Being specific about geographic information crowdsourcing: a typology and analysis of the Missing Maps project in South Kivu

Benjamin Herfort
GIScience Chair, Heidelberg University, Germany
herfort@stud.uni-heidelberg.de

Melanie Eckle
GIScience Chair, Heidelberg University, Germany
eckle@stud.uni-heidelberg.de

João Porto de Albuquerque
Centre for Interdisciplinary Methodologies, University of Warwick, UK | Dept. of Computer Systems/ICMC, University of Sao Paulo, Brazil | GIScience Chair, Heidelberg University, Germany
j.porto@warwick.ac.uk

ABSTRACT

Recent development in disaster management and humanitarian aid is shaped by the rise of new information sources such as social media or volunteered geographic information. As these show great potential, making sense out of the new geographical datasets is a field of important scientific research. Therefore, this paper attempts to develop a typology of geographical information crowdsourcing. Furthermore, we use this typology to frame existing crowdsourcing projects and to further point out the potential of different kinds of crowdsourcing for disaster management and humanitarian aid. In order to exemplify its practical usage and value, we apply the typology to analyze the crowdsourcing methods utilized by the members of the Missing Maps project developed in South Kivu.

Keywords

Crowdsourcing, Classification, Digitization, Conflation, OpenStreetMap

INTRODUCTION

Crowdsourced geographic information has become a new information source for disaster risk management and humanitarian aid in recent years. The efforts during the earthquake that hit Haiti in 2010 have shown that information from Social Media, via SMS and from OpenStreetMap can complement official information. In some cases, when official information is missing at all, these new sources become even more relevant (Horita, Degrossi, Assis, Zipf, & de Albuquerque, 2013). Since the Haitian earthquake new organizations and groups such as the Humanitarian OpenStreetMap Team (HOT), the Standby Task Force, the GEOCAN consortium or the Digital Humanitarian Network break new grounds on how to organize and incorporate the work force of volunteers for disaster management and humanitarian aid. Their work shows, that crowdsourcing geographical information also demands tools that allow the collaborative and coordinated work of thousands of people.

Nevertheless, recent research often focuses on crowdsourcing as a data source and equates it with terms like volunteered geographic information or user generated geographical content. As a result, crowdsourcing activities are commonly distinguished with respect to their different platforms or data sources such as Twitter, Flickr, OpenStreetMap or Wikimapia.

In contrast, this paper utilizes a definition of crowdsourcing as a method and not as a data source. We therefore understand that crowdsourcing is, as defined by Quinn & Bederson (2011), the act of taking a job traditionally performed by a designated agent (usually an employee) and outsource it to an undefined, generally large group of people in the form of an open call.

There is some recent work that attempts to classify the different roles that can be played by the crowd in disaster management (Liu, 2014) and how they fit in the different phases of the disaster management cycle (Horita et al., 2013). In scientific research there is a strong link between the fields of crowdsourcing geographic information and citizen science. Haklay (2013) proposes four different levels of participation and engagement in citizen science projects: crowdsourcing, distributed intelligence, participatory sensing and extreme citizen science. However, what is currently missing is an overview about the specific tasks related to the creation of geographic information by the crowd, i.e. what are the tasks related to the production or enrichment of geographic information that can be 'outsourced' to the crowd. To fill this gap, this paper attempts to develop a typology of geographical information crowdsourcing. Furthermore, we use this typology to frame existing crowdsourcing projects and to further point out the potential of different kinds of crowdsourcing for disaster management and humanitarian aid. Finally, in order to exemplify its practical usage and value, we apply the typology to analyze the crowdsourcing methods utilized by the members of the Missing Maps project developed in South Kivu.

TYPOLOGY

In this section, we present a typology of geographical information crowdsourcing. In the following we distinguish three levels of crowdsourcing: classification, digitization and conflation. For each of these levels we will provide information on purpose, subtasks and complexity. Furthermore, we review and describe existing crowdsourcing projects to exemplify each category. An overview of these issues can be found in table 1.

Levels	Description	Complexity	Examples in the extant literature
1: classification	The process of assigning predefined attributes (values /categories) to existing geographical information	low	Chan, Munro, & Schnoebelen (2013), Vieweg, Castillo, & Imran (2014), Imran, Castillo, Lucas, Meier, & Rogstadius (2014), Ostermann (2015)
2: digitization	The process of creating new geographic objects based on existing geographic information.	medium	Barrington et al. (2011), Hillen & Höfle, (2015), Arcanjo, Luz, Fazenda, & Ramos (2016)
3: conflation	The process of integrating existing geographic information representing the same real-world object into a consistent representation	high	Soden, Budhathoki, & Palen (2014)

Table 1. Typology of geographical information crowdsourcing

The typology proposes 'classification' as the first level of crowdsourcing geographical information, i.e. the assignment of an additional attribute to an existing piece of geographic information. This attribute (also refered to as 'tag' or 'label') usually indicates a category and/or property of the geographic object to which it is attached. This task has a low level of complexity and thus bears the potential to be executed by a large number of volunteers, since it does not demand too much in terms of experience or skills. Classification tasks are usually based on the content of a single information source. Furthermore, this first level of crowdsourcing geographical information has the potential to be combined and augment automated classification methods, such as those based on machine learning techniques.

Several past projects used crowdsourcing to classify geographical information. In the aftermath of hurricane Sandy hitting the US east coast in 2012, for instance, CAP and FEMA used an online crowdsourcing system for damage assessment. Non-expert volunteers were asked to evaluate the level of damage present in aerial images captured after the hurricane hit the coastline (Chan et al., 2013). Through the MicroMappers crowdsourcing platform, volunteers can support disaster management by reading and labeling tweets based on the information they contain (Vieweg et al., 2014). Imran, Castillo, Lucas, Meier, & Rogstadius (2014) use such crowdsourced classification of social media messages to train supervised classifiers that are applied to a message stream. A similar approach that combines automated and human classification is pursued by Ostermann (2015). The author uses crowdsourcing for the supervision of machine learning classification and regression tasks in order to process geo-social media streams.

Digitization is the second level of crowdsourcing geographical information in our typology. Digitization describes the process of creating new geographic objects, based on existing geographical information. Such spatial objects, representing real-world objects or features, have a topology (which can be points, lines or areas) and must be located in space (e.g. using a reference coordinate system). This level requires some experience and knowledge in using mapping tools or GIS software. Crowdsourced digitization of geographical information is also usually based on the contents of a single information source. Although the automated extraction of geographic objects and features from geographical information is a field of vast research, the results of automated approaches often require detailed knowledge about the characteristics of the features and the given input datasets.

Several past projects relied upon crowdsourcing techniques to digitize geographical information. After the 2010 Haiti catastrophe, crowdsourced digitization of buildings was used for the rapid analysis of damage caused by the earthquake based on satellite imagery (Barrington et al., 2011). Hillen & Höfle (2015) present a crowdsourcing approach to digitize buildings from earth observation data incorporating the reCAPTCHA concept. The ForestWatcher citizen science project asked users to draw polygons for regions that show deforestation (Arcanjo et al., 2016). The OpenStreetMap community and the Humanitarian OpenStreetMap Team play a major role in the domain of crowdsourced digitization of geographic information. In the field of social media digitization, e.g. creation of a point object, can base on the content of an image or text message. Ostermann (2015) point out that even if more information comes geo-referenced digitization of social-media streams will remain necessary because content location and the information's origin may be different.

Conflation is proposed as the third level of crowdsourcing geographical information, and it consists of the integration of geographic information representing the same real-world object into a consistent representation. In the domain of geographic information science conflation refers to this process of combining geographic information from overlapping sources, whereas other disciplines may also refer to this process using the terms fusion or integration. This task has a high complexity and requires more advanced knowledge and skills. Nevertheless, the conflation of geographical information shows great potential and value within the fields of disaster management and humanitarian aid. Since this task is based on the content of several information sources it requires the interpretation of the context of the information. Automated information fusion still faces many limitations, especially when such different data sources and types are concerned as those used in disaster management.

Only few projects have used crowdsourcing techniques to conflate geographical information so far, and we could find only one example for the literature for this level of crowdsourcing. Soden et al. (2014) describe the work of the open cities Kathmandu project, which has developed and implemented a successful approach to geographic data collection and community building. Despite this lack of projects and research studies that actually incorporate the crowdsourced conflation of geographic information, several studies and authors perform the conflation of different information types and sources and emphasize its importance. Schnebele, Cervone, & Waters (2014) show that the conflation of authoritative and non-authoritative data can augment flood extent mapping and the evaluation of transportation infrastructure. By conflating damage assessment layers along with authoritative high-resolution road network layer, road damage can be identified at the street level. In general, on-the-ground information will be generated using different devices, tools and will be recorded in different data formats. For instance, American Red Cross (2015) uses Field Papers[1], phones, large-format printed maps and GPS devices in their field-mapping campaigns. Eckle & de Albuquerque (2015) point out that local OSM contributors know their area of interest and rely upon local knowledge, while the sole basis for remote mapping, e.g. through the OSM Tasking Manager, is often satellite imagery. Conflating all these information sources could provide a huge benefit for disaster management and strong contribute towards improving situation

[1] http://wiki.openstreetmap.org/wiki/Field_Papers, 28.02.2016

awareness.

The typology presented in this section reveals three core principles of crowdsourcing geographical information for disaster management.

1. Different levels of crowdsourcing are suitable to solve different kinds of information extraction problems. In situations where there is no need to process information on a feature level, classification tasks are the best choice.
2. The benefit of using crowdsourcing methods increases with the higher levels. The conflation of geographical information is, in the general case, difficult using automated methods, and will strongly overload the resources of humanitarian organizations when using their own staff.
3. The integration of all crowdsourcing levels can help to harness the power of volunteers in a way that everyone can contribute according to his interests, skills and experience.

MISSING MAPS – MAP SOUTH KIVU PROJECT

Missing Maps is a humanitarian project founded in November 2014 by the American Red Cross, British Red Cross, Humanitarian OpenStreetMap Team and Doctors Without Borders. The goal of the project is "to map the most vulnerable places in the world". This tasks is addressed by collaborating with the OSM and HOT team which they are supporting with technological skills, tools and outreach. The OSM material can then be used by local and international organizations as well as the local public to enable emergency planning as well as disaster response.[2]

The Map South Kivu was established as part of the Missing Maps project. By creating base map information of South Kivu, Democratic Republic of Congo, this project supports the response by international and local organizations. Within the project different levels of geographical information crowdsourcing are utilized and integrated. This section provides deeper insights into this workflow and will discuss potentials as well as limitations.[3]

Workflow

In this section the overall mapping workflow used in the Map South Kivu project will be presented. Table 2 provides an overview. A detailed description of this workflow can be also found in the OpenStreetMap Wiki.[3] The workflow consists of the different levels of geographic information crowdsourcing, classified according to the typology presented in the previous section. Several tasks are assigned to each level. Each task requires one or more input data sources. To organize and structure the efforts of the volunteers some tasks use specific crowdsourcing tools like PyBossa or OSM Tasking Manager. The next sections explain the tasks of each of level of crowdsourcing.

Level	Tasks	Input Data	Tools
1: classification	imagery quality analysis	satellite imagery	PyBossa
	assess inhabited area	satellite imagery, imagery quality raster	PyBossa
2: digitization	trace roads and residential areas	satellite imagery, (inhabited area raster)	OSM Tasking Manager
	trace buildings	satellite imagery, residential area features	OSM Tasking Manager
3: conflation	capture field data	OpenStreetMap basemaps	Field Papers
	add village names	OSM features, Field Papers,	-

[2] http://wiki.openstreetmap.org/wiki/Missing_Maps_Project, 28.02.2016
[3] http://wiki.openstreetmap.org/wiki/Map_South_Kivu_tasking, 28.02.2016

		GPS tracks	
	add attributes to buildings and highways	OSM features, Field Papers, GPS tracks	-
	update geometry	satellite imagery	PyBossa

Table 2. Map South Kivu Workflow

Level 1: classification

The first step of the workflow consists of a classification of satellite imagery. When mapping South Kivu aerial imagery functions as a basic input. However two problems occur when using this data. Firstly, for some areas in South Kivu there is no aerial imagery coverage using Bing Maps or dense clouds hide the ground's surface. Secondly, land use and distribution of human settlements are heterogeneous such as some areas are covered with dense forest among a few villages and settlements appear.

Two tasks were implemented using the PyBossa framework. For the first task volunteers were asked to classify aerial imagery in four categories: no clouds, some clouds, many clouds, no imagery. The results of this first task function as an input for the second classification task. In this case, volunteers were asked whether they can see settlements or roads in the satellite imagery and classify accordingly. Since these tasks demand no experience in using GIS tools or local knowledge about the area of interest, they can be done by a huge amount of volunteers and even within a short time period. Figure 1 depicts the interface implemented.

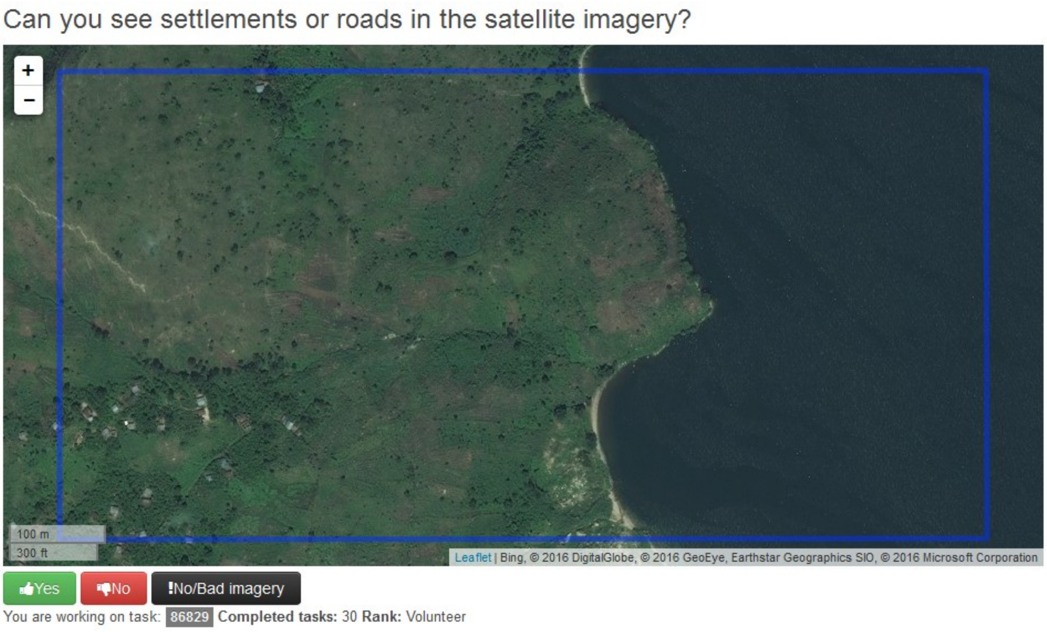

Figure 1. Crowdsourcing imagery classification using PyBossa

Level 2: digitization

The results of the imagery classification function as an input for the second crowdsourcing level. In this step the real world objects are populated into the OpenStreetMap database. Therefore the tiles that have been determined to show no cloud coverage as well as human settlements or roads were added to the OSM Tasking Manager. This tool was developed by the Humanitarian OpenStreetMap Team to facilitate the coordination of volunteers.

The overall digitization process is divided in two sub-tasks. First, roads and residential areas will be traced using satellite imagery. Once roads and residential areas in a certain area have been mapped, these features function as an input for a building mapping task. Each task is hereby provided with a description of the project context, a list of features the volunteers are asked to digitize and links to training material. For this project, specific material to support the mappers in the identification, digitization and moreover the right attribution of the road network and the residential areas was developed. These materials show common road structures and the various building structures which can be found in South Kivu. In different examples these features and different appearances according to the different satellite imageries are presented. With a large majority of the mappers contributing to the project remotely and therefore not being familiar with local features in South Kivu, these trainings provide support for the mappers and moreover ensure a certain level of data quality.[4] Experienced OSM mappers furthermore review the work of the volunteers using an integrated validation function in the Tasking Manager.

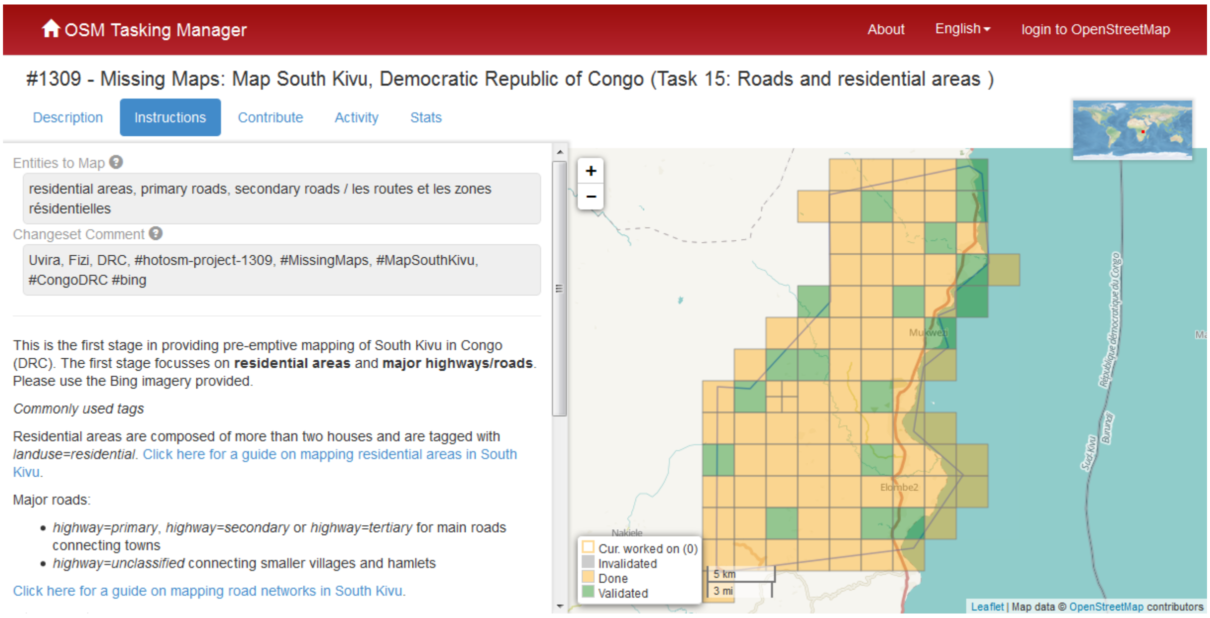

Figure 2. Crowdsourcing Digitization using OSM Tasking Manager

Level 3: conflation

After completing the remote data collection, the contributions should be validated and expanded in the field. Satellite imagery allows volunteers all around the globe to add information about visible features. However, attributes such as building functions and road names need to be added by people with local knowledge or on the ground. Also, further information on geographic features such as village names, also in native languages, can only be captured from local knowledge. Furthermore, map data needs to be updated and validated.

These challenges could be addressed with crowdsourcing tasks of level 3, since the geographic data produced in levels 1 and 2 could be conflated with different information sources to enrich the information produced by the crowd so far. For instance, by looking at satellite imagery captured in two different time stamps side-by-side, volunteers could conflate the geographic information contained in both images and compare them with the information contained in the OSM database. In this manner, OSM could be kept as an up-to-date information source for disaster management for a sustained period of time.

Nevertheless, in the current state of the project, this level of crowdsourcing is not yet used. One of the main reasons for this is the lack of supporting tools to organize and structure conflation tasks for crowdsourcing. In this manner, conflation tasks today can only be executed manually by a few highly experienced and skilled volunteers, who employ advanced GIS tools and knowledge to compare and conflate different information sources.

[4] http://wiki.openstreetmap.org/wiki/RemoteMappingGuideSouthKivu, 28.02.2016

DISCUSSION AND CONCLUSION

In this paper, the Map South Kivu was used as an exemplary crowdsourcing project in which the task of mapping a whole region was initially approached. This project was analyzed using the typology of geographic information crowdsourcing proposed in this paper, so as to shed light on how crowdsourcing tools and techniques were used in this particular project, as well as on the potentials and limitations of this usage.

A research question which still needs to be addressed is the quality of the microtasking classification of satellite imagery, that was made without reference data. If the crowdsourced approach can be shown to provide a sufficient level of data quality, this workflow could be applied to numerous further regions are currently lacking detailed maps. Thereby, other vulnerable regions could be easier provided with humanitarian aid. Apart from that, map data enables development. Crowdsourcing projects do not only profit from the local knowledge, they can moreover support community engagement as the local public is given a chance to participate in the data acquisition. This offers many opportunities especially for citizens in countries without access to official map material.

With the presented providing a great potential for adaptation, the combination of crowdsourced tasks could be applied to various regions with heterogeneous given conditions and requirements. While the Pybossa tool and the OSM Tasking Manager were well suited to address classification and digitization tasks, there is a lack of crowdsourcing tools to support the conflation of geographical information. This is an important topic to be addressed in future research endeavors.

REFERENCES

American Red Cross. (2015). Case Study : Lessons learned from community mapping in urban and rural areas in East Africa and Bangladesh, (September), 1–6. Retrieved from http://osmlab.github.io/basket/assets/Missing_Maps_community_mapping_sep16.pdf

Arcanjo, J. S., Luz, E. F. P., Fazenda, Á. L., & Ramos, F. M. (2016). Methods for evaluating volunteers' contributions in a deforestation detection citizen science project. *Future Generation Computer Systems*, *56*, 550–557. doi:10.1016/j.future.2015.07.005

Barrington, L., Ghosh, S., Greene, M., Har-Noy, S., Berger, J., Gill, S., … Huyck, C. (2011). Crowdsourcing earthquake damage assessment using remote sensing imagery. *Annals of Geophysics*, *54*, 680–687. doi:10.4401/ag-5324

Chan, J., Crowley, J., Elhami, S., Erle, S., Munro, R., & Schnoebelen, T. (2013). Aerial Damage Assessment Following Hurricane Sandy. Retrieved from http://idibon.com/wp-content/uploads/2013/05/Sandy_After_Action_Report1.pdf

Eckle, M., & de Albuquerque, J. P. (2015). Quality Assessment of Remote Mapping in OpenStreetMap for Disaster Management Purposes. *Proceedings of the ISCRAM 2015 Conference - Kristiansand, May 24-27*, 1–8.

Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. *Crowdsourcing Geographic Knowledge*. Retrieved from http://link.springer.com/chapter/10.1007/978-94-007-4587-2_7

Hillen, F., & Höfle, B. (2015). Geo-reCAPTCHA: Crowdsourcing large amounts of geographic information from earth observation data. *International Journal of Applied Earth Observation and Geoinformation*, *40*, 29–38. doi:10.1016/j.jag.2015.03.012

Horita, F. E. A., Degrossi, L. C., Assis, L. F. F. G., Zipf, A., & de Albuquerque, J. P. de. (2013). The use of Volunteered Geographic Information and Crowdsourcing in Disaster Management: a Systematic Literature Review. In *Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago Illinois, August 15-17, 2013* (pp. 1–10). Atlanta, GA, USA: AIS.

Imran, M. ., Castillo, C., Lucas, J. ., Meier, P., & Rogstadius, J. . (2014). Coordinating human and machine intelligence to classify microblog communications in crises. In *ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management* (pp. 712–721). Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84905845537&partnerID=40&md5=fef2f1230eb28cf458a2696d329b9007

Liu, S. B. (2014). Crisis Crowdsourcing Framework: Designing Strategic Configurations of Crowdsourcing for the Emergency Management Domain. *Computer Supported Cooperative Work (CSCW)*. doi:10.1007/s10606-014-9204-3

Ostermann, F. (2015). Hybrid geo-information processing : Crowdsourced supervision of geo - spatial machine learning tasks. In *Proceedings of the 18th AGILE International conference on geographic information science* (p. 4).

Quinn, A. J., & Bederson, B. B. (2011). Human computation. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*, 1403. doi:10.1145/1978942.1979148

Schnebele, E., Cervone, G., & Waters, N. (2014). Road assessment after flood events using non-authoritative data. *Natural Hazards and Earth System Science*, *14*(4), 1007–1015. doi:10.5194/nhess-14-1007-2014

Soden, R., Budhathoki, N., & Palen, L. (2014). Resilience-Building and the Crisis Informatics Agenda: Lessons Learned from Open Cities Kathmandu. *Proceedings of the 11th International ISCRAM Conference – University Park, Pennsylvania, USA, May 2014*, (May), 1–10. Retrieved from https://www.cs.colorado.edu/~palen/palen_papers/SodenBudhathokiPalen-ISCRAMKathmandu.pdf

Vieweg, S., Castillo, C., & Imran, M. (2014). Integrating Social Media Communications into the Rapid Assessment of Sudden Onset Disasters. *Social Informatics*, 444–461. doi:10.13140/2.1.3437.9524