has existed for many decades in military warfare.[12,13] Therefore, the strategies for countering false information with a malicious agenda in the online environment by either coordinated efforts of humans or bots could be informed by the offline environment as well.[14,15]

Problem

Virality of Social Media

By Catherine Graham, Humanity Road

After the April 2015 earthquake in Nepal, a Facebook post described 300 houses in Dhading needed aid. The post was shared over 1,000 times, reaching over 350,000 people within 48 hours. The originator of this message was trying to find help for Ward #4's villagers via social media. Facebook statistics show that the average user has 350 contacts, meaning this one message was viewed by approximately 350,000 Facebook users. A week before the viral post, this need had already been shared on quakemap.org, a crisis-mapping database built by online volunteers and managed by Kathmandu Living Labs. On May 7, Helping Hands (a humanitarian group) was notified, and by May 11, Ward #4 received much-needed food and shelter. While the late Facebook post was meant to be helpful, the need had already been taken care of. This short example demonstrates that sharing outdated information can waste resources and risk lives.[1]

One of the biggest challenges public safety agencies and organizations face is how to reduce or eliminate the spread of false information, especially as public demands for a response from these authorities increases. Social media can distribute news faster and to a wider audience than traditional news sources. However, that also means the potential for misinformation, false information and rumors to spread and go viral is high.[16,17] A factor that may impede first responders' ability to mitigate and minimize the spread of misinformation, rumors and false information is the decreasing public trust in government, media and nongovernmental organizations (NGOs). While 2017 was a low point in terms of credibility of the media, the 2018 Edelman's Trust Barometer showed trust in journalism jumped five points and trust in social media platforms dipped two points. In addition, the credibility of "a person like yourself" — often a source of news and information on social media — dipped to an all-time low in the study's history. While this paper is focused on social media, responder agencies should be aware that many people still get their news from television, which serves as an additional resource to counter false information.[18]

---

[12] Whaley, B. "Toward a General Theory of Deception." *The Journal of Strategic Studies*. 1982. 5(1), 178-192.

[13] Holt, T. *The Deceivers: Allied Military Deception in the Second World War*. Simon and Schuster, 2010.

[14] A computer program that performs automatic repetitive tasks. <https://www.merriam-webster.com/dictionary/bot>.

[15] For future reading on this whole section, see Manheim, Jerol. *Strategy in Information and Influence Campaigns: How Police Advocates, Social Movements, Insurgent Groups, Corporations, Governments and Others Get What They Want*. Routledge, 2010.

[16] Incorrect or misleading information. <https://www.merriam-webster.com/dictionary/misinformation>.

[17] Madhusree Mukerjee. "How Fake News Goes Viral – Here's the Math." *Scientific American*, July 14, 2017. <https://www.scientificamerican.com/article/how-fake-news-goes-viral-mdash-heres-the-math/>.

[18] Pew Research Center. "Pathways to News." July 7, 2016. <http://www.journalism.org/2016/07/07/pathways-to-news/>.

Solving the problem of how to reduce or eliminate the spread of false information requires an understanding of the following questions.

- What are the causes of misinformation, rumors or false information, and what are its characteristics?
- How does false information spread?
- What are best practices to counter the spread of false information?

This paper builds on real-world case studies of several incidents to explain and investigate answers to the aforementioned questions.

Causes and Spread

In social media, misinformation, rumors and false information are most often caused by four underlying issues, which are detailed more fully below:[19]

1. Incorrect information - intentional versus unintentional;
2. Insufficient information;
3. Opportunistic disinformation; and
4. Outdated information.

Incorrect Information

Incorrect information can be caused by situations where the true situation is difficult to confirm. Radiation in Japan was a good example. After the meltdown at the Fukushima Daiichi Nuclear Power Station in March 2011, many rumors circulated regarding appropriate safety precautions, such as whether people should evacuate, the possibility of food and water shortages, and whether there would be additional radioactive releases (this example also illustrates insufficient information).

Figure 1: Official Tweet from Fairfax County Government addressing fake Twitter account.

Incorrect information and rumors can also be caused by individuals who wish to create confusion. One example is when fake accounts are created that impersonate an official account. Fairfax County, Virginia was proactive during a winter storm in January 2014 as its school system was faced with many fake accounts announcing incorrect closures (Figure 1, above). Government and schools worked together to actively advise people where to find official information.

[19] Humanity Road Rumor Management Team Training, June 20, 2016.

Figure 2: Tweet from user @ComfortablySmug sharing a rumor.

Another example comes from Hurricane Sandy in October 2012. Twitter user @ComfortablySmug began spreading several rumors via social media, including that the New York Stock Exchange Building was flooded, Con Edison was preemptively shutting off power in New York City, and all bridges going to and from Manhattan were being sealed off (Figure 2, below). Additionally, digitally altered pictures of sharks swimming in the streets, screenshots from the movie *The Day After Tomorrow* and other dramatic pictures from past storms proliferated on social media.[20] Incorrect information can also be malicious (see the Motivation section above), as with online conspiracy theorists harassing survivors of the Las Vegas mass shooting in October 2017.[21]

Insufficient Information

When information is slow to emerge on circumstances surrounding an event, rumors can start rapidly. Insufficient information can be a result of several factors, such as: not having clearance to release the data, lack of a designated official for that information, or a belief that information must be complete to release and therefore intentionally withheld. Confusion continues to arise when official channels do not release information fast enough, provide information updates in the right social media and traditional media channels, or the population is unaware of or does not trust the official source for that information. The public will generally follow and amplify official information when they can access information they believe. This happened after the Nepal earthquake in 2015. When there is a new emerging situation that can be confusing, agencies will open their channels of information (such as a conference bridge for volunteers and partners), which can be critical to avoiding mistakes in information management.

[20] DHS S&T Virtual Social Media Working Group. "Lessons Learned: Social Media and Hurricane Sandy", June 2013, p. 22. <https://www.dhs.gov/publication/lessons-learned-social-media-hurricane-sandy>.

[21] Lois Becket and Sam Levin. "U.S. Gun Violence Spawns a New Epidemic: Conspiracy Theorists Harassing Victims." *The Guardian*, November 28, 2017. <https://www.theguardian.com/us-news/2017/nov/28/us-guns-mass-shootings-hoax-conspiracy-theories>.

Opportunistic Disinformation

Opportunistic disinformation occurs when predatory individuals attempt to capitalize on a particular event or incident.[22] Opportunistic misinformation generally falls into one of two categories: revenue-generating and financially incentivized, or malicious and politically incentivized.

Revenue-generating disinformation attempts to hijack the attention of social media users from a particularly newsworthy happening, and redirect their attention for commercial purposes. A phishing scam or spammer may mimic a pre-existing website and redirect the user to a sales pitch or other ad.[23] This technique is similar to malware that operates by hijacking a browser and redirecting traffic to an alternate website. Scammers capitalize on a popular hashtag and use click-throughs to boost viewer statistics on a website, or encourage the purchase of a specific product or service unrelated to the original hashtag. An example of this is an article that circulated after a 2014 Sicilian earthquake to supposedly provide news, however, the article was referencing a 1908 earthquake.

Malicious disinformation is typically politically motivated, and can be even more challenging to both identify and counter. Studies that have examined the volume, timing and location (e.g., tracked IP addresses, associated time zone and geo-tagged posts) of this category of social media posts indicate an intent to cause harm and disrupt the standard flow of truthful information during a specific event or incident. During Hurricane Harvey in 2017, a rumor spread on Twitter that officials were asking shelter-seekers about their immigration status. Also in 2017 after Hurricane Irma, a rumor surfaced that survivors would receive generators from the federal government. While some posts could be attributed to innocent mistakes, the scope and velocity of amplification seems to indicate an intentional rebroadcasting of disinformation with the intent to frighten vulnerable members of the local communities and weaken their trust with government entities offering essential aid.[24] Similar behavior was demonstrated during response to a train derailment in DuPont, Washington, in December 2017. Emergency managers and social media specialists noticed an immediate surge of propaganda articles assigning blame for the derailment to an anti-fascist group, despite no evidence supporting this claim.

Outdated Information

Today's media environment relies heavily on being first with information. When crisis rumors start to surface, novice and experienced users alike will scour the internet, often posting images of the initial returns from their search without first verifying the date or accuracy of the data they are sharing. This happens most often with users sharing photos from past disasters in a hurry as evidence of a disaster, which is often believed as being true as the phrase "pictures or it didn't happen" have permeated social

[22] False information deliberately and often covertly spread (as by the planting of rumors) in order to influence public opinion or obscure the truth. <https://www.merriam-webster.com/dictionary/disinformation>.
[23] A scam by which an internet user is duped (as by a deceptive e-mail message) into revealing personal or confidential information which the scammer can use illicitly. <https://www.merriam-webster.com/dictionary/phishing>.
[24] Cory Nealon. "False Tweets During Harvey, Irma Under Scrutiny by UB Researchers." University of Buffalo News Center, September 28, 2017. <http://www.buffalo.edu/news/releases/2017/09/044.html>.

media users' mindsets. Secondarily, older articles describing a past incident can resurface when reposted, and publication dates are changed automatically through re-posting. Some examples include:

- Old photos that are tagged and used for a new event. A user tweeted a photo from 2010 Nashville, Tennessee, flooding and incorrectly tagged it as Houston flooding in 2016. Using www.tineye.com, the photo was verified as from Nashville. The user was notified, deleted the tweet and replaced it with another photo.
- One of the most famous examples of an old picture (Figure 3) being shared erroneously is that of the Bitterroot fire complex in Montana in July 2000, which was one of *Time* magazine's photos of the year. This photo titled "Elk Bath" shot by John McColgan has been shared in every major wildfire since.[25]

Figure 3: "Elk Bath" photo taken in 2000, which has been erroneously shared and attributed to other wildfires.

As recently as 2016, this fire was used to represent a fire burning in Tennessee. The rumor was corrected by KRTV, correctly assigning the location and date the photo was captured.

[25] Photo courtesy John McColgan, Alaska Fire Service from NASA's Earth Observatory website: https://earthobservatory.nasa.gov/IOTD/view.php?id=843.

Case Studies

Incident Name: 2014 South Napa Earthquake

by Jennifer Lazo, City of Berkeley [CA] Emergency Services
Rumor Type: Phishing/Spammers Spreading Malicious Information

On August 24, 2014, at 3:20 a.m., a 6.0-magnitude earthquake struck the area of American Canyon and Napa, California. The earthquake shook awake many residents in the Bay Area and provoked a nearly instant social media response, particularly on Twitter. As information about the earthquake became available online, the hashtags #NapaQuake and #NapaEQ were broadly used by people in the affected area and those responding to the earthquake. Popular hashtags often serve as a useful way to find and sort relevant, actionable information during disasters, and the South Napa Earthquake was no exception.

Problem:
Soon after the earthquake, the top earthquake related hashtags began trending locally and across the world. As hashtags become popular on Twitter, "spammers" and "trolls" (i.e., people who sow discord on the internet, including through social media platforms) target those hashtags in an effort to have a broader audience for their unrelated message. In the case of #NapaQuake, a particularly disturbing "hashtag hijacking" took place. For much of the first days of the earthquake response, a significant portion of the tweets on the most popular response hashtags contained graphic pictures of dead bodies from unrelated events. The main subject of the hijacking tweets were accusations of U.S. military misconduct with images of people being tortured or horribly mangled bodies being included as evidence. This was shocking content for social media monitors who were used to dealing with more standard disaster response tweets, not inflammatory and graphic material.

It was unclear initially whether Twitter was taking any organizational approach to removing the inappropriate tweets. However, some techniques used by social media monitors were effective at blocking the worst of the content. One approach was to filter out geolocated and geographic information that did not seem relevant to the incident. Most of the graphic tweets originated from outside of the United States, so adding a simple geographic filter to tweets made it easier to filter out the hijacking tweets. Over time, it seemed that some of the tweets were also being deleted on the back-end by Twitter.[26] Newer algorithms on Twitter may hide these inappropriate tweets, but it is unclear if those filters may also pick up and hide disturbing but relevant and useful content that social media monitors need to function effectively, such as images of bodies left outside during Hurricane Katrina or anger directed at a response organization for alleged dysfunction.

[26] "Hijacked #NapaQuake Hashtag Turns Up Images Of Islamic Militant Slogans & Dead U.S. Soldiers." August 25, 2014. <http://sanfrancisco.cbslocal.com/2014/08/25/hijacked-napaquake-hashtag-turns-up-images-of-islamic-militant-slogans-dead-u-s-soldiers-james-foley-isis-ferguson-world-cup/>.

Best Practices:

The tactic of location-based filtering of unrelated tweets remains a best practice for those attempting to monitor social media during disasters, but in recent years, the spammers and trolls on social media sites have used different methods and tactics to overtake hashtags and cause confusion. Agencies should be aware that filtering by only location-based tweets can suppress local information originating from devices that are not geolocation enabled. A Georgia Tech study conducted in 2012 indicated that less than 1.4 percent of all content on Twitter is geolocated. [27] Another study conducted by Humanity Road and Arizona State University on 2012 Hurricane Sandy data indicated that there is a potential significant decline in geolocation data during weather events.[28]

Incident Name: 2016 Louisiana Floods

by Amy Greber, American Red Cross via email on July 5, 2017, and LaVondra Dobbs, ViaLink Louisiana via phone call on July 12, 2017

Rumor Type: Incorrect Information

[27] Ryan Gomba. "What Percentage of Tweets are Geotagged?" January 30, 2012. <https://www.quora.com/What-percentage-of-tweets-are-geotagged-What-percentage-of-geotagged-tweets-are-ascribed-to-a-venue>.

[28] Morstatter et al. "Finding Eyewitness Tweets During Crises." Arizona State University. <http://www.public.asu.edu/~fmorstat/paperpdfs/lang_loc.pdf>.

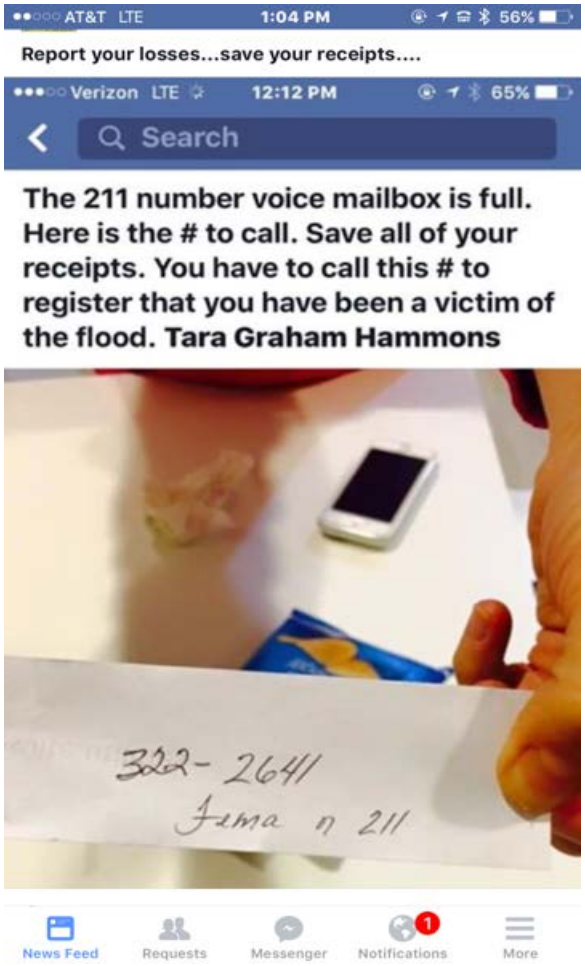


Figure 4: Inaccurate Facebook post claiming 2-1-1 would help with disaster claim forms.

Problem:

ViaLink Louisiana, a 2-1-1 provider, found itself overwhelmed with calls following the March 2016 floods in Louisiana. ViaLink noticed multiple inaccurate Facebook messages and posts (Figure 4) that went viral and contributed to the number of calls. After FEMA declared a disaster, the calls kept coming. In addition to the continuing Facebook messages and posts, FEMA was also giving out the incorrect information and referring people to 2-1-1 for claim assistance (the information was later corrected).

In a similar situation, during the response to Louisiana's summer floods in 2016, the American Red Cross was confronted with multiple rumors and misinformation on social media related to its shelter policies and food distribution. False claims spread especially rapidly through new video tools, such as Facebook Live, and threatened to erode the public's trust and support, as well as eclipse the personalized care and outreach that the organization was providing through social media.

To dispel rumors and also raise awareness of critical resources, the Red Cross published a blog entitled, "Top Questions About Louisiana Flood Relief" that was shared 2,175 times. The Red Cross, its digital volunteers and other online supporters used it to spread information among their personal networks, as well as to confront critics.[29] The Red Cross Social Engagement team also created a secret Facebook group where they could funnel important updates, flag urgent issues and collaborate among a larger public affairs team; any further outcomes were then updated on the blog. At the local level, the Red Cross communications team in Louisiana created regular informational videos from the organization's Louisiana warehousing site, where they provided situational updates for those seeking help and support – and also as a way to combat misinformation. The first on-the-ground video update alone received 447,000 views.

Best Practices:

Actively publish frequent updates to help promote transparency and control the message.

[29] Digital volunteers as applied to emergency management and disaster recovery is a group of trusted agents that can lend support via the internet to those on-site who may otherwise be overwhelmed by the volume of social media data generated during a disaster.

Using trained digital volunteers can help shepherd affected people to critical resources and spread reliable information online.

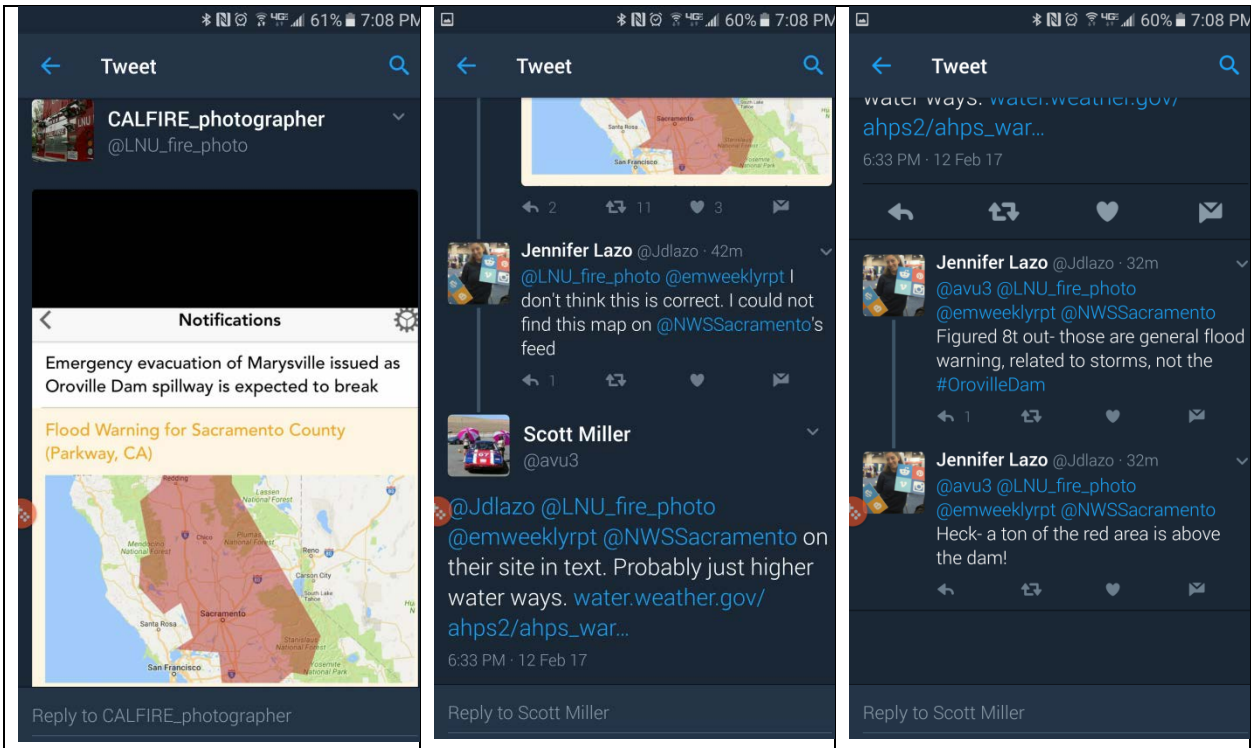
Incident Name: 2017 Oroville Dam Evacuation

by Mary Jo Flynn, Sacramento County Office of Emergency Services (OES) via email on January 3, 2018
Rumor Type: Insufficient Information

Due to rapidly eroding earth at the site of the emergency spillway that was compromising the integrity of a small portion of the side of Oroville Dam, an evacuation order was sent out to residents in the nearby communities and downstream of the dam to evacuate immediately. In the chaos of a rapid mobilization and evacuation of thousands of people, a significant amount of misinformation spread rapidly.

Problem:
The National Weather Service (NWS) Sacramento was distributing hydrology information impacting the rivers, dams and levees within the northern portion of the Sacramento Valley. The area was receiving significant rainfall accumulation from repeated Atmospheric Rivers causing widespread flooding. Prior to the Oroville Dam evacuation, the NWS Sacramento station distributed a tweet that showed the area of flooding from expected rainfall. This mapped image encompassed all of Sacramento County and areas that included Oroville Dam. This image was distributed by @LNU_fire_photo (Figure 5A) just after 6p.m. local time with information regarding the evacuation of Oroville Dam, leading those who saw the image to believe that all of Sacramento County was under an evacuation order. Even though @JDLazo was correcting the errant tweet (Figures 5B and 5C) at 6:30 p.m. local time, calls began flooding Sacramento County 9-1-1 dispatch in the hours that followed. At 8:50 p.m., staff from Sacramento County EOC completed a Facebook Live Video and a Periscope video to correct misinformation. Almost immediately, news media including television and iHeartRadio began quoting the live feeds in their broadcasts and the videos were widely circulated on social media.

Best Practices:
This rumor was believable due to the easily misconstrued flood map during the evacuation, and people were generally starved for information. The best practice of having support from a digital volunteer helped tremendously in minimizing the rumor; however, once it took hold, people's genuine concern caused increased stress and anxiety, and resulted in calls to 9-1-1. The use of live streamed video was another best practice in this situation by getting the right information to the right people at the right time. Because of previous live broadcasts, television and radio news media followed Sacramento County OES's social media feeds to share pertinent information. Pre-established partnerships with 3-1-1 and 2-1-1, as members of the Joint Information Center, were receiving the same social media feeds and information directly to provide to their callers. The radio stations broadcasted the audio portion of the live video feeds, and television stations utilized their anchors to quote information from the Sacramento County EOC. After the broadcasts, the County's 9-1-1 dispatch center reported calls regarding Oroville had stopped and their call volume had returned to normal.



A B C

Figure 5: Twitter conversation during Oroville Dam crisis discussing misinterpretation of Sacramento County flood map.

Suggested Best Practices

Agencies should first determine their comfort level with the various tactics suggested in this section. It will likely vary from agency to agency and will also depend on technology, training and policy issues. Agencies should focus on the information sharing challenge or goal, not the tool (there are many possible tools to use).

Here are some recommended goals for agencies to keep in mind when using any of the suggested best practices below to counter misinformation, rumors and false information:

- Improve the quality of communication through double-verification of information;
- Remove ambiguity and uncertainty caused by misinformation, rumors and false information;
- Reduce alert fatigue and the risk of "cry wolf" scenarios;
- Seek ground truth as opposed to assumptions;
- Be swift with releasing accurate information or acknowledging the situation to help the agency work with a network of truth amplifiers and establish credibility early on; and
- Determine relevance of various social media information.

Best Practices – People

- Work with local television and radio news media (traditional media) to disseminate useful information and correct bad information. Pre-establish partnerships or relationships with stations or reporters that can be called upon in a disaster or emergency. While this paper is focused on social media and increasingly more people are turning to this medium for their news, a recent Pew report estimates 50 percent of people still get their news via television broadcasts, compared to 43 percent who get their news online.[31]

Hurricane Harvey and the Oregon VOST

By Elizabeth King, University of Washington Emergency Management

Local Texas emergency operations agencies and FEMA Region VI experienced a tremendous amount of social media traffic in the wake of Hurricane Harvey. A FEMA Digital Reservist suggested that FEMA reach out and request support from a VOST in three mission areas. The first mission involved tracking and delivery of large donations, and the second mission addressed tracking the recruitment of volunteers and their locations. The Oregon VOST provided a daily listening report to the FEMA Region VI Mass Care Public Information Officers. Among other things, they discovered a volunteer group using a Google Sheet that was publicly available to collect personally identifiable information, which was shared in the listening report. The third mission of the Oregon VOST was tracking scams for donations, including spoofing the American Red Cross and webpages that redirected funds for victims to other funds unrelated to Hurricane Harvey. Multiple VOSTs worked together to provide much-needed support and not duplicate efforts.[30]

- Use Virtual Operation Support Teams (VOSTs) or other digital volunteers to monitor social media, identify rumors and report back to officials so they can work to correct it. Prepare workflows, practices and activation procedures. The SMWGESDM's Operationalizing Social Media report offers guidance in this area.
- Create partnerships through mutual aid agreements with FEMA, Red Cross, or other agencies and organizations that have the necessary skills, personnel and systems to identify rumors and misinformation. These partnerships could also be collaborative efforts among local, state and federal agencies, as well as jurisdiction-to-jurisdiction (e.g., local partners amplifying facts on behalf of other municipalities if and when requested).
- Identify and leverage trusted crowd sources or influencers, such as on-ground/on-scene users and emerging influencers to propagate critical 'good' information perceived by the crowd and engage them to disseminate rumor correction information. For example, the National Voluntary Organizations Active in Disasters network is on-the-ground and can be trusted to provide key information because the organizations work closely with the government.
- Train and exercise first responders and digital volunteers to spot misinformation, rumors and false information, and when and how to respond to bad information. Develop rules of engagement or a concept of operations document for when or when not to respond. One way

[30] Information provided by Elizabeth King, University of Washington Emergency Management via email on January 31, 2018.

[31] Pew Research Center, "Americans' Online News Use is Closing in on TV News Use." September 7, 2017. <http://www.pewresearch.org/fact-tank/2017/09/07/americans-online-news-use-vs-tv-news-use/>.

to research questions from the public is to search social media platforms with the event name and a question mark (for example "flood?").

Best Practices – Processes

- Use the Joint Information System (JIS) to coordinate public information efforts among jurisdictions and agencies, and for standing up a Joint Information Center (in-person or virtual) to facilitate the operation of the JIS.
- Prepare pre-scripted messages and choose or create a hashtag specific to place, disaster, agency, etc.[32] Be consistent and provide useful and actionable information.
- Prepare draft visuals, including graphics, photos and videos, for potential disaster scenarios. Archive these in such a way that they are rapidly accessible and can be quickly modified to meet the needs of a situation. As an example, a tornado warning graphic (Figure 6) might include simple instructions on what to do.

Figure 6: Sample tornado safety graphic.

- Actively tweet or post to correct misinformation, rumors or false information. Use hashtags such as #mythbuster, #RUMOR or #IncidentNameFact in posts and redirect back to official sources. Use checkmark emojis or big red Xs on images. "Share back" some of the misinformation (labeled as such) with the general public while the events are still relevant to directly address inaccurate tweets or posts.
- Ensure older information is appropriately labeled and not re-circulated as new:
    - For critical information, continue updating a single Facebook post or existing news story, rather than starting a new one.
    - Use the pinned post features on Facebook and Twitter so critical information remains at the top of the news feed.
    - List known conspiracy sites and consider creating lists of bad actors in order to have awareness of their posts.
- Verification tactics:
    - Reference this infographic to do a quick fact check of the information.
    - Verify the legitimacy of information disseminated on social media platforms using two or more sources, which could be websites (Factcheck.org or Snopes.com), social media

[32] Document provided by Emergency Operations Coordinator Mary Jo Flynn, Sacramento OES.

platforms, eyewitnesses or first responders (ground truth). Compare questionable claims or posts to content being disseminated by official sources.

- Acknowledge the rumor when appropriate. Statements such as “We are aware of the rumor about [Topic].” Then use the opportunity to:
  - State the facts; and
  - Direct readers to sources of correct information.
- Conduct reverse image searches using Google and TinEye. These may also be done via this shortcut in Tweetdeck.
- Use a geofence and/or Twitter searches for locations to mine and help separate real and false information (false information often comes from outside the disaster-affected area).

- Consider a cross-entity social media campaign as a complement to more official resources (incorporating something like a "verify2x" hashtag so that it could also be shared by partner organizations).

Best Practices – Technology

- Set up a central website as a one-stop shop/portal similar to the Federal Emergency Management Agency’s (FEMA) rumor control pages (Figure 7 is a recent example from Hurricane Harvey). Have a generic page ready to go in case of disaster, so an agency can turn it on and start publishing. Embed a list of social media accounts that provide information in an emergency (i.e., local police, fire, city, 311, etc.). This way, the most up-to-date information is always available on that page.

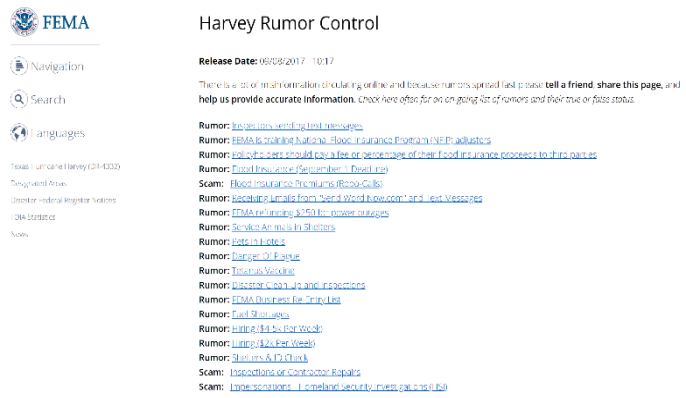


Figure 7: FEMA’s Hurricane Harvey rumor control page.

  - If a rumor contains information that provokes a strong emotional response, it is probably a troll. Agencies should be careful not to directly engage the troll, as they will often go through great lengths to increase what they are doing. Instead, attack the content of the rumor through acknowledgment or posting to a rumor webpage.
- Consider using Tweet hyperlinks embedded within webpages and blog posts that encourage sharing of factual information. A sample tweet might contain part or all of the message and easily allow a user to share.
- Use live video (Periscope, Facebook Live or YouTube) to correct information with empathy and accuracy. An example is this video from the Oroville Dam crisis where rumors were addressed and corrected.

- Monitoring software, such as Hootsuite or Tweetdeck, and more powerful tools, such as Dataminr, Liferaft or Meltwater, can help agencies identify false information and watch for influencers. This white paper does not endorse a particular product; however, such tools can be a helpful addition, taking into account an agency's circumstances (i.e., cost, human resources, etc.)

Additional Considerations

With the best practices given above, agencies should be aware of the following additional considerations when choosing to counter or correct misinformation, rumors and false information.

There are some risks to engagement, including:

- Accidentally amplifying the false message;
- Repeating the false message, even with a correction, can lead to more people believing it; and
- Correcting efforts can "backfire," i.e., challenging a story may cause people who already believe the false information to believe it even more.

Furthermore, the following factors should be considered:

- Geography: If the rumor is a local rumor, agencies should address it. If the rumor is not spreading among the locals, agencies and responders may want to keep it on their radar for their situational awareness but not address it.
- Volume: If the rumor is at a low volume, keep an eye on it. If the rumor spikes and becomes high-volume, respond with a message highlighting the true facts that does not repeat the false story.
- Subjects: For stories around collaborations and humanitarian response, highlight the positive side of the story to indirectly counteract the effect of negative rumors.

Challenges

Below are some challenges first responder agencies may encounter when countering misinformation, rumors and false information.

Legal Issues: Agencies do not want to incur legal liability when correcting information and rumors. They are recommended to consult their legal department or general counsel's office in their jurisdiction to create a framework to cover an agency's action, which can be challenging, as the legal system is not always keeping up with the pace of rapidly advancing technology. In addition, agencies should consider the need for continued engagement with the public once an issue has been identified and how to best handle that so they do not incur additional liability.

Lack of buy-in from executive staff and decision makers: Agencies need to establish clear rules of engagement for responding to rumors and false information. As part of establishing these rules of engagement, agencies should engage with executive staff and decision makers (including the legal office, see above) to ensure that everyone is on the same page regarding the rules of engagement.

Privacy: The SMWGESDM's previous report on using social media for situational awareness and decision-making discussed privacy issues around social media, so they will not be repeated here.

Funding: Some of the best practices recommended in this paper require funding allocations to support training, personnel and technology. If an agency chooses to implement a particular social media monitoring technology, for example, they will require funding to support acquiring and using that technology.

Conclusion

As rumors, misinformation and false information will continue to circulate, they cannot be entirely eliminated. Agencies can leverage the above proactive and preemptive measures to lessen the risks during disasters and emergencies as a result of misinformation, rumors and false information. Some of the measures detailed in this report include mutual aid and partnerships with credentialed digital volunteers, pre-scripting messages, verification tactics, setting up a centralized webpage and more.

Agencies should consider testing and exercising with rumors, misinformation and false information to help them determine which best practices will work best for their audience. The SMWGESDM's previous report on incorporating social media into exercises offers how-to guidance.

Social media is a continually changing topic, and while the tactics discussed in this paper are relevant now, the landscape continues to evolve. In the future, the authors of this paper may add to this paper or create an external living document of references and resources that may be relevant for first responder agencies.