Social Haystack: Dynamic Quality Assessment of Citizen-Generated Content during Emergencies

THOMAS LUDWIG, CHRISTIAN REUTER, and VOLKMAR PIPEK,
Institute for Information Systems, University of Siegen

People all over the world are regularly affected by disasters and emergencies. Besides official emergency services, ordinary citizens are getting increasingly involved in crisis response work. They are usually present on-site at the place of incident and use social media to share information about the event. For emergency services, the large amount of citizen-generated content in social media, however, means that finding high-quality information is similar to "finding a needle in a haystack". This article presents an approach to how a dynamic and subjective quality assessment of citizen-generated content could support the work of emergency services. First, we present results of our empirical study concerning the usage of citizen-generated content by emergency services. Based on our literature review and empirical study, we derive design guidelines and describe a concept for dynamic quality measurement that is implemented as a service-oriented web-application "Social Haystack." Finally, we outline findings of its evaluation and implications thereof.

Categories and Subject Descriptors: H.5.2 [**User Interfaces**]: User-Centered Design

General Terms: Design

Additional Key Words and Phrases: Social media, information quality, quality assessment, emergency management, citizen-generated content, design case study

ACM Reference Format:
Thomas Ludwig, Christian Reuter, and Volkmar Pipek. 2015. Social haystack: Dynamic quality assessment of citizen-generated content during emergencies. ACM Trans. Comput.-Hum. Interact. 22, 4, Article 17 (June 2015), 27 pages.
DOI: http://dx.doi.org/10.1145/2749461

1. INTRODUCTION

Human Computer Interaction (HCI) has been an evolving field for a number of years and has seen the progressive development of its interests from its original concerns with the usability of computer interfaces towards a more generic set of issues grounded in the recognition that computer use is more likely to be group- or team-oriented than previously; that "usefulness" as opposed to usability is grounded in a range of social and organizational features, that the social media constitute a radically new problem set, and finally that new methods and new concepts may be necessary to understand these complex issues more clearly. Allied to this has been a developing recognition that local conditions may substantially affect the way in which work is done and thus the kind of IT support which may turn out to be valuable [Normark and Randall 2005]. This change of emphasis has included, for instance, ethnographic

The study was conducted within the research project EmerGent, which was funded by a grant of the European Union (FP7 No. 608352).
Authors' addresses: T. Ludwig, C. Reuter, and V. Pipek, Institute for Information Systems, University of Siegen; emails: {thomas.ludwig, christian.reuter, volkmar.pipek}@uni-siegen.de.

DOI: http://dx.doi.org/10.1145/2749461

approaches to the study of complex, real-world organizational settings and a focus on the notion of "practice" as it relates to the management of heterogeneous work tasks in local real-world settings [Schmidt 2014]. One central interest over time which reflects this gradual shift has been the use of information, incorporating such matters as information history, information relevance, personal information use, information visualization and so on [e.g., Card et al. 1991; Wexelblat and Maes 1999]. In the following, we examine some themes which exemplify the "turn to the social" and which underpin some, at least, of this interest in the usefulness of information, and which can be understood through careful examination of cases grounded in real-world problems. Specifically, we examine information use in and through the social media by groups of people in the context of disaster management.

In November 2013, typhoon Haiyan killed approximately 10,000 people and forced nearly one million people to leave their homes. Floods in the affected regions rendered rapid assistance from abroad almost impossible. One year earlier, in October 2012, hurricane Sandy turned New York into a disaster area. Approximately 18 months earlier, the 2011 tsunami in Japan, triggered by the Tohoku Earthquake, cost nearly 16,000 lives (not taking into account the long-term damages caused by the nuclear disaster in Fukushima) and half a million people had to be accommodated in temporary shelters. These large-scale emergencies of the recent years show that people all over the world are regularly affected by such events. It has been argued, in our view correctly, that in the context of emergency management, information is the most valuable resource for coordination and response work [Marino et al. 2012]. The uncertain character of emergencies, nevertheless, challenges the rapid provision of information for all organizations involved [Turoff et al. 2009].

It has also long been known that "information quality" is of paramount importance within emergency services work. Moreover, research has established the need for efficient and effective information and communication technologies (ICT), which enable the assessment and communication of accurate and dynamic information in highly emergent situations [see, e.g., Bharosa et al. 2008; Ley et al. 2014]. Turoff et al. [2004], for instance, draw requirements for emergency management systems, which include "establishing and supporting confidence in a decision by supplying the best possible up-to-date information". In addition to classical emergency services such as police, fire departments, civil defense, medical services, operators of critical infrastructures, local citizens and volunteers are increasingly becoming involved in current response practices, primarily through social media [Homeland Security 2013; Jennex 2012; Reuter et al. 2013]. Social media is defined as a "group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content" [Kaplan and Haenlein 2010].

In preparation for Hurricane Sandy in 2012, the US emergency relief organization Federal Emergency Management Agency (FEMA) requested that people should avoid calling or texting with phones, but instead should post their status on Facebook. Shortly after the hurricane abated, the status "We are ok!" was the most-posted Facebook status. During the 2013 European floods, 157 Facebook groups were created in Germany with the keyword "Flut" (English: Flood) in which citizens organized and coordinated private relief efforts [Reuter et al. 2015]. The "Person Finder" was set up by Google in response to the 2010 Haiti earthquake and has been used to bring searchers and missing people together [Yates and Paquette 2011]. Such examples show that important citizen-generated content on social media is already available and its amount is set to increase in the future. Finding appropriate high-quality information in social media is, however, still challenging [Homeland Security 2013]. It is also, as yet, unclear whether local and national circumstances will impact on the facility with which these problems can be managed. There is also an obvious need for some comparison

of circumstances to judge the degree to which possible solutions can be generically applied. Our study examines how can emergency services be supported in extracting the relevant citizen-generated information from the entire scope of social media in the organizational contexts we describe?

After presenting the related work (Section 2) and our overall research approach (Section 3), we will outline the results of our qualitative empirical study, in which we explored the emergency services' attitude towards citizen-generated content, the use of social media in their current work practices and its perceived quality (Section 4). From this prestudy, we uncovered design challenges and derived technical implications that focus on a subjective (in accordance with the individual user's needs) and dynamic (customizable in real time) assessment of information. The filtering of social media content should, we suggest, stimulate a high-quality and appropriate information basis for sense-making and crisis response activities (Section 5). We introduce the resulting prototype of a web-based application: "Social Haystack" is expected to provide the relevant high-quality information support (Section 6). Results of its evaluation will be presented in Section 7. Finally, we will discuss our findings and draw relevant conclusions and design guidelines for assessing and filtering citizen-generated content from social media during emergencies (Section 8).

2. RELATED WORK ON INFORMATION QUALITY OF CITIZEN-GENERATED CONTENT

Understanding the need for cooperation between official emergency services, the public as affected citizens and volunteers has become a vibrant concern in the research fields of CSCW and HCI. It is particularly interesting as it stands at a juncture of several different research interests, including how information is used by large, heterogeneous populations, how responses to dynamic, emergent crisis situations are managed, what the role of the social media might be, and what kind of interface can mediate and support effective and efficient responses. In this section, we review the related work in these areas and identify a research gap in the current state-of-the-art.

2.1. Information Quality in Crisis Management

The terms, "disaster, crisis, catastrophe, and emergency management are sometimes used synonymously and sometimes with slight differences, by scholars and practitioners" [Hiltz et al. 2011]. Regardless, within the field of crisis management, as we will call it, information is always the basis for decisions [Bharosa et al. 2008]. If decisions are based on low-quality information, the probability of making bad decisions increases immediately [Fisher and Kingma 2001], and equally the need for high-quality information rises in accordance with the importance of the decision and action outcomes [Friberg et al. 2011]. As a consequence for crisis management, supplying high-quality information to decision makers is a critical factor for success of actions and the overall emergency response work [Bharosa et al. 2008]. According to Wang and Strong [1996], information quality can be defined as the extent to which information can be used by those who apply it. It must correspond to the expectations and requirements of the respective user [Fisher et al. 2011] and must be of high value to them [Eppler 2006] depending on their context [Christofzik and Reuter 2013]. It represents the applicability of information for effective and efficient decision-making in different situations [van de Walle and Turoff 2008]. Therefore, information quality has to be viewed subjectively [Ludwig et al. 2013].

The requirements that information should fulfill are already addressed by Eppler [2006] as well as by Wang and Strong [1996]. On the basis of empirical studies, the latter determined 15 different dimensions of information quality clustered into the categories *intrinsic data quality* (credibility, accuracy, objectivity, reputation), *contextual data quality* (value-added, relevancy, up-to-datedness, completeness,

appropriate amount of data), *representational data quality* (interpretability, ease of understanding, representational consistency, concise representation), and *accessible data quality* (accessibility, access security). They claim that the existence of universal criteria for information quality can never replace the contextual weighting of the requirements as well as subjective evaluation of quality by the respective information user.

When emergency situations arise, the demand for precise, valid, up-to-date and reliable information is obvious [Friberg et al. 2011]. Having said that, decision makers must find a compromise between the need for urgency and the need for accuracy. Good decision-making needs to be both timely and accurate and therefore depends on the quality of available information [Shamoug et al. 2012]. The dynamics of emergencies can lead to changes in information needs [Longstaff 2005] as well as information "offers" [Bharosa et al. 2009]. It is necessary to constantly update and re-evaluate the decision makers' overview of a situation as well as measures taken and strategies chosen. In this highly dynamic environment, there is a real danger that the existing information basis does not reflect the actual unfolding situation [Bharosa et al. 2008]. To cope with the (possibly rapidly changing) dynamics of a crisis, the information base must always be up-to-date [Chen et al. 2008]. Nontransparency can only be overcome by clear, objective, and reliable context information, as information overload from diverse sources can lead to confusion [Friberg et al. 2010]. O'Reilly [1980] studied how the amount of information affects the overall quality of the decisions made. He found, surprisingly, that participants who thought they were overloaded with information were happier with the information they received than participants who thought they were receiving inadequate information but the decision-making was better in the latter group. Both "lack of information" and "information overload" have negative effects, but: *"what one perceives as information overload, may be perfectly manageable to the other"* [Mulder et al. 2006], reinforcing the view that information quality is characterized by a high degree of subjectivity. In crisis situations, the complete fulfillment of all requirements is unlikely to be entirely achievable because accurate and consistent information is rare [Palen et al. 2011]. The paramount need, therefore, is to provide a means of information coordination which allows both some kind of overview and information which is up-to-date enough for a user's context.

2.2. Citizen-Generated Information in Crisis Management

Individuals and groups getting together to form emergent and temporary organizations for improvised relief and rescue activities is hardly a new phenomenon [Stallings and Quarantelli 1985; Wachtendorf and Kendra 2006]. However, the emergence of social media has arguably changed the type of involvement we now see [Palen and Liu 2007]. Citizens now use social media not only to receive information from authorities, but also to communicate with family and friends as well as to coordinate relief activities with each other [Reuter et al. 2012]. In these situations, the search for information constitutes a kind of peer-to-peer collective behavior [Starbird et al. 2012], in which citizens use both official and unofficial sources to obtain an up-to-date overview of the situation they are concerned with [Qu et al. 2009].

Social media have become immensely valuable, as sources of information, as a result of its public availability. Whether intentional or not, information can be communicated to a large population of friends, volunteers or authorities [Reuter et al. 2015]. It therefore constitutes a ready-made knowledge base, albeit one which lacks any formal organization [Ludwig et al. 2015]. Emergency services already rely on such information: Interviews with representatives of large international disaster response organizations show that "emergency responders already operate with less than reliable, or "good enough," information in offline practice, and that social media data is useful to

responders, but only in specific crisis situations"; however, this data is mostly typically shared "only within their known community and extended network" [Tapia and Moore 2014]. According to Starbird et al. [2010], in their analysis of Twitter use, the most valuable are "generative tweets" which describe the "facts on the ground". Such tweets usually have an autobiographic and narrative character and are typically published by locally based private citizens. Locally affected citizens can, then, be a valuable source of information for the emergency services because they provide unique and up-to-date contextual information. Moreover, they possess knowledge of local geographic or cultural features which are relevant to official crisis response but which otherwise might go unrecognized [Starbird et al. 2012]. During extreme situations, however, citizens publish more material related to threats compared to other events [Gaspar et al. 2014] and also can be indiscriminate in what they report, often communicating information that is only relevant to a specific set of interests, or is about "less important" crisis related topics.

Besides providing information, citizens aggregate and validate it at the same time [Palen and Liu 2007]. To support this process Purohit et al. [2014] suggest "a supportive technology that recognizes the existing capabilities of the informal response community to identify needs (seeker behavior) and provide resources (supplier behavior)". Integrating citizen-generated information might not only enable access to a new information source for contextual information for sense making [Wu et al. 2013] and map-based decision-making [Convertino et al. 2011], but also can mean that information available through crowdsourcing mechanisms is more thoroughly distributed across a wider population and is more timely. Nevertheless, as with all information resources, some selection and sorting is necessary for it to be maximally useful. One existing concept is the selection and assessment of social media data by the social media users themselves. An example is "VOST" (Virtual Operations Support Teams) in the US (http://vosg.us/history/). These teams are established to make use of new communication technologies and social media tools so that trusted agents can lend support via the Internet to those on-site, who may otherwise be overwhelmed by the volume of data generated online during a disaster.

2.3. The Quality of Citizen-Generated Information

Approaches for filtering and ranking techniques in the searching of social media content already exist. Shankaranarayanan et al. [2012] mention that "[some] quality dimensions are still applicable to social media data, others are not due to the nature of social media data. It also led us to conclude that some dimensions such as believability and relevance will gain in stature as important quality dimensions for social media data." Agichtein et al. [2008] present a graph-based model of contributor relationships and combine it with content- and usage- based features to exploit community feedback (such as links; explicit quality ratings from members) to automatically identify high-quality content in "Yahoo! Answers". Answer length, the number of "thumbs up," and the non-stop word overlap with the questions were also identified as indicators. One particular feature—the answer length—dominates other features. However, with Tweets, length is rather limited. Uysal and Croft [2011] therefore present a personalized tweet ranking method, leveraging the use of retweet behavior, to bring more important tweets forward based on author, tweet, content or user.

Here, the main challenge is to filter the crisis-relevant information out of the total amount of citizen-generated information (first main challenge—C1) and to evaluate the quality of this pre-filtered information suitably, in a way that a subjective ranking of filtered information can be achieved (second main challenge—C2). A number of public, scientific or commercial applications have been developed for coping with citizen-generated content from social media for emergency management, for

Table I. Related Approaches and Systems

Approach	C1: filter crisis-relevant information	C2: evaluate the quality of prefiltered information with subjective ranking
Brandwatch (http://www.brandwatch.com)	Based on customizable search and filtering queries	Not addressed (algorithmic measure of sentiment, influence, customizable dashboard, manual categorization capabilities)
ESA [Yin et al. 2012]	Based on customizable search and filtering queries; Twitter only	Not addressed (static algorithms of burst detection, text classification, topic clustering, geotagging)
Geofeedia (http://geofeedia.com)	Location-based information filtering, further refinement by keyword, user, day, hour	Not addressed (quantitative analytics and visualizations)
TweetDeck [Twitter 2014]	Based on simple search queries (keyword or hashtag), refinement by quantitative filtering criteria; Twitter only	Not addressed (based on search queries)
TwitInfo [Marcus et al. 2011]	Based on simple search queries (keyword or hashtag), map and timeline filtering; Twitter only	Not addressed (static measure of sentiment)
SensePlace2 [Robinson et al. 2013]	Based on simple search queries (keyword or hashtag), map and timeline filtering; Twitter only	Not addressed (place links, tag cloud, tree view of extracted location entities)
Hootsuite (https://hootsuite.com/)	Based on customizable search and filtering queries	Not addressed (algorithmic measures of sentiment, demographics, brand growth, influence)
Sproutsocial (http://sproutsocial.com/)	Based on customizable search and filtering queries	Not addressed (influence, barely customizable, rather static and quantitative)
Mibazaar [Liu and Palen 2010]	Based on hashtags	Not addressed
Tweak the Tweet [Starbird and Stamberger 2010]	Based on a specific syntax	Not addressed (display of all messages based on the hashtag)
Ushahidi [Okolloh 2008]	Based on the crowd	Based on the crowd, not individually
Twitcident [Abel et al. 2012b]	Based on customizable search and filtering queries	Not addressed (customizable queries, but no adjustable ranking)

example, Brandwatch, ESA, Geofeedia, TweetDeck, TwitInfo, SensePlace2, Hootsuite, and Sproutsocial (Table I).

Some specific relevant tools that try to address the mentioned main challenges are Tweak the Tweet, Ushahidi and Twitcident. All provide valuable functionality but, as we will argue, none provides—as yet—approaches to the problem of the subjective and dynamic character of information assessment, filtering and coordination by differently situated users. Introducing *Tweak the Tweet*, Starbird and Stamberger [2010] suggest a prescriptive tool based on hashtags to increase the value of Twitter as an information source. The syntax recommends the user to phrase the tweet in a machine-readable structure and enables the rapid detection of those tweets which are relevant to the users in their current context. C1 is met. Nevertheless, the results of the search request may exceed the capability of the user, who is often under enormous time pressure. C2 is not explicitly addressed by the tool design. The assessment of quality is left up to the user, as is detecting the qualitatively most valuable information. The open source platform *Ushahidi* allows crisis-relevant information from various social media to be captured, aggregated, and visualized. Ushahidi supports quality evaluation undertaken by the

"crowd" (C1). However, there is no differentiation into separate quality characteristics (C2). Compared to the approaches mentioned previously, the added-value of *Twitcident* is provided here by the semantic enrichment of the tweets by so-called Named Entity Recognizers (NER), on which the semantic filter and the faceted search are based. By evaluating the efficiency of the semantic filter and faceted search, Abel et al. [2012b] show that semantic filtering yields more than twice as many accurate results as the keyword-based search. This approach focuses on C1; C2, however, is disregarded. The result set is sorted according to a statically implemented quality algorithm which does not allow any adaptation of the situational, subjective information needs of the user: "The ranking of the tweets that match a query is a research problem of its own" [Abel et al. 2012a].

To summarize: Although promising, the existing approaches for integrating citizen-generated information into official emergency management currently do not support dynamic quality assessment during emergencies. Especially the subjective quality evaluation of prefiltered data with regard to the individual actor is still challenging and not addressed. Also other tools, as mentioned in Table I, do not address C2 appropriately.

3. RESEARCH APPROACH

Our objective, then, is to examine the potential of citizen-generated content from social media as a useful information resource in crisis response work by addressing two research questions:

(1) How can emergency services be supported in extracting the relevant citizen-generated information from the entire corpus of social media content?
(2) How can a dynamic, context-dependent quality assessment of citizen-generated content from social media be supported by IT?

Our argument is that we first need to understand, as best we can, the situations in which information gathering, situation assessment, and decision-making practices of all relevant stakeholders involved in emergency management take place. This is a non-trivial problem since—by definition—we cannot predict the occurrence of emergencies. We use the *design case study* [Wulf et al. 2011] approach, which consists of the three phases (1) empirical study of current work practices in the field, (2) the development of innovative ICT artifacts related to the empirical findings, and (3) the evaluation of their usability and appropriation in practice. This research framework is inspired by Lewin's *action research* as "comparative research on the conditions and effects of various forms of social action and research leading to social action" that uses "a spiral of steps, each of which is composed of a circle of planning, action, and fact-finding about the result of the action" [Lewin 1958]. This, as has been argued, is a methodology appropriate for research in HCI [Hayes 2011]. More specifically, because of the fact that our work is based on observational practices that can be thought of as ethnographic, the approach can be characterized as "ethnographic action research" [Hearn et al. 2008]. In our case "planning" is equivalent to the empirical study of the given practices, "action" focuses on the design and use of suggested ICT artifacts, and the evaluation leads towards "fact-finding about the results of the action". Such an interpretation is also consistent with Hevner and Chatterjee's suggestion [2010] to integrate action research with design research. A *design science approach*, they suggest, is comprised of the design of an artifact for a relevant problem combined with rigorous evaluation methods for the design [Hevner et al. 2004].

Our entire study was conducted in two counties in Germany in the years 2010–2014. County A is densely wooded, hilly, and rural, whereas county B consists of ten expanding urban communes. In both regions, we focused on several affected organizations: Public strategic administration (crisis management group, operations management),

public operative administration (e.g., police, fire department), and citizens. According to our overall research approach, we first present the findings of our empirical study into the potential of, and obstacles related to, the integration of citizen-generated content followed by the derivation of design challenges and the development of Social Haystack—a system for the subjective and dynamic assessment of citizen-generated content related to our empirical findings. Finally, we present our evaluation and the implications thereof.

4. EMPIRICAL STUDY: THE USE OF SOCIAL MEDIA FOR SITUATION ASSESSMENT IN EMERGENCIES

4.1. Methodology

The empirical studies we described were embedded in a scenario framework describing a windstorm with many smaller connected incidents and energy breakdowns. The scenario was developed together with actors from the police and fire departments, county administration, and an energy network operator. The purpose of the scenario was to be able to create a common understanding of an occurring emergency quickly and therefore it helped to increase validity and comparability in our interviews. First, we conducted observations to acquire knowledge about practical work in interorganizational crisis management. The observations were conducted in a control center during a normal working day (observation time: 9h), in the crisis management group and the operations management team during a crisis communication training (4h) as well as at a major cultural event with about 400,000 visitors (6h). Besides observations, we conducted 5 interorganizational group discussions in workshops (labeled W1–W5) to understand communication practices of interorganizational crisis management.

We further conducted 38 individual interviews (labeled I01–I22+I24; IM1–IM16) with actors from the participating organizations. During our interviews we aimed for a cross section of organizational and hierarchical units ranging from a (lower level) Head of Section to a (high level) Head of Control Center. By focusing on all management levels, our aim was a comprehensive overview of the entire organization and the work practices entailed in different roles. A Head of Section, for instance, is responsible for an assessment of his/her section; a Head of Control Center, for instance, is responsible for the assessment of an entire operation (or more operations that run in parallel). The lower the hierarchy, the lower is the responsible area of assessment. Each interview lasted between 1 and 2h and followed a guideline, which was separated into three parts. The first part focused on the participants' role, qualification and work activities under normal conditions. The second part covered the participants' tasks during emergencies in our scenario framework (covering decision making, collaboration, results, errors, stress, overview, changes). The third part covered applied information and communication systems and perceived problems with these tools (problems, unsecure information, citizens, awareness, and adaptations). Within this third part one main focus lies on the role of citizens in emergency management and the impact of citizen-generated content from social media, which covers the main part of our analysis within this article. Other parts of the empirical work, that focus on intraorganizational collaboration of emergency services as well as a detailed description of the participants were already published in Reuter et al. [2014].

Group discussions and all interviews were audio recorded and transcribed for subsequent data analysis. The analysis of the data was based on the *grounded theory* approach [Strauss 1987]. We chose this systematic methodology to discover insights about the work practices through the analysis of data. We did not approach the application field with predefined categories, but used the open coding associated with grounded theory to derive categories from empirical data by careful reading in question and

aggregation of categories that seemed "similar" through an inductive process. To be able to use this methodology, the transcripts were open coded and the agents' statements were divided first into text modules and later into categories. The knowledge previously acquired in the literature study was used to heighten *theoretical sensitivity* [Strauss 1987]. Part of the approach involves *theoretical sampling*, meaning that the contents and units to be studied are selected according to the theory which emerges during analysis. Within our data analysis we have classified the divided text modules into the seven categories (1) *technology usage*, (2) *situation illustration and construction*, (3) *information quality, quantity and trustworthiness,* (4) *communication practices,* (5) *cooperation and collaboration,* (6) *debriefing and learning*, (7) *citizen involvement*. After that, we identified the specific challenges regarding the use of social media for situation assessment during emergencies.

4.2. Results

4.2.1. Citizens and Social Media are already Part of Crisis Management Practices. It is current practice in operations of professional crisis management to consult civil experts from the general public (I10). This is due to their knowledge and routine in specific fields:

> There are special cases where you really need basic understanding or knowledge, but there are also cases where you can draw on the knowledge and skills of others because it is their 'daily bread'. (I11)

The consequences of an emergency are, in large part, unforeseeable and as such the knowledge relevant to dealing with these consequences may not be predictable. Emergency services sometimes have to fall back on citizens' knowledge and their qualifications. Citizens are therefore already involved directly in operations, at least to a degree:

> I'm always grateful to have a chimney sweep on hand when there is a chimney fire. He knows exactly what needs to be done. I'm the one who does the job, but you can't get any better information, and that's why I like to encourage other people to help. (I11)

Moreover, the Internet, with all its applications, is already an indispensable source of information for the professional services, especially in situation assessment (I14). Useful location information can often be retrieved from the web much faster than it can be provided by the professional operatives on-site (I15). Contributions from journalists, including photographic information, (I02, I04) and user-generated content from social media are already used (I15):

> Twitter is so often faster with information than police radio or cell phones. If people tweet live. So I look all the time, I always run Twitter to see what people are writing. The tweets often include pictures or videos about current events which can be found on the Internet. (I02)

In addition to their primary work practices, the units often use social media for enhancing information:

> Our crime fighters have already been very successful against the drug business. The [dealers] are sometimes that stupid – they use Twitter quite openly to arrange their meetings. (I02)

Besides Twitter, also other sources of information retrieval like Google or Facebook are used, but the main focus remains on Twitter:

> Other than tweets, I can't really say that we pick anything up continuously. Unless it comes from a Google search – that someone has got a live ticker running, for example. That's a popular one. (I02)

Our respondents consistently referred to Twitter as a source of useful information, and do so because as I02 said, it is "*often faster*". However, and this needs to be stressed, social media content is only suitable as a source of data when used extremely cautiously, due to the highly subjective and heterogeneous perceptions of events as well as the unclear legal situation in respect to handling the information from social media sources. Despite this, social media content is often used in an ad hoc manner with special notes and annotations (I14). It should also be noted that, in some cases, expectations about the accuracy and usefulness of information are low. We should clarify this is mainly a function of the point at which such information is being interrogated—fairly early in the unfolding event—when almost by definition the situation will be unclear and confusing. The participants nevertheless agree that "*new digital media are increasingly influencing the type of processing we use, and they are influencing the nature of the information collected as well*" (I02).

4.2.2. Quality of Social Media Content from the Perspective of Emergency Services. As a result of the great mass of data produced, the quality, credibility and integrity of citizen-generated content in social media are generally assessed as being relatively low (W2):

> Of 300 entries, 290 are a waste of time. You can get what you need from ten messages. But you have to read all of them. It would be very helpful if there was an automatic pre-selection (I02). Sometimes there is a huge gap between the facts understood on the phone and the actual situation on-site. (I04)

Of course, how to determine the 10 essential posts out of 300 and how to extract them is the problem:

> Who evaluates it and who looks at the whole picture? Many pictures are likely to show just a storm. [. . .] It would be nice if you had a filter to separate what is important from what is unimportant [. . .]. But unfortunately we do not have such intelligent systems. (I03)

At the moment, no technological options for the filtering or assessment of user-generated content are systematically integrated into the current practices of emergency services. It is, therefore, a matter for individual actors. As they report, given the problem of information "overload," such filtering is nontrivial:

> Each of us can only absorb a certain amount of information. When you see so many live images appearing, you want to distinguish the important from the unimportant, but how? You really need to turn on 3–4 filter settings, so you only get what is most likely the most important information. (I01)

4.2.3. Mass of Information Impedes the Measurement of Information Quality. Despite the evident fact that some of the information available from social media is potentially useful, assessment of the quality of the information alongside the mass of information available is considered to be an organizational bottleneck (W2) and judgment about information requires coordinated effort. At the moment, firefighters, police, and energy operators are overwhelmed with the volume of effort, both human and technical, demanded of them (W3). They have in the first instance to fulfill their primary tasks, such as planning immediate operations, allocating machinery and units on-site or operating the situation map. In emergencies, time and staff shortages emerge quickly, which exacerbate the situation.

The mismatch between the flood of information from social media and the capability to assess it runs through the entire emergency situation. Therefore, it is currently not feasible that "*100 teenagers take loads of photos each, and five people are expected to assess them all*" (W2):

> If there are only three of four colleagues up here and we are heading the operation, I just do not have the time to deal with external [new and unknown] stuff myself. (I02)

If emergency services are to be supported in extracting citizen-generated content from social media, a number of salient features need to be recognized. These include firstly, and unsurprisingly, that, proper assessment of the information is absolutely essential, but that it is a very demanding task (I15). The demand emanates from the fact that situations evolve fast, a number of complex tasks have to be managed at the same time, and operators experience "information overload." This is exacerbated by the fact that citizen-generated content is of variable quality and relevance. Because of the sheer speed with which it is made available, however, it remains potentially useful.

5. A CONCEPT FOR DYNAMIC QUALITY ASSESSMENT IN SOCIAL MEDIA

5.1. Challenges and Derived Design Implications for Social Media Assessment

Our empirical results show that citizens are already an important part of the overall emergency management process. Officials try to integrate information from various social media into their current work, especially for situation assessment, but the amount of information is overwhelming and therefore the respective quality of information hampers appropriate handling. Moreover, in the circumstances we describe, limited resources and legal restrictions mean that finding relevant and useful information is even more problematic. The degree to which these resource limitations affect emergency service work elsewhere remains unclear. Taken together, the wealth of information and limited investigatory resources, along with the fact that needs vary from situation to situation and from individual to individual (I02) means that disentangling the useful from the irrelevant is akin to finding a "needle in a haystack."

Actors frequently describe their struggle with the quality assessment of social media content. It is, they argue, almost impossible to specify needed information in advance, which leads to a demand for a dynamic as well as subjective information assessment of citizen-generated content. As already indicated, the existing literature indicates some of the generic information needs of the emergency services. Together with our empirical work on the specific circumstances in which such needs are locally mediated, Table II indicates the findings and main challenges for the assessment of citizen-generated content as well as the design implications we derived for the conceptual architecture of our approach.

Information quality is clearly context dependent. It differs not only from situation to situation but varies according to the particular stage of an evolving situation (*C1: context-dependency*). Coping with unexpected situations and ad hoc assessment practices often requires rapid improvisation on the basis of information of varying quality. For example, a tweet "Big fire ahead, two kids injured!" has no relevance for emergency services if it contains no details about a location or time. We therefore focus on gathering the existing metadata of social media content (like author, time, and so on) as well as using NER to structure information. Due to the subjectivity of information quality and amount, the needs of the individual user must also be addressed (*C2: subjectivity*). Therefore, instead of implementing a predefined setting for information quality, we must allow the user to dynamically specify and adjust values directly within a situation.

Table II. Literature- and PreStudy-based Design Implications

No.	Literature/Empirical Findings (related to Sections 2 and 4)	Existing Challenges for Social Media Content Integration	Design Implications
Challenge 1: context-dependency	Information quality is context-dependent [Fisher and Kingma 2001; Wang and Strong 1996] and differs from situation to situation (I02)	Gathering as much context information as possible to enrich information as much as possible	Using the existing meta data of social media information as well as Named Entity Recognizer to extract further information
Challenge 2: subjectivity	Information quality and amount is subjective [Eppler 2006; Ludwig et al. 2013] and personal capacity of processing differs from individual to individual (I01, I02)	Instead of implementing information quality and amount in advance, the user should specify it directly within a situation	Enabling the individual user to dynamically specify and adjust the amount and needed quality of social media information
Challenge 3: quantity	Although many messages are useless, emergency services must read them all (I02) and they have no options to filter what content is important (I01, I03)	Decrease the amount of information and simplify the search for relevant social media information	Enabling the individual user to filter the results of information from social media by ranking as well as sorting the extracted information from social media
Challenge 4: trust	Trustworthiness of data is often questionable [Tapia and Moore 2014]	Make calculated trustworthiness of social media information transparent to individual users	Enabling the individual user to access the trustworthiness of results by presenting percentages of quality fulfillments
Challenge 5: locations	Citizens provide knowledge about local locations [Starbird et al. 2012]; locations can be retrieved faster from the web than provided by officials on-site (I15)	Extract and integrate geographical information of social media into situation assessment practices	Enabling an automatic extraction of geographical data from social media information by using Named Entity Recognizer
Challenge 6: aggregation	Citizen aggregate and validate information while providing it [Palen et al. 2011]	Enable easy aggregation and validation of social media information from citizens	Displaying dissemination of social media content to the individual user as indications for their importance [Uysal and Croft 2011]
Challenge 7: distribution	Needed information is often distributed (I02), for example, within different social media services [Reuter et al. 2015].	Enable a cross social media search to get as more potentially important data	Implementing a search of information for different social media services

As the head of a police control center mentioned, the citizen-generated content of social media is overwhelming but, in the absence of any existing filters, the police effectively have to somehow get through all of it (*C3: quantity*). For that reason, the search for relevant social media data must be simplified in a way that the information flow is made manageable. A design implication is therefore to enable the user to filter the results of data from social media and to rank the search results according to his/her specific individual needs.

A further, though related, issue is the trustworthiness of social media information. Emergency services often struggle with the credibility of authors and the dissemination of rumors (*C4: trust*). A challenge is therefore to make kinds of trustworthiness more transparent to the individual user. As already shown, information quality is subjective and users need to be able to decide if a piece of information is trustable or not for their specific purposes. A first step towards assessing the trustworthiness is therefore to present the users with details about the results, especially the conformity of the provided result with the users' information demands, as articulated in the dialogue.

One of the obvious benefits of citizen-generated content is that it can provide knowledge about location much faster than on-site emergency service units can (*C5: locations*). Such geographical information is important for early situation assessment practices, so an automatic extraction of locations needs to be supported. We aim at gathering of geographical data with the support of NER to extract locations not only within the metadata but also through text analysis.

As Palen et al. [2011] have already shown, citizens aggregate and validate information while providing it (*C6: aggregation*). A challenge is currently how emergency services can easily comprehend the aggregated and validated citizen-generated content. As Uysal and Croft [2011] argue, retweets or likes as well as external links [Ha and Ahn 2011] can be indicators to the importance of information as people usually forward important or interesting contents. Such indicators must be presented to professionals for assessment and weighting. In addition, the information relevant to a specific situation is often distributed to different social media services (*C7: distribution*). The search for information must therefore be enabled across different social media services simultaneously to allow an overall view of available information.

To summarize the design implications, our concept must support users in searching and analyzing content from various social media sources according to their individual requirements and needs. This subjective element means that we could not automatically determine the types and sources which should be used. This decision, we felt, should be passed to the users themselves. To address the flood of information which can lead to information overload, the user needs to be able to select possible source platforms; to select a specific geographic area; and to filter the results dynamically according to individual requirements. People use the same social media in crises as in everyday life [Jennex 2012]; citizen-generated content can therefore normally be found on these media. The most popular services—Facebook, Google+ and Twitter—provide an API to select content, which is a precondition for the use of its content on an independent, self-contained platform. The integration of various platforms seems to be necessary because of their more or less prominence.

5.2. Semantic Enrichment of Citizen-Generated Content by NER

After using and testing specific keywords to search for content on the selected social media services, it became apparent that most social media content does not provide sophisticated data relating to the location in their metadata. For example, only 10.3% of all Twitter users worldwide have the geo-location enabled and therefore around 90% of all tweets do not contain location information within the metadata (http://www.beevolve.com/twitter-statistics/). Abel et al. [2012a, 2012b] show, with their

prototype *Twitcident,* the advantages of semantic enrichment with NER. Natural Language Processing (NLP) tools, a subset of NER, aim to extract semantic information from unstructured texts using linguistic concepts such as part-of-speech or grammatical structures [Kao and Poteet 2007]. These are able to recognize locations (LOC), organizations (ORG), people (PER), and miscellaneous information (MISC) from unstructured data. In addition, geocoding services such as the Google Geocoder can deliver coordinates for locations (e.g., city names or addresses).

Different NERs are currently available that try to help identifying semantic information from unstructured data: The *Stanford Natural Language Processing Group* (http://www-nlp.stanford.edu/ner/) provides a Java-based Open-Source NER. Faruqui and Pad [2010] developed two German language models for Stanford NER: DeWac uses the web as a corpus, but only considers .de domains [Baroni et al. 2009]; HGC (huge German corpus) is based on the Stuttgart University Newspaper corpus, consisting of newspapers and legal texts (http://ims.uni-stuttgart.de). The commercial service AlchemyAPI (http://alchemyapi.com) provides NLP-services via http. We used the noncommercial trial version of Alchemy with a limit of 10,000 requests per day.

The NERs mentioned are trained using well-written text like news or web data, which are different from citizen-generated content within social media and therefore may not work well when directly applied to social media posts. Ritter et al. [2011] present a system called T-NER, which is a special NER for analyzing Twitter messages, but it focuses only on the Twitter sphere and is not open source. The available NERs (even nonoptimized for social media information) that are presented here have various advantages and disadvantages in dealing with the quality of content and with nonEnglish language. To optimize the results, we conducted a benchmark with about 130 crisis-relevant posts containing the word "Brand" (German word for "burning"). During this benchmarking process, we manually tagged locations, organizations, people and other information. Then we used the NER tools DeWac, HGC and Alchemy to compare the results. Based on the results, we focus for our latter analysis on the four different categories "Locations" (LOC), "Organizations" (ORG), "Personal names" (PER), and "Miscellaneous" (MISC). Taking the results of the benchmark into account we decided to combine the existing NERs as follows:

$$\text{LOC} = \{\text{DeWac}_{\text{LOC}} \cup \text{HGC}_{\text{LOC}} \cup \text{Alchemy}_{\text{LOC}}\}$$
$$\text{ORG} = \text{Alchemy}_{\text{ORG}}$$
$$\text{PER} = \text{DeWac}_{\text{PER}}$$
$$\text{MISC} = \{\text{DeWac}_{\text{MISC}} \cup \text{HGC}_{\text{MISC}} \cup \text{Alchemy}_{\text{MISC}}\} / \{\text{LOC} \cup \text{ORG} \cup \text{PER}\}$$

5.3. Criteria for Quality Measurement of Citizen-Generated Content

Based on the criteria (especially intrinsic data quality) derived by Wang and Strong [1996] and on existing metadata in social media, we selected five quality criteria for personal user settings in our approach. The other quality criteria cannot be specified by the users, and instead their importance is calculated internally. To what extent these five individually adjustable, and the internally calculated, quality criteria correspond to the users' understanding of information quality will be evaluated later on.

—*Link*: This criterion indicates whether a post contains a link. Due to limitations like the 140 characters of a tweet, links are often used to share further content. "Link" refers to the contextual data quality criterion "value-added" [Wang and Strong 1996] and it especially applies to photos or videos. Ha and Ahn [2011] show that "the existence of external links in a tweet moderates the impact of argument quality. The greater the number of positive responses relative to negative responses, the higher the quality of a given argument on users' attitudes toward received tweets".

—*Credibility*: The number of follower/friends is used as an indicator of the author's "reputation" [Wang and Strong 1996]. Ha and Ahn [2011] show that perceived credibility of a source strongly influences peoples' information sharing behavior. Furthermore, Kane and Ransbotham [2012] show that the quality of information provided by well-connected contributors within social media is perceived of as having higher value.
—*Up-to-datedness*: This criterion refers to when the posts were published and can be compared to the data quality criterion "timeliness" [Wang and Strong 1996]. The latest entry of a result set is weighted with a value of 100, whereas the earliest is given a value of 0. The calculation of the degree of fulfillment between the two extremes is achieved by a linear function.
—*Dissemination*: The number of, for instance, retweets in Twitter or shares in Facebook and Google+ serves as an indicator of the dissemination of social media content. As already shown, posts and retweets are used by the "crowd" in such ways that mainly high-quality posts are disseminated and a collective filter emerges. For calculation, a tweet of a result set featuring the most retweets is rated 100, a tweet which has not been retweeted is rated 0 [Wang and Strong 1996]. We would suggest that this is also a reasonable measure of "believability," because unreliable information is unlikely to be retweeted for long.
—*Quality of coordinates*: The validity and accuracy of coordinates often differ [Gao et al. 2011]. With regard to coordinates in citizen-generated content, we must consider two dimensions of locations. The first is the location of the social media post itself, which is provided as metadata and expresses where a message was sent from. The second dimension is the location and coordinates which can be found directly in the content itself, which often refers to the logical content of a message. Both do not necessarily match. We defined the following ranking of quality: A coordinate is most exact when transmitted via mobile device within the content. As part of the crisis-related analysis of social media, Terpstra et al. [2012] state that GPS-coordinates are contained in not more than 1% of all tweets. If a coordinate indicated as the place of residence on the author's profile, a published post will not necessarily be about or sent from the corresponding location. The validity lessens if this place of residence in the profile only exists in textual form and has to be localized by a geocoding service first. This criterion refers to "accuracy" [Wang and Strong 1996].

5.4. Dynamic Quality Measurement of Citizen-Generated Content

As mentioned earlier, Wang and Strong's [1996] other criteria, especially *representational data quality* (interpretability, ease of understanding, representational consistency, concise representation) and *accessible data quality* (accessibility, access security) are going to be supported by dynamic quality measurement and in the presentation of the results. To measure the multidimensional construct *information quality,* we must first identify the requirements for high-quality citizen-generated content in crisis situations. Our quality score Q for content m is calculated according to following formula. For every quality criterion i the product of weight w_i and degree of fulfillment $f_i(m)$ is computed and then added up. This sum is divided by the sum of the weights of all quality criteria. $Q(m)$, w_i as well as $f_i(m)$ lie within the range between 0 and 100.

$$Q(m) = \frac{\sum_{i=1}^{n} w_i \, f_i(m)}{\sum_{i=1}^{n} w_i}.$$

m = Citizen-generated content $Q(m)$: Quality score of m with $0 \leq \mathrm{Q(m)} \leq 100$
n = Amount of quality criteria $\mathrm{w_i}$=Weight of the quality criteria i with $0 \leq \mathrm{w_i} \leq 100$
$f_i(m)$ = Degree of fulfillment of criteria i with $0 \leq f_i \leq 100$

Fig. 1. Search for social media entries.

The specific characteristics of each source platform have to be considered for the quality rating. The function f_i for calculating the degree of a fulfillment can vary. In Twitter, for instance, the number of retweets can be understood as an indicator of the popularity of a tweet, YouTube provides numbers of positive and negative user ratings, but Facebook only provides "likes." The indicators for the same quality criterion therefore differ in their value as well as in their semantics. In the case of a tweet, the function f_{i1} for calculating the degree of fulfillment of the criterion "popularity" merely uses the indicator "number of retweets." In the case of a YouTube video, the numbers of positive as well as of negative user ratings are considered as indicators. For Facebook only the "likes" are relevant. The platform-specific definition of the function $f_i(m)$ must therefore necessarily be rendered as precisely as possible. Now, we should state at this point that such measures will not provide infallibly accurate and up-to-date material. In a situation, however, in which the current state of play is such that there are no guarantees at all of accuracy and timeliness, we would argue that a "satisficing" approach of this kind gives a much better result than was previously available, if not a perfect one.

6. IMPLEMENTATION OF THE APPLICATION "SOCIAL HAYSTACK"

To examine the support of dynamic and subjective quality assessment of citizen-generated content in emergencies, we created a prototype Social Haystack, as a service-oriented client–server-application. The client is a common web application using HTML, CSS, JavaScript and jQuery, and the server was programmed with the .NET framework. The data is stored in an MSSQL database and exchanged as JSON.

The search view is intended for the entering of search requests (Figure 1). On the left-hand side, the display elements of the geo search can be seen. These comprise a map, a text field to display and enter the radius of the search area and a button to reverse the geo search. On the right-hand side there are checkboxes for choosing the source platforms and a text field to enter the search string. The geo search is activated by double-clicking on the map, in which the search area is represented by a red shaded circle, the center and radius of which can still be changed afterwards. After selecting the location area, social media source and keywords, the data retrieved is saved in

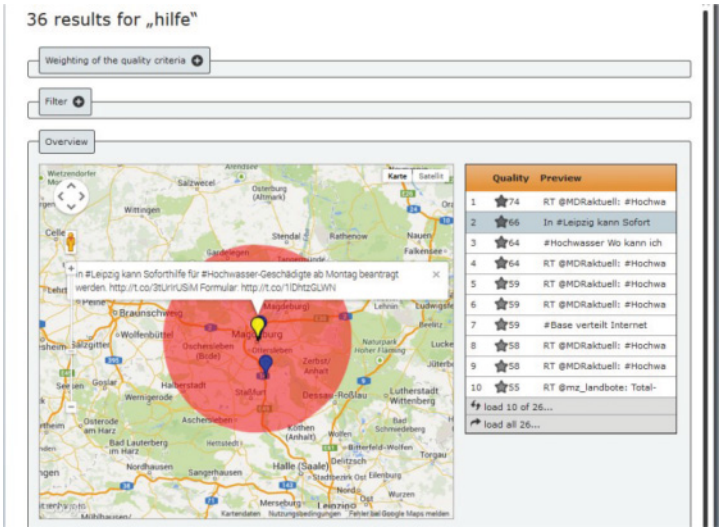

Fig. 2. Display of results.

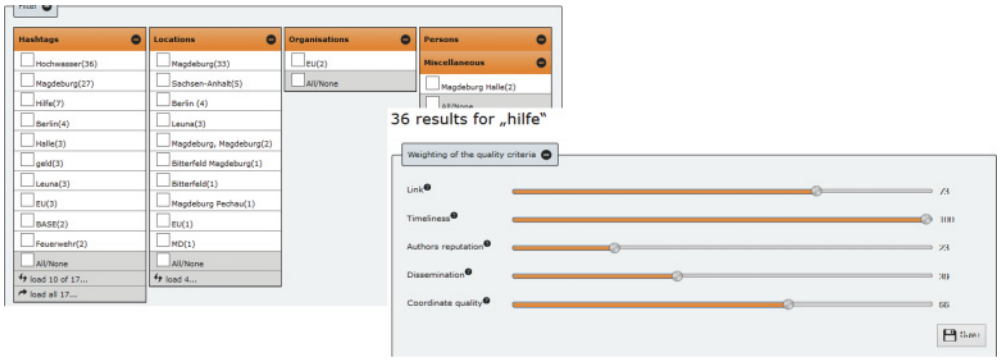

Fig. 3. Dynamic filtering the search results (left) and weighting of the criteria (right).

OpenSocial format (http://opensocial.org/). OpenSocial provides open standards in the context of social media and aims to break down technical barriers between different systems and provide interoperability. Using OpenSocial provides an opportunity for reuse in external or additional systems.

After a successful search process, the user is referred to the "result" view (Figure 2). This consists of an overview field, (initially collapsed) sliders for weighting the quality criteria and the filter. In our case, we searched for "hilfe" (German word for *help*) and found 36 matching results. According to the geographic relevance of the results such as texts, images or videos, they are displayed on a map. The content in the results table is graded according to quality, so that the most valuable information has the highest position. The numeric quality score from 0 to 100 is additionally visualized by a star icon. Content with a quality score of 0 is marked by an empty star, whereas a quality score of 100 is marked by a star which is filled in.

The respective quality criteria can be specified by using the aforementioned sliders (Figure 3, right). Tooltips provide a short description of the respective criterion. Here one can see the five previously specified quality criteria "link," "timeliness," "author's

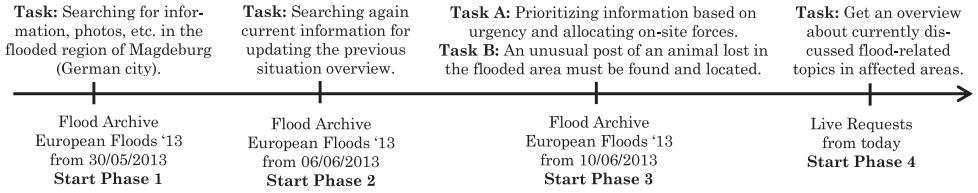

Fig. 4. Timeline of Flood Archive datasets and tasks during evaluation.

reputation," "dissemination," and "coordinate quality." On the right side of the slider the current weight is displayed, for example, the timeliness has the most relevance for the user (weight = 100). After the user clicks on "save", the tweets are re-evaluated and the result view is reloaded. During this weighting the keywords and NER results are checked against a stored synonym list and abbreviations as well as slang are spelled in full based on a slang dictionary (http://www.noslang.com/dictionary/). On opening up the collapsed filter tab (Figure 3, left) all keywords that were recognized by the NERs are displayed. They are grouped by "hashtags," "locations," "organizations," "persons," and "miscellaneous" and sorted by the number of times they occurred in the posts. The user has the option of checking or unchecking recognized keywords to narrow down the results.

7. EVALUATION

7.1. Methodology

To evaluate our overall concept and Social Haystack, we conducted evaluations with a total of 20 users. We aimed at testing both usability, for which we used a group of students, and practice relevance, for which we enlisted professionals in the field of crisis management. In the second instance, we were primarily concerned with identifying whether and in what ways the application would be used, and what difficulties in use might be encountered. The philosophy behind the evaluation process was derived from the notion of "situated evaluation" [Twidale et al. 1994] in which qualitative methods are used to draw conclusions about real-world use of a technology using domain experts. The aim here is not to measure the relationship between evaluation goals and outcomes but to derive subjective views from experts about how useful and relevant the technology might be in use. The evaluations took place on the basis of a scenario to simulate special characteristics of crisis situations for the participants. Although our system had been fully implemented, IT security regulations as well as privacy and documentation concerns of the emergency services prevented us from having an in-use and real-world evaluation.

The scenario was based on an actual flood caused by the German river "Elbe" in mid-2013. A dataset with about 80,000 tweets (collected via Tweet Archivist) identified by the hashtag #hochwasser (German word for "flood") and Facebook posts (collected manually by scanning the emerged Facebook groups) and Google Maps' situation maps (collected manually by taking screenshots of the public maps) was collected in the period from 30/05/2013 to 28/06/2013. To use our evaluation data, we integrated a new fake source platform "Flood Archive", which accesses the recorded flood contents (in addition to the "live" data of social media). To enable the simulation of the chronological sequence of the archived data, we implemented a backend option for selecting a time period that in turn automatically changed time-linked social media content from the archive. We broke the entire flood into four different phases. The timeline of our evaluation and each phase of the "Flood Archive" datasets are presented in Figure 4. Each evaluation lasted about 60min per participant during which the participants were asked to *think aloud*

[Nielsen 1993] and were audio-recorded. The participants received a rough overview of the research field and the issues we were concerned within our study and then the scenario and their upcoming tasks were introduced. The tasks were linked to the chronological sequence of four phases. For each phase we updated the dataset from the archived original data. So, each phase had its own dataset of social media information. After describing the tasks, the participants were asked to fulfill them by using Social Haystack.

The main purpose of separating the entire scenario to four phases was to test the potential of Social Haystack for different and heterogeneous work practices in an emergency. In *phase 1* the participants had to search for information in the flooded region of Magdeburg (a German city) to obtain a situation overview. The purpose of this first phase was to identify how Social Haystack might be used for a rough situation assessment under time-pressure with minimal knowledge of the emergency occurring. Within the following *phase 2* the users were tasked with updating their rough situation overview by obtaining the latest content. The aim of phase 2 was to examine how emergency services might use the tool for detailed assessments of an emergency if they have enough time for refining keywords, locations or other parameters. Within *phase 3*, two different tasks had to be achieved: Task A was allocating on-site forces by searching for help requests and then deciding which request the on-site units should respond to in case of a real emergency. In task B, an unusual post about an animal lost in the flood area has to be found and located, for which only rough data is available. The purpose of this phase was to evaluate the potential of Social Haystack for deriving concrete prioritizations of actions based on the situation assessments. The final *phase 4 (searching for interests)* takes place months after the flood. The participant was asked to get an overview (i.e. relevant to their current interests) of the flood-related topics that are currently being discussed in the area which was previously affected. For this task we used the live search instead the "Flood Archive". The aim of phase 4 was to uncover potential use cases of the tool if users are not assigned with tasks. In the case of this last task it is hard to measure the results in any other way than through interviews, given the subjective nature of the possible interests in question, and what kind of "overview" would be regarded as satisfactory in the light of those interests. Subsequent to the conduct of the scenario phase, open interviews concerning impressions and perceptions of Social Haystack and therefore the relevance of our subjective and dynamic quality assessment concept in practice were conducted.

7.2. Results

The recorded interviews of the evaluation were transcribed for subsequent data analysis. To analyze our data, we used a qualitative content analysis approach, which is appropriate when prior theory and knowledge can be drawn on (from the literature study), but stays open to unexpected themes and issues. Only at a later stage findings are related back to existing theory. Within the analysis, as indicated, we wanted to uncover the potential of approaches like Social Haystack with regard to its usability and its usefulness in relation to practice. We describe the results according to our derived categories in the following.

7.2.1. Platform Independent Information Search. During emergencies, important citizen-generated information is spread over various social media which constitutes something of a difficulty when searching for relevant material: *"It is not easy to put everything together, because no one has all the information" (E19)*. This divergence between the different platforms as well as the fact that officials are not always registered in all of them, or are not familiar with them, can lead to them not assessing information that

eventually proves relevant. A subjective and dynamic approach makes it possible to search for data across multiple platforms and to filter information simultaneously:

> So, I think the advantage is obvious, you can find out very quickly what people are saying about a certain topic in the different networks (E15). So basically, you have built a bridge. (E18)

The interlayer is based on a kind of service-oriented concept, which makes it possible to connect further social media during runtime and, as a consequence, integrate their information into the search too, *"because the public networks are used very frequently, it is even better to combine them"* (E17).

This does not mean all possible useful sources are integrated here, as was pointed out to us: *"I could image integrating certain news sites"* (E16). News sites are already being used during emergencies because they are often the only way to obtain on-site photos. In addition to news sites, selection of preselected special interest groups (e.g., on Facebook) or authors (e.g., Twitter) is possible:

> So that you can choose which groups you want to search for information. (E16)

Possibilities for direct contact with authors were also mentioned, as with E4's comment: "*I would call the guy and ask him how many helpers he needs*". Regarding the post "*Pechau [city] needs help!*", E5 says: "*This post is so general that I have no option but to contact him*" (E5). It would be desirable for authors who contribute valuable data regularly to be "subscribed to" without using the wider platform to any great extent (E1).

7.2.2. Presentation of Search Results. Even if the presentation of search results is preselected, a large amount of data is still available. This large amount of data results in the acquisition of important data being slowed down and in consequence the emergency services receive it too late (E7). To address this issue, certain search parameters become necessary:

> Now, if I get 300,000 search results and they are all quite similar, then I would be interested in using the extended search parameters to find exactly what I want. (E11)

Furthermore, the number of results can be reduced by defining a limit in the first instance. In this respect, one participant defines a limitation of 20–25 entries:

> 20 is an appropriate number, maybe 25. But everything exceeding this is too overwhelming. I would have liked the first page to show me the first 20 items and the option to click for further items. (E10)

Of course, this predefined constraint significantly narrows the information range. Therefore, a certain balance between the amount of information and information quality must be struck. Limiting the number of presented results and giving users the option to view them all in detail could be an initial step. In addition to the influence which the users exert on the measurement of information quality, they perceive a need to be able to understand why the search results are sorted as they are.

> I find it important that the user knows exactly what has been measured and how it has been measured... and not only that something has been measured. (E11)

This is the reason why the system provides users with information on the exact measurement regarding each criterion, exhibiting the degree of fulfillment of a criterion (with regard to all other criteria).

When I use the cursor to hover over each of these results, I can see how they all fit together, because the highest one then has [. . .] 'Credibility: 100%'. (E17)

7.2.3. Importance of Individual Information Quality Criteria. "Up-to-datedness" is seen as the most important quality criterion by almost all of the participants. The more up-to-date a help request is, the more urgent it is considered to be (E3; E1). For E5, up-to-datedness is important because *"I want to help where it is most critical at the moment"* (E5). "Dissemination" is an important quality criterion for many as well:

Maybe you could derive the importance from the number of retweets or the number of people affected. (E6)

For E5, dissemination can be a criterion for the urgency of a help request: *"Maybe you could assume that something which is severe is retweeted more often"* (E5). Additionally, dissemination is seen as a criterion for particularly relevant topics (E1). For E6, the "author's reputation" is an indicator of the reliability of a tweet:

. . . and maybe the author's reputation, whether or not I can believe them. (E6)

However, the connection between the number of followers and reputation has also been questioned (E4). E1 states that the author's reputation is difficult to understand for users who do not know much about social media. The quality criterion "link", which refers to the existence of a link in the content, is not understood by E3. For E6 this criterion is a "*Yes/No-matter*" (E6) which is not suitable to be adjusted using a slider. Many participants, such as E4, do not understand "quality of coordinates": *"OK, sounds complicated"* (E4). For E3, this criterion is not comprehensible either. E6 states that it is difficult to judge what the system is doing here. However, after reading the tooltip, E1 remarks:

That means precisely how it is shown on the map. He thinks this criterion is important in order to know exactly where it is. (E1)

7.2.4. Usability of Dynamic Information Quality Measurement. The clarity of the setting options and quality criteria was more or less universally praised:

This has definitely been kept very clear and expedient (E15). It's really fantastic that you can have such an influence on grading the results. (E12)

Our prototype has not only made it possible to weight and to rank the Internet-based data, but has done this in a way which is simple to learn for the user. However, a large number of setting options poses certain obstacles when getting started and dealing with the application, because users must first acquaint themselves with the different options (E7). However, such overhead seems mainly to be limited to the first use:

This only impeded me at the beginning or when I looked at it for the first time. (E9)

The effort entailed in cases of nonregular usage of a technology is a common problem already known in security research [Kyng et al. 2006]. Tools designed for emergencies fall into this category. Time is a critical factor in the saving of human lives and therefore the barrier to using the system should be as low as possible so as to minimize overhead. This is why many setting options are a blessing and a curse at the same time:

There are too many things that I have to adjust. [. . .] If I don't have enough time for this, it will overload me (E19). The suggestion was therefore made that quality criteria *[. . .] should be adjusted so you can find your way more quickly (E12).*

Nevertheless, the question still remains how perfect standardization can be achieved given highly heterogeneous needs. It seems that different situations require different quality criteria. It means the need for quality criteria does not only depend on the current emergency but on each individual within its organization and task to be accomplished:

> Well, it's not important for me but maybe there are people who find it important" (E19). "A professional fire or police officer will probably need even more options. (E8)

7.2.5. Further Potential for Integrating Social Haystack. The particular advantage of our prototype lies in increasing efficiency, under time and quality constraints, with which information flows can be coordinated by the tool, thus reducing the overload problem whilst also maximizing the degree of high quality and relevant information. For our participants, it was obvious that such a tool serves only insofar as it complements their current practices. However, they saw the merit in additional information given earlier and beyond reports produced on-site by professional units. Complementary information from social media, in the main, they saw as facilitating the emergence of an even more comprehensive picture of the current situation:

> If something is happening somewhere, the forces on-site can now make use of information from a much larger number of people than only from the emergency services, who must go here and there to check the water level. This way, I get information that is usually much more difficult to obtain. (E13)

To review the information provided by social media, visualization of geographic information on a situation map, we feel, must be enabled. Situation maps are already established in the current practice of emergency services and are available in the control room as well as on-site.

> I can see directly on the map where the information has been sent from. Then I can see immediately if there is a river nearby or maybe I can concentrate on a certain location and can determine an accumulation of information at certain places. (E14)

This spatial distribution of information has advantages in capturing the magnitude of an incident. But herein lays a certain disadvantage: It is essential that information which can be displayed on a situation map, has a geo-reference—either via social media directly or automatic determination by the NER. If no geographic meta-information is available, the location will not be shown on the map. As a consequence the user might miss some information and would thus receive a limited picture of the situation:

> It wasn't obvious to me that there could be other information not containing any geographic position. (E12)

8. DISCUSSION AND CONCLUSION

In recent years, research in the area of crisis management has tended to focus on two separate issues: The first is the problem of how to coordinate response in what are temporary organizational and interorganizational arrangements and increasingly require what Luhmann [1996] call "systems of systems" and which support the development of advanced HCI techniques for more effective utilization of high-quality information. The complexity of such arrangements has generated increasing focus on supporting emergent interoperability between heterogeneous agencies and their information systems [Ley et al. 2014]. The second relates to the way in which, besides official emergency services and authorities, ordinary citizens are getting increasingly involved in response

work [Reuter et al. 2012; Starbird et al. 2012]. That is, and outside of official organizational arrangements, the peer-to-peer culture celebrated by Shirky [2008] can have powerful effects on information gathering and use. Our article contributes a view of how a dynamic and subjective quality assessment of citizen-generated content could bring these two agendas together to support the work of emergency services.

It is obvious that information is the basis for human decision-making and action, and equally obvious that such information should be as accurate and timely as is realistically possible. If decisions are based on information of lower quality, the probability of unsatisfactory or even harmful results increases. The need for high-quality information rises in accordance with the importance of decisions and actions to be taken. In the context of a crisis, inappropriate quality can, in extreme cases, mean the difference between life and death.

To evaluate information with respect to its quality, it is necessary to adapt the abstract concept of information quality to the field of crisis management. Here, especially, time is such an important constraint that one may have to "satisfice." The tradeoff between quality and time is critical and, moreover, depends on the very subjective, context-dependent and situational character of information quality and its assessment during crisis response. Current approaches and systems try to implement this character through static quality algorithms or, in the case of Ushahidi [Okolloh 2008], delegate the evaluation of the content to the "crowd." However, on the one hand, approaches which focus only on automatically processing do not satisfy the dynamic needs of official and professional crisis management and the potential negotiation processes entailed. On the other hand, approaches that either just focus on human mechanisms for assessing citizen-generated content often overload the emergency services personnel or those approaches that try to engage the external crowd for evaluation of information are not always suitable to meet the very situational and context-dependent character of an emergency.

The prototype developed in our study contributes a proof of concept to existing work by emphasizing the subjective and context-dependent nature of information quality. The information requirements of a specific user do not necessarily coincide with those of a static algorithm or a community of users whose background is generally not the field of crisis management. The evaluation of Social Haystack has shown that the most serious challenge lies in the identification of appropriate quality criteria. Although, for example, "up-to-datedness" is understood as an important criterion, only a few users could imagine what the criterion "quality of coordinates" might be. One participant also questioned the relation between the number of subscribers an author has and his/her reputation. Our evaluation demonstrates again the need for dynamic, context-dependent quality assessment of citizen-generated content in social media. More importantly, and given the kinds of organizational and interorganizational heterogeneity we have identified, we demonstrate that support systems for use during emergencies need to allow context-dependent and subjective weighting and filtering of quality criteria. This seems, on the evidence of our evaluation, to address many of the difficulties associated with collecting and assessing citizen-generated information from heterogeneous social media.

Based on the results of our proof-of-concept, we were able to outline five lessons learnt and to present guidelines for designing approaches that deal with the assessment and handing of citizen-generated content from social media during emergencies that should be of interest for other HCI designers or software developers:

(1) *Fit of information*: We argue that, in dynamic situations of this kind, the "fit" of information to specific tasks is more important than generic assessments of information quality. The fit of information can be understood as the quality information has for a certain purpose/goal.

(2) *Subjectivity*: Although the roles of emergency services are bound by a common goal, the needs of individuals in a variety of organizational roles need to be addressed. We cannot specify in advance how many social media entries or which weighting of different quality criteria is needed. Each user must be able to formulate his/her individual qualitative information needs which can then be assessed by a tool.
(3) *Cross social media search*: The diversity of different social media services and platforms requires a cross social media approach that enables information search independently of the source platform. Although emergency services mainly focus on Twitter and Facebook, other services providing photos or videos, for instance Instagram or Flickr, must be mined as well to get as much relevant information as possible.
(4) *Traceability of results*: Although approaches like those we tested with Social Haystack help in assessing the relevance of information by ranking as well as sorting mechanisms, the results must always be traceable for the user. The composition of results must always be transparent.
(5) *NERs for social media*: We argue combining different available NER adds value to the analysis of social media information, especially when using languages different from English, where these services are not as appropriate. Locations in particular can be derived reliably and enhance location metadata of social media messages which is otherwise often missing.

One limitation of our overall approach is that social media, in the context we examined, is currently not part of the official work and task structures (which is about to change), and is only informally used for getting an (unofficial) situation overview. In general, in Germany emergency services do not necessarily monitor social media systematically, because it is not required by the current laws and therefore by the organizational structures of public authorities—which depend on such laws. The officials are therefore not very experienced in using social media for situation assessment practices. Other limitations with regard to our prototype itself are that the used NERs do not work as well with the German language as with the English language, and the functionality for weighting criteria can overload individual users. Therefore, an adequate degree of subjectivity and automatic configuration must be found. A possible approach might be to use previously saved configurations of criteria and/or recommending those to other users. Misinterpretation remains a problem, even with our concept. Citizens using social media are more likely technophiles. This will not be true for all people affected and we cannot assume an accurate assessment of an overall situation will be possible. Furthermore the sample of our evaluation was rather small ($n = 20$), and was limited to a flood scenario in Germany. Moreover, Social Haystack has not yet been evaluated in operational use during an actual disaster, when there might be hundreds of users and hundreds of thousands of potentially relevant social media posts per hour.

A long-term evaluation of our tool in the context of our running research project will likely be able to show the effects of a subjective and context-dependent nature of information quality on a larger level. At the same time, and as we have been at pains to point out, our goal was not perfection but improvement. We have no doubt our design guidelines constitute a way forward for more accurate information use in dynamic contexts like emergencies than is currently possible. With regard to the results of our study, our future work will—besides including a long-term evaluation—focus on training the NERs based on data sets of current emergencies and trying to use those trained NERs for extracting more semantic data from social media. We will further try to extend our approach to other—nonemergency related—application fields.

ACKNOWLEDGMENTS

We like to thank Jonas Winter for his valuable support in the early phase of the prototype's conception as well as Dave Randall and Starr Roxanne Hiltz for their editorial support.

REFERENCES

F. Abel, C. Hauff, and R. Stronkman. 2012a. Semantics + Filtering + Search = Twitcident exploring information in social web streams categories and subject descriptors. In *Proc. Hypertext.* Milwaukee, USA, 5–8.

F. Abel, C. Hauff, and R. Stronkman. 2012b. Twitcident: Fighting fire with information from social web streams. In *Proc. WWW.* New York, 5–8.

E. Agichtein, C. Castillo, and D. Donato. 2008. Finding high-quality content in social media. In *Proc. WSDM.* Atlanta, USA, 183–193.

M. Baroni, S. Bernardini, A. Ferraresi, and E. Zanchetta. 2009. The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43, 3, 209–226.

N. Bharosa, M. Janssen, and J. Lee. 2008. Adaptive information orchestration: architectural principles improving information quality. In *Proc. ISCRAM.* Washington, 556–565.

N. Bharosa, J. Lee, and M. Janssen. 2009. Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises. *Information Systems Frontiers* 12, 1, 49–65. DOI:10.1007/s10796-009-9174-z

S. K. Card, G. G. Robertson, and J. D. Mackinlay. 1991. The information visualizer: An information workspace. In *Proc. CHI.* ACM New York, NY, USA, 181–186.

R. Chen, R. Sharman, H. R. Rao, and S. J. Upadhyaya. 2008. Coordination in emergency response management. *Communications of the ACM* 51, 5, 66–73. DOI:http://dx.doi.org/10.1145/1342327.1342340

D. Christofzik and C. Reuter. 2013. The aggregation of information qualities in collaborative software. *International Journal of Entrepreneurial Venturing (IJEV)* 5, 3, 257–271.

G. Convertino, H. M. Mentis, A. Slavkovic, M. B. Rosson, and J. M. Carroll. 2011. Supporting common ground and awareness in emergency management planning. *ToCHI* 18, 4, 1–34.

M. J. Eppler. 2006. *Managing Information Quality*. Springer, Berlin, Heidelberg, New York.

M. Faruqui and S. Pad. 2010. Training and evaluating a German named entity recognizer with semantic generalization at Stuttgart. In *Proc. KONVENS.* Saarbrücken, Germany, 129–133.

C. Fisher, E. Lauria, S. Chengalur-Smith, and R. Wang. 2011. *Introduction to information quality*. AuthorHouse, Bloomington, Indiana, USA.

C. W. Fisher and B. R. Kingma. 2001. Criticality of data quality as exemplified in two disasters. *Information & Management* 39, 2, 109–116. DOI:http://dx.doi.org/10.1016/S0378-7206(01)00083-0

T. Friberg, S. Prödel, and R. Koch. 2010. Analysis of information quality criteria in a crisis situation as a characteristic of complex situations. In *Proc. IQ.* Little Rock, AR.

T. Friberg, S. Prödel, and R. Koch. 2011. Information quality criteria and their importance for experts in crisis situations. In *Proc. ISCRAM.* Lisbon, Portugal.

H. Gao, G. Barbier, and R. Goolsby. 2011. Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems* 26, 3, 10–14. DOI:http://dx.doi.org/10.1109/MIS.2011.52

R. Gaspar, S. Gorjão, B. Seibt, L. Lima, J. Barnett, A. Moss, and J. Wills. 2014. Tweeting during food crises: A psychosocial analysis of threat coping expressions in Spain, during the 2011 European EHEC outbreak. *IJHCS* 72, 2, 239–254. DOI:10.1016/j.ijhcs.2013.10.001

S. Ha and J. Ahn. 2011. Why are you sharing others' Tweets?: The impact of argument quality and source credibility on information sharing behavior. In *Proc. ICIS.* Shanghai, China, 1–11.

G. R. Hayes. 2011. The relationship of action research to human-computer interaction. *ToCHI* 18, 3, 1–20. DOI:http://dx.doi.org/10.1145/1993060.1993065

G. N. Hearn, J. A. Tacchi, M. Foth, and J. Lennie. 2008. *Action research and new media: Concepts, Methods and Cases*. Hampton Press, New York, USA.

A. Hevner and S. Chatterjee. 2010. *Design Research in Information Systems: Theory and Practice*. Springer, New York, USA.

A. Hevner, S. T. March, J. Park, and S. Ram. 2004. Design science in information systems research. *MIS Quarterly* 28, 1, 75–105.

S. R. Hiltz, P. Diaz, and G. Mark. 2011. Introduction: Social media and collaborative systems for crisis management. *ToCHI* 18, 4, 1–6.

Homeland Security. 2013. Lessons learned: Social media and Hurricane Sandy - virtual social media working group and group DHS first responders. Retrieved June 7, 2015 from http://www.ghinternational.com/docs/DHS_VSMWG_Lessons_Learned_Social_Media_and_Hurricane_Sandy_Formatted_June_2013_FINAL.pdf.

M. E. Jennex. 2012. Social media – truly viable for crisis response? In *Proc. ISCRAM*. Vancouver, Canada, 1–5.

G. Kane and S. Ransbotham. 2012. Codification and collaboration: Information quality in social media. In *Proc. ICIS*. Orlando, Florida, USA, 1–16.

A. Kao and S. R. Poteet. 2007. *Natural Language Processing and Text Mining*. Springer, London.

A. M. Kaplan and M. Haenlein. 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons* 53, 1, 59–68. DOI:http://dx.doi.org/10.1016/j.bushor.2009.09.003

M. Kyng, E. T. Nielsen, and M. Kristensen. 2006. Challenges in designing interactive systems for emergency response. In *Proc. DIS*p. 301. ACM Press, New York, USA, p. 301. DOI:http://dx.doi.org/10.1145/1142405.1142450

K. Lewin. 1958. *Group Decision and Social Change*. Holt, Rinehart and Winston, New York, USA.

B. Ley, T. Ludwig, V. Pipek, D. Randall, C. Reuter, and T. Wiedenhoefer. 2014. Information and expertise sharing in inter-organizational crisis management. *JCSCW* 23, 4–6, 347–387.

S. B. Liu and L. Palen. 2010. The new cartographers: crisis map mashups and the emergence of neogeographic practice. *Cartography and Geographic Information Science* 37, 1, 69–90.

P. H. Longstaff. 2005. Security, Resilience and Communication in Unpredictable Environments such as Terrorism, Natural Disasters and Complex Technology. Harvard University, Cambridge, UK.

T. Ludwig, C. Reuter, and V. Pipek. 2013. What you see is what I need: Mobile reporting practices in emergencies. In *Proc. ECSCW* (181–206). Springer, Paphos, Cyprus.

T. Ludwig, C. Reuter, T. Siebigteroth, and V. Pipek. 2015. CrowdMonitor: Mobile crowd sensing for assessing physical and digital activities of citizens during emergencies. In *Proc. CHI*. ACM Press, Seoul, Korea.

N. Luhmann. 1996. *Social Systems* (1st. ed.). Stanford University Press, Stanford, California, USA.

A. Marcus, M. Bernstein, and O. Badar. 2011. Twitinfo: Aggregating and visualizing microblogs for event exploration. In *Proc. CHI*. Vancouver, Canada.

T. B. Marino, B. S. do Nascimento, and M. R. S. Borges. 2012. GIS supporting data gathering and fast decision making in emergencies situations. In *Proc. ISCRAM*. Vancouver, Canada.

I. Mulder, H. de Poot, C. Verwij, R. Janssen, and M. Bijlsma. 2006. An information overload study: Using design methods for understanding. In *Proc. OZCHI*. 245–252.

J. Nielsen. 1993. Usability engineering. Morgan Kaufmann, San Francisco, California, USA. DOI:http://dx.doi.org/10.1145/1508044.1508050

M. Normark and D. Randall. 2005. Local expertise at an emergency call centre. In H. Gellersen, K. Schmidt, M. Beaudouin-Lafon, and W. Mackay (Eds.), *Proc. ECSCW*. Springer, Paris, France, 347–366.

C. A. O'Reilly. 1980. Individuals and information overload in organizations: Is more necessarily better? *Academy of Management Journal* 23, 4, 684–696. DOI:http://dx.doi.org/10.2307/255556

O. Okolloh. 2008. Ushahidi, or "testimony": Web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action* 59 (January), 65–70.

L. Palen and S. B. Liu. 2007. Citizen communications in crisis: Anticipating a future of ICT-supported public participation. In *Proc. CHI. San Jose*, ACM Press, USA.

L. Palen, S. Vieweg, and K. M. Anderson. 2011. Supporting "Everyday Analysts" in safety- and time-critical situations. *The Information Society* 27, 1, 52–62. DOI:http://dx.doi.org/10.1080/01972243.2011.534370

H. Purohit, A. Hampton, S. Bhatt, V. L. Shalin, A. P. Sheth, and J. M. Flach. 2014. Identifying seekers and suppliers in social media communities to support crisis coordination. *JCSCW* 23, 4–6, 513–545.

Y. Qu, P. F. Wu, and X. Wang. 2009. Online community response to major disaster: A study of Tianya forum in the 2008 Sichuan earthquake. In *Proc. HICSS*. Waikoloa, USA.

C. Reuter, O. Heger, and V. Pipek. 2013. Combining real and virtual volunteers through social media. In *Proc. ISCRAM*. Baden-Baden, Germany, 780–790.

C. Reuter, T. Ludwig, M.-A. Kaufhold, and V. Pipek. 2015. XHELP: Design of a cross-platform social-media application to support volunteer moderators in disasters. In *Proc. CHI*. ACM Press, Seoul, Korea.

C. Reuter, T. Ludwig, and V. Pipek. 2014. Ad hoc participation in situation assessment: Supporting mobile collaboration in emergencies. *ToCHI* 21, 5, 26:1–26:26.

C. Reuter, A. Marx, and V. Pipek. 2012. Crisis management 2.0: Towards a systematization of social software use in crisis situations. *IJISCRAM* 4, 1, 1–16.

A. Ritter, S. Clark, and O. Etzioni. 2011. Named entity recognition in tweets: An experimental study. In *Conference on Empirical Methods in Natural Language Processing*. Edinburgh, UK, 1524–1534.

A. Robinson, A. Savelyev, S. Pezanowski, and A. M. MacEachren. 2013. Understanding the utility of geospatial information in social media. In *Proc. ISCRAM*. Baden-Baden, Germany, 918–922.

K. Schmidt. 2014. The concept of "Practice": What's the point? In *Proc. COOP*. 427–444. Nice, France.

A. Shamoug, R. Juric, and S. Paurobally. 2012. Ontological reasoning as a tool for humanitarian decision making. In *Proc. ISCRAM*. Vancouver, Canada.

G. Shankaranarayanan, B. Iyer, and D. Stoddard. 2012. Quality of social media data and implications of social media for data quality. In *Proc. ICIQ*. Paris, France, 311–325.

C. Shirky. 2008. *Here Comes Everybody: The Power of Organizing Without Organizations*. Penguin Books, London, UK.

R. A. Stallings and E. L. Quarantelli. 1985. Emergent citizen groups and emergency management. *Public administration review* 45 (Special Issue), 93–100.

K. Starbird, G. Muzny, and L. Palen. 2012. Learning from the crowd: Collaborative filtering techniques for identifying on-the-ground twitterers during mass disruptions. In *Proc. ISCRAM*. Vancouver, Canada.

K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg. 2010. Chatter on the red: What hazards threat reveals about the social life of microblogged information. In *Proc. CSCW*. 241–250.

K. Starbird and J. Stamberger. 2010. Tweak the Tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. In *Proc. ISCRAM*. Seattle, USA.

A. L. Strauss. 1987. *Qualitative Analysis for Social Scientists*. Cambridge Press, Cambridge, UK.

A. H. Tapia and K. Moore. 2014. Good enough is good enough: Overcoming disaster response organizations' slow social media data adoption. *JCSCW* 23, 4–6, 483–512.

T. Terpstra, A. de Vries, R. Stronkman, and G. L. Paradies. 2012. Towards a realtime Twitter analysis during crises for operational crisis management. In *Proc. ISCRAM*. Vancouver, Canada, 1–9.

M. Turoff, M. Chumer, B. van de Walle, and X. Yao. 2004. The design of a dynamic emergency response management information system (DERMIS). *JITTA* 5, 4, 1–35.

M. Turoff, S. R. Hiltz, C. White, L. Plotnick, A. Hendela, and X. Yao. 2009. The past as the future for emergency planning and response. *IJISCRAM* 1, 1, 12–28.

M. Twidale, D. Randall, and R. Bentley. 1994. Situated evaluation for cooperative systems. In *Proc. CSCW*. Chapel Hill, North Carolina, USA, 441–452.

Twitter. 2014. TweetDeck. Retrieved July 23, 2014 from https://about.twitter.com/products/tweetdeck.

I. Uysal and W. B. Croft. 2011. User oriented tweet ranking: A filtering approach to microblogs. In *Proc. CIKM*. Glasgow, Great Britain, 2261–2264.

T. Wachtendorf and J. M. Kendra. 2006. Improvising disaster in the city of Jazz: Organizational response to Hurricane Katrina. Retrieved March 25, 2015 from http://understandingkatrina.ssrc.org/Wachtendorf_Kendra/.

R. Wang and D. Strong. 1996. Beyond accuracy: What data quality means to data consumers. *JMIS* 12, 4, 5–33.

B. van de Walle and M. Turoff. 2008. Decision support for emergency situations. *ISeB* 6, 3, 295–316.

A. Wexelblat and P. Maes. 1999. Footprints: History-rich tools for information foraging. In *Proc. CHI*. 270–277.

A. Wu, G. Convertino, C. H. Ganoe, J. M. Carroll, and X. L. Zhang. 2013. Supporting collaborative sense-making in emergency management through geo-visualization. *IJHCS* 71, 1, 4–23. DOI:http://dx.doi.org/10.1016/j.ijhcs.2012.07.007

V. Wulf, M. Rohde, V. Pipek, and G. Stevens. 2011. Engaging with practices: Design case studies as a research framework in CSCW. In *Proc. CSCW*. ACM Press, Hangzhou, China.

D. Yates and S. Paquette. 2011. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *IJIM* 31, 1, 6–13. DOI:http://dx.doi.org/10.1016/j.ijinfomgt.2010.10.001

J. Yin, S. Karimi, B. Robinson, and M. Cameron. 2012. ESA : Emergency situation awareness via microbloggers. In *Proc. CIKM*. 1–3.

Received September 2014; revised March 2015; accepted March 2015