Trust, but verify: social media models for disaster management

Amisha M. Mehta Associate Professor, QUT Business School, Queensland University of Technology, Australia, **Axel Bruns** Professor, Digital Media Research Centre, Queensland University of Technology, Australia, and **Judith Newton** Researcher, Digital Media Research Centre, Queensland University of Technology, Australia,

A lack of trust in the information exchanged via social media may significantly hinder decision-making by community members and emergency services during disasters. The need for timely information at such times, though, challenges traditional ways of establishing trust. This paper, building on a multi-year research project that combined social media data analysis and participant observation within an emergency management organisation and in-depth engagement with stakeholders across the sector, pinpoints and examines assumptions governing trust and trusting relationships in social media disaster management. It assesses three models for using social media in disaster management—information gathering, quasi-journalistic verification, and crowdsourcing—in relation to the guardianship of trust to highlight the verification process for content and source and to identify the role of power and responsibilities. The conclusions contain important implications for emergency management organisations seeking to enhance their mechanisms for incorporating user-generated information from social media sources in their disaster response efforts.

Keywords: agency theory, emergency management, natural disasters, social media, trust

Introduction

Trust plays a critical role in communication, exchange, and decision-making. Typically, it is built over time by drawing on personal sources of information that legitimise and reduce risk in trusting others. Impersonal trust, or trust without a personal history (Shapiro, 1987), is a commonly accepted process that operates in relationships between and among citizens and institutions such as government and the mass media. For instance, a journalist acts on behalf of readers or viewers to identify and investigate the story, verify facts and sources, and share knowledge that is likely to influence opinion and decision-making (Mehta, 2007).

The context of disaster management provides an interesting lens with which to examine trust, which assumes new meaning as people use information to make swift and critical decisions

about how to protect lives and property. Critical and operationally relevant information is now commonly exchanged via social media and other digital platforms that are well-suited to delivering data directly to citizens and emergency management organisations and to encouraging exchanges between them. In such environments, however, information may not pass through the usual process of verification, challenging the way in which both citizens and governments accept, process, and act on it.

Disasters create a unique space in which to consider how the concept of time can challenge traditional models of trust in communication. Indeed, recent research following earthquakes in Chile showed how disasters established an opportunity to build interpersonal trust in communities (Dussaillant and Guzmán, 2014). This paper, drawing on the unique environment presented by disasters, investigates the intersection between the management of natural disasters, trust, and social media to understand better how trust is verified online. This process builds on existing work that categorises the functionality of social media (Jung and Moro, 2014; Houston et al., 2015) by assessing the verification of information during disasters using a guardianship of trust model. This approach highlights the mechanisms that guide and assure trust and influence power in relationships between individuals and emergency services organisations.

To achieve this task, this study defines trust both from a theoretical perspective and in the context of disaster. The next section outlines the methodology used to evaluate three broad approaches utilised by emergency management organisations to validate social media content during disasters. Three models for social media and trust are then presented and explored. Finally, the paper sets out some implications for social media management during disasters.

Trust during disasters

Trust, in the context of disasters, is an important precursor to risk perception and risk acceptance (Poortinga and Pidgeon, 2003), making it an essential component of risk and crisis communication. While there are many definitions of trust, a broad one is that it is a belief that, in the absence of any evidence, things will work out (Gambetta, 1988). Mayer, Davis, and Schoorman (1995, p. 712) offer a more specific one: the 'willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party'. This definition, drawn from management research, underlines how

trust traditionally is understood as a dyadic process between two parties: a trustor or trusting party; and a trustee or party to be trusted (Mayer, Davis, and Schoorman, 1995). Trust, though, also functions in more complex relationship models composed of multiple parties and varying levels of power.

Trusting beliefs are based on three characteristics—ability, benevolence, and integrity—which comprise the trustworthiness of a person or organisation and are used to evaluate the characteristics and actions that lead to trust (Mayer, Davis, and Schoorman, 1995; Schoorman, Mayer, and Davis, 2007). Perceived ability is defined as the skills and competencies that enable influence within a particular domain (Mayer, Davis, and Schoorman, 1995; Tomlinson and Mayer, 2009). Perceived benevolence is the extent to which the 'trustee is believed to want to do good to the trustor' (Mayer, Davis, and Schoorman, 1995, p. 718). Perceived integrity refers to the trustee's adherence to standards, principles, and values relevant to the trustor (Mayer, Davis, and Schoorman, 1995; Tomlinson and Mayer, 2009). When ability, benevolence, and integrity are perceived to be high, high levels of trustworthiness follow (Mayer, Davis, and Schoorman, 1995). Disaster researchers also recognise the various aspects of trust, identifying competence, vested interest, and scepticism (Haynes, Barclay, and Pidgeon, 2008), which constitute trust and which influence judgements and behaviours related to it.

Trust under uncertainty and time pressures

Ideally, trust builds over time, as institutions, organisations, and people engage and generate knowledge and develop confidence in each other. Trust can be a taken-for-granted part of exchange or transactional relationships, yet often it is tested in situations such as natural disasters. The uncertainty and time pressures of disasters can work together to challenge existing trust and to create new pathways for it. While emergency management organisations are frequently highly trusted agencies, disasters present multiple uncertainties that can limit their expertise and can lead community members to value alternative sources (Haynes, Barclay, and Pidgeon, 2008).

Disasters also bring together rapidly a number of official and unofficial stakeholders who may not have any pre-existing relationships with each other or familiarity with events (Paton, 2007). In such a setting, people may judge trust intuitively by drawing on similarities in perceived values rather than undertaking systematic assessments (Poortinga and Pidgeon,

2004, p. 1476). This kind of trust preferences connections with those who have similar social identities and evaluations of the circumstances (Poortinga and Pidgeon, 2004), which may conflict with the advice of emergency management organisations.

The foundations of impersonal trust

Owing to time pressures and uncertainty, impersonal trust may replace personal trust as a driver of evaluating information during disasters. Impersonal trust exists when people or institutions lack common personal links, a history, or even the potential for ongoing exchange (Lane, 1998, p. 12). Impersonal trust may be situation-specific and short term among institutions and people, but it can also become a means to underwrite personal trust (Luhmann, 1988). Shapiro (1987, p. 634) describes the process and the importance of impersonal trust:

> Impersonal trust arises when social-control measures derived from social ties and direct contact between the principal and agent are unavailable . . . when faceless and readily interchangeable individual or organizational agents exercise considerable delegated power and privilege on behalf of principals who can neither specify, scrutinize, evaluate, nor constrain their performance.

There are a number of mechanisms available to institutions, such as emergency management organisations and the media, to exhibit swiftly impersonal trust to the general public, but knowledge of how ordinary people develop and demonstrate impersonal trust among each other and to emergency services agencies is scarce. Agency theory offers some insights to fill this gap. Here, trust is based on an exchange or network of relationships. In principal-agent relationships, all parties work towards outcomes from an impersonal basis, and principals can invest authority or responsibility in agents to act on their behalf (Shapiro, 1987, p. 626). In trusting agents, principals bear risks because they cannot observe the agent's efforts in completing required tasks (Shapiro, 1987; Beccerra and Gupta, 1999; Zech, 2001; Miller and Whitford, 2002). As a result, principals trust agents to 'bridge the barriers of direct physical access to information and property' (Shapiro, 1987, p. 627).

Trust between the community and emergency management organisations in a time of disaster is crucial as it enables these organisations to independently respond to, protect, and minimise harm to the community they serve. Not only does this apply to impersonal trust in the sense of a physical response, but also in relation to social media communication. Without any direct

or personal contact, the community, as principals, must trust the information and warnings provided by emergency services, as they act as organisational agents.

Vulnerability in situations of impersonal trust led Shapiro (1987, p. 635) to introduce the concept of guardians of trust, 'a supporting social-control framework of procedural norms, organisational forms, and social-control specialists, which institutionalise distrust', on which principals rely. Traditional guardians of trust are institutions such as government or professional associations, as well as agents and private entrepreneurs (Shapiro, 1987, p. 635). For instance, professional bodies use membership or accreditation to monitor the performance of members and to hold those in positions of trust to account (Shapiro, 1987). To achieve positive outcomes within the principal-agency relationship, Eisenhardt (1989, as cited in Creed and Miles, 1996, p. 23) focuses on the use of 'controls and incentives to combat the purported tendency of agents to engage in suboptimal and/or self-serving behaviours'.

Social media and communication during disasters challenge some of the theoretical assumptions of principal-agent relationships, creating opportunities for the evaluation and production of ongoing trust. Social media make visible transactions that enable principals to monitor information posted by organisational agents, through multi-way communications with other social media users, crowdsourced information, self-regulation, and debunking of posts (Mendoza, Poblete, and Castillo, 2010; Veil, Buehner, and Palenchar, 2011). This can be quite problematic in some cases for emergency management organisations as the likelihood of building trust relationships with individual users across wide geographical areas where disasters have the potential to strike is limited. In addition, it restricts emergency management organisations' ability to trust information coming from their online community without validation of content.

At the same time, the perceived authority and expert knowledge of emergency management organisations can facilitate the building of a trust relationship between the online social media community and related organisations. This trust may be strengthened further when organisations share a history with their online networks (Veil, Buehner, and Palenchar, 2011). For example, the interaction of the Queensland Police Service with its online community during the devastating flooding in the northeast state of Australia in January 2011 (Bruns et al., 2012) generated enough impersonal trust to sustain ongoing network growth and multi-

way communication in subsequent times of crisis and routine operations. Social media thus provide a channel for organisations to transform impersonal trust into interpersonal trust via a shared history, sustained social media interactions, and the emerging reputation with the online community that is gradually forged.

Trust and emergency management processes and policy

Research shows that, in some instances, emergency management organisations have been hesitant to trust and act on user-reported information from social media platforms, owing to the inability to validate message content within a very short timeframe (Crowe, 2012; Haddow and Haddow, 2014). This positioning frequently limits social media to serving as tools for dissemination, rather than as platforms for systematic information exchange and usage (Lindsay, 2011). Five years ago, Gao, Barbier, and Goolsby (2011) pointed out that knowledge management during disasters was emergent, highlighting the potential for trust, crowdsourcing, and effective and efficient responses (Yates and Paquette, 2011, p. 13). At the same time, the dangers of social media as knowledge management tools also were noted, including accuracy of information exchanged, malicious uses, technological limitations, administrative costs, and privacy (Lindsay, 2011).

This paper analyses the ways in which trust is produced, guarded, or verified on social media during disasters as a means to inform emergency management processes and policy. Specifically, it explores the potential mechanisms by which emergency management organisations may come to trust the information originating from the wider community of social media users in general, and may identify especially trustworthy individual users with whom it would be valuable to develop a closer relationship built on interpersonal trust.

Methods and data sources

The observations presented in this paper build on a multi-year research project conducted at the Centre for Emergency and Disaster Management, Queensland University of Technology, and funded by the Australian Research Council. The project, entitled ‘Social Media in Times of Crisis’, reviewed the social media strategies of Australian emergency management organisations during a number of recent natural disasters, with a focus on Facebook and especially on Twitter. The mixed-methods quantitative and qualitative analysis of social media activities by emergency management organisations and their followers (see, for example, Bruns et al., 2012) was complemented by the placement of a long-term researcher

with the social media team of the Queensland Fire and Emergency Services (QFES), enabling participant observation of its social media activities, including during a number of recent disasters, such as cyclones and storm-related flooding throughout the state's annual summer storm season between November and April.

In addition, the research project developed a nationwide network of contacts across the Australian emergency services sector, concentrating particularly on media and social media officers in emergency services organisations and the fire and police services. Drawing on this network, the project convened a number of roundtables and workshops to gather information, experiences, and opinions on the use of social media in crisis communication and disaster management, and to gauge the need for further research and policy initiatives in this field. A policy proposal seeking to enhance recognition of and support for social media teams within emergency management organisations is currently being developed.

This paper presents a synthesis of the observations made through these research activities. As well as secondary sources, it draws chiefly on analysis of observable social media activities on Facebook and Twitter, on participant observation of and follow-up interviews with QFES social media team personnel, and especially on several workshop and roundtable discussions with staff from a range of Australian emergency services organisations. The bulk of this data gathering was conducted on condition of anonymity to enable staff to speak freely without fear of repercussions from higher levels of management. To protect the informants, the viewpoints are presented as a composite, idealised picture that refrains from citing individual statements and from referring directly to the specific approaches of any one of the organisations involved in the study.

Models for social media use

From the observations across an array of emergency management organisations, one can distinguish three broad models for the positioning of social media use in such entities, each pointing to a different process of developing and verifying trust in social media users and the crisis-relevant information they share (see Table 1). The three models as presented here serve as ideal types; in their practical application in emergency management, combinations of these approaches also can be found. Furthermore, each of the models reflects a different conceptualisation of the role of social media within the overall emergency management and

crisis communication effort, and thus necessitates a change in the organisational placement of staff charged with operating the social media activities.

Although the focus of the research is on the use of social media during natural disasters, it is likely that these models also apply to other human-made disasters and crises. Similarly, while the research engaged exclusively Australian emergency management organisations, these models should also be of relevance to bodies in comparable developed nations, and potentially beyond.

Table 1. Three models for online trust in disasters

Social media model	**Verification process**	**Trust relationship**
Intelligence gathering to detect and track disaster patterns.	Individual updates not considered; decisions based on aggregate patterns in large datasets.	Traditional principal-agency model with emergency services organisations holding power; limited ability for individuals to monitor process.
Quasi-journalistic verification of the content and source of individual messages.	Content and source verification based on evaluation by emergency services staff; small-scale, manual information processing.	Multiple dyadic relationships involving interpersonal trust principles with emergency services assuming power; ability for individuals to monitor and be involved in the process.
Crowdsourcing of verification processes.	Content and source verification based on evaluations by communities of volunteers; fast, distributed information processing.	Networked relationships drawing on impersonal and interpersonal trust; power and monitoring capability shared among users; may result in inaccuracy depending on user intentions.

Source: authors.

Intelligence gathering

The first trust approach treats social media as a source of intelligence on the situation in the crisis area during an emergency. This model seeks both to discover previously unknown emergencies by detecting unusual spikes in social media activity that reference key crisis-related terms, such as earthquake, fire, flood, or storm, and to track known events by following established markers, such as the Twitter hashtags that emerge in response to such

situations. Such detection and tracking centrally utilises programmatic means of discovery and evaluation, including real-time social media analytics that assess volumetric, network, and sentiment patterns in the incoming data stream. Where feasible, these are combined, especially once indications of a potential crisis have been detected, with further manual evaluation and oversight—cf. Stieglitz et al. (2014) on the overall concept of social media analytics.

The social media content identified through this intelligence gathering approach is then further correlated with other data sources available to the emergency management organisation, including reports from emergency staff on the ground as well as background data on crisis-affected locations. This means that social media data are not considered to be accurate and reliable per se, and that individual updates on the local situation are not examined separately. Instead, they are utilised in aggregate form and correlated with other data sources, which often similarly exist in aggregate form, to paint a broader picture of the current situation. This focus on data *patterns* rather than on individual data *points* also removes any inherent need to engage directly with social media users as the sources of these data points (for example, to ask them for clarification on their updates). As a result of this purely observational and largely automated approach, this intelligence gathering model also represents the fastest and least resource-intensive approach to extracting information from social media feeds, at least once the technical infrastructure for the real-time tracking and analysis of such feeds has been established.

In this model, then, the staff of the emergency management organisation who operate the social media intelligence gathering tools represent the only principal actors in the information verification process. Social media users are treated as agents, but are observed from an aggregate, bird's-eye perspective rather than engaged as trusted partners in the information verification process. Indeed, their updates are never verified as such, with confidence in the situational patterns emerging from the social media feeds being established only because of the correlation of such patterns with other sources of information (which themselves may also not be trusted entirely).

While this use of social media in intelligence gathering is still in its formative stages, emergency management organisations are beginning to develop processes to identify and validate intelligence found on social media platforms. Tools that analyse social media

messages using data mining techniques and natural language processing are able to provide authorities with enhanced situational awareness in a time of disaster. This on-the-ground and near real-time information is crucial to emergency management organisations to enable informed decisions to be made in relation to response and recovery efforts (Yin et al., 2012, pp. 52–53). For the present purposes, however, it addresses the challenge of developing trust in individuals and groups of social media users largely by bypassing it altogether: only the collective, aggregate patterns of social media activity among very large numbers of users are taken into account by such approaches, and even then often only as early indicators of an emerging crisis. Any further and more fine-grained engagement with the social media content underpinning any sudden spikes in activity requires different, less automated approaches, and hence builds on different mechanisms for establishing trust in the information provided by users.

Quasi-journalistic verification

The second major model for using social media-derived information in emergency management, quasi-journalistic verification, shares some traits with the intelligence gathering approach. However, it operates at a smaller scale and with considerably more direct intervention from the organisation's social media staff. This model bases its core activities on long-established journalistic practices of source verification, which, at least in their orthodox formulation, require any unit of information to be confirmed by at least two independent sources before it becomes fit to publish. The social media staff in emergency management organisations who operate under this model therefore tend to take a more hands-on approach to monitoring social media feeds (utilising technical assistance from social media analytics applications to identify indicators for current and emerging crisis situations, but then tracking manually the feed of updates relating to such events), and focus less on the raw patterns of relevant social media activity than on individual items of information appearing in these feeds. Much like journalists, they mostly use their own judgement, based on professional experience, brief background searches, and common sense, in separating trustworthy from untrustworthy updates (and, over time, reliable from unreliable social media users).

Emergency management organisations have been slow to trust information received from social media platforms during disasters. A study by San Su, Wardell, III, and Thorkildsen (2013, p. 50) of emergency management organisations in the United States and their social media use found that more than 75 per cent of these agencies would not act on social media

information unless it was verified by another response agency or trusted source (see also Haddow and Haddow, 2014). However, while this concern about the accuracy of social media content is persistent and perhaps even justified, the increasingly widespread use of social media during a crisis pushes emergency management organisations into considering how to integrate information collection and verification strategies into emergency management policy and planning (Sutton, Palen, and Shklovski, 2008, p. 7).

In pursuing this second trust approach, then, social media staff follow the journalistic principle of independent verification wherever possible: they seek to receive confirmation of apparent situational updates from the crisis area from at least two social media users on the ground who have reported the situation independently of each other (such as in original tweets that do not retweet or @reply to each other), and ideally to correlate such information emerging from social media with reports from emergency management staff in the area. Such verification may also involve direct engagement with social media users through public @replies or non-public direct messages on Twitter, or their equivalents in other channels—that is, they assume an active role in information gathering and verification, rather than merely serving as the passive monitors of information streams seen in the previous model.

Such active verification results in greater levels of confidence in the information extracted from social media sources than was possible in the first trust approach, even without correlation with other sources available to the emergency management organisation, and over time generates institutional knowledge of reliable users in specific areas who may be called upon again in the future to provide situational information crisis updates. The social media staff in the emergency management organisation continue to act as principals in the information evaluation process, therefore, but social media users become more active agents in their own right and are invited to engage more directly in two-way conversations with the principals, rather than merely being observed from a distance by them. More elaborate, manual verification processes, though, also come at a cost: they are more time- and resource-intensive, and they are unable to process very large volumes of social media updates with the speed that may be necessary during a disaster. A combination of automated intelligence gathering and manual verification approaches may be required, therefore, to address these limitations.

From a trust studies perspective, this quasi-journalistic approach to verification attempts as much as possible to fast-track emergency services staff members' interpersonal trust evaluations of previously unknown social media users. For instance, while emergency services staff can appraise briefly the users' social media track record and gather publicly available information on them (from search engines), this information must also be subjected to personal assessment. By examining other users, emergency services staff rely to a considerable extent on their 'gut feeling', which is developed through sustained professional practice in emergency management and personal practice in actively using social media themselves.

By imposing the journalistic criterion of independent verification by multiple unrelated sources, this approach also attempts to minimise further negative repercussions from possible misjudgements. In this way, the medium becomes both an information source and the conduit for building and monitoring trust. The performance of individual social media users as information sources subsequent to the initial trust judgement may also confirm or challenge the staff member's perception of a user as a trustworthy source. This could eventually lead to the development of a more permanent personal trust relationship lasting to (and perhaps beyond) the conclusion of the crisis.

Crowdsourcing

The third trust approach is crowdsourcing. The development of interpersonal trust relationships between emergency management staff and social media users is time-consuming and attention-intensive, and hence rarely is feasible in acute crisis contexts beyond a small number of clearly important social media users acting as key information sources. However, within social media, which by definition are designed around the establishment and maintenance of *social* relationships between users, a substantial and complex network of interpersonal trust relationships already exists among users that predates the emergence of any disaster that requires the attention of emergency management organisations. Rather than attempting to establish quickly trust relationships with a significant number of social media users during a disaster, emergency management organisations may tap into existing information verification processes that are occurring constantly and naturally between users in social media environments.

Consequently, this model for gathering and verifying crisis-related social media information departs more substantially from the other two approaches. Here, social media (or more broadly, internet) users are more directly involved in the verification process themselves, rather than merely serving as sources of information on the current situation in the crisis area. This has the potential to speed up the verification process, as a larger base of participants may now be available to support that process than may be available even in a very well-resourced emergency management organisation. Yet, it can also create doubt about the reliability of this verification process, as a greater number of self-nominating volunteer participants with unknown levels of expertise in evaluating situational information are now involved in the endeavour. Given the uncertainties associated with this approach, the crowdsourcing model has been trialled to date only in a handful of crises, and often outside of, but sometimes with the implicit or explicit support of, relevant emergency management organisations. To the extent that such trials have resulted in demonstrably positive outcomes, a further mainstreaming of crowdsourcing approaches across emergency management organisations may follow in the future.

The crowdsourcing model borrows from the other two approaches. First, it builds in part on the automated intelligence gathering method by utilising similar tools to detect and acquire situationally-relevant social media updates in real time, and it may engage in some degree of aggregation and preselection to identify those updates that are most important and likely to be evaluated further by the crowdsourcing community. Second, it follows up on this initial detection and filtering stage by openly inviting volunteers to assess the potential trustworthiness and reliability of individual social media updates based on their own common-sense assessment. Next, it aggregates these multiple evaluations for each individual update (and in more sophisticated systems, also for each unique user in the dataset, where the user has posted more than one situational update) to generate confidence scores that point to those updates considered as most important and most trustworthy. Ranked in order of confidence, and filtering out the updates with confidence scores below a certain threshold, these evaluated social media updates may then be fed back into the internal processes of the emergency management organisation, where they can be correlated with situational and background information from other sources.

This approach, then, treats emergency management staff and social media participants alike as both agents and principals in the process of information verification. In facilitating the

crowdsourcing process, staff are able to set a range of general parameters for how to proceed (for instance, the selection of social media updates to evaluate, and the criteria for evaluation), but otherwise they depend on the voluntary contributions of a sufficiently large and motivated group of participants as a fundamental requirement for the success of the exercise. An emergency management organisation's ability to attract the user base necessary for this approach therefore also depends crucially on its standing in the general community: users will be motivated to participate in the crowdsourcing effort only if they feel that their effort is recognised by the organisation, and that it makes a difference in the management of the emergency at hand.

Crowdsourcing became prominent during the Haitian earthquake in January 2010, where volunteers monitored traditional media and social media sources to uncover critical information, such as about where people were trapped, about building collapses, and about medical emergencies, and then plotted this information on a live map that was available online. Emergency services then used this critical information in their disaster response planning (Haddow, Bullock, and Coppola, 2014, pp. 158–159). Similar crisis mapping approaches, often using the Ushahidi Maps platform,[1] were pursued in subsequent disasters. The Fukushima Daiichi Nuclear Plant meltdown that occurred as a result of the Tohoku earthquake in Japan in March 2011 sparked a different, more long-term crowdsourcing effort: following the disaster, the Government of Japan and the country's media initially played down its severity and did not provide the community with accurate information on the levels of radiation that ensued. In response to such misinformation, members of the public used social media to recruit and mobilise volunteers who were trained to measure radiation levels themselves and then to record and post these findings using Google Maps.[2] The results of this collective effort forced the government to become more transparent and to furnish the community with greater detail about the levels of radiation caused by the meltdown (Appleby (2012), as cited in Haddow, Bullock, and Coppola, 2014, pp. 158–159).

Crowdsourcing was used similarly in numerous ways during Hurricane Sandy in the US in October 2012 to provide information on the location of shelters, road closures, damage caused, and essential services that were still in operation, as well as to advise communities on

[1] See https://www.ushahidi.com (last accessed on 28 July 2016).

[2] See https://www.google.com/maps (last accessed on 28 July 2016).

the availability of desk space, electricity, and internet connectivity (UN-SPIDER, n.d., pp. 2–8). This widespread use of crowdsourcing led Haddow, Bullock, and Coppola (2014, p. 149) to claim that the employment of social media during Sandy was instrumental in driving change in the way governmental agencies use social media for information dissemination, rumour control, and connecting people in a crisis.

Towards more effective integration of social media information

The steps towards drawing on information sourced from social media in emergency management, as described by the three models, and the necessary approaches to developing trust in social media users and their information, to date remain fledgling at best, however, and there is still considerable reluctance on the part of many emergency management organisations to engage fully with the complex, changeable, and (to them) often still very alien spaces of contemporary social media. Indeed, the development of the crowdsourcing approach (model three above) is driven in significant part also by this reluctance, as social media users and non-governmental organisations assume the initiative to explore new uses for social media in crisis communication while key governmental emergency management bodies dither. Even if such third-party and user initiatives have led to the creation of a number of valuable tools and techniques, including platforms such as Ushahidi Maps, the present situation is unsustainable and ultimately even dangerous, as it may dilute emergency response efforts and establish parallel, conflicting information and action structures that could cost lives.

If emergency services accept, as this study argues they must, that social media can be an important source of situational information during a crisis, then they must begin the process of incorporating social media information in their day-to-day emergency management processes. As demonstrated above, this involves tackling, crucially, the question of trust. The three models for approaching the verification of social media information outlined in this paper each provide a different answer to addressing this question, and these answers are not necessarily mutually exclusive; as seen, broad aggregation of user activity patterns under the intelligence gathering model may be combined with a more in-depth evaluation of individual social media posts (and of the users posting them), which may be either conducted by staff members employing quasi-journalistic techniques or performed through broad-based crowdsourcing processes. Which model, or which combination of models, is to be employed may differ from crisis to crisis (crowdsourcing volunteers may only make themselves

available for major disasters, for example), but it depends too on the capacity of individual emergency management organisations and their staff to deploy specific verification techniques.

Such organisational positioning is necessarily also directly related to technical capabilities and organisational structures, then. The intelligence gathering model, for instance, inherently relies on the availability of advanced social media monitoring tools (and of the personnel trained to use them), whereas the quasi-journalistic model may require less technological investment, but is more work-intensive and hence relies on the availability of greater staff numbers. A combination of both (to identify broad user activity patterns and select key messages and users for additional, manual follow-up), while significantly more powerful, further amplifies these needs, and requires an organisational structure that avoids any barriers between passive information monitoring and active user engagement.

Such supportive conditions are far from guaranteed, as many emergency services draw clear organisational boundaries between those of their internal units dealing respectively with incoming intelligence and information and with outgoing advice and media releases; indeed, because of the relative novelty of social media monitoring methods and tools, such efforts are sometimes still outsourced entirely to third-party services and contracted social media analysts. This constitutes a highly problematic organisational model, as the institutional barriers between such external contractors and internal staff must necessarily limit effective information flows. However, more positive and constructive approaches do exist: for instance, the Country Fire Authority in Victoria, Australia, has enhanced its processes of verifying social media communication by co-locating the personnel who monitor, verify, use, authorise, and release information via social media channels (Anderson, 2012, pp. 7–8). Similarly, in several US emergency management organisations, public information officers work closely with senior management to facilitate the swift release of approved information via social media channels (Hughes, 2012, p. 159).

Finally, crowdsourcing approaches require considerable technical skills and investment in technology—even if a number of industry-standard platforms such as Ushahidi Maps are now readily available—as well as significant staff time for the facilitation of constructive crowdsourcing practices. Limited acceptance by government of crowdsourcing as a legitimate form of community involvement, and, in some countries and contexts, underlying

distrust of the motives of governmental authorities by citizen participants, mean that third parties at arms' length from governmental agencies themselves continue to be considerably more visible as facilitators of crowdsourcing efforts during a crisis than official emergency management organisations. Such shortcomings might be able to be addressed in the near term simply by emergency services tapping into the outcomes of crowdsourcing efforts even if they did not facilitate such processes themselves; but this requires, too, the development of inter-organisational trust relationships between governmental and non-governmental services, of course.

Conclusion

This brief exploration of three different models for the verification of social media-sourced information by emergency management organisations has revealed a complex network of trust relationships. These determine the extent to which and the processes via which such information may come to be incorporated in crisis response efforts:

- impersonal and interpersonal trust by emergency management organisations in social media platforms as such, in social media communities, and in individual social media users;
- trust by users and communities in emergency management staff, organisations, and the governmental structures of which they are a part;
- trust among governmental and non-governmental organisations to collaborate effectively in the disaster relief effort; and
- even personal and impersonal trust between individual staff and operational units within emergency management organisations themselves.

Each of these trust relationships has a tangible impact on the possibility and the efficiency of information flowing from social media sources into emergency management decision-making processes.

The verification models outlined, although clearly distinct from one another, each have the potential to affect a subset of these trust relationships, at least to some extent, and thus to improve the quality of the information on which crisis response decisions are based. It is important to stress, however, that this is not necessarily an argument for substantially greater

or wholesale inclusion of social media content in such decisions: the purpose of verification is to work out what (or whom) to trust as much as what *not* to trust, and the verification processes sketched out here may well lead to the dismissal of large volumes of user-provided information at times. This exclusionary process, though, is just as important as the highlighting of specific social media-derived updates for inclusion. Indeed, recent history shows that identification of rumours currently circulating on social media, resulting from the verification process, also provides an important opportunity for emergency services to address directly and counteract such misinformation. During the Queensland floods of January 2011, the state's police service successfully deployed the #mythbuster hashtag on Twitter to respond to current rumours and to avoid any negative repercussions (Bird, Ling, and Haynes, 2012, p. 31; Bruns et al., 2012, p. 8). In other words, it used #mythbuster to encourage actively community *mis*trust in specific information and in the users circulating it. A similar principle of demonstrating trust in and between community members was adopted by the US-based Centers for Disease Control and Prevention during the Influenza A (H1N1) pandemic of 2009–10, with community members self-correcting information online (Veil, Buehner, and Palenchar, 2011).

Such active operationalisation of the existing social media community's trust in an emergency organisation to discredit the information circulated by untrustworthy sources remains an exceptional case, but it does demonstrate the powerful potential of strong trust relationships between citizens and governmental agencies. Yet, even in cases where emergency management organisations have still to develop such long-term, mutually beneficial trust relationships, there is a range of promising models and mechanisms for the evaluation and verification of user-generated content. These may be employed in real time to establish at least a level of provisional trust in the situational updates shared by social media users.

Correspondence

Dr Amisha Mehta, QUT Business School, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia. Telephone: +61(0)7 3138 1798; e-mail: a.mehta@qut.edu.au

References

Anderson, M. (2012) *Integrating Social Media into Traditional Emergency Management Command and Control Structures: The Square Peg into the Round Hole?* Paper presented at the 'Emergency Media and Public Affairs Conference', Melbourne, Australia, 6–8 May 2012.

Beccerra, M. and A.K. Gupta (1999) 'Trust within the organization: integrating the trust literature with agency theory and transaction cost economics'. *Public Administration Quarterly.* 23(2). pp. 177–203.

Bird, D., M. Ling, and K. Haynes (2012) 'Flooding Facebook – The use of social media during the Queensland and Victorian floods'. *Australian Journal of Emergency Management.* 27(1). pp. 27–33.

Bruns, A., J. Burgess, K. Crawford, and F. Shaw (2012) *#qldfloods and @QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods*. ARC Centre of Excellence for Creative Industries and Innovation, Brisbane.

Creed, W.E.D. and R.E. Miles (1996) 'Trust in organizations: a conceptual framework linking organizational forms, managerial philosophies, and the opportunity costs of controls' In R.M. Kramer and T.R. Tyler (eds.) *Trust in Organizations.* Sage Publications, Thousand Oaks, CA. pp. 16–38.

Crowe, A. (2012) *Disasters 2.0: The Application of Social Media Systems for Modern Emergency Management.* CRC Press, Boca Raton, FL.

Dussaillant, F. and E. Guzmán (2014) 'Trust via disasters: the case of Chile's 2010 earthquake'. *Disasters.* 38(4). pp. 808–832.

Gambetta, D. (1988) 'Can we trust trust?'. In D. Gambetta (ed.) *Trust: Making and Breaking Cooperative Relations.* Basil Blackwell, Oxford. pp. 158–175.

Gao, H., G. Barbier, and R. Goolsby (2011) 'Harnessing the crowdsourcing power of social media for disaster relief'. *IEEE Intelligent Systems.* 26(3). pp. 10–14.

Haddow, G.D., J.A. Bullock, and D.P. Coppola (2014) *Introduction to Emergency Management.* Fifth edition. Butterworth-Heinemann, Waltham, MA.

Haddow, G.D. and K.S. Haddow (2014) *Disaster Communications in a Changing Media World.* Second edition. Butterworth-Heinemann, Waltham, MA.

Haynes, K., J. Barclay, and N.F. Pidgeon (2008) 'The issue of trust and its influence on risk communication during a volcanic crisis'. *Bulletin of Volcanology.* 70(5). pp. 605–621.

Houston, J.B. et al. (2015) 'Social media and disasters: a functional framework for social media use in disaster planning, response, and research'. *Disasters.* 39(1). pp. 1–22.

Hughes, A. (2012) *Supporting the Social Media Needs of Emergency Public Information Officers with Human-Centered Design and Development.* PhD dissertation. University of Colorado at Boulder, CO. http://eric.ed.gov/?id=ED551650 (last accessed on 28 July 2016).

Jung, J.Y. and M. Moro (2014) 'Multi-level functionality of social media in the aftermath of the Great East Japan Earthquake'. *Disasters.* 38(S2). pp. S123–S143.

Lane, C. (1998) 'Introduction: theories and issues in the study of trust'. In C. Lane and R. Bachmann (eds.) *Trust Within and Between Organizations.* Oxford University Press, New York, NY. pp. 1–30.

Lindsay, B.R. (2011) *Social Media and Disasters: Current Uses, Future Options, and Policy Considerations.* Congressional Research Service, Washington, DC.

Luhmann, N. (1988) 'Familiarity, confidence, trust: problems and alternatives'. In D. Gambetta (ed.) *Trust: Making and Breaking Cooperative Relations.* Basil Blackwell, Oxford. pp. 94–107.

Mayer, R.C., J.H. Davis, and F.D. Schoorman (1995) 'An integrative model of organizational trust'. *Academy of Management Review.* 20(3). pp. 709–734.

Mehta, A. (2007) 'Trust in a time of crisis: the mass media as a guardian of trust'. In V. Bakir and D.M. Barlow (eds.) *Communication in the Age of Suspicion.* Palgrave Macmillan, Basingstoke. pp. 155–166.

Mendoza, M., B. Poblete, and C. Castillo (2010) *Twitter under Crisis: Can we Trust what we RT?.* Paper presented at the 'First Workshop on Social Media Analytics (SOMA '10)', Washington, DC, 25 July 2010. http://snap.stanford.edu/soma2010/papers/soma2010_11.pdf (last accessed on 28 July 2016).

Miller, G.J. and A.B. Whitford (2002) 'Trust and incentives in principal-agent negotiations'. *Journal of Theoretical Politics.* 14(2). pp. 231–267.

Paton, D. (2007) 'Preparing for natural hazards: the role of community trust'. *Disaster Prevention and Management: An International Journal.* 16(3). pp. 370–379.

Poortinga, W. and N.F. Pidgeon (2003) 'Exploring the dimensionality of trust in risk regulation'. *Risk Analysis.* 23(5). pp. 961–972.

Poortinga, W. and N.F. Pidgeon (2004) 'Trust, the asymmetry principle, and the role of prior beliefs'. *Risk Analysis.* 24(6). pp. 1475–1486.

San Su, Y., C. Wardell, III, and Z. Thorkildsen (2013) *Social Media in the Emergency Management Field*. CAN Analysis and Solutions, Alexandria, VA.

Schoorman, F.D., R.C. Mayer, and J.H. Davis (2007) 'An integrative model of organisational trust: past, present, and future'. *Academy of Management Review.* 32(2). pp. 344–354.

Shapiro, S.P. (1987) 'The social control of impersonal trust'. *American Journal of Sociology.* 93(3). pp. 623–658.

Stieglitz, S., L. Dang-Xuan, A. Bruns, and C. Neuberger (2014) 'Social media analytics: an interdisciplinary approach and its implications for information systems'. *Business and Information Systems Engineering.* 6(2). pp. 89–96.

Sutton, J., L. Palen, and I. Shklovski (2008) 'Backchannels on the front lines: emergent uses of social media in the 2007 southern California wildfires'. In F. Fiedrich and B. Van de Walle (eds.) *Proceedings of the 5th International ISCRAM Conference*. Washington, DC, 4–7 May 2008. pp. 1–9. http://iscramlive.org/portal/node/2236 (last accessed on 28 July 2016).

Tomlinson, E.C. and R.C. Mayer (2009) 'The role of causal attribution dimensions in trust repair'. *Academy of Management Review.* 34(1). pp. 85–104.

UN-SPIDER (United Nations Platform for Space-Based Information for Disaster Management and Emergency Response). n.d. *Hurricane Sandy and the Crowdsourcing Community.* http://www.un-spider.org/sites/default/files/Leson_0.pdf (last accessed on 28 July 2016).

Veil, S.R., T. Buehner, and M.J. Palenchar (2011) 'A work-in-process literature review: incorporating social media in risk and crisis communication'. *Journal of Contingencies and Crisis Management.* 19(2). pp. 110–122.

Yates, D. and S. Paquette (2011) 'Emergency knowledge management and social media technologies: a case study of the 2010 Haitian earthquake'. *International Journal of Information Management.* 31(1). pp. 6–13.

Yin, J., A. Lampert, M. Cameron, B. Robinson, and R. Power (2012) 'Using social media to enhance emergency situation awareness'. *IEEE Intelligent Systems.* 27(6). pp. 52–59.

Zech, C. (2001) 'An agency analysis of church–pastor relations'. *Managerial and Decision Economics.* 22(6). pp. 327–332.