Åbo Akademi

OPEN ACCESS

This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Promoting Situation Awareness: Usability of a Social Media Tool for Journalists and Crisis Managers

Backholm, Klas Henrik; Högväg, Joachim; Lindholm, Jenny; Knutsen, Jørn; Westvang, Even

Published in:
International Journal of Information Systems for Crisis Response and Management

DOI:
10.4018/IJISCRAM.2018010103

Published: 01/01/2018

Link to publication

Please cite the original version:
Backholm, K. H., Högväg, J., Lindholm, J., Knutsen, J., & Westvang, E. (2018). Promoting Situation Awareness: Usability of a Social Media Tool for Journalists and Crisis Managers. *International Journal of Information Systems for Crisis Response and Management*, *10*(1), –. https://doi.org/10.4018/IJISCRAM.2018010103

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This document is downloaded from the Research Information Portal of ÅAU: 26. Apr. 2024

International Journal of Information Systems for Crisis Response and Management
Volume 10 • Issue 1 • January-March 2018

Promoting Situation Awareness:
Usability of a Social Media Tool for Journalists and Crisis Managers

Klas Backholm, Åbo Akademi University, Vasa, Finland
Joachim Högväg, Åbo Akademi University, Vasa, Finland
Jenny Lindholm, Åbo Akademi University, Vasa, Finland
Jørn Knutsen, The Oslo School of Architecture and Design/Voy, Oslo, Norway
Even Westvang, Bengler AS, Oslo, Norway

ABSTRACT

Crisis communicators and journalists need stable structures to handle social media content in emergencies, but struggle with information overload. The usability of a tool intended to support information gathering was investigated by conducting two usability tests (low- and high-fidelity prototypes) with journalists. The aims were to investigate how well the design reflected users' general mental models of emergency work, and how it responded to the specific requirements set by work in high-stress surroundings. Tests were conducted in a laboratory. Participants understood the main prototype concepts, but struggled with time-consuming tasks, for instance, those related to saving content or evaluating information quality. To provide good situation awareness - and to fit in with user expectations - a system should gather information from several social media outlets and allow for varying possible user modes. However, system designers need to carefully balance between including necessary features and avoiding tasks that require complex manual actions.

KEYWORDS

Crisis, Crisis Management, Emergencies, Information Gathering, Journalism, Situation Awareness, Social Media, Usability

DOI: 10.4018/IJISCRAM.2018010103

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

INTRODUCTION

A rampage truck attack, a devastating earthquake, or a sudden flooding - during an emergency[1], information about the ongoing event is spread rapidly through a complex network of interacting communication channels (Latonero & Shklovski, 2011; Simon, Goldberg, & Adini, 2015). Via social media such as Facebook, Twitter or Instagram, content is posted and may be merged with new facts or rumors. Parts of this information are integrated into traditional mass media products or forwarded by direct interpersonal communication or reposted in social media in new forms. Soon, it becomes difficult for crisis communicators to handle the vast amount of content and estimate the trustworthiness of seemingly central information (Eriksson, 2012). This may have serious effects on the perception of a situation and undermine a deeper understanding of the problem; or even cause additional emergency damage (Coombs, 2015).

Technical systems can provide support for crisis communicators who need to handle vast information flows in emergencies (Ruggiero & Vos, 2014). However, gathering social media with such systems is a complex process, requiring an understanding of data discovery, collection, preparation and analysis (Imran, Castillo, Diaz, & Vieweg, 2015; Stieglitz, Mirbabaie, Ross, & Neuberger, 2018). In this study, the work behind developing the prototypes of a new tool to support the gathering of social media information is presented (Sections 4-7). The tool is intended for users who work professionally with gathering crisis-related information, such as authorities, first response rescuers, NGOs, and news organizations.

The tool itself was designed with a user-centered approach, the aim being to optimize it according to how users can, want, or need to use it (ISO 9241, 2010; Wallach & Scholz, 2012). The main focus of this article is to investigate the usability of two prototypes of this tool (Sections 5-6). The results from the usability tests conducted in a laboratory setting with one intended user group, news journalists, are presented. The sample was limited to journalists as the prototype developer wished to work with one clearly defined group during the first parts of the development process.

This study investigates how well the prototype tools fit in with journalists' existing mental models, and how the design should be improved to both support work tasks and an overall awareness of an unfolding emergency (Sections 2-3, 7). This study contributes to the field of IT support for emergency management by providing knowledge on how to design social media information gathering systems that reflect the specific requirements set by emergencies and emergency-related communication practices.

Understanding Social Media Information Gathering in Emergencies

Social media ecosystems are based on four elements, one of them being the possibility for content datafication (Van Dijk & Poell, 2013). This means that social media platforms produce data that may be collected and used by different actors for various purposes. In a crisis communication context, datafication enables external actors to use technology in order to gather user-generated content about an unfolding emergency, and to distill it into more useful forms that may support efficient actions or responses (Pirolli & Russell, 2011). While the specific purpose of gathering information will vary between organizations, the practical processes related to the information gathering are similar (Coombs, 2015; Ruggiero & Vos, 2014).

Scholars have documented several examples of the usefulness of such data gathering for emergency communicators. During Hurricane Sandy, rescuers used Facebook to continuously monitor, listen in, and respond to the public's concerns about how damaged specifically-affected areas were, and when evacuated residents were allowed to return home (Hughes et al., 2014). Ludwig et al. (2015) showed several examples of how emergency managers should monitor social media to identify and possibly benefit from emerging self-organizing groups, digital volunteers, who in an improvised manner start coordinating support efforts via social media platforms. Brandtzæg, Luders, Spangenberg, Rath-Wiggins, and Folstad (2016) highlighted how news organizations can benefit from social media information in their production of emergency news content.

When designing a system intended to support information gathering processes, it is vital that designers have an understanding of users' cognitive mental models (ISO 9241, 2010). There are several views on mental models of relevance for information systems design (Rutherford & Wilson, 1991). Mental models represent a user's view of processes or tasks, such as how a specific work task should be carried out (Johnson-Laird, 1983). When needed, such as when trying out a new system, the user retrieves relevant models from memory, and applies them to the situation at hand. Models are based on the user's gathered experiences and knowledge and are continuously reformulated if new relevant information is acquired. Rouse and Morris (1986) specifically see mental models in relation to a system's design and define them as user mechanisms specifically for understanding and describing system factors such as function, form, and purpose.

In this study, these two main views of mental modeling are not seen as contradictory, but rather as focusing on different aspects of user cognitive processes. When a new system is introduced, a user will naturally start constructing a mental model of this system, while simultaneously comparing it with pre-existing mental models of similar systems and general expectations of work tasks. By implementing a repetitive usability test strategy with intended user groups throughout the system development process, and including a pre-design mapping of central needs, designers can address discrepancies between the design, users' existing general work task expectations, and users' emerging mental models of the system (Hertzum & Clemmensen, 2012; ISO 9241, 2010; Wallach & Scholz, 2012). Accordingly, an understanding of more general mental models of work tasks is central in the early stages of usability testing, while information on how the user constructs a model related to the new system becomes more important during later tests.

In addition to understanding user mental models and needs, a system for emergencies needs to correspond to contextual requirements. For instance, during the acute phase of an emergency, the user is expected to carry out tasks at a fast pace, and several surrounding actors and other moving parts within the system and around the user will simultaneously require attention. Furthermore, the vast amounts of available social media content contribute to a possible "information overload". The user is also experiencing the emergency and must balance potential personal distress or worry with workplace expectations. All this may add to an increased user stress level, and affect cognitive mechanisms such as information processing, decision-making, and communication negatively (Edland & Svenson, 1993; Mendl, 1999). To avoid adding to the user's stress level, a system needs to have a clear and straightforward design that leaves little room for (mis)interpretation.

In this context, the situation awareness (SA) framework is useful (Endsley, 2009; Salmon et al., 2008). A high level of SA is to experience that you are in control and understand what is happening while using a system, and what this means to you and your ongoing work tasks (Endsley, 2009; Yin et al., 2012). This naturally also contributes to a better sense-making of the overall situation. To support good SA, systems design should support three levels of situational understanding: perception, comprehension and projection (Endsley & Jones, 2004). Perception is about perceiving relevant elements of information in the system, for example via visual or auditive cues. Comprehension is to understand what the information means for ongoing work tasks and related goals, by fitting pieces of information together and prioritizing them. Projection is to predict what the collected information will mean for decisions made in the near future.

Usability tests of systems intended for emergency use need to carefully consider both the practical design and contextual factors that may affect SA. For example, technical factors such as unclear feature placement or instructions, or user fatigue due to vast workloads, may affect working memory or how decisions are made (Endsley & Jones, 2004; Pirolli & Russell, 2011; Salmon et al., 2008). The SA and mental model frameworks partly overlap each other. However, SA is specifically focused on high-stress situations and systems design, and how contextual factors in such scenarios affect user cognitions.

News Journalists and Social Media Information Gathering

Currently, the first attempts at technological innovations to support journalistic information gathering are taking form (Diakopoulous et al., 2012; Schwartz et al., 2015). Scholars have provided overviews, for instance in the Verification Handbook (Silverman, 2014), including systems either specifically designed for journalists or intended for a broader user group. However, there is currently little documentation about whether journalists actually use the tools listed in such overviews (Brandtzæg et al., 2016).

When scholars have investigated what journalists expect from an information gathering system, several central needs have emerged. A tool should both monitor emerging events and identify information during ongoing events (e.g. sudden turns or new trends; Brandtzæg et al., 2016; Schwartz et al., 2015). The tool should integrate content from several social media, and handle varying forms of content (text, pictures, videos; Brandtzæg et al., 2016; Schifferes et al., 2014). It should include visualizations of gathered content and provide good functionality across varying forms of technical equipment (Backholm et al., 2017; Schwartz et al., 2015).

A tool should also allow the user to define the type of content that is tracked or presented, such as filtering features for content as well as for sources (Diakopoulous et al., 2012; Schifferes et al., 2014). Furthermore, it should give the user a good understanding of how automatic search algorithms are carried out, as this creates a better understanding of how reported system results are produced (Schwartz et al., 2015). Finally, samples have expressed concern about including too many functions. This should be avoided since it will affect the tool usability negatively (Backholm et al., 2017).

Aim of the Study

The main aim was to test the usability of a tool intended to support crisis communicators' social media information gathering processes. This was done by conducting two usability tests of low- and high-fidelity tool prototypes with one intended user group, news journalists. Based on the theoretical framework and previous related work, two research questions were formulated:

RQ1: How should a social media information gathering tool be designed to correspond to news journalists' mental models about how to carry out information gathering tasks in general, and during emergencies in particular?

RQ2: How should a social media information gathering tool be designed to enhance emergency-situation awareness for news journalists?

Below, the tool prototypes, test methodology, and central results are described separately for each test. In the concluding part of the article, the possible implications for crisis communicators in general are discussed.

USABILITY TEST 1 – A SEMI-INTERACTIVE WIREFRAME PROTOTYPE

A pre-design mapping of users' existing mental models about social media information gathering in emergencies, and needs related to this, was conducted with European news journalists. The results have been presented in detail in a previous study (Backholm et al., 2017). The mapping showed that most journalists work with an extensive range of different types of assignments, including but not limited to emergencies. Therefore, a tool intended for journalistic work needs a functionality that applies to both daily and emergency use.

Tool features should focus on supporting information gathering, but not attempt to automatize content validation processes. Journalists would not trust a system to evaluate information quality. Furthermore, each journalist applies his or her own methods for finding information, and the tool should reflect this variability. Thus, to support daily and emergency work, a tool should collect information

Table 1. An overview of the usability testing process applied in the study. A pre-design mapping of news journalists' user needs (part A) influenced the design of tool Prototype 1 (part B). The results from the usability test of this prototype (part C) influenced the design of Prototype 2 (part D), which was evaluated in a second usability test (part E).

A. Pre-design mapping of user needs (27 European news journalists)[1]
Method: Interviews with journalists; shadowing journalistic work tasks during crisis scenario Main results: A social media information gathering tool should automatically collect information across platforms and formats into one workspace It should offer several options for manual handling of content (e.g., what to search for, how to sort or visualize results) It should not automate validation of collected content – but provide visualizations of detailed source parameters important for manual validation purposes It must be relevant for both emergency and everyday (e.g., monitoring social media for potentially newsworthy content) news work
B. Tool prototype 1[2]
Prototype functionality: A semi-interactive wireframe version. Users can navigate static images by clicking on predefined areas Central prototype features: Two work modes, monitor and story In monitor, the user follows various topics (e.g., trusted sources or geographic area) In story, the user collects content related to an ongoing story (e.g., an emergency) Story mode has three tabs. Feed tab includes all found content related to the story; selected tab includes content of specific interest for the story, that is moved manually from feed to selected tab by the user; Tasks tab includes listed tasks related to the story for the user to carry out, created by the same or other users
C. Usability test on tool prototype 1 (17 news journalists)[3]
Method: Laboratory test including interviews, observations, and surveys Main results: General first impression of prototype mainly positive – one participant did not understand the main point with the prototype Journalists would work in monitor mode when searching for newsworthy content in everyday work, and in story mode during an emergency Problems in story mode with moving content from feed to selected tab Problems in story mode with understanding the main idea with the tasks tab
D. Tool prototype 2[4]
Prototype functionality: A close-to working version with a reactive wireframe that responds to user input and uses live data Central prototype features: Keeps the two work modes, monitor and story. More features added, and styling of components and color scheme updated after first prototype Users can create new stories and monitor, add and modify content Due to results from the first usability test, the selected tab in story mode was renamed to saved tab Due to results from the first usability test, the tasks tab in story mode was removed entirely
E. Usability test on tool prototype 2 (15 news journalists)[5]
Method: Laboratory test including interviews, observations, and surveys Main results: First impression was positive - the new prototype was logical and easy to approach Journalists saw content feed overview and modification options as useful Same problems as in previous test with moving content from feed to saved tab in story mode remained Problems in story mode with understanding new detailed features (e.g., how to manually add location or evaluation to pieces of content)

[1]Detailed results have been published in a previous study byBackholm et al. (2017)

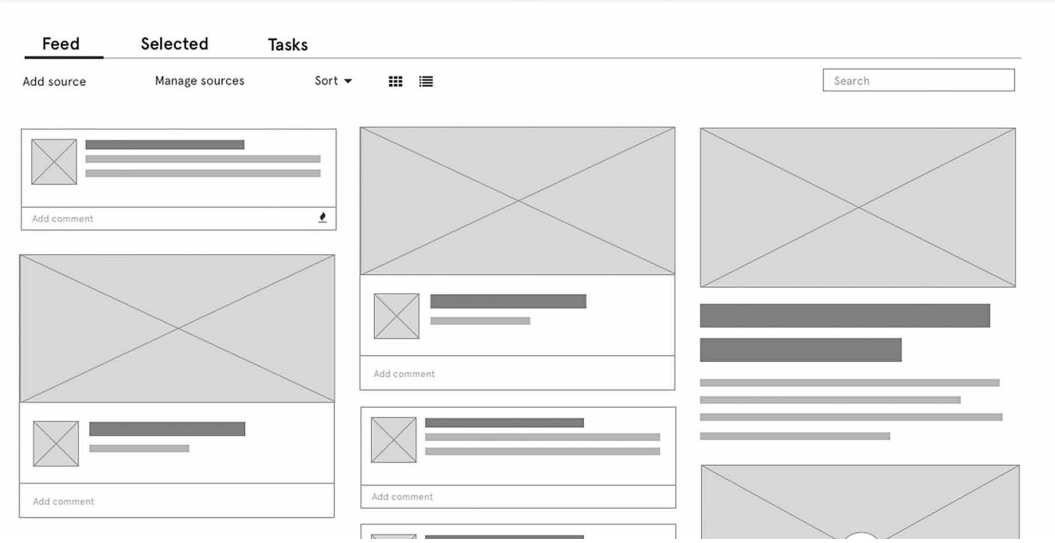
[2]Tool prototype illustrated inFigure 1

[3]Conducted in April 2016

[4]Tool prototype illustrated inFigure 3

[5]Conducted in January-February 2017

Figure 1. The semi-interactive wireframe prototype used in Test 1. The story/feed tab, where the user collects content from social media and web sources, is shown in the figure. In the test, the user could navigate static images by clicking on predefined areas.



from across social media outlets into the same feed. It should include several possibilities to modify or limit searches, for instance according to type of format, time or geographical location. To further support personal needs, the tool should both visualize content in a general feed and allow for sorting and visualization according to the user's preferences. The results are summarized in part A of Table 1.

The first prototype was a semi-interactive wireframe version (Figure 1), where users could navigate static images through clicking on predefined areas. The tool was centered around the concept of a story. A story is a field of interest that the user is currently working on, like a breaking news incident.

In a story, the journalist may collect content from a range of different sources (e.g. tweets, Instagram images, or news articles). Users may also subscribe to a range of different feeds or searches that they set up (e.g. a hashtag, a geographical location tag, or an RSS-feed). Each story contains a feed, much like a Facebook feed, where content from these subscriptions will appear – the newest appearing on top. From here, the user can sift through the incoming feed, getting an overview of the content.

When users notice content that is of particular interest for the story they are working on, they may save that specific object in a separate tab called "selected". When selecting, the user can add some metadata to the object: attach the object to a collaborative task manager, rate how trustworthy and important it is for the ongoing story, assign a geographical location, or add a comment.

Under the selected tab, the user is given a range of ways to visualize selected content: a timeline view, presenting objects on a map, or displaying in accordance with the evaluation given. The story view also includes a task tab, where users can create and communicate (instant messaging) tasks related to the story, such as "please check this source in more detail".

In addition to the story view, the prototype incorporated a feature called "monitor". This is a view where a user will typically set up a range of searches or subscriptions that don't pertain to a specific story. Envisioned usage would be to monitor a specific area or a list of trusted sources on various topics. This feature is a more general space where one may discover new stories rather than focusing on what you are currently working on[2].

Sample and Method

The main focus of the test was to collect information about users' first impressions of the prototype design, how well users understood the logic behind included features, and whether the design fit in with users' mental models of how to carry out information gathering tasks (Table 1). All data was handled in accordance with ethical guidelines. The test procedures have received an ethical approval from the Board of Ethics at Åbo Akademi University.

Forty news journalists were approached, of whom 17 chose to participate (11 females, 65%; age 27−60, $M = 38$). The final sample had between 3 and 35 years of work experience ($M = 14$). Inclusion criteria included to currently work as a journalist, have a working experience of at least one year, and using the most common social media platforms to some degree as a part of one's work. Eleven journalists worked mainly with newspapers, five with radio, three with television, and 15 with the web (participants could indicate more than one media format). Participants used between one and six social media platforms for general work and leisure ($M = 3.3$), and one to four platforms for information searching at work ($M = 2.2$). In total, eight platforms were mentioned by the sample, Facebook, Instagram, and Twitter being the most common. One journalist used a third-party social media validation or verification tool at work.

Data was collected during two weeks in April 2016 in a laboratory setting. The test was conducted in two rooms, one furnished as a living room (parts 1−2 and 4, see descriptions below) and one as an office with a stationary computer (part 3). The tool prototype was shown on a full HD 24-inch computer screen. Data from one participant was collected per session, and each session took between 1−2 h. Two researchers participated during each test.

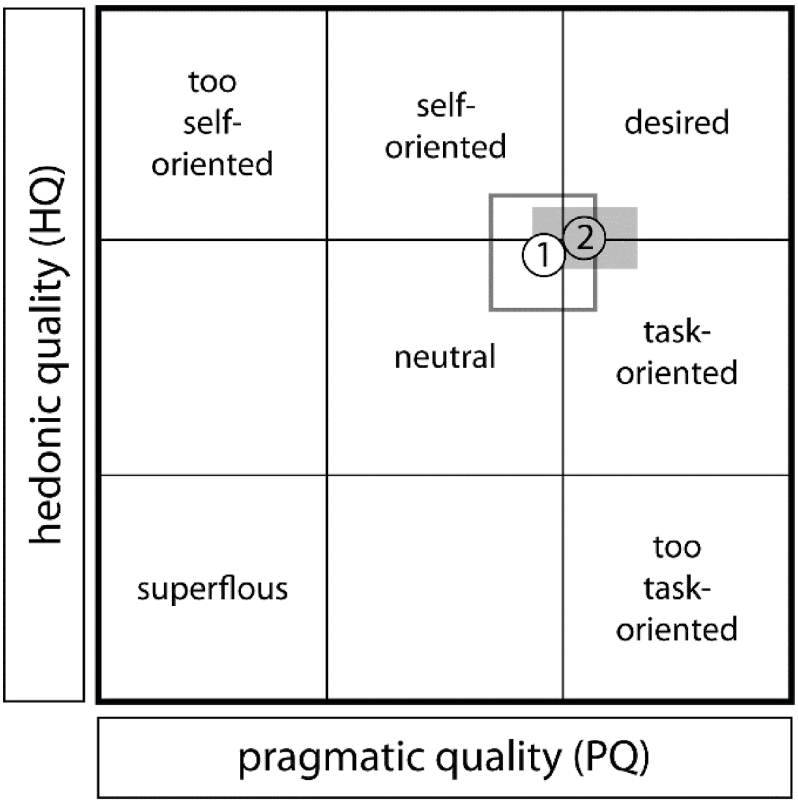
The test procedure consisted of four parts. First, the study purpose and structure as well as participant rights were introduced. Second, a semi-structured interview was conducted, with the purpose of gathering demographic information and level of domain expertise. Third, the participant was guided through the tool prototype, using a set path of tasks. One researcher guided the participant, while another observed. The tasks consisted of 34 structured steps, divided into two main categories:

1. **Meaning category:** questions about the first impression of the presented content, without clicking (19 steps): "What do you think *location* means?"
2. **Click category:** structured navigation through presented content (15 steps): "Click on *location*, what happened? What can you do here?"

A concurrent think aloud process was applied, and researchers scored task comprehension using three levels: "pass" (no problems), "struggle" (some problems, but completed the task) or "fail" (did not complete the task). Criteria for reaching each level of understanding had been defined prior to data collection and tested in a pilot study. Fourth, another semi-structured interview was conducted[3]. The aim was to discuss first general impressions of the prototype as a whole. To support the discussion, a stimulated recall procedure was applied. The tool prototype was shown on a screen and the participant was encouraged to point out features that could be improved.

After the interview, the participant completed the AttrakDiff 2 survey (Hassenzahl, 2006). The survey consists of pairs of words that represent opposite extremes of usability-related dimensions. The participant grades each word pair using scores 1−7, where 4 represents a neutral level. Two subscales were used: the perceived pragmatic qualities subscale (PQ), which indicates how easy the user finds it to manipulate a product, and the perceived hedonic quality-stimulation (HQ-S) subscale that focuses on how stimulating the product feels. The remaining parts of the AttrakDiff were not relevant for this prototype, since they focus on how visually pleasing and to what degree a user identifies with a product – factors that are difficult to evaluate in a rudimentary prototype.

Figure 2. The Attrakdiff results from prototype tests 1–2 illustrated according to scale dimensions provided by Hassenzahl (2006). The group mean from the first test is indicated with number 1, and confidence intervals within the transparent rectangle. Test 2 is indicated with the corresponding number, and confidence intervals within the semi-transparent gray rectangle. Note: N = 15 in both tests. Possible group mean range 1–7, the score 4 represents a neutral level. The HQ-axis consists of the HQ-S and HQ-I subscales, and the PQ-axis of the PQ subscale. Results from Test 1 are based on the PQ and HQ-S subscales: PQ M = 4.92, 95% CI [4.44, 5.41]; HQ-S M = 4.85, 95% CI [4.39, 5.29]. Results from Test 2 are based on three subscales: PQ M = 5.20, 95% CI [4.71, 5.69]; HQ-S M = 4.98, 95% CI [4.69, 5.27]; and HQ-I M = 5.13, 95% CI [4.82, 5.44].



Main Results

The key feature (Table 1) to include social media content from several platforms and present it in a way that can be modified according to personal needs, was well received. Twelve participants reported a mainly positive first impression of the prototype. One participant struggled with understanding the prototype as a whole, what it was meant for and how this was relevant for oneself.

Most participants saw themselves as working mainly in monitor mode when scanning for ideas for journalistic stories on an ordinary workday, and in story mode when working a specific story (e.g. an emergency). The AttrakDiff[4] indicated that the sample saw the rudimentary prototype as relatively easy to manipulate and somewhat stimulating to use (Figure 2).

Participants understood specific tasks to an acceptable degree; nine participants passed 19 of in total 34 tasks without a struggle, and 13 passed eight of the tasks. None of the tasks caused significant confusion in all participants; but on the other hand, every task was confusing for more than one participating journalist.

One of the main prototype challenges was that most participants had difficulties understanding the workflow between two of the story tabs (feed/selected). The feed tab, including social media content about the user's chosen topic was, in itself, relatively easy to understand. However, the function to

select interesting content from the feed and add it to a separate selected tab, indicated by a plus icon that opened a pop-up window, was unclear. Many thought that the function would merely open up the content on the feed tab in more detail.

The content visualization options (timeline/location/evaluation) in the selected tab also caused problems. Participants mainly had a good understanding of what the options represented. However, many thought they were sorting all content from the feed, for instance according to a timeline, not content they had manually selected. In other words, participants understood what the visualization features were supposed to do but missed the main underlying idea that the presented content had previously been handpicked by the user.

The third story tab, tasks, was the most problematic. The main function of this tab was to create and communicate tasks with colleagues. Thus, this tab introduced a new way of thinking in the prototype. Until now, participants had mainly focused on individual tasks, but in this tab, the main feature had to do with work distribution and communication. Participants struggled with varying issues, ranging from not understanding the overall function of the feature, to details such as who sends/receives tasks or what tasks consist of. The tab also provoked frustration since newsrooms have established systems for communicating tasks.

Some issues related to participants' more general expectations about journalistic work also provoked discussion. Six participants had a negative impression of the feature to evaluate content in the selected tab. Discussions focused, for example, on whether participants evaluated trustworthiness of content or of the social media source, and that evaluating information would require too much time. However, participants saw that if such issues could be sorted out, an evaluation function would be a useful tool feature.

USABILITY TEST 2 – A CLOSE-TO WORKING PROTOTYPE

The second usability test was carried out nine-ten months after the first one. The second prototype iteration was a close-to working prototype with a reactive interface that responds to user input and makes use of live data. This entails that the users would be able to create new stories and monitor, add and modify content. The list of sources that worked with live data included Twitter and RSS feeds. Aesthetically, the prototype keeps the general two-column layout from the wireframe prototype but added a new color scheme and styling of the components (see Figure 3).

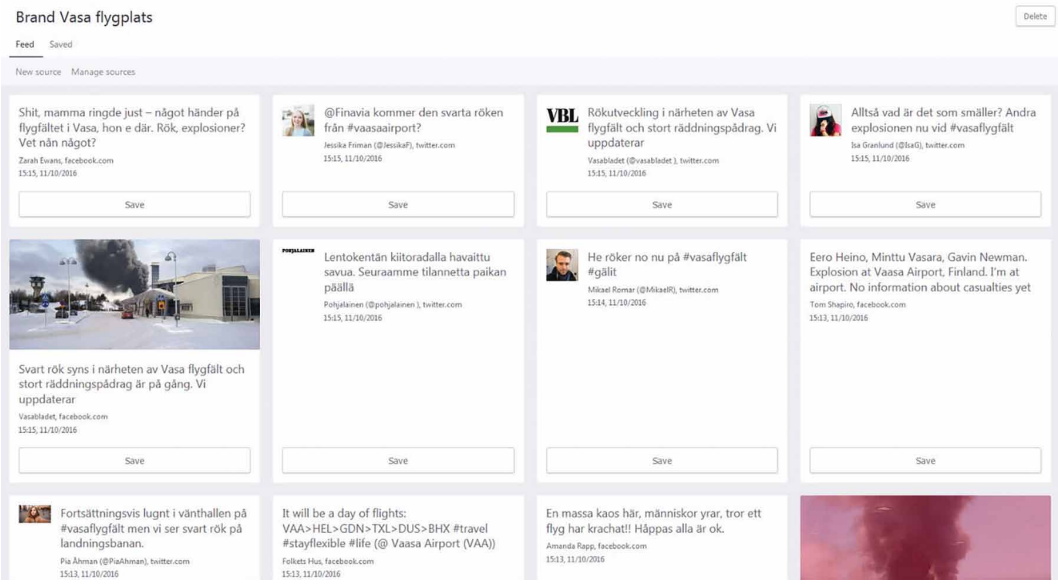
The main effort when developing the second prototype was to get the necessary infrastructure in place (i.e. data and user handling) and to transform static wireframes into programmed and interactive user interfaces. Based on the results from the first usability test, central prototype changes included removing the story/task tab and renaming the story/selected tab to story/saved. In this prototype, saved items in the feed were marked with a red opaque overlay and a checkmark.

Sample and Method

The main aim of the test was to identify usability problems within specific prototype features in more detail. An additional aim was to look at prototype development from the first prototype. Thus, participants needed to have an understanding of the first version, and all participants from the first round were contacted. Out of the 17 possible participants, 15 chose to participate (9 females, 60%; age in sample 28–61, $M = 40$). Dropouts were due to relocation and sick leave. Participant demographic information had not changed since the first test.

Data was collected in a laboratory and a similar design as in the previous test was used. A laboratory was deemed as the most suitable setting since the prototype still did not have enough functionality to be tested in a newsroom. Two researchers participated during each test. A test session consisted of three parts. Firstly, the structure of the study and participant rights were introduced, and possible alterations to demographic information were collected. The participants were asked how

Figure 3. The close-to working prototype with a reactive interface used in Test 2. The story/feed tab, including social media updates from a fake emergency scenario, a fire at the local airport, is shown in the figure.



much they remembered from the first prototype and then given a brief introduction to the structure of the current version, to ensure that all participants had the same information.

Secondly, each participant was seated in front of a computer screen and asked to complete tasks in a predefined order (36 tasks in total). As the second prototype had more functionality and responded better to user input, the two test step types implemented in the previous test ("meaning" and "click" categories) were further refined. Two categories were added to collect more data on how participants understood the collected content and what features were intended for. Participants thus responded to four types of questions/tasks:

1. **Meaning category:** questions phrased as "what do you think is the meaning of x?"
2. **Click category:** specific prototype tasks to carry out.
3. **Source category:** questions concerning where the presented information originates from.
4. **Function category:** questions concerning whether the test participant understood a feature.

In this test, a retrospective think aloud procedure was implemented, so as not to disturb task completion[5]. Researchers observed the task completion and graded participant performance in a similar fashion as in Test 1: "pass" (no problems), "struggle" (some problems, but completed the task) or "fail" (did not complete the task). The criteria for reaching each level had been defined prior to data collection and tested in a pilot study.

For the test, a simulation of a real-life emergency, a fire at the regional airport, was developed. This consisted of a premade dataset of social media content[6] that was played back during the test sessions – each piece of content appearing at specific time intervals. However, as a simulation carried out in a laboratory setting cannot reflect all aspects of a real emergency in a realistic way, the simulation formed a backdrop during test completion and was not a central factor for reaching the test aims. Thus, participants were not dependent on following each update or creating a thorough understanding of the unfolding emergency. Instead, usability tasks were carefully planned so that they could be carried out independently of how far the emergency scenario had developed. The role

of the emergency scenario in relation to task completion was explained to the participants in detail before the test.

Thirdly, a structured interview was conducted with the participants with the aim of pinpointing positive and negative feedback and to get an overall impression of the prototype. After the interview, each participant filled out the AttrakDiff 2 questionnaire. All subscales (pragmatic qualities PQ; hedonic quality-stimulation HQ-S; hedonic quality identity HQ-I; and attractiveness, ATT) were included in this test. HQ-I represents how much the user identifies with a product, while ATT is about how good the product looks. PQ and HQ-S are explained in the section about Test 1.

Main Results

The first general impression was that this prototype was easy to approach and felt more logical than the previous one. Participants saw the social media content feed overview as well as the visualization options the prototype provides (e.g. timeline, location) as very useful. Several participants highlighted that this prototype seemed to require less initial learning than is usually the case with new systems.

Eleven participants saw that the prototype would be a useful addition to their journalistic toolkit. Approximately half of those saw this tool as useful specifically during emergency-related assignments, while the rest discussed how it could be applied to daily work tasks. The AttrakDiff showed that the sample was quite pleased with this prototype version and saw it as well-functioning, but that there is still room for improvement (Figure 2).

In this test, participants completed 36 tasks. Eight participants passed 32 tasks without a struggle, and 12 passed 26 of the tasks. Six tasks did not cause significant confusion for anyone in the sample. None of the tasks caused confusion in all participants.

The tasks where most participants struggled/failed were about detailed features related to advanced search or visualization functions. Such features included, for example, adding content from a webpage link or searching for content with a combination of several hashtags. These results were also reflected in interviews, where participants generally described difficulties with more detailed features, while the overall tool logic and the main features worked well. Mentioned difficulties included keeping track of the options, pop-up windows, and features that require manual input or a lot of focus on content interpretation.

For example, the option to add an evaluation to saved content caused problems, as evaluating content trustworthiness when saving required too many interface clicks. Also, the evaluation scale was seen as too detailed. Journalistic work, especially during emergency assignments, would not allow for such time-consuming activities. The location feature, visualizing saved content on a map, made participants unsure about where the location information came from. In both features, participants also wanted more clearly visible information about evaluation/location data sources and timestamps or similar facts about content history.

Roughly two thirds of the sample struggled with noticing or following which pieces of content in the story/feed tab they had saved to the ongoing story/saved tab. Some participants thought they had deleted rather than saved the chosen content, while others lost track of which content they were about to save. Observations showed that these problems were caused by unclear visualizations of selected content and inconsistencies in the navigational schemes between the feed and saved tabs.

To conclude, the main needs for improvement in the second usability test focused on detailed usability issues, rather than main prototype functionality. Thus, the goals of the tool development process between tests had been reached. However, several areas were still in need of improvement, mainly fine-tuning features related to navigational schemes, manual input (e.g. adding information or moving content around) or visualizations of saved content (e.g. how content location or evaluation was presented).

CONCLUSION

The datafication aspect of social media ecosystems creates opportunities for actors to collect user-generated content and use it for varying purposes (van Dijk & Poell, 2013; Pirolli & Russell, 2011). Thus, emergency communicators can benefit from implementing information gathering systems that support them with handling the vast amounts of social media data. However, in high-stress situations, it is crucial that such systems work well and do not become an obstacle for users (Stieglitz et al., 2018). This study contributes to the field by introducing a new tool intended for professional social media information gathering during emergencies. Moreover, the study identifies factors that are especially important for promoting a user-friendly design and illustrates how these factors can be investigated in a laboratory setting.

In the study, two usability tests of tool prototypes were conducted in a laboratory setting with one intended user group, news journalists. Throughout the tests, two research questions guided the process. The first question was stated as "How should a social media information gathering tool be designed to correspond to news journalists' mental models about how to carry out information gathering tasks in general, and during emergencies in particular?". Journalists work on a daily basis with newsgathering tasks that include monitoring social media and the web to find newsworthy content and to check the trustworthiness of potentially important information (Silverman, 2014). In an emergency, journalists carry out similar tasks, but with an increased work pace and more social media information to evaluate (Brandtzæg et al., 2016; Schifferes et al., 2014). The study results showed that to reflect journalists' expectations, a tool should gather social media content across platforms into one single workspace to save valuable time and allow for a large degree of user choices when, for example, searching for content. Most test participants saw the tool prototype as useful for their daily as well as emergency work, once the problems identified in the study are sorted out.

One problematic feature, in relation to participants' mental models of how systems could be used in emergency work, was to manually select and save interesting posts. A majority struggled with understanding the action to "move" posts from the ongoing story/feed to the saved tab, as this feature did not fit well with their existing information gathering workflow. Thus, creating an understanding of the user groups' workflow in an emergency, by mapping user needs or usage of existing systems, are crucial strategies for identifying where specific features should be placed in the system (ISO 9241, 2010; Wallach & Scholz, 2012).

Furthermore, when including a feature with a different work logic, such as the task feature in Prototype 1 that switched user actions from information gathering to distributing/communicating work assignments, additional user effort and cognitive capacity are needed. Designers need to carefully consider the necessity of such features.

The evaluation feature of saved content was a much-debated function in the tests. Users wanted such a function, as they are expected to manually evaluate gathered information, but they also saw challenges. The usefulness of a scale for evaluation is limited due to individual differences in scoring behavior, and the criteria for different scores should therefore be clearly understood by all users. There is a need for more research on best practices for how to implement data verification processes into information support systems.

The second research question focused on the contextual requirements of an emergency and was stated as "How should a social media information gathering tool be designed to enhance emergency situation awareness for news journalists?". In an emergency, the social media content flow can be overwhelming, especially when information from several platforms is channeled into one feed. Emergency work also sets strict time constraints and demands rapid responses from crisis managers and media organizations (Boin et al., 2005; Coombs, 2015). The study test samples reacted negatively to the vast number of pop-up windows and alternatives in the prototype, since these may cause confusion and thus affect all three SA levels (perception, comprehension, and projection) negatively, and since these features require manual user input. On the other hand, users wanted more information,

for instance about content location time stamps and history. Therefore, the number of visual features in an emergency system, as well as the manual work effort required by each included task, should be carefully balanced and kept to a minimum where possible (Backholm et al., 2017; Endsley, 2009).

This was perhaps most emphasized in relation to the evaluation feature of saved content. In a stressful situation, the task of evaluating the importance and trustworthiness of specific posts can require considerable cognitive attention and take too much time. Study participants saw problems with effort used to comprehend each piece of information and to project what the evaluation decision means for the coming work (Endsley & Jones, 2004). To simplify this, another way of evaluating content would be to base it on sources only, rather than on specific posts (e.g. "what do my trusted sources say in their posts about a specific topic?").

Users also saw that features supporting teamwork, such as a communication option or logs of previous actions taken by other users in the system (Salmon et al., 2008) would be helpful for their workflow and SA, if factors disturbing the usability in the current prototype could be improved. Participants wanted additional features related to SA perception and comprehension (Endsley & Jones, 2004), such as a clearly visible log of other users' actions, indicating who has done what, and when. Thus, consistent navigational schemes and easily understandable icons that clarify and visualize what is going on are central in a tool intended for work in high-stress situations.

Some general challenges related to research on and implementation of social media information handling systems are worth highlighting. To promote emergency SA and efficient tool usage, users and organizations need a basic understanding of what happens both inside the tool and in social media in general. The first part is about having an insight into how the tool gathering algorithms work, and how they can be manipulated by the user (e.g. "what happens in the tool when I ask it to conduct this search?"; Schwartz et al., 2015). The second part has to do with the nature and limitations related to each social media platform. For instance, who uses it, how is information posted on the platform, or who has access to the information. For example, some social media are private and only accessible to certain people, which may make team sharing of content problematic. Furthermore, social media platform limitations and preferences change over time, and companies such as Facebook are increasing regulations regarding how researchers and third-party applications can access their data.

In order to achieve good usability and user understanding of a tool that allows for several possible user modes, clearly specific organizational and individual user strategies about the aims with the system are needed. For instance, an organizational strategy for the tool tested in this study could state that "we use story mode to create a systematic overview about how our organizational response to the emergency is being handled in social media, and in order to identify hot topics that we need to react to in our own social media communication". The content of such strategies should vary from organization to organization, as organizations have varying goals with collecting social media information (Coombs, 2015; Hughes et al., 2014; Ludwig et al., 2015). The individual user strategy, then, is more about how to apply organizational guidelines to the personal workflow.

Some limitations need to be considered when interpreting the study results. As the tool prototype was tested exclusively on journalists, thus, the interpretability of the results across user groups is limited. The tool applicability should be investigated in other user groups as well (ISO 9241, 2010), and we intend to carry out additional usability tests with other crisis communicators. Furthermore, tests were conducted in a lab setting, and should ideally be combined with tests in real-life scenarios and settings (Schwartz et al., 2015). Further tests in office settings are needed, to investigate how the tool is implemented individually and within teamwork, and with existing tools and systems. And finally, the results from Test 2 should be interpreted in light of the fact that the same journalists participated in both tests. In the second test, journalists naturally had a basic understanding of the prototype, which would not have been the case if introducing a new sample. All limitations are due to the practical restraints and choices made during the test process.

To conclude this discussion, our suggestions of factors that should be carefully considered when designing a user-friendly social media information gathering tool for emergencies include taking the following general levels into account:

1. How should this system support the intended user groups, and which features already exist in other implemented systems?
2. Which features are the intended user groups ready to trust a tool to handle (e.g. information gathering procedures), and which need to remain as manual tasks (e.g. evaluating content quality)?
3. How should any included features balance between automated tasks and manual input, in order to best respond to workflow requirements and without affecting user cognitive attention and overall emergency SA negatively?
4. Which constraints are set by the situations where the tool will be used (system and contextual requirements in emergencies compared to daily work)?
5. Which constraints are set by the social media platforms or other channels that content for the tool is taken from?

ACKNOWLEDGMENT

This research was supported by the Norwegian Research Council [grant number 233975/H20]; the Högskolestiftelsen i Österbotten Foundation; and the Svensk-Österbottniska Samfundet Foundation.

REFERENCES

Backholm, K., Ausserhofer, J., Frey, E., Grøndahl Larsen, A., Hornmoen, H., Högväg, J., & Reimerth, G. (2017). Crises, rumours and reposts: Journalists' social media content gathering and verification practices in breaking news situations. *Media and Communication*, *5*(2), 67–76. doi:10.17645/mac.v5i2.878

Boin, A., 't Hart, P., Stern, E., & Sundelius, B. (2005). *The politics of crisis management. Public leadership under pressure*. New York: Cambridge University Press. doi:10.1017/CBO9780511490880

Brandtzæg, P., Luders, M., Spangenberg, J., Rath-Wiggins, L., & Folstad, A. (2016). Emerging journalistic verification practices concerning social media. *Journalism Practice*, *10*(3), 323–342. doi:10.1080/17512786.2015.1020331

Coombs, W. T. (2015). *Ongoing crisis communication*. Thousand Oaks, CA: Sage.

Diakopoulous, N., De Choudhury, M., & Naaman, M. (2012). Finding and assessing social media information sources in the context of journalism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Austin, TX (pp 2451–2460). ACM. doi:10.1145/2207676.2208409

Edland, A., & Svenson, O. (1993). Judgement and decision making under time pressure. In O. Svenson & A. J. Maule (Eds.), *Time pressure and stress in human judgement and decision making* (pp. 27–40). New York, NY: Springer. doi:10.1007/978-1-4757-6846-6_2

Endsley, M. R. (2009). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situational awareness analysis and measurement* (pp. 3–29). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. R., & Jones, D. G. (2004). *Designing for situation awareness. An approach to user-centered design* (2nd ed.). Boca Raton, FL: CRC Press.

Eriksson, M. (2012). On-line strategic crisis communication: In search of a descriptive model approach. *International Journal of Strategic Communication*, *6*(4), 309–327. doi:10.1080/1553118X.2012.711403

Hassenzahl, M. (2006). Hedonic, emotional and experiential perspectives on product quality. In C. Ghaoui (Ed.), *Encyclopedia of human computer interaction* (pp. 266–272). Hershey, PA: IGI Global. doi:10.4018/978-1-59140-562-7.ch042

Hertzum, M., & Clemmensen, T. (2012). How do usability professionals construe usability? *International Journal of Human-Computer Studies*, *70*(1), 26–42. doi:10.1016/j.ijhcs.2011.08.001

Hughes, A. L., St. Denis, L. A., Palen, L., & Anderson, K. M. (2014). Online public communications by police & fire services during the 2012 Hurricane Sandy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing System* (pp. 1505–1514). Toronto: ACM. doi:10.1145/2556288.2557227

Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. *ACM Computing Surveys*, *47*(4), 67. doi:10.1145/2771588

International Standard Organization. (2010). ISO 9241. (2010). *Ergonomics of human-system interaction—Part 210: Human-centered design for interactive systems.* Geneva: International Standard Organization.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge: Cambridge University Press.

Latonero, M., & Shklovski, I. (2011). Emergency management, Twitter, and social media evangelism. *International Journal of Information Systems for Crisis Response and Management*, *3*(4), 1–16. doi:10.4018/jiscrm.2011100101

Lindholm, J., Högväg, J., & Backholm, K. (2018). *Using biometric and affective data to investigate the usability of an information validation tool for crisis situations.* Manuscript in preparation.

Ludwig, T., Reuter, C., Siebigteroth, T., & Pipek, V. (2015). CrowdMonitor: Mobile crowd sensing for assessing physical and digital activities of citizens during emergencies. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 4083–4092). Seoul: ACM. doi:10.1145/2702123.2702265

Mendl, M. (1999). Performing under pressure: Stress and cognitive function. *Applied Animal Behaviour Science*, *65*(3), 221–244. doi:10.1016/S0168-1591(99)00088-X

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis. A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.

Pirolli, P., & Russell, D. M. (2011). Introduction to this special issue on sensemaking. *Human-Computer Interaction*, *26*(1−2), 1–8. doi:10.1080/07370024.2011.556557

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*(3), 349–363. doi:10.1037/0033-2909.100.3.349

Ruggiero, A., & Vos, M. (2014). Social media monitoring for crisis communication: Process, methods and trends in the scientific literature. *Online Journal of Communication and Media Technologies, 4*(1), 105–130. Retrieved from http://www.ojcmt.net/articles/41/415.pdf

Rutherford, A., & Wilson, J. R. (1991). Models of mental models: An ergonomist – psychologist dialogue. In M. J. Tauber & D. Auckermann (Eds.), *Mental models and human-computer interaction* (pp. 39–59). Amsterdam, Netherlands: Elsevier. doi:10.1016/B978-0-444-88602-6.50006-2

Salmon, P. M., Stanton, N. A., Walker, G. H., Baber, C., Jenkins, D. P., McMaster, R., & Young, M. S. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science*, *9*(4), 297–323. doi:10.1080/14639220701561775

Schifferes, S., Newman, N., Thurman, N., Corney, D., Goker, A. S., & Martin, C. (2014). Identifying and verifying news through social media: Developing a user-centred tool for professional journalists. *Digital Journalism*, *2*(3), 406–418. doi:10.1080/21670811.2014.892747

Schwartz, R., Naaman, M., & Teodoro, R. (2015). *Editorial algorithms: Using social media to discover and report local news.* Paper presented at the Ninth International AAAI Conference on Web and Social Media, Oxford, UK. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10593/10523

Silverman, C. (Ed.). (2014). *Verification handbook. A definitive guide to verifying digital content for emergency coverage*. Maastricht: European Journalism Centre. Retrieved from http://verificationhandbook.com/

Simon, T., Goldberg, A., & Adini, B. (2015). Socializing in emergencies – a review of the use of social media in emergency situations. *International Journal of Information Management*, *35*(5), 609–619. doi:10.1016/j.ijinfomgt.2015.07.001

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, *39*, 156–168. doi:10.1016/j.ijinfomgt.2017.12.002

Van Den Haak, M. J., De Jong, M. D. T., & Schellens, P. J. (2003). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*, *22*(5), 339–351. doi:10.1080/0044929031000

van Dijk, J., & Poell, T. (2013). Understanding social media logic. *Media and Communication, (1)*, 2-14. doi:10.17645/mac.v1i1.70

Wallach, D., & Scholz, S. C. (2012). User-centered design: Why and how to put users first in software development. In A. Maedche, A. Botzenhardt, & L. Neer (Eds.), *Software for people* (pp. 11–38). Berlin: Springer. doi:10.1007/978-3-642-31371-4_2

World Health Organization. (n.d.). *Humanitarian Health Action. Definitions: Emergencies.* Retrieved from http://www.who.int/hac/about/definitions/en/

Yin, J., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2012). Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, *27*(6), 52–59. doi:10.1109/MIS.2012.6

ENDNOTES

[1] In the article, emergency is used as an umbrella term for a state of collective disruptancy, caused for instance by a natural disaster or man-made violence that is usually of a sudden and unexpected nature.

An emergency is limited to a specific area and time period, and requires extraordinary response actions (Boin et al., 2005; World Health Organization, n.d.).

2 The prototype concepts "monitor" and "story" are also relevant for other groups of crisis communicators. For such organizations, the story concept should be understood as a space where you are working with gathering information from an identified ongoing emergency.

3 Conceptual frameworks, as described in the transcendental realism approach (Miles, Huberman, & Saldana, 2014) were used to analyze data.

4 Due to missing data, the AttrakDiff results from two participants were not included in the analysis.

5 For instance, Van Den Haak, De Jong, and Schellens (2003) have shown that concurrent think aloud procedures may affect completion of complex tasks negatively. Furthermore, participants' biometric data (skin conductance, facial responses, and eye tracking) was also collected during the test, and concurrent think aloud procedures may affect biometric data quality. Detailed results regarding biometric data are beyond the scope of this study, and have been reported in detail elsewhere (Lindholm, Högväg, & Backholm, 2018).

6 The content consisted of written and picture content, posted by fake eye witnesses, the public, authorities, and media organizations: "Something is happening here but what? @ [Regional] Airport (fake Instagram update by citizen at airport, illustrated with picture of black smoke at airport runway). Content was created by the authors and tested in a pilot study; in total 88 posts, 31 Facebook, 52 Twitter, and 5 Instagram updates.

APPENDIX 1

Short article summary: The article investigates how to promote situation awareness in emergencies through the help of a technical innovation. The study introduces a tool that will support journalists and crisis managers with gathering information on social media. Two laboratory usability tests were conducted to identify if the user's mental models fit the product design. Results show that such a system needs to gather information from several social media channels and allow for varying, personalized modes. Users understood main tool concepts, but some features, such as saving content or manually evaluating information quality, were too time consuming. The study contributes to the field by identifying factors important for promoting a user-friendly design in emergencies. For instance, to avoid affecting user cognitive attention, the design needs to carefully balance between automated tasks and manual input.

Klas Backholm is Coordinator of the Mass Communication Programme at Political Science with Mass Communication, Åbo Akademi University (Finland). His research areas include (a) usability-testing of technical innovations for journalism, and (b) the psychological wellbeing of journalists after work-related crisis exposure.

Joachim Högväg works at Experience Lab, Åbo Akademi University, Finland. Högväg does research in experimental psychology, cognitive psychology and cognitive science.

Jenny Lindholm holds a PhD in political science and works as a lecturer and researcher at Åbo Akademi University (Finland). Her research interest is how crises and disasters affect society, as well as experimental and laboratory research.

Jørn Knutsen is Interaction Designer and Associate Professor at the Oslo School of Architecture and Design, Norway (Arkitektur og -designhøgskolen i Oslo).

Even Westvang is one of the founders of Sanity.io (Oslo, Norway), specializing in concept development, user interface design and data visualization.