Crowdsourcing Linked Open Data for Disaster Management

Jens Ortmann[1], Minu Limbu*[1], Dong Wang[1], and Tomi Kauppinen[1]

Institute for Geoinformatics, University of Muenster, Germany
(jens.ortmann, wang.dong, tomi.kauppinen)@uni-muenster.de
minulimbu@gmail.com

Abstract. This paper shows how Linked Open Data can ease the challenges of information triage in disaster response efforts. Recently, disaster management has seen a revolution in data collection. Local victims as well as people all over the world collect observations and make them available on the web. Yet, this crucial and timely information source comes unstructured. This hinders a processing and integration, and often a general consideration of this information. Linked Open Data is supported by number of freely available technologies, backed up by a large community in academia and it offers the opportunity to create flexible mash-up solutions. At hand of the Ushahidi Haiti platform, this paper suggests crowdsourced Linked Open Data. We take a look at the requirements, the tools that are there to meet these requirements, and suggest an architecture to enable non-experts to contribute Linked Open Data.

1 Introduction

The world has recently seen a number of environmental disasters, both natural and man-made. The most severe ones in the last two years were the earthquake that hit Haiti in January 2010 and the earthquake that hit Japan in March 2011. In both cases, information technologies contributed to increasing global awareness of these disasters. Modern communication channels and services have enabled people around the world to spread information, thereby changing the landscape of geographic information.

Crucial parts of disaster management are the acquisition, assessment, processing and distribution of information. In the mentioned disasters, so-called crowdsourced [9] information was massively generated. Crowdsourced information is information that is generated by a large heterogeneous crowd: People in the disaster-struck area share reports online; people all over the world help to

* Minu Limbu has worked more than four years with the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) in the field of Information Coordination/Management and was in Haiti for the International Federation for Red Cross and Red Crescent Movement (IFRC) Shelter Cluster immediately after the earthquake in January 2010.

process and distribute information. Crowdsourcing has proven to be an efficient approach to quickly generate huge amounts of near real-time data on a given situation [12,15]. We conducted a short survey targeting disaster management experts that revealed persisting problems of data processing and information integration. Relief organizations working in the disaster-struck area were unable to cope with unstructured and unconfirmed reports. They simply lack the time to integrate these data into existing information systems. This is not to say that crowdsourced data was not used at all. Yet, its potential is far bigger than the actual impact it made, especially in the early phase of a disaster response.

This paper describes an approach based on Linked Open Data [5] to alleviate the integration problems of crowdsourced data and to improve the exploitation of crowdsourced data in disaster management. We suggest to engage people in processing unstructured observations into structured RDF[1]-triples according to Linked Open Data principles. Thereby, a substantial part of the integration problem is left to be solved by the crowd.

Hence, the goal of the efforts described in this paper is nothing less than to allow the use of linked open crowdsourced data for disaster management. This will increase the impact of crowdsourced data in disaster management and help humanitarian agencies to make informed decisions.

In Sect. 2 we shortly tell the story so far of crowdsourced data in disaster management and of Linked Open Data in disaster management. We then (in Sect. 3) walk through the envisioned architecture and discuss the requirements. Finally, the conclusion is drawn in Sect. 4.

2 The Story So Far

2.1 Crowdsourcing in Disaster Management

In this subsection we summarize the results of a small survey that we conducted with 14 experts in disaster management, as well as the results of Zook et al. [15].

Our survey comprised seven multiple-choice questions and one free text field for comments. Two questions assessed the participants' background, five questions targeted their experience with crowdsourcing services. The questionnaire was answered by experts from, including but not limited to, agencies like the United Nations, Red Cross, non-governmental organizations and donor communities. We asked the participants about their awareness of certain platforms that allow the crowd to contribute or process information. Fig. 1 shows the results for Twitter[2], Google Map Maker[3] (GMM), Open Street Map[4] (OSM) and Ushahidi[5]. The two mapping services GMM and OSM are know by about half the participants and also largely used when known. OSM and GMM provide

[1] Resource Description Framework (RDF), see `http://www.w3.org/RDF/`
[2] `http://www.twitter.com`
[3] `http://www.google.com/mapmaker`
[4] `http://www.openstreetmap.org/`
[5] see for example the Ushahidi platform for Haiti: `http://haiti.ushahidi.com`

structured map-based information and an immediate visualization. In the case of the Haiti Earthquake these two platforms were a big success in terms of contributions and still a success in terms of use [15]. However, there exist compatibility problems between OSM and GMM and with other Geographic Information Systems (GIS) [15]. One of the authors witnessed first-hand how PhD students at the geography department at the University of Buffalo (USA) mapped damaged buildings in their desktop GIS, creating layers that were compatible with their research group's disaster response efforts.

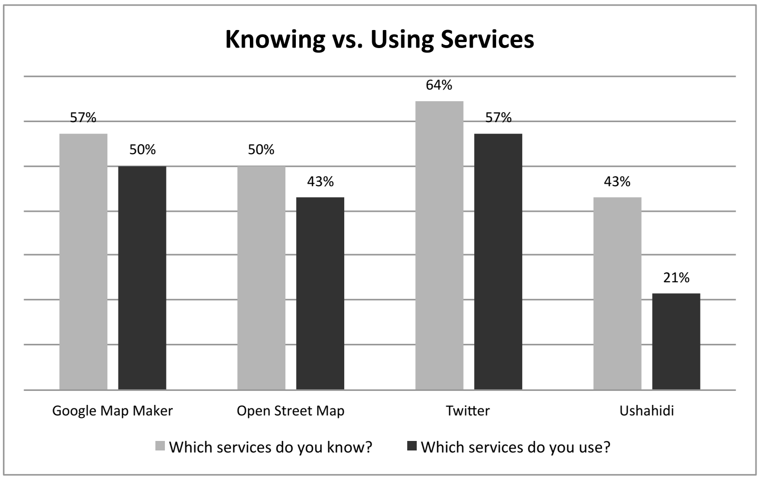

Fig. 1. Bar chart comparing knowledge of a service with its usage. All values are percentages of the 14 participating experts.

Twitter and Ushahidi provide natural language reports along with structural elements to capture formal aspects like time, location and tags or categories. Twitter is widely used to broadcast short messages. The survey responses surprisingly indicate that disaster managers use Twitter more than other platforms. However, one particular expert stated that Twitter is used mostly to broadcast information rather than capturing tweets from the crowd. Hence, the indicated usage does not necessarily reflect the efficiency of Twitter as information source.

Ushahidi is the most recent service. It is remarkable that nearly half the survey participants are aware of it. Unfortunately, Ushahidi shows the biggest drop from knowing to using the service. Out of six experts that knew Ushahidi only three answered that they actually use it. It is worth noting that one of the experts also commented that during early stages of a disaster, humanitarian workers are "too busy to think about such platforms during emergency response". Another disaster expert noted that crowd information have heterogeneous and incompatible formats making it very difficult to integrate them into humanitarian agency specific information systems.

Morrow et al. [10] identified a general inconsistency between aggregated data in Ushahidi and requirements of relief organizations. Our survey revealed that many information managers see problems mostly in uncertainty (57% of the experts), trust (50%), and semantic[6] problems (50%). On the upside, disaster managers acknowledged that crowdsourced information is collected near real-time and for free. Furthermore, experts see the greatest need for improvement in filtering of information (64%), training of volunteers (57%), ranking of relevant information (50%), structural compatibility (43%) and compliance with standard terminologies (36%).

Currently, two worlds of information infrastructures exist in parallel. On the one hand, relief organizations have their own information systems. On the other hand, ad-hoc information infrastructures emerge in the social web, which are mostly fed by crowdsourced data. Integration across these two worlds is only possible manually, and in the case of Haiti a full integration did not take place. Yet, despite the flaws of crowdsourced information, many disaster managers are willing to learn how crowdsourced information can be efficiently and effectively integrated into decision making processes.

2.2 Linked Open Data

In this section we outline the choices made to create Linked Open Data[7] and give examples. Linked Open Data [2,5] is about using web techniques to share and interconnect pieces of data by following a few simple principles. Linked Open Data builds on Semantic Web technologies, wherein data is encoded in the form of <subject,predicate,object> RDF-triples. In the case of Haiti Data, real-time information about a crisis such as Twitter messages, news, articles and weather forecasts are identified, accessed and linked through URIs.

In disaster management, authorities like the relief organizations try to publish their data to people. Usually, these data come in non machine-understandable formats as PDF or CSV, or stored in different information systems. In some sense, these data are all Open Data, as they can be found and used somehow, but they are hard to access without knowing how the CSV-files should be interpreted. Since Linked Open Data is used as a framework of loose integration of data [7], it provides a method to make data really open by interlinking the data sources flexibly in the Semantic Web. In this work, we used the Ushahidi reports from Haiti as input and triplified them into RDF to make the information in the report accessible and linked.

It is considered that the World Wide Web has an abundance of data resources related to disaster management, either authoritative data possessed by a relief organization or crowdsourced data in the social web. The advantages of Linked Open Data as easy manipulation and loose integration make it a potential way to interlink the observed data from volunteers to existing systems.

[6] we asked whether they see "Difficulties in interpreting the information".
[7] We do not distinguish between Linked Data and Linked Open Data here.

3 The Lifecycle of a Crowdsourced Emergency Report

In the lifecycle of a human observation we identified three different personas. There is Jean[8], the local observer, who is immediately affected by the earthquake. Then, there is Desiree[9], a web-user with some knowledge of Linked Open Data. Finally, there is Paul[10], the information manager working for a relief organization.

Jean, the Local Observer. The people affected by the earthquake can be seen as a set of human sensors [8]. Equipped with mobile phones or access to the internet, this set of sensors turns into one big human sensor network.

Jean lives in Jacmel, a town in the department Sud-Est in Haiti. To communicate his observations there exist several channels. Maybe the most well-known channel is Twitter. With hashtags (like #haiti), the Twitter message can be tagged. In recent disasters people used hashtags to mark their messages as related to the disaster. This facilitated the triage of Twitter messages. Other examples of communication channels are websites that allow reporting or special emergency report numbers that mobile providers made available. To send this message Jean needs the following:

1. A device to access a communication network.
2. A communication network to access services.
3. A service that allows sharing human observations.

We have hardly any influence on meeting Jean's needs, only the service can be provided externally. There are specific platforms such as Ushahidi, but also more general solutions like Twitter or Facebook can be used.

Jean sends a message via Twitter, to report shortage of food, water and medication. Jean wrote the message in French Creole, his native language:

> Nou bezwen aide en urgence nou nan place en face palais justice la .gen anpil ti bb, nou bezwen dlo , mangé , médicament[11]

Reporting Platform. Jean's message is broadcast on Twitter. Based on the location information that Twitter provides, the message is identified as coming from Haiti. A volunteer from the Ushahidi Haiti team enters the message into the system [10]. Desiree, who knows some French Creole and English, translates the message:

> We need help.We are located at the place in front of the Palais de Justice There' s a lot of babies.We need water,food and medication.

[8] Jean can be found at `http://personas.mspace.fm/wiki/Jean`
[9] Desiree can be found at `http://personas.mspace.fm/wiki/Desiree`
[10] Paul can be found at `http://personas.mspace.fm/wiki/Paul`
[11] This is a message taken from Ushahidi, it can be found at: `http://haiti.ushahidi.com/reports/view/3815`

The message can be interpreted in terms of the Ushahidi schema. Ushahidi uses ten fields to describe a report. The fields contain the message, a title, the date, the location (place name and coordinate) and categories. The categories are defined by the Ushahidi Haiti team, each report can be in one or more categories. The message above was put into the categories "Water Shortage", "Food Shortage" and "Hospital Operating". Additionally, Ushahidi uses two fields to flag approval and verification. The reporting platform Ushahidi makes the reports available as CSV file and as RSS feed. Additionally, Ushahidi publishes an interactive map of reports online.

Paul, the Information Manager. Paul, an Information Manager from an emergency cluster[12] is overwhelmed by the requirement to acquire and process information for both local office and the headquarters.

Information managers from a humanitarian agency struggle to gather ground reality information. To support timely informed decision making process and save more lives, the primary task of Paul as humanitarian cluster information manager is to gather information about the following fundamental questions [14]:

- What type of disasters occurred when and where?
- How many people are affected/have died/are injured?
- Which humanitarian agency is currently working in the region?
- Which agency is doing what kind of humanitarian response, where and when?
- What are the most urgent life saving humanitarian needs and what are the gaps that need urgent attention?

However, even though Paul is aware of the Ushahidi service for Haiti, Paul simply does not have the time to integrate Ushahidi's CSV files or RSS feeds into his information system. Furthermore, he struggles with the meaning of categories and the problem of trusting this new and rather unknown information source.

The lifecycle up to now reflects the situation in Haiti in January 2010. Fig. 2 wraps up the lifecycle of an emergency report mediated through Ushahidi.

Linked Open Data. To overcome Paul's problems, the authors transformed the Ushahidi Haiti dataset into Linked Open Data. A simple Java program based on JENA[13] read the CSV file and wrote the RDF graph. The triples employ standard vocabularies like Dublin Core[14] where possible. When there are no vocabularies to express required information, we created new ones[15]. An example of a triplified report is depicted in Listing 1.

[12] Emergency Clusters eg. Food, Shelter or Health `http://www.humanitarianinfo.org/iasc/pageloader.aspx?page=content-focalpoint-default`
[13] see `http://jena.sourceforge.net/`
[14] The Dublin Core Metadata Element Set: `http://dublincore.org/documents/dces/`
[15] cf. `http://observedchange.com/moac/ns/` and `http://observedchange.com/tisc/ns/`

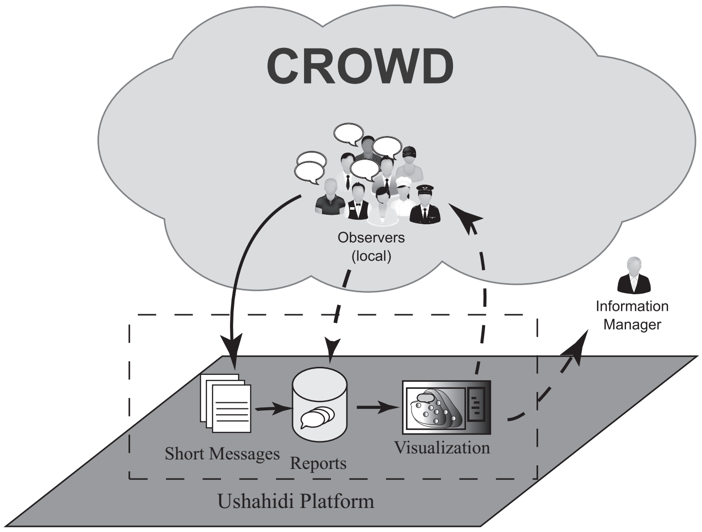

Fig. 2. A report by local observers is entered into the Ushahidi database and subsequently considered in map-visualizations. An information manager has access to the reports and to the visualizations, but can not integrate these sources into his information system. The dashed lines inidcate channels that are not fully utilized.

Listing 1. The tripleset for an Ushahidi Report in Turtle form. We left out the prefix specification here.

```
@base <http://haiti.ushahidi.com/reports/view> .
</3815> a </UshahidiReport> ;
      tisc:temporalDistance 9 ;
      dc:subject moac:WaterShortage ;
          moac:FoodShortage ;
          moac:HospitalOperating ;
      dc:coverage "Coordinates are for Palais de Justice in
          Jacmel" ;
      geo:latitude 18.233695 ;
      geo:longitude -72.536217 ;
      sioc:content "We need help.We are located at the place in
           front of the Palais de Justice There\\\\' s a lot of
           babies.We need water,food and medication. \\t

Nou bezwen aide en urgence nou nan place en face palais
    justice la .gen anpil ti bb, nou bezwen dlo , mang?, m?
    dicament Time:2010-01-21 19:58:06" ;
      moac:verified false ;
      moac:approved true ;
      dc:title "Water, Food, Medicine Needed at Palais de
          Justice, Jacmel" ;
      dc:date "2010-01-21T19:58:00-05:00"^^xsd:dateTime .
```

The reports are made available online on a website and can be inserted into an etherpad[16] for collaborative editing. The RDF content of the etherpad can be retrieved directly from other websites through the export plain text function.

Desiree, the Web User. Desiree studies Geoinformatics, and tutors a course on Linked Open Data. On Twitter she heard about the needs of the people in Haiti. She also heard of the information integration problems and of our effort to crowdsource Linked Open Data. She downloads a report and starts editing it. She detects that the category "Hospital Operating" was assigned falsely, she modifies the respective triple to point to the category "Medical Equipment And Supply Needs". She then introduces two new triples. Desiree's new triples are shown in Listing 2.

Listing 2. Two additional triples that enrich the original tripleset of report 3815.

```
</3815> tisc:locatedAt <http://www.geonames.org/3723779/jacmel
    .html> .
    dc:subject  <http://dbpedia.org/page/Jacmel> .
```

These triples lead to more information about Jacmel. Among others, the population figures and the current mayor are available there. With her contribution Desiree helps to extend the triples with information that cannot be harvested automatically from Ushahidi. Furthermore, the introduction of an additional crowdsourcing layer for processing adds some quality control.[17] The misclassification of medical needs to "hospital operating" in this example can have fatal consequences in practice. There are the following requirements on Desiree's task:

4. Access to the Internet.
5. Access to triples with the basic information.
6. Basic knowledge about triples, linked data and RDF.
7. Basic knowledge about disaster management and its terminologies.
8. A service to edit and validate the triples.
9. Vocabularies to edit and create triples.

We take the access to the Internet for granted. Desiree is not in Haiti but processes the report remotely. Access to the triples can be ensured through a website that makes the reports available in triple format. Desiree needs some basic knowledge to create RDF triples. This limits the crowd. However, many crowdsourced processing tasks require certain skills. We think that given an initial set of triples and a short overview of basic turtle syntax, the addition of triples in turtle format is not expected too much of a proficient web user. The point to consider is the additional interpretation that is made by us here when creating the basic set of triples. To describe the report, we use predicates and

[16] see `http://etherpad.org/` for more information.

[17] Desiree's tasks involves translation between natural languages and from a natural to a formal language. This can be considered human computation [13]. In fact, many of the tasks described here as crowdsourcing may lay at the intersection of crowdsourcing and human computation as described in [13].

objects that come from shared vocabularies. There exist several well-established vocabularies like Dublin Core or SIOC to describe the formal aspects of the report, but to annotate the content Desiree needs domain specific vocabularies or ontologies. To our knowledge, no vocabularies or ontologies exist that specifically allow describing observations made by victims of disasters. The Ushahidi team came up with a taxonomy of incidents. To describe these incidents in triple format we translated the Ushahidi categories into an RDF vocabulary. However, as also [6] pointed out, further work is required to establish vocabularies and ontologies for disaster management. The lack of vocabularies in this domain constitutes one obstacle to information integration in crisis management. Etherpads are one option to edit triples collaboratively. Etherpad is an open source text editing service. It is easy to handle and the content, e.g. the triples, can be directly accessed from a static URL that returns the content of the etherpad as plain text[18]. Etherpads also allow creating different versions that can be individually referenced. However, so far Etherpads only allow writing plain text, there is no syntax highlighting or validation of triples. To validate the edited tripleset Desiree has to resort to another service, for example sindice inspector[19].

Hence, editing and validating triples is possible, but there is no integrated service that meets all requirements. Furthermore, necessary vocabularies are still missing.

Visualization. With the reports given in triple form we can take the next step. We have set up a website[20] that uses SIMILE Exhibit[21] to visualize the triples in different forms, for example as timeline, on a map or as thumbnails. Fig. 3 shows a screenshot of a map with report locations.

The illustration shows only one way of using the reports in Linked Open Data. There are not only many more tools for analysis and visualization available, but also further Linked Open Data can be integrated with the Ushahidi reports. Linked Open Data is supported by a growing community of users. The processing, analysis and visualization of Linked Open Data can be seen as an additional crowdsourcing task. However, this requires knowledge of web frameworks that exceeds the typical web user's expertise.

Paul, the Information Manager. Given the reports in Linked Open Data, Paul can access the reports far easier. The linked structure allows for an easier filtering of information, by category, by time, by location and so on. Paul works in an emergency cluster for medical support, the change that Desiree made to the report is important when he decides to distribute the resources he has.

Within 24 hours of the disaster, Paul needs to provide evidence based "humanitarian needs and gaps" figures in a disaster affected area. The chances for

[18] for example `http://pad.ifgi.de/ep/pad/export/haiti-ushahidi-report3815/latest?format=txt`
[19] see `http://inspector.sindice.com/`
[20] see `http://www.observedchange.com/demos/linked-haiti/`
[21] see `http://www.simile-widgets.org/exhibit/` for more information.

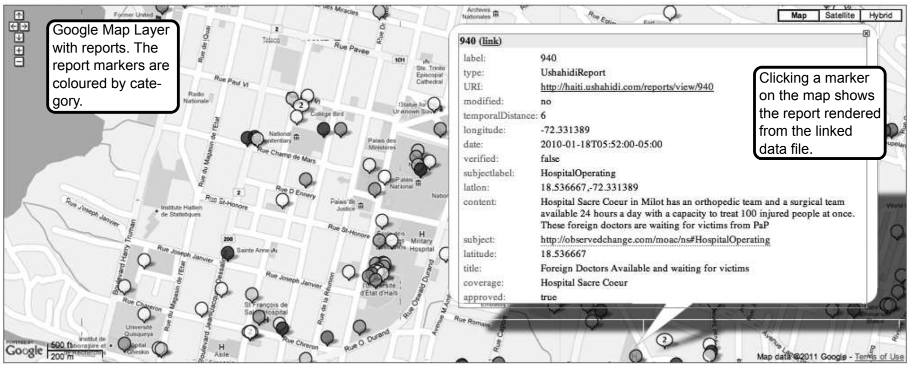

Fig. 3. A screenshot of the SIMILE Map Widget. The map illustrates the reports, coloured by category. A click on the marker shows the full tripleset for the report.

a quick fact finding rapid assessment [11] are very low. The reports by local observers are coming in nearly immediately after the earthquake.

For instance, Paul finds out that he can identify at least what the urgent life saving humanitarian concerns are and where. As information manager Paul knows and understands that trust and quality of crowd information are a concern. However, such quick access to crowd information and visualization within the first 2-3 days of a disaster can help better plan the fact finding mission or possible future assessments to further confirm the humanitarian needs and possible intervention in the areas. Paul's basic requirements are:

10. A portal that gathers and integrates crowdsourced information sources.
11. A way to connect his information system to information sources in RDF.

We have exemplary set up a facet-based browsing facility to access the triples we used at `http://www.observedchange.com/demos/linked-haiti/`. The idea is that it is easily possible to extend the portal and to connect it to further information sources. However, the problem remains unsolved on the side of Paul's humanitarian information system. The problem has been recognized [6] and first solutions (e.g., [1]) are under development.

Summary. The new lifecycle of Jean's reports is illustrated in Fig. 4. An additional step of crowdsourcing Linked Open Data refines the data. Linked Open Data allows not only Paul to access the data according to well-established principles, but it also enables members of a world-wide Linked Open Data community to create mash-ups of various sources and tools.

4 Conclusion

This paper suggested crowdsourcing Linked Open Data as the next step towards a full exploitation of crowdsourced information in disaster management. Due to

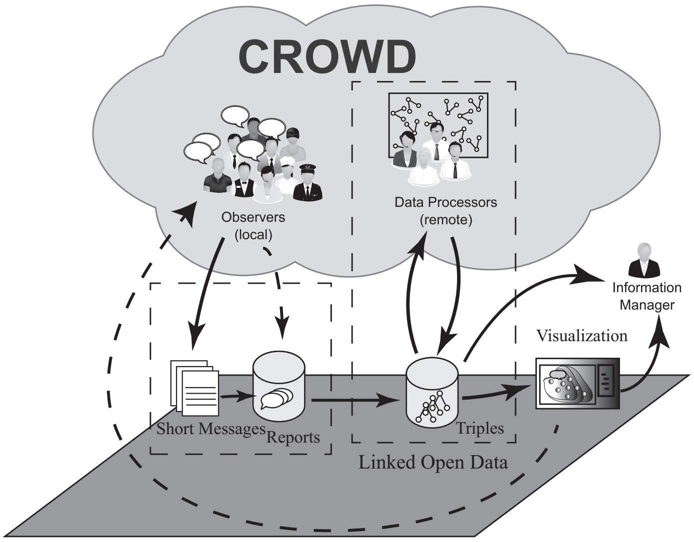

Fig. 4. Workflow including crowdsourced Linked Open Data. An intermediate crowdsourcing task of enriching triples is implemented now. The dashed lines indicate channels that are available but not at the focus of this paper.

the amount of information required and the short time available, crowdsourcing is a promising candidate for disaster management. Linked Open Data serves as common exchange format. Its simplicity and the large community behind it make it well suited for crowdsourcing efforts. The suggested approach has the potential to solve the problems of structural and semantic interoperability [4]. Nonetheless, the issues of uncertainty and trust remain unsolved in our example.[22] However, in the response phase immediately after a disaster occurs, when there is no information available, trust and uncertainty issues of crowdsourced information are accepted. Furthermore, to the authors' knowledge, no standardized and formalized vocabulary for disaster management exists. Finally, it is up to the relief organizations to make their systems and data ready for Linked Open Data.

Acknowledgements

This work has been partly supported through the International Research Training Group on Semantic Integration of Geospatial Information by the DFG (German Research Foundation, GRK 1498), by the China Scholarship Council (CSC) and through the Erasmus Mundus Master Program in Geospatial Technologies (contract no. 2007-0064/001/FRAME MUN123).

[22] see [3] for an example of a trust an reputation system for human observations of water availability in dry regions.

References

1. Babitski, G., Bergweiler, A., Grebner, O., Oberle, D., Paulheim, H., Probst, F.: Soknos using semantic technologies in disaster management software. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) The Semanic Web: Research and Applications. Lecture Notes in Computer Science, vol. 6644, pp. 183–197. Springer Berlin / Heidelberg (2011)
2. Berners-Lee, T.: Linked Data. W3C (2006), http://www.w3.org/DesignIssues/LinkedData
3. Bishr, M.: A Trust and Reputation Model for Evaluating Human Sensor Observations. Ph.D. thesis, University of Muenster, Germany, Muenster, Germany (2011), url = http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-6032/diss_bishr.pdf
4. Bishr, Y.: Overcoming the semantic and other barriers to gis interoperability. International Journal of Geographical Information Science 12(4), 299–314 (1998)
5. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data The Story So Far. International Journal on Semantic Web and Information Systems 5(3), 2–9 (2009)
6. Di Maio, P.: An open ontology for open source emergency response systems, see http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.93.1829
7. Feridun, M., Tanner, A.: Using linked data for systems management. In: NOMS'10. pp. 926–929 (2010)
8. Goodchild, M.: Citizens as sensors: the world of volunteered geography. GeoJournal 69(4), 211–221 (2007)
9. Howe, J.: The rise of crowdsourcing. Wired magazine 14(6), 1–4 (2006)
10. Morrow, N., Mock, N., Papendieck, A. Kocmich, N.: Independent evaluation of the ushahidi haiti project. Tech. rep., The UHP Independent Evaluation Team (2011)
11. OCHA: Assessment and classification of emergencies (ace) project. Tech. rep., United Nations Office for the Coordination of Humanitarian Affairs (2009)
12. Okolloh, O.: Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. Participatory Learning and Action 59(1), 65–70 (2009)
13. Quinn, A.J., Bederson, B.B.: Human computation: a survey and taxonomy of a growing field. In: Proceedings of the 2011 annual conference on Human factors in computing systems. pp. 1403–1412. CHI '11, ACM, New York, NY, USA (2011)
14. UNICEF: Emergency Field Handbook: A guide for UNICEF staff. The United Nations Childrens Fund (UNICEF) (2005), url = http://www.unicef.org/publications/files/UNICEF_EFH_2005.pdf
15. Zook, M., Graham, M., Shelton, T., Gorman, S.: Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. World Medical & Health Policy 2(2), 2 (2010)