International Journal of Disaster Risk Reduction 42 (2020) 101367

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: http://www.elsevier.com/locate/ijdrr

Exploring the usefulness and feasibility of software requirements for social media use in emergency management

S. Roxanne Hiltz [a], Amanda Lee Hughes [b,*], Muhammad Imran [c], Linda Plotnick [a], Robert Power [d], Murray Turoff [a]

[a] New Jersey Institute of Technology, USA
[b] Brigham Young University, USA
[c] Qatar Computing Research Institute, Hamad Bin Khalifa University, Qatar
[d] Commonwealth Scientific and Industrial Research Organisation, Australia

ARTICLE INFO

Keywords:
Social media
Emergency management
Crisis informatics
Software requirements
Delphi method

ABSTRACT

Social Media (SM) contain a wealth of information that could improve the situational awareness of Emergency Managers during a crisis, but many barriers stand in the way. These include information overload, making it impossible to deal with the flood of raw posts, and lack of trust in unverified crowdsourced data. The purpose of this project is to build a communications bridge between emergency responders and technologists who can provide the advances needed to realize social media's full potential. We employed a two round Delphi study survey design, which is a technique for exploring and developing consensus among a group of experts around a particular topic. Participants included emergency managers, researchers, and technologists with experience in software to support the use of SM in crisis response, from many countries. The study topics are described, and results are presented for both Round 1 (N = 36) and Round 2 (N = 29) of the study, including a ranked list of the top 16 useful features. The top four features include: viewing SM data as classified by geographic location with map-based display; viewing SM data as generated by categories of users; dynamically extracting emerging information; and automatically processing SM images to identify relevant ones.

1. Introduction

Social Media (SM) contain a wealth of information that could improve the situational awareness of Emergency Managers (EMs) during a crisis [1,2] but agencies are often reluctant to use SM, especially to gather unverified crowdsourced data [3,4]. There is also a disconnect between the technologists who design and build automatic data processing systems to analyze SM data and the EMs, the users, who could benefit from the systems. EMs do not know what is technically possible while technologists do not know what EMs want. Moreover, most existing systems rely only on a single SM data source (e.g., Twitter or Facebook) [5]. However, combining informative signals from multiple SM data sources could be useful in several ways, e.g., determining the trustworthiness of SM data [6,7], obtaining missing information [8], etc. Furthermore, these diverse SM data sources produce different content types. For instance, Flicker is best for images, YouTube for videos, whereas Twitter and Facebook are good for both text and images. Diversity in content types not only brings complementary information [7], it is also useful to gain contextual understanding [9]. Despite the fact that complementary information in the form of either images or videos is readily available on many SM platforms, many past efforts to build automated systems for disaster response and management only focused on the textual content available on SM [5].

The purpose of this project is to build a communications bridge between emergency responders (for both governments and NGOs) working in Emergency Management and technologists who can provide the advances needed to realize SM's full potential. Moreover, we aim to solicit software requirements beyond the use of a single SM data source and single content type. To build this bridge, we employ a Delphi study, which is a technique for exploring and developing consensus among a group of experts around a particular topic (in this case, SM use in emergency management). Our study includes emergency responders and technologists (researchers and academics) from many domains and nations. We take an international approach because we hope to inform

* Corresponding author.
E-mail addresses: roxanne.hiltz@gmail.com (S.R. Hiltz), Amanda_Hughes@byu.edu (A.L. Hughes), mimran@hbku.edu.qa (M. Imran), Linda.Plotnick@njit.edu (L. Plotnick), robert.power@data61.csiro.au (R. Power), murrayturoff@gmail.com (M. Turoff).

https://doi.org/10.1016/j.ijdrr.2019.101367
Received 15 August 2019; Received in revised form 11 October 2019; Accepted 12 October 2019
Available online 23 October 2019

software that works globally, and thus we need to discover important differences in requirements among different nations. The final product of the study is a set of software requirements to better enable EMs to gather organized information from SM that is rated for trustworthiness and useable for decision making, as summarized in the Conclusion section. Our overarching goal is to inform technologists of EM needs and to help EMs understand the feasibility of, and possible solutions to, meeting those needs.

In the sections that follow, we review the prior literature on software to support the use of social media by emergency managers. A brief description of the Delphi method is followed by the methods used in this Delphi study. The Results section discusses systems to be included in a comprehensive platform for gathering, organizing, and preparing social media posts for use by Emergency Managers, plus a ranked list of the desirable features for such new systems.

2. Literature review

Social Media have become ubiquitous and are often used by emergency management organizations to disseminate information [4]. However, prior studies [4,10] suggest that there is great hesitation by emergency management (EM) organizations to use SM to collect data from the public during an emergency. This hesitation can create missed opportunities to improve situational awareness [11] or find actionable information [12]. Additionally, the public often expects EMs to respond to their SM queries [13] which cannot be done if EM agencies do not accept and use public SM posts.

2.1. Software enhancements for social media use in emergency management

There are many barriers to the use of social media for emergency management, both technical and organizational. Social media are examples of "socio-technical" systems [14,15]; their use and effectiveness are determined not only by the features and quality of the systems themselves, but also by social context factors such as user attributes (e.g., skills and training, or the lack of them), and by organizational norms and policies and resources [16]. This is especially true of the use of SM as part of an emergency response management system, when the information that could be retrieved and used is generated by the public and communicated via a public commercial system, rather than by trusted information systems under organizational control [4]. Some of these barriers can only be overcome through organizational changes, but others could be dealt with through software pre-processing systems—especially those having to do with issues of information overload (thousands to millions of tweets and posts on a specific disaster; too much information arriving too fast and in a disorganized manner) and trustworthiness of SM posts [3,17,18].

Many studies describe possible software enhancements and systems that could improve the usability and usefulness of social media for disaster management (e.g., Refs. [18–20]). We also note that many technologies for managing social media in emergency contexts currently exist (e.g., Ushahidi [21], AIDR [22], Social Haystack [23], TweetDeck [24], and Hootesuite [25]). See Refs. [5,26,27] for more extensive surveys of these systems. A great many of these technologies for processing social media data incorporate both human efforts (e.g. crowdsourcing to create training sets for machine learning) and automated tasks [28]. Many of these systems are also limited in that they only support a single social media platform, they have limited analytical capabilities, or they have only been deployed in a research setting [5]. There are no universally adopted systems in this space; use depends on context, system features, user expertise, funding for purchasing software, and a willingness to adopt new (and often experimental) technologies. Some of the challenges facing software developers of these systems include real-time parsing of brief and informal messages, determining information credibility, and prioritizing useful information [29,30]. A primary goal of this paper is to build upon and learn from existing tools, systems, and applications so we can design and develop the next generation of technology.

In a survey of 477 U S. county-level EMs [31], several of the specific potential software enhancements described in Ref. [27] that could help overcome the limitations of SM for EMs were rated with the result being that they are perceived as highly useful. These enhancements included the ability to view SM data as generated by geographical location on a map-based display and categorization of SM data by sub-events (e.g., rescue issues). However, that study had several limitations: participants were limited to U.S. county level EMs and many potential enhancements that have subsequently appeared in the literature were not included. The current study aims to overcome these limitations by including software developers and researchers related to SM in EM from many countries and types of organizations, and by adding features from systems described in recent studies and proposed and vetted by the study participants.

2.2. The Delphi method

A Delphi study consists of two or more rounds of structured written exchanges between anonymous experts with different types of expertise relevant for a topic [32,33]. It was developed in the 1950s to obtain expert input on a particular problem while allowing the participants to remain anonymous. The expert panels receive questionnaires (mailed or online) and answer them at a time and place convenient to them. The technique is particularly useful in cases where the expert panel is dispersed over a wide geographic area, and for complex, multi-disciplinary problems [34].

The Delphi method uses a series of questions over two or more rounds where each participant may suggest additional items for consideration in subsequent rounds. An important aspect of a Delphi is that experts are generally asked not only to rate various alternatives or issues or items, but also to explain their rating in comments. The comments often surface the underlying reason for disagreement on ratings, such as one person assuming that a related development is likely, while another expert assumes that it is not. After each set of questionnaires is completed, a facilitator summarizes the experts' inputs and then distributes the summary with another round of questions. The results are annotated to show which category(ies) of expert provided each feedback item. Thus, although anonymity is maintained, expert participants know the domain of the person providing feedback. This is helpful when the experts evaluate the feedback. The experts can then revise their answers from the first round in response to the ratings and comments of others and add ratings of newly surfaced items or issues.

3. Research method

Our objective is to foster a dialogue and exchange of knowledge among software developers, emergency managers, and researchers on SM, from different nations and types of organizations. Thus, we aimed to obtain a representative "spread" of diverse opinions from different types of stakeholders, rather than a random or representative sample of any particular population.

3.1. Expert recruitment

In recruiting possible participants, we cast a wide net, using our professional networks. This included sending invitation messages to researchers who had published papers in the social media tracks of recent meetings sponsored by the Information Systems for Crisis Response and Management (ISCRAM) community; sending emails to lists of U.S. emergency managers; posting on LinkedIn discussion boards for EMs; personal messages to software developers; and requests for those initially invited to suggest other experts.

Following approval of the study by an Institutional Review Board

(which included the consent form, screening questionnaire, and Round 1 questions of the Delphi), the initial invitation to participants was to complete a screening questionnaire. We used email to send the prospective participants a link to a Recruitment survey to build our expert panel prior to the actual Delphi. The Recruitment survey asked for information such as years of experience in EM and in software development for SM in EM, to make sure the volunteers qualified as "experts." Since it was stated that the Delphi rounds of the study might take between 30 min and 2 h per round, depending on how many comments a participant made, the response rate may have been affected. There were 48 valid respondents to the Recruitment survey who indicated they were willing to participate out of over 400 prospective participants who received invitation messages or may have seen one on a discussion board.

We reviewed the responses to the Recruitment survey to confirm that the respondents were qualified by their expertise and experience to participate as members of the Delphi panel of experts. Those few respondents (N = 9) to the Recruitment survey who were deemed unqualified either failed to complete the survey beyond the consent or did not have experience as a researcher, practitioner, and/or academic in the focus area of use of social media for emergency management. We maintained a list of the qualified respondents and then sent email to only qualified respondents with the link to the Round 1 Delphi questionnaire. The Recruitment survey took, on average, 4 min for the respondents to complete.

3.2. Delphi questionnaires

Once the panel was formed, our Delphi process consisted of two rounds of surveys with feedback to the experts from each.

The Round 1 questionnaire was a mix of qualitative and quantitative questions focused on four broad topic areas of participant demographics and experience, social media platforms, tool features, and recommendations. These topic areas were determined from the literature survey (particularly [4,27]) and through iterative discussion within the research team. Once these broad topic areas were determined, they were refined and expanded into the specific survey questions, again through iterative discussion. In Round 1, we asked participants to rate the usefulness and feasibility of a list of possible tools and features (provided by the research team) as well as to suggest others and to comment on software requirements. We were mindful of the trade-off between obtaining comprehensive answers from respondents and the time taken for them to complete the survey. Thus, the questions were structured as a mix of closed questions (using rating scales) and optional open-ended questions to provide further detailed information about the reasoning behind each rating. The Round 1 questionnaire is included as Appendix 1. We deployed the questionnaires using the Survey Monkey® system. Thirty-six respondents completed the survey. Round 1 took place in late November 2018 through January of 2019 and took respondents an average of 36 min to complete.

For the second round, in early 2019, the results were fed back, including new suggestions and repeated questions from Round 1, to try to reach greater agreement. Answering the repeated Round 1 questions was optional so that if an expert had a change of mind, it could be reflected in the Round 2 results. Twenty-nine of the original respondents also completed Round 2. The new suggestions that were rated in Round 2, based on suggestions from Round 1, are shown in Table 3 below. Data were analyzed using SPSS®. Our analysis also includes a selection of representative quotes that explain the reasons for ratings.

4. Round 1 results

4.1. Characteristics of the expert panel respondents

Thirty-six of the invited and referred experts responded to Round 1 of the Delphi study. A little less than half (16) are from the U.S., with 5 from Spain, 3 from New Zealand, two each from Germany and Nepal, and one each from several other countries in Europe, Asia, the Middle East, and Canada. The participants are diverse and well balanced among different types of background experiences, as shown below in Tables 1 and 2.

Note that participants could choose more than one work experience domain. To further explore these domains, we created mutually exclusive and comprehensive categories by categorizing participants as shown in Table 2:

To look for relationships between work experience domain and opinions, the respondents were re-classified into three groups: Creators, Users, and Both. *Creators* include researchers and software developers (39%). *Users* are the practitioners (19%). *Both* (42%) are those who have domain experience in both creating and using. Although the proportion of "pure users" is lower than we would have liked, the majority of the experts did have practitioner experience.

In terms of relevant years of work experience related to social media use for EM, only 3 had 2 years or less; 10, 3–5 years; 15, 6–10 years; and 8, over 10 years. Given the recency of the emergence of this field of expertise, this is a very experienced set of experts. Most participants are between 20 and 49 years old, which is consistent with the years reported working in the field. They are 69% male, which reflects the gender makeup of the field.

4.2. Results

4.2.1. Systems to be included

Many of the existing prototype systems for collecting, categorizing, and analyzing Social Media posts for disaster work with only one Social Media platform, most often Twitter or Facebook [5]. The feasibility of creating systems for a single platform thus has been repeatedly demonstrated. However, it is not reasonable to expect EMs to use different software aids for different social media platforms and somehow integrate this information; 61% of the experts say it is not useful to have a system that uses only one SM platform. However, if the software analysis system could handle both Facebook and Twitter, then a substantial majority say it would be quite useful, as shown in Fig. 1 below:

Next, the experts were asked what the "major" platforms would be for their country, if they were to have a system that handles and integrates information from "all major" social media platforms. The results are shown in Fig. 2.

After Twitter and Facebook—mentioned by almost all—Instagram was checked by 21 of the 36 participants, WhatsApp by 18, YouTube by 14, Flickr by 10, and several other platforms suggested by one to three. Among the platforms mentioned as "other", Reddit, Pinterest, Discord, and Snapchat were on top. Note that each respondent had a specific but different list of systems that should be included for their country. The results of the question on the usefulness of a system that could integrate "all major" social media systems for a country are shown in Fig. 3. Over 70% of the experts said this would be "very useful."

Though this design requirement—to include all major SM platforms—is considered very important, the feasibility of doing so is questioned by many of the experts, as shown in Fig. 4. Thus, the panelists identified an important design challenge for the research community.

Many participant comments highlight the types of challenges posed in building such a system, of which the following two quotes reflect frequent themes:

Table 1
Work experience is in the domain of (check all that apply).

Software Design	NGO EM	Local gov't EM	State gov't EM	National gov't EM	Research
18	7	13	8	9	22

Table 2
Work experience domains.

	Frequency	Percent
Software Design only	2	5.6
Research only	7	19.4
Practitioner only	8	22.2
Software design and practitioner	4	11.1
Software design and research	6	16.7
Practitioner and research	3	8.3
Software design, practitioner and research	6	16.7
Total	36	100

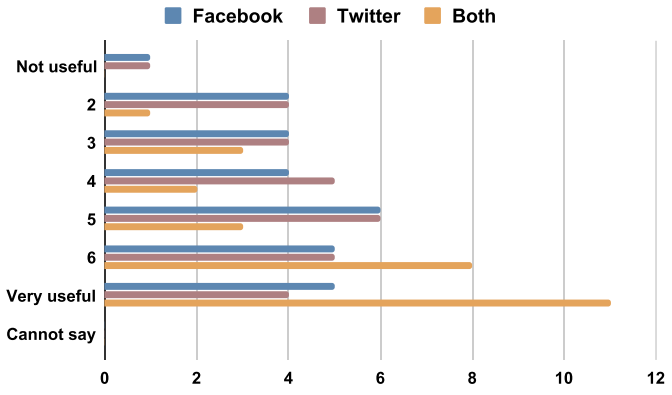


Fig. 1. How useful to an EM would a system be that handles: only Facebook, only Twitter, both Facebook and Twitter posts, but not other systems?

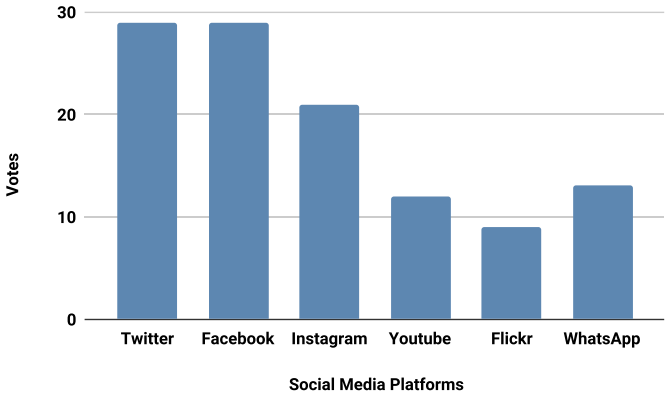


Fig. 2. Suppose the aim is to have a system that handles and integrates information from "all major" SM platforms. What SM platforms should be included in the category of "major" for your country/location?

> *"There are several aspects to developing the SM platform that integrates multiple social platforms. One of the primary issues that I struggle with is the restrictions applied on the data – while Twitter is one of the best options, Facebook has a lot of content that is more promising – however, accessing FB is a massive struggle given the platform restrictions. Most often, creating public FB pages that many people 'like' and contribute to is one of the very few means of getting data but is highly restrictive, time consuming and potentially biased."*

> *"I think it would be incredibly useful to have a system that integrated multiple social media platforms, especially if it could be tailored to a local area. However, there are a lot of challenges. First, one has to identify the most commonly used platforms (which is in flux). Second, there are many interoperability issues. Each social media platform has a different API, different types of data (and associations between those data), different privacy settings, different terms of services, and different ways of representing user data and messages."*

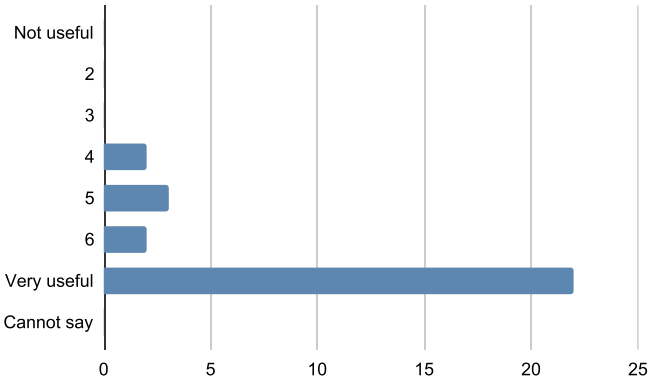


Fig. 3. How useful would it be to have a real-time system that handles and integrates information from "all major" SM platforms customized for a particular country?

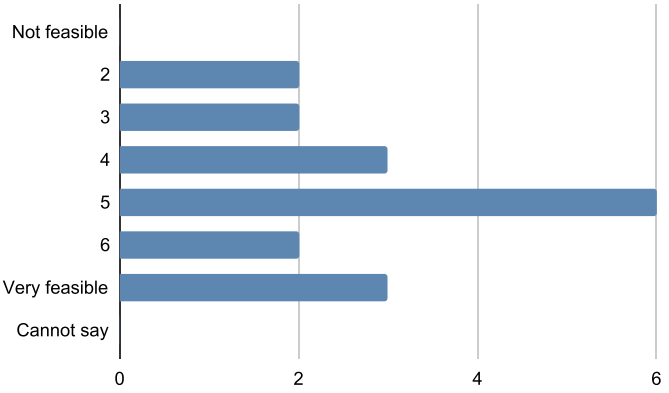


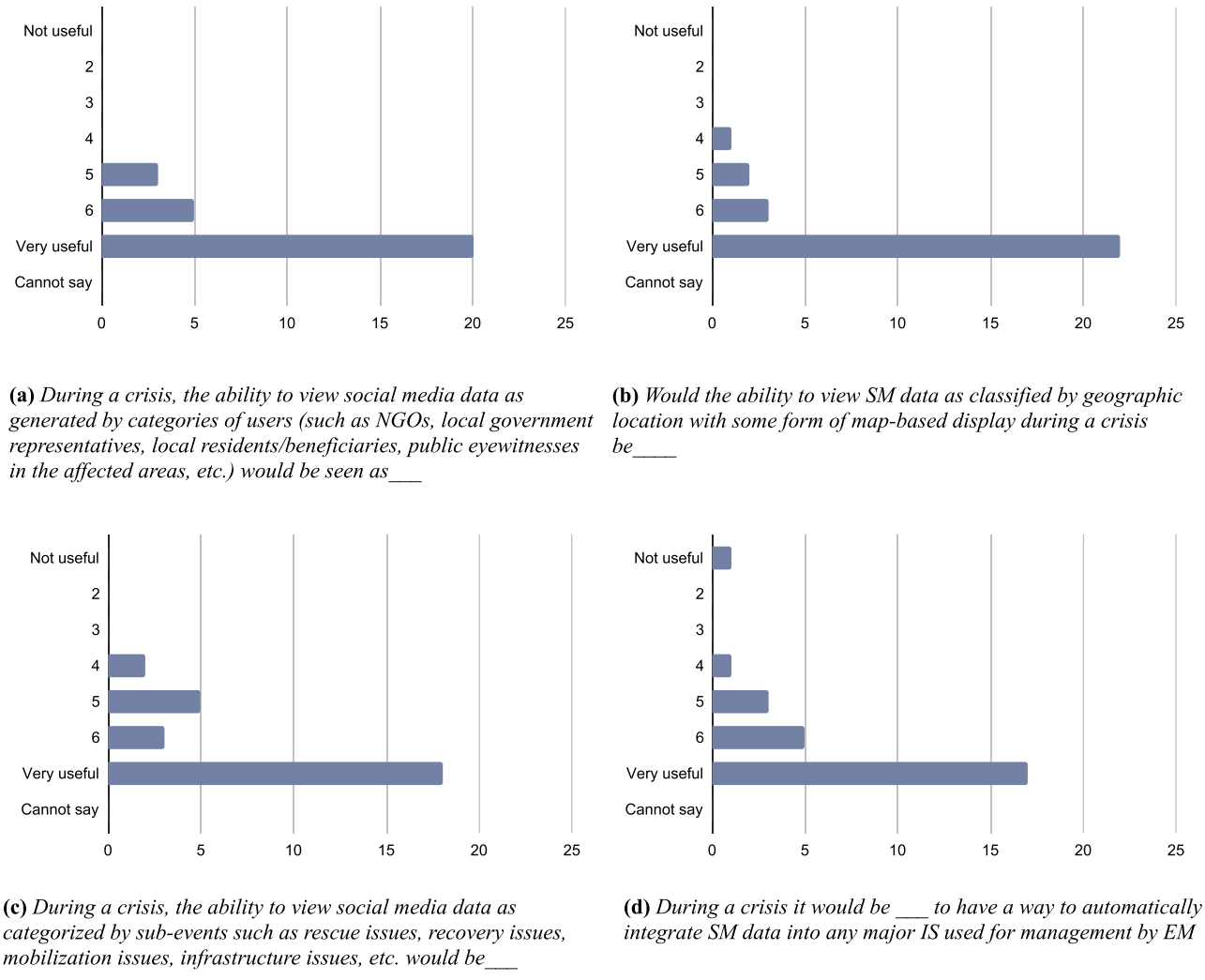
Fig. 4. How feasible is it to implement a real-time system that handles and integrates information from "all major" SM platforms customized for a particular country?

4.2.2. *Useful features*

The usefulness and feasibility of specific features or capabilities that could or should be included in an ideal SM analysis system for the future were rated on semantic differential scales; the scales range from 1 (not useful/not feasible) to 7 (very useful/very feasible). Fig. 5 shows a few important features and their ranking by our participants. First, a system that identifies and categorizes different types of users on social media—such as eyewitnesses or NGOs—is considered very useful, as shown in Fig. 5 (a). A non-trivial issue of social media data is the lack of geographical information. Fig. 5 (b) highlights the importance of SM data classified by geographical location on a map. The identification and categorization of sub-events—i.e., issues caused by a major disaster from SM data—is perceived as very useful in Fig. 5 (c). Finally, the integration of a SM-based system into an existing information system of an organization is also seen as very useful, see Fig. 5 (d).

In addition to textual messages, social media platforms are a rich source of other content types such as images and videos. Several research studies have demonstrated the usefulness of images posted on SM during disasters and emergency events [35–37]. We asked several questions to our participants regarding different uses of SM images for disaster response. Specifically, we asked about the relevancy of images for detecting damages, assessing the severity of damage, and identifying injured, trapped, and displaced people. Fig. 6(a–d) shows the ranking obtained for these questions. Overall, there is agreement among our participants that images on SM are very useful.

In the rank ordered results shown below in Fig. 7, the means are reported with the "cannot say" responses removed. As can be seen in Fig. 7, all except the last three potential features (combining text and



(a) *During a crisis, the ability to view social media data as generated by categories of users (such as NGOs, local government representatives, local residents/beneficiaries, public eyewitnesses in the affected areas, etc.) would be seen as___*

(b) *Would the ability to view SM data as classified by geographic location with some form of map-based display during a crisis be____*

(c) *During a crisis, the ability to view social media data as categorized by sub-events such as rescue issues, recovery issues, mobilization issues, infrastructure issues, etc. would be___*

(d) *During a crisis it would be ___ to have a way to automatically integrate SM data into any major IS used for management by EM*

Fig. 5. Useful Features in a SM data processing system.

image content or dealing with only Twitter or only Facebook), the means are well above 5 on the 7-point scale, so most of the potential features are rated as highly useful. In addition, the standard deviations are relatively small, indicating that the experts had a consensus about the usefulness of the features.

A clear pattern emerges in the results: the usefulness of features is rated higher than the feasibility of developing or having the features (not shown). Statistical significance is not calculated but it appears that while there is enthusiasm for the features, there is less confidence that they can be implemented.

Design choices for integrating SM data into existing IS systems used by EMs need further exploration. As one expert explains:

> *"I am on the fence about this. In terms of utility, having "yet another system" is not great for responders. However, one can gain a lot more utility from the information with custom features that may only be available in a separate platform. In terms of feasibility, you can pipe in information to many other systems as long as there is an API with minimal technical restrictions (x tweets per hour). But as mentioned before, you can't always perform a full complement of analyses in another platform. So the question is do you simply bring the raw data into the primary EM platform (or even another analytics platform) to combine with other tools/data or do you bring in the products (analyses - maps, graphs, reports, etc.)? The latter may be problematic if responders want to conduct follow-up analyses. Which will be more useful? Not sure."*

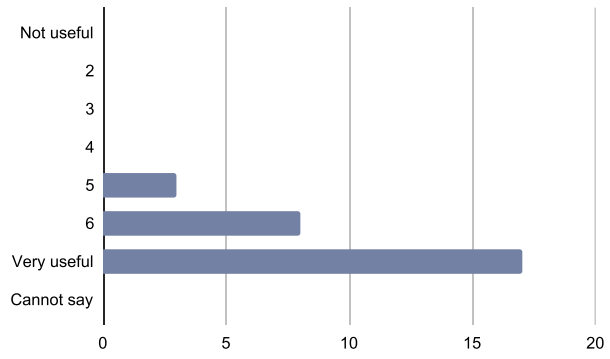
On a question about which information categories would be useful to organize for emergency managers, we received participants' ratings as shown in Fig. 8. The top three categories include *infrastructure damage reports*, *urgent needs of affected people*, and *injured people reports*.

4.2.3. Do ratings of usefulness and feasibility differ by the domain of expert experience?

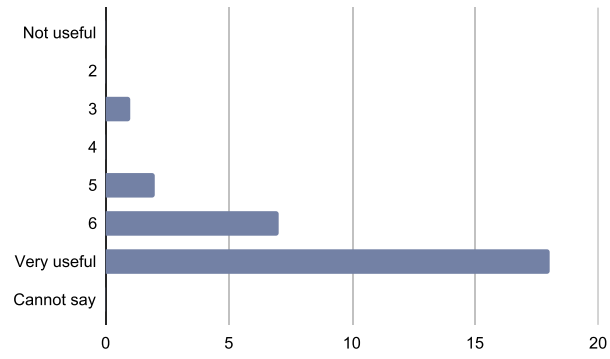
We divided the experts into three groups: Creators, Users, and Both. We examined if the ratings on usefulness and feasibility of features differs amongst these groups. Because the data are not normally distributed, we used non-parametric Kruskal-Wallis tests. Non-parametric tests will show if there are significant differences but will not show where they are. So, if a significant difference was found, we then tested the feature responses with ANOVA with Tukey's post-hoc test. We are able to do that because while ANOVA (a parametric test) is not robust with non-normal data and thus may not find a significant difference that exists, it will not have false positives. That is, it will not find a significant difference where there is none.

The Tukey's post-hoc test showed a significant difference between how Creators and Users rated the usefulness of a system that only deals with Twitter posts: Creators rate the usefulness higher than Users. This type of result is important as, ultimately, Creators need to fulfill the needs of the Users through their designs of SM devices.

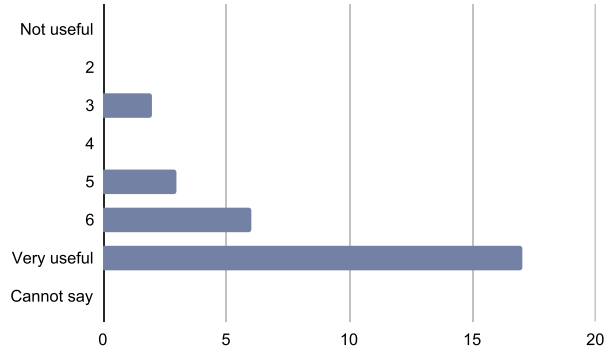
Significant differences appeared between groups for the questions, "*How feasible is it to implement a real-time system that handles and integrates*



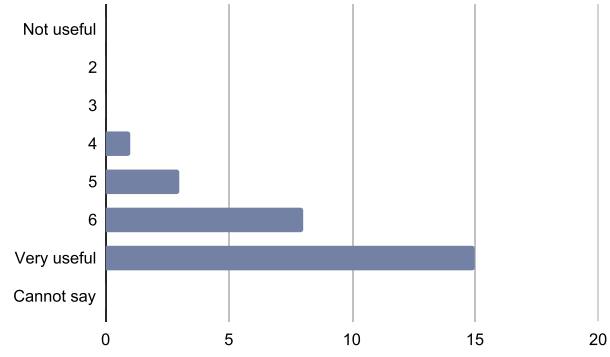
(a) *A system that can automatically process SM images to identify relevant ones to a disaster event would be____*



(b) *A system that can automatically process SM images to identify whether or not an image shows damage would be_____*



(c) *A system that can automatically process SM images to assess the severity of damage shown in an image would be____*



(d) *A system that can automatically process SM images to identify injured, trapped, or displaced people would be____*

Fig. 6. Various features related to social media image processing and usefulness for emergency management.

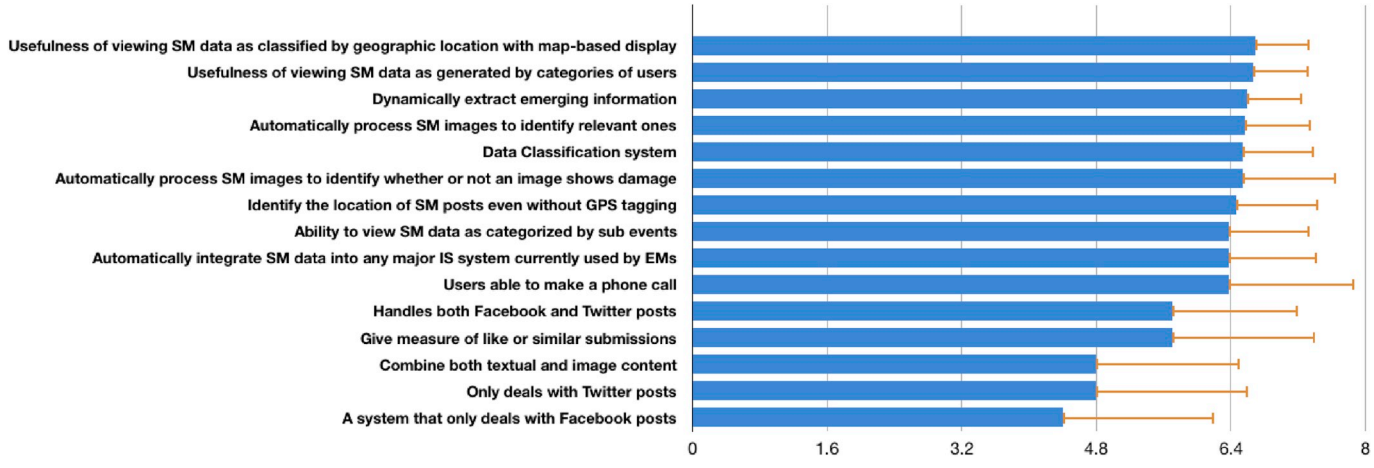


Fig. 7. The usefulness of features (means).

information from "all major" social media platforms customized for a particular country?" and *"A system that could dynamically extract emerging/unanticipated types of information from SM data during a disaster would be __ feasible."* However, ANOVA was unable to determine which groups were significantly different from one another for these two questions.

For the question, *"A system that can combine both textual and image content on SM to enhance situational awareness would be __ feasible,"* Users thought it was more feasible than did the Creators. The same pattern occurred for the item, *"A system that can combine both textual and image content on SM to enhance situational awareness would be __ feasible."* It is possible this is because Creators may have a better understanding, in general, of design and implementation feasibility than do Users.

Most of the features were rated similarly for usefulness and feasibility by the groups of experts based upon domains of experience. This finding suggests that the chasm between practitioners and creators

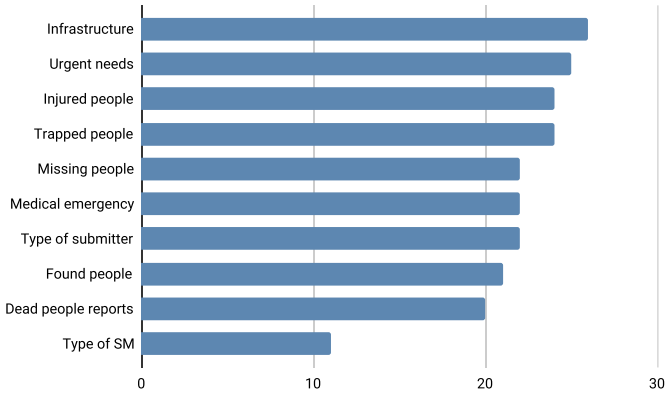


Fig. 8. Please check all types of information categories you think would be useful to organize information for EMs.

(researchers and technologists) is not as wide as we had feared. Thus, it seems possible to have the various experts come together in consensus to develop a ranked list of needed and feasible features.

4.2.4. Differences by U.S. vs. other countries

We also performed Mann-Whitney U tests to ascertain if the ratings of usability and feasibility of features varied by whether the expert was from the U.S. or not. The only significant difference was found for the feasibility of a system that can rate the trustworthiness of posts. The U.S. respondents rated the feasibility significantly lower than did the respondents from other countries (Mann-Whitney $U = 21.5$, $p = .017$). That only one test showed significant differences is promising. It suggests that future feature design will have global usefulness.

5. Round 2 results

For Round 2 of the Delphi study, suggestions from participants in Round 1 about issues or features that we might have missed were culled and edited into a new set of questions. These new items are shown in Table 3 below, with the labels for these variables to be used subsequently shown in bold.

The frequency distributions as well as means and standard deviations for agreement with these items are shown in Table 4, with the items rank ordered by means.

Note that when calculating the mean, only the responses on the semantic differential scale were included. That is, "Cannot say" responses were included for frequencies, but not when the mean was calculated. As can be seen, the highest rating related to the new considerations is that dealing with the need for flexibility in classifications of posts, including being able to use multiple categories. However, the mean is lower than those for most of the features included in Round 1. This is a good sign that our first round likely included the major software considerations that are important to developers and emergency managers.

Comments about flexibility include the following, which help to define the specifications needed:

> *"Working with SM on a regular basis, I think a quality of this kind of data is that categories are only sticky to a point, thus flexibility of inspection and classification will be needed for many information tasks."*

The lowest ranked of the new suggestions is including calls for help from Smart Home devices rather than just humans. The comments that shed light on this controversial suggestion include:

> *"Smart Home devices serve a different purpose compared to social media for crisis response. They should probably be kept separated."*

> *"Calls for help must go to people who actually coordinate emergency response, e.g. 911 dispatch centers or corporate call centers responsible for contacting 911. Emergency managers are often not appropriate end-users for systems supporting emergency response (especially during periods of stability when EOCs are inactive)."*

> *"I think including smart home devices is an interesting idea, but there are a lot of privacy concerns that would need to be worked through."*

Table 3
New Round 2 Questions. Responses ranged from Strong disagree (1) to Strongly agree (7) with a Cannot say (8) option.

To what extent do you agree with the following statements?
a. security/privacy **API restrictions** on social media platforms are a serious issue affecting the feasibility of creating software to automatically identify and analyze posts during disasters.
b. In classifying posts by categories, it would be important to have the flexibility to change categories and apply multiple categories. (**Flexibility**)
c. I would encourage initiatives aimed at building and sharing large scale datasets. **(share data)**
d. "**Rumors**" should be one of the tags or folders for classification of posts.
e. Displays of results should include "**dynamics**".
f. A system should enable an EM to **direct**ly **contact** a user who has posted on social media.
g. Future systems must take **ethical issues** into account.
h. Social media should be combined in real-time with remote **sensing data**.
i. Calls for help from **Smart Home** devices should be included in future systems.

5.1. Comparison of round 1 and round 2 ratings

Round 2 also repeated all of the questions from Round 1. Wilcoxon Matched Pairs tests were performed to test for significant differences between responses to the same question. Only the comparison of the usefulness of Facebook is significant, with the respondents finding Facebook less useful in Round 2. This can probably be attributed to the unfavorable news stories about Facebook that occurred in the interim, including revealed incidents about privacy issues and "fake news" on that platform in particular. Because there are no significant differences between Round 1 and Round 2 except for that question, and because our number of responses was a bit higher for Round 1, in consolidating results we used the means from Round 1 for the items repeated on both rounds.

Combining the results of Rounds 1 and 2, the top 16 features for priority development, as rated by perceived usefulness, are found in Table 5:

6. Discussion: summary, limitations, and conclusion

In this paper, we summarize findings from a two round Delphi Study designed to understand how software can better support the SM needs of emergency managers. Experts in the study identified many useful features for managing SM information during crises, but it seems that at least some of these features may be difficult to develop. For example, a system that could monitor multiple social media platforms was rated as highly useful by experts, but the feasibility of developing such a system was seen as challenging. Thus, the feasibility of developing these capabilities must be considered in combination with their usefulness.

We presented a prioritized list of 16 features that experts would find useful in software to support social media use in emergency management. This list will help software developers focus their attention on features that would be most useful for emergency managers. We note that some of the features on this list are logically linked; for instance, one would need better algorithms to identify geographic location for a post, to make the top-ranked feature of map-based displays very useful, and to integrate SM data into systems currently used by EMs. In explaining the latter link, one expert notes, "*I do not think this is very feasible mainly because a majority of SM data is not geocoded. Implementing a generic location tagger is a difficult task especially when dealing with multiple languages.*" In addition, it would make sense for a team to treat all desired features for automatic analysis of images rather than text, as a package,

Table 4
Responses to new round 2 questions (N = 29).

	Strongly disagree	2	3	4	5	6	Strongly Agree	Cannot say	Mean	Standard Deviation
Flexibility	3.4	0	0	0	6.9	27.6	58.6	3.4	6.36	1.22
Share Data	0	3.4	0	3.4	13.8	10.3	65.5	3.4	6.32	1.22
Ethical issues	3.4	0	0	6.9	3.4	20.7	58.6	6.9	6.26	1.38
Sensing data	3.4	3.4	0	3.4	13.8	24.1	41.4	10.3	5.88	1.56
Dynamics	3.4	0	0	6.0	20.7	27.6	34.5	6.9	5.81	1.36
API restrictions	3.4	0	6.9	3.4	10.3	31.0	34.5	10.3	5.77	1.53
Rumors tag	0	6.9	0	13.8	3.4	24.1	37.9	13.8	5.76	1.56
Contact User	0	3.4	3.4	24.1	10.3	17.2	34.5	6.9	5.48	1.50
Smart Home Calls	0	3.4	10.3	10.3	6.9	24.1	20.7	24.1	5.32	1.59

Table 5
Top 16 features for priority development.

1. Viewing SM data as classified by geographic location with map-based display
2. Viewing SM data as generated by categories of users
3. Dynamically extracting emerging information
4. Automatically processing SM images to identify relevant ones
5. Data Classification system
6. Automatically processing SM images to identify whether or not an image shows damage
7. Identifying the location of SM posts even without GPS tagging
8. Handling and integrating information from all major SM platforms
9. Ability to view SM data as categorized by sub events
10. Automatically integrate SM data into any major IS system currently used by EMs
11. View SM data as categorized by subevents
12. Process images to assess severity
13. Flexibility in categorization
14. Automatically process SM images to identify injured, trapped, or displaced people
15. Identify actionable information
16. Data processed to rate trustworthiness

and the same is true for the various aspects of a data classification system.

6.1. Limitations

One limitation of this study is that we had a limited number of participants from a relatively small number of countries. They represented the creators (software designers and researchers) and the users (Emergency Managers) who have studied software systems and their use by EMs. It would have been ideal to include more users, and we did contact hundreds through email, but we had limited success recruiting busy EMs to take part in the survey. Given that there is substantial "red tape" involved in adopting new systems to manage emergencies [3], it would also be good to incorporate elected officials or other policy makers as stakeholders, perhaps in a subsequent study in which they are given the results and asked to comment on barriers to implementation that they see from their perspective.

Another limitation is that we had to restrict the number of software features and issues that could be included in the Delphi study, so as to not add to the questionnaire length and thereby decrease the completion rate. Additional issues that could be explored in a subsequent study, include the following:

- What security requirements are needed for software applications that monitor social media?
- Who should own such applications and how will they be maintained?
- Who will have access to the information produced by these applications and what processes should be followed when granting or denying access?

Moreover, since the main focus of this study was to understand useful software-related features and their feasibility, we did not study the issues associated with SM data such as sample-selection bias, privacy, veracity. However, we remark that sometimes SM data alone is not suitable to draw conclusions. Important insights and signals from different SM platforms should be learned and analyzed together with other data sources (e.g., field assessment reports if available) to perform triangulation, fill-in missing gaps (data completeness), and improve data veracity. A sensor-fusion approach can help achieve these objectives [38].

6.2. Discussion and future work

This study demonstrates that there is much room for improvement—given a long list of useful features—regarding systems that support social media use in emergency management. Much work has focused on attempting to develop the features identified in this study, such as geographically mapping social media data [39], assessing disaster damage from social media images [36], and categorizing social media data by categories of users [40,41]. Yet it is clear there is much left to do. The purpose of this study was to help researchers, practitioners, and technologists better understand needs, so future efforts can be more focused.

A logical next step would be to form collaborative working groups on sets of features by holding sponsored (expense paid) workshops for software developers to organize a plan of work and a division of labor, which could then be continued by online communication and sharing data. As mentioned in the Limitations section, it would also be good to incorporate policy makers into a subsequent phase of the project, perhaps as a new Delphi round, or as participants in a workshop.

Acknowledgments

Authors are listed in alphabetical order. We are grateful to all of the participants in this study, who gave generously of their time.

Appendix 1

The Round 1 questionnaire consisted of 64 questions organized into eight sections focused on the themes of: consent (2 questions), participant emergency management experience (1), social media platforms (9), desirable features (2), additional features (14), other features (28), recommendations (3) and participant demographics (5). There were 39 mandatory and 25 optional questions with a mix of 28 free text responses, one check box for multiple responses, three check boxes that also included a free text "other" option and 32 radio buttons for a single response. A full copy of the questionnaire is available from the authors upon request.

The questionnaire begins with details of the co-investigators and references the principle institution conducting the research before asking for the participants name as shown below (Fig. 1).

Fig. 1. Participant name question.

Then an overview of the research and participant instructions are provided, as shown below, followed by a question to obtain explicit consent from the participant as shown in Fig. 2 below.

Purpose

The purpose of our project is to build a communications bridge between emergency responders for both governments and NGOs (non governmental organizations) working in Emergency Management and technologists and researchers who can provide the advances needed to realize SM's (social media) full potential. This will create a synergy that will lead to a set of software requirements to better enable EMs (emergency managers) to gather an organized set of information from SM postings that is rated for trustworthiness and useable for decision making.

Duration

My participation in this study will last for approximately 4 months (2 surveys several months apart), each requiring 1–2 h to read background material and then answer thoroughly.

I have been told that my participation in this research is important for the success of the research and that the results of this research study are expected to produce the following benefits to society and for me as a subject.

Benefits for society and the subject

I have been told that the benefits are:

For society: recommendations and software requirements that will lead to more effective use of social media in crisis management, thus better protecting lives and property.

For you: opportunity to exchange information and ideas with peers about future needed software systems related to social media and emergency management.

Procedures

I have been told that, during the course of this study, the following will occur:

1. Potential participants will be asked to answer a short questionnaire indicating their interest and describing their relevant expertise.
2. From these applicants, a balanced panel of experts will be invited to participate in two rounds of online questionnaires (known as a "Delphi" design of a study, which consists of two or more rounds of structured written exchanges between anonymous experts with different types of expertise relevant for a topic.) Our topic is the desirability and feasibility of a number of software enhancements for managing social media input to emergency managers. Our 8 categories of experts include software developers, researchers, government emergency managers, and NGO emergency managers, from the U.S. and internationally.
3. Results of Round 1 will be fed back to participants, with Round 2 asking for re-ratings.
4. Participants will receive a draft report of findings prior to public release and invited to comment.

Participants

I will be one of about 50 participants in this study.

Exclusions

I will inform the researcher if any of the following apply to me:

1. You must be at least 18 years of age.
2. You must be fluent in written English.

Risks/discomforts

I have been told that the study described above may involve the following risks and/or discomforts:

As an online participant in this research, there is always the risk of intrusion by outside agents (i.e. hacking) and, therefore the possibility of being identified exists.

There also may be risks and discomforts that are not yet known.

I fully recognize that there are risks that I may be exposed to by volunteering in this study which are inherent in participating in any study; I understand that I am not covered by ⟨redacted⟩'s insurance policy for any injury or loss I might sustain in the course of participating in the study.

Confidentiality

I understand that confidential is not the same as anonymous. Confidential means that my name will not be disclosed if there exists a documented linkage between my identity and my responses as recorded in the research records. Every effort will be made to maintain the confidentiality of my study records. If the findings from the study are published, I will not be identified by name. My identity will remain confidential unless disclosure is required by law.

Note 1: You will be assigned an ID based on this form; the file that matches the names and ID number will be kept separately and in a secure location; thus the data file that contains your answers will not be identifiable by name.

Note 2: If you agree, the final report on the study will list your name in a list of expert participants; if you do not explicitly agree, your name will not be included as a contributor to the research.

Right to refuse or withdraw

I understand that my participation is voluntary and I may refuse to participate, or may discontinue my participation at any time with no adverse consequences. I also understand that the investigator has the right to withdraw me from the study at any time.

Individual to contact

If I have any questions about my treatment or research procedures, I understand that I should contact the principal investigator at: ⟨redacted⟩.
If I have any additional questions about my rights as a research subject, I may contact: <redacted⟩.

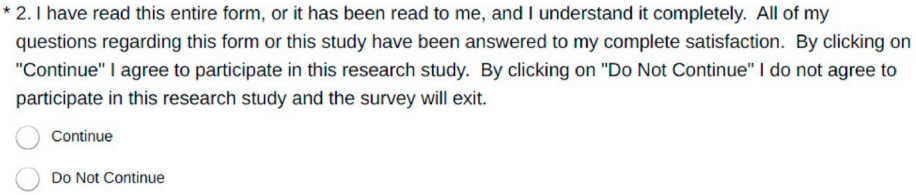


Fig. 2. The Consent question.

The figures below show a selection of questions from the survey instrument. The question in Fig. 3 corresponds to the results shown in Tables 1 and 2 above. Note the asterisk (*) next to the question number (3 in this case) indicates that the question is mandatory. This is an example of a check box question where the participant can select multiple responses.

Fig. 3. Participant work/experience.

Fig. 4 shows the first four questions in the social media section of the survey where each is a radio button allowing only a single response. This section is preceded by an introduction describing the purpose of the survey in terms of identifying software enhancements as follows:

Our overarching goal is to make social media more useable for emergency managers (EMs) to gather and assess information from the public, to aid in "situational awareness" and decision making. In the research literature, there are many systems described that take the form of gathering information from social media (SM) posts during a crisis and processing it by organizing it, displaying it, and/or assessing its validity, etc. The purpose of this study is to gather expert wisdom about the potential usability and feasibility of a number of such possible software enhancements, in order to arrive at a set of requirements to guide future developments. We begin in this round with your ratings of an initial set of possible improvements that are already specified in the current literature, and then ask you to add additional improvements you think will help EMs more effectively use SM. Please supply as many comments as you can in the text boxes, in terms of additional ideas or reasons for your ratings.

If you do not have experience with software engineering, you should check "cannot say" when estimating feasibility of a possible enhancement.

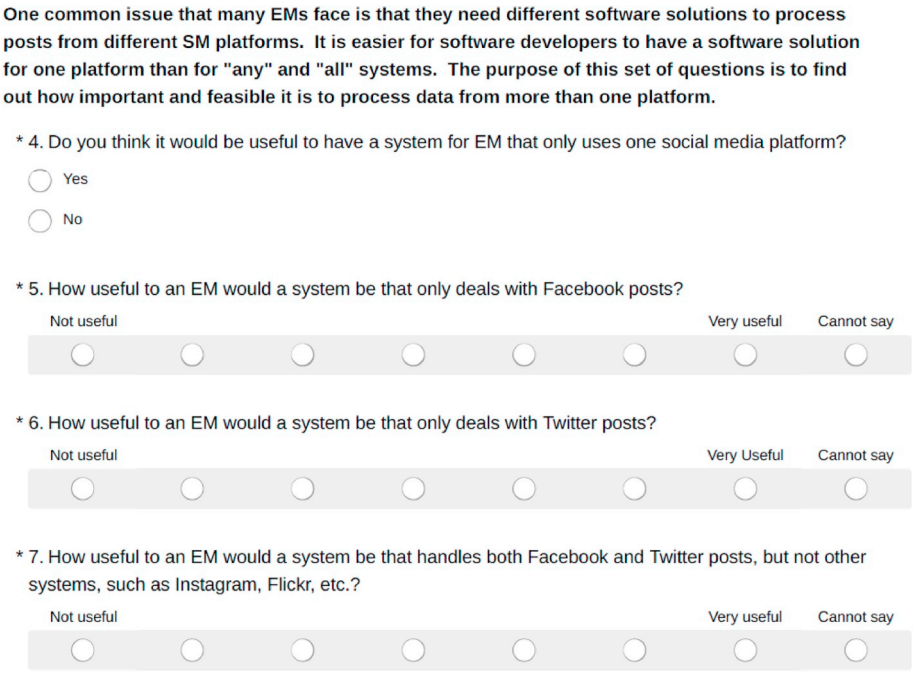


Fig. 4. Example Social Media Platform questions.

Fig. 5 shows the two "Desirable Features" questions while Fig. 6 shows the first four questions from the "Additional Features" section. Questions 15 and 17 in Fig. 6 show that some questions have multiple responses.

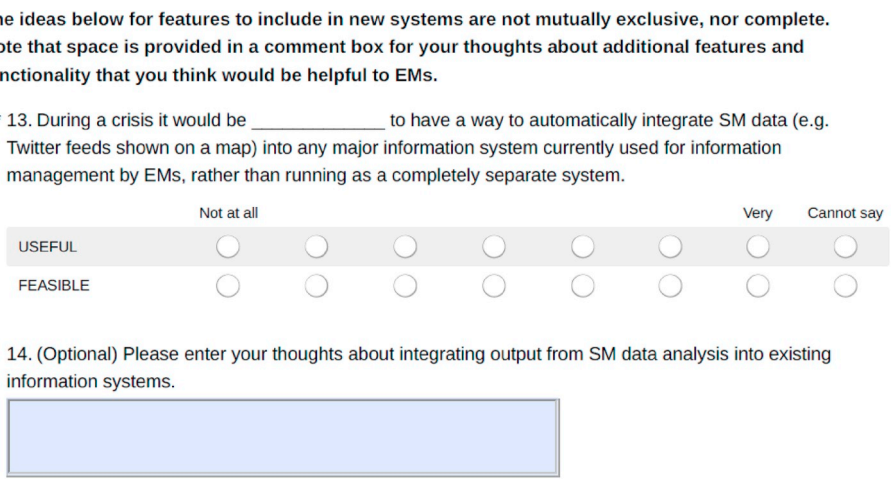


Fig. 5. Desirable Features questions.

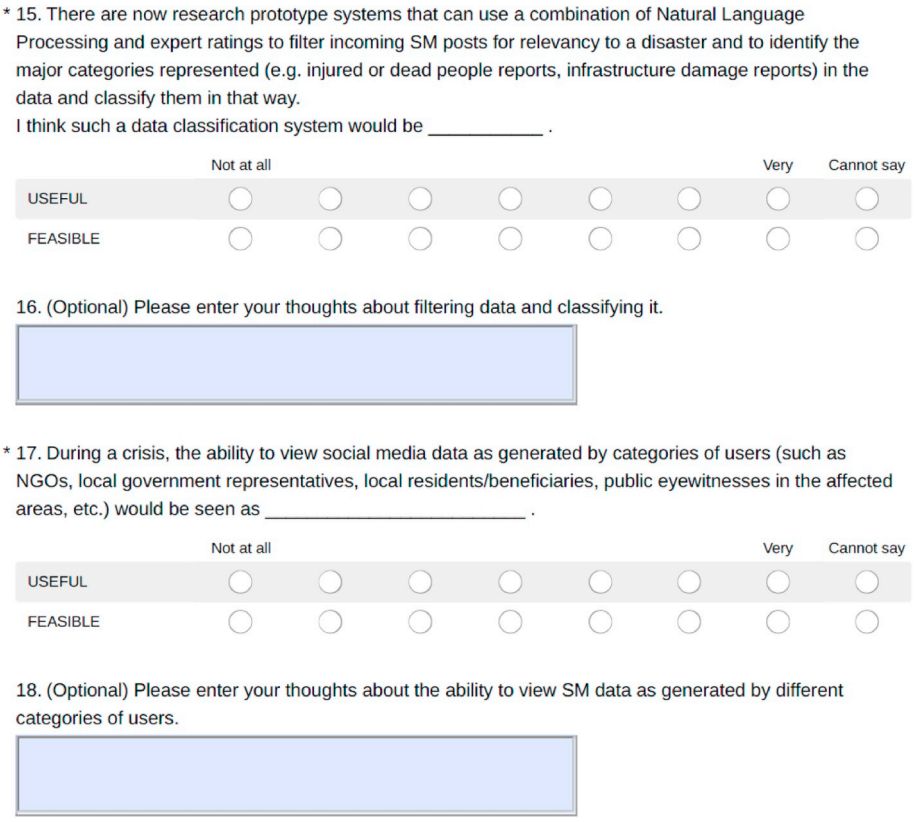


Fig. 6. Example Additional Features questions.

Fig. 7 shows the first four questions from the "Other Features" section.

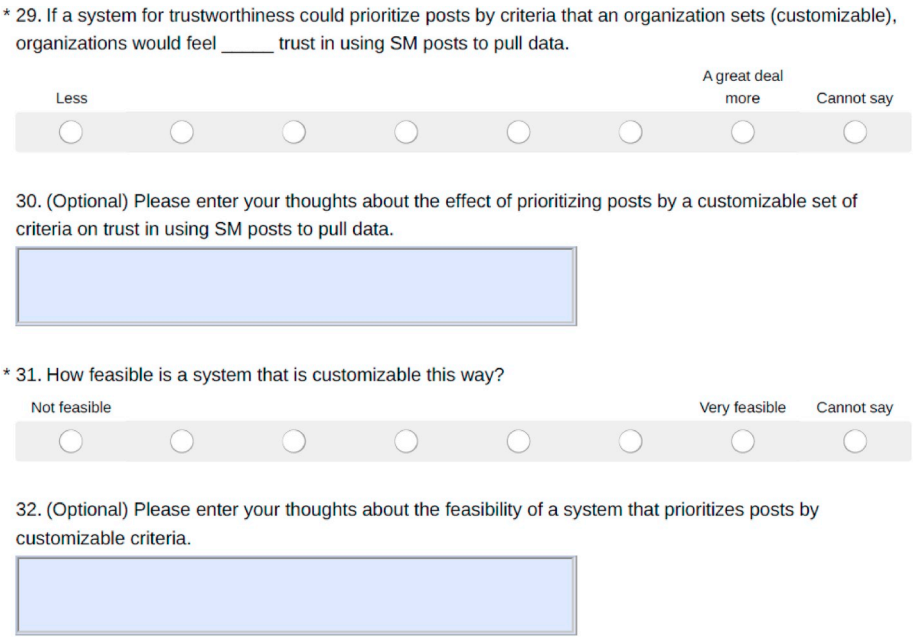


Fig. 7. Example Other Features questions.

Fig. 8 shows the three "Your Recommendations" questions and Fig. 9 the participant demographics questions.

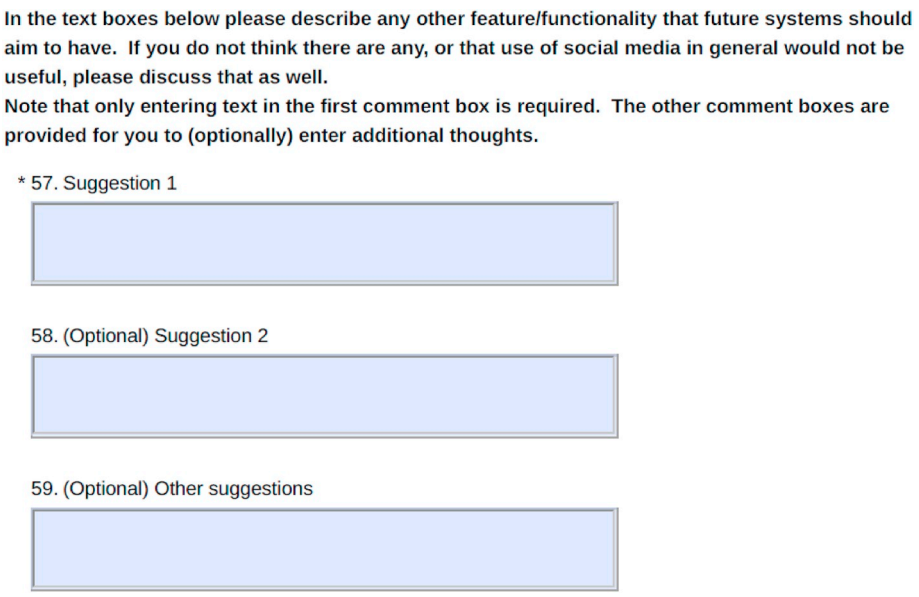


Fig. 8. The three Your Recommendations questions.

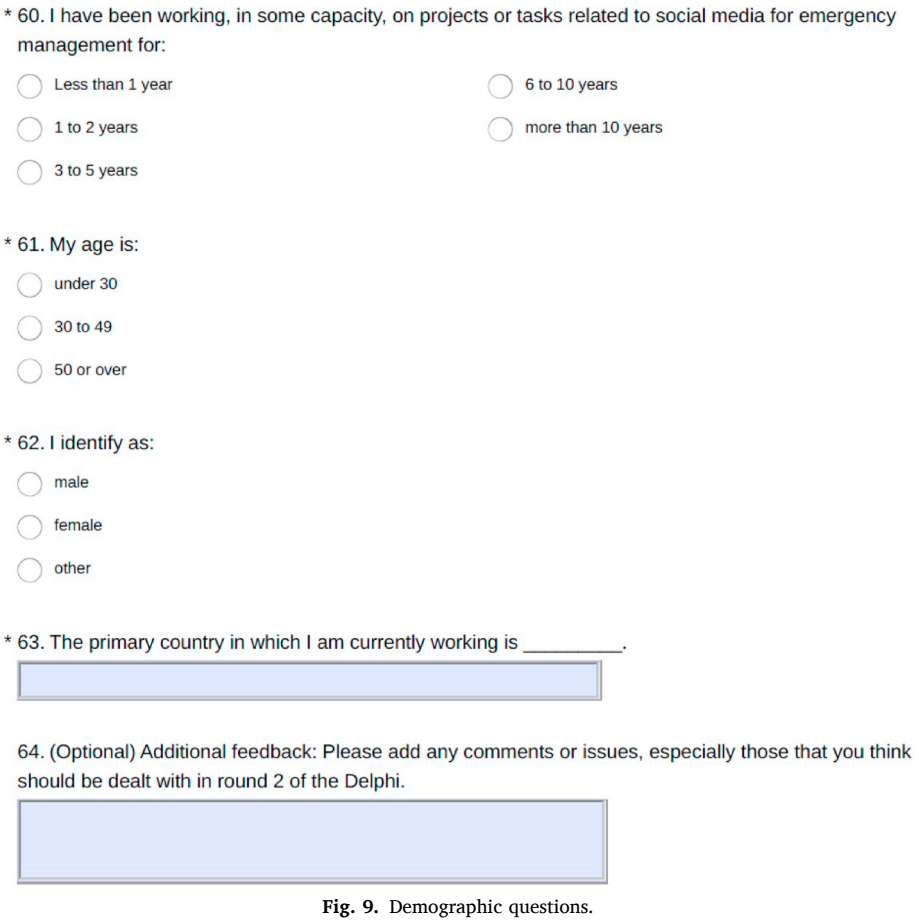


Fig. 9. Demographic questions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijdrr.2019.101367.

References

[1] R. Power, B. Robinson, J. Colton, M. Cameron, Emergency situation awareness: twitter case studies, in: Proceedings of Information Systems for Crisis Response and Management in Mediterranean Countries, ISCRAM-Med 2014, Toulouse, France, 2014.

[2] L. Palen, A.L. Hughes, Social media in disaster communication, in: Havidan Rodrigues, Joseph E. Trainor, William Donner, Antonio Paniagua Guzman (Eds.), Handbook of Disaster Research, second ed., Springer, 2018, pp. 497–518.

[3] L. Plotnick, S.R. Hiltz, J.A. Kushma, A. Tapia, Red tape: attitudes and issues related to use of social media by U.S. county-level emergency managers, in: Proceedings of the ISCRAM Conference, Kristianstad, Norway, May 24-27, 2015.

[4] L. Plotnick, S.R. Hiltz, Barriers to use of social media by emergency managers, Homeland Secur. Emerg. Manag. 13 (2) (2016). January 2016.

[5] C. Reuter, A.L. Hughes, M.-A. Kaufhold, Social media in crisis management: an evaluation and analysis of crisis informatics research, Int. J. Hum. Comput. Interact. 34 (4) (2018) 280–294, https://doi.org/10.1080/10447318.2018.1427832.

[6] W. Sherchan, S. Nepal, C. Paris, A survey of trust in social networks, ACM Comput. Surv. 445 (2013) 4.

[7] G. Panteras, S. Wise, X. Lu, A. Croitoru, A. Crooks, A. Stefanidis, Triangulating social multimedia content for event localization using Flickr and Twitter, Trans. GIS 19 (5) (2015) 694–715.

[8] J. Mahmud, J. Nichols, C. Drews, in: Where is this Tweet From? Inferring Home Locations of Twitter Users, vol. 12, ICWSM, 2012, pp. 511–514.

[9] K. Bontcheva, D. Rout, Making sense of social media streams through semantics: a survey, Semantic Web 5 (5) (2014) 373–403.

[10] A.L. Hughes, L. Palen, The evolving role of the public information officer: an examination of social media in emergency management, J. Homel. Secur. Emerg. Manag. 9 (1) (2012), https://doi.org/10.1515/1547-7355.1976.

[11] S. Vieweg, A.L. Hughes, K. Starbird, L. Palen, Microblogging during two natural hazards events: what twitter may contribute to situational awareness, in: Proceedings of the ACM 2010 Conference on Computer Human Interaction, 2010, pp. 1079–1088.

[12] H. Zade, K. Shah, V. Rangarajan, P. Kshirsagar, M. Imran, K. Starbird, From situational awareness to actionability: towards improving the utility of social media data for crisis response, in: Proceedings of the ACM Human-Computer Interaction, vol. 2, CSCW, 2018, pp. 195:1–195:18.

[13] L. Petersen, L. Fallou, P. Reilly, E. Serafinelli, Public expectations of social media use by critical infrastructure operators in crisis communication, in: Proceedings of the 14th International Conference on Information Systems for Crisis Response and Management, Albi, France, 2017.

[14] A. Cherns, The principles of sociotechnical design, Hum. Relat. 29 (8) (1976) 783–792, https://doi.org/10.1177/001872677602900806.

[15] C.W. Clegg, Sociotechnical principles for system design, Appl. Ergon. 31 (5) (2000) 463–477.

[16] R. Power, J. Kibell, The social media intelligence analyst for emergency management, in: Proceedings of the 50th Hawaii International Conference on System Sciences, 2017.

[17] A.H. Tapia, K. Bajpai, B.J. Jansen, J. Yen, Seeking the trustworthy tweet: can microblogged data fit the information needs of disaster response and humanitarian relief organizations, in: Proceedings of the 8th International ISCRAM Conference, Lisbon, Portugal, 2011.

[18] A.L. Hughes, R. Shah, Designing an application for social media needs in emergency public information work, in: Proceedings of the 19th International Conference on Supporting Group Work (GROUP 2016), ACM, NY, 2016, pp. 399–408.

[19] T. Terpstra, A. de Vries, R. Stronkman, G.I. Paradies, Towards a realtime twitter analysis during crisis for operational crisis management, in: Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM 2012),Vancouver, BC, 2012.

[20] M.M. Maresh-Fuehrer, R. Smith, Social media mapping innovations for crisis prevention, response, and evaluation, Comput. Hum. Behav. 54 (2016) 620–629, https://doi.org/10.1016/j.chb.2015.08.041.

[21] O. Okolloh, Ushahidi, or 'testimony': web 2.0 tools for crowdsourcing crisis information, Particip. Learn. Action 59 (1) (2009) 65–70.

[22] M. Imran, C. Castillo, J. Lucas, P. Meier, S. Vieweg, AIDR: artificial intelligence for disaster response, in: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, 2014, pp. 159–162.

[23] T. Ludwig, C. Reuter, V. Pipek, Social Haystack: dynamic quality assessment of citizen-generated content during emergencies, ACM Trans. Comput. Hum. Interact. (TOCHI) 22 (4) (2015), https://doi.org/10.1145/2749461, 17:1–17:27.

[24] Twitter, TweetDeck, 2019. Retrieved October 3, 2019 from, https://tweetdeck.twitter.com/.

[25] Hootesuite, Hootesuite, 2019. Retrieved October 3, 2019 from, https://hootsuite.com/.

[26] D. Pohl, Social media analysis for crisis management: a brief survey, IEEE STCSN E-Lett. 2 (1) (2013).

[27] M. Imran, C. Castillo, R. Diaz, S. Vieweg, Processing social media messages in mass emergency: a survey, ACM Comput. Surv. 47 (4) (2015). Article 67, July 2015.

[28] M. Imran, P. Mitra, C. Castillo, Twitter as a lifeline: human-annotated twitter corpora for NLP of crisis-related messages, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016) (23-28), European Language Resources Association (ELRA), Paris, France, 2016.

[29] M. Imran, C. Castillo, F. Diaz, S. Vieweg, Processing social media messages in mass emergency: survey Summary, in: Proceedings of WWW 2018, April 23-27, Lyon, France, 2018.

[30] R. McCreadie, C. Buntain, I. Soboroff, TREC incident streams: finding actionable information on social media, in: Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management, Valencia, Spain, 2019.

[31] L. Plotnick, S.R. Hiltz, Software innovations to support the use of social media by emergency managers, Int. J. Hum. Comput. Interact. 34 (2018) 4.

[32] H. Linstone, M. Turoff, The Delphi Method: Techniques and Applications, Addison-Wesley, Reading, MA, 1975.

[33] H. Linstone, M. Turoff, Delphi: a brief look backward and forward, Technol. Forecast. Soc. Chang. 78 (9) (2011) 1712–1719.

[34] A.H. Hendela, M. Turoff, S.R. Hiltz, J.L. Fjermestad, A risk scenario for small businesses in Hurricane Sandy type disasters, in: Proceedings of the Hawaii International Conference on System Sciences, 2017.

[35] S.B. Liu, L. Palen, J. Sutton, A.L. Hughes, S. Vieweg, In search of the bigger picture: the emergent role of on-line photo sharing in times of disaster, in: Proceedings of the Information Systems for Crisis Response and Management Conference, ISCRAM 2008), Washington, DC, 2008.

[36] D.T. Nguyen, F. Ofli, M. Imran, P. Mitra, Damage assessment from social media imagery data during disasters, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, 2017, pp. 569–576.

[37] F. Alam, F. Ofli, M. Imran, Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of Hurricanes Harvey, Irma, and Maria, Behav. Inf. Technol. 0 (0) (2019) 1–31, https://doi.org/10.1080/0144929X.2019.1610908.

[38] G. Rogova, P. Scott (Eds.), Fusion Methodologies in Crisis Management: Higher Level Fusion and Decision Making, Springer, 2016.

[39] D. Sui, M. Goodchild, The convergence of GIS and social media: challenges for GIScience, Int. J. Geogr. Inf. Sci. 25 (11) (2011) 1737–1748, https://doi.org/10.1080/13658816.2011.604636.

[40] H. Purohit, A. Hampton, S. Bhatt, V.L. Shalin, A.P. Sheth, J.M. Flach, Identifying seekers and suppliers in social media communities to support crisis coordination, Comput. Support. Coop. Work 23 (4–6) (2014) 513–545, https://doi.org/10.1007/s10606-014-9209-y.

[41] A. Chauhan, A.L. Hughes, Providing online crisis information: an analysis of official sources during the 2014 carlton complex wildfire, in: Proceedings of the 35th International Conference on Human Factors in Computing Systems (CHI 2017), 2017.