Using Linked Data and LOD2 stack in Emergency Management

Vuk Mijović, Lydia Kraus
University of Belgrade, Mihajlo Pupin Institute
Volgina 15
11060 Belgrade, Serbia
+381(11)6773149

Abstract – *This paper investigates how Linked Data can be applied to emergency management. Intended to IT engineers and safety experts, this paper shows how Linked Data is used today in emergency management systems, as well as how LOD2 stack, a collection of tools for managing the life-cycle of Linked Data, can be used to improve the learning curve of new personnel and increase situational awareness during emergencies.*

Keywords - Emergency management, Linked Open Data.

1. INTRODUCTION

With numerous crises and emergencies taking place within the last years, such as natural hazards and terrorist attacks, the European Commission has identified the need to push the performance of research in the emergency management field. Thus, within the Sixth and Seventh Framework Program (FP6 and FP7), many projects addressing the field of emergency/disaster/crisis management were/are undertaken [1]. These projects use a variety of approaches to solve different issues that have to be handled in emergency management, such as collaboration between stakeholders, provision of communication infrastructure, dealing with uncertain information and sharing of a common picture of the emergency situation.

Emergency management is a complex field, with a lot of different active and passive stakeholders that need to perform different tasks in a short time. The emergency management process consists of four main phases: prevention, preparedness, response and recovery. During disaster/emergency response, responders from different organizations such as emergency medical services, fire fighting services, police, local/regional/state authorities, NGOs and volunteers have to collaborate. However, these organizations have different communication and coordination systems and organizational structures, providing obstacles in collaboration [1]. Also, the stakeholders often lack situational awareness and don't share a common picture of the situation.

It was already recognized that Linked Open Data offers a high potential to improve siuational awareness and collaboration during disaster and emergency response [2]. Therefore, we will examine how these goals can be achieved by using the LOD2 stack, which comprises a number of tools for managing the life-cycle of Linked Data, including:

- CKAN - registry or catalogue system for datasets or other "knowledge" resources
- SILK - a framework that supports data publishers in setting explicit RDF links between data items within different data sources
- DBPedia - a community effort to extract structured information from Wikipedia and to make this information available on the Web
- OntoWiki - enables intuitive authoring of semantic content, fosters social collaboration aspects by keeping track of changes, allowing to comment and discuss every single part of a knowledge base
- PoolParty - a thesaurus management system and a SKOS editor for the Semantic Web including text mining and linked data capabilities
- LinkedGeoData - an effort to add a spatial dimension to the Web of Data / Semantic Web

This paper identifies different categories of linked data based EM (Emergency Management) information systems in Section 2, then gives the analysis of current approaches to using Linked Data in emergency management in Section 3, reports on the opportunities to use some of the LOD2 stack components in Section 4, and finally gives conclusions in Section 5.

2. CATEGORIES OF LINKED DATA BASED EM INFORMATION SYSTEMS

Through an analysis of the use of linked data in emergency management information systems, we identified four kind of emergency management systems in which linked data is already used:

1. EM information systems transforming data gathered through a conventional EM platform to RDF, however the transformation is not done on-the-fly
2. EM information systems where linked data is generated, enriched and linked on-the-fly leading to a mash up of private and public data for increased situational awareness
3. EM information systems based on the full LOD lifecycle, including not only linking but as well enrichment of information through semantic information (e.g. ontology or taxonomy) and integration of past and newly generated data
4. Social media based EM Information Systems; direct transformation of Twitter messages to RDF, analysis, enriching and linking of the RDF messages, provision of mash up

3. ANALYSIS OF DIFFERENT SYSTEMS

System described in [2] (Ortmann, Limbu, Wang & Kaupinen 2011) - The system described in [2] belongs to category 1. It is based on the Ushahidi emergency management platform [3] which gathers data from responders and people on scene. From the article it seems that this data is at the time of gathering not in RDF format. The Ushahidi platform was used during Haiti earthquake and the so collected data was transformed to RDF and matched to an ontology.

System described in [4] (Van Leeuven, 2010) - Another application of linked data in emergency management is given in [4]. The described system belongs to category 2. The author describes the application of linked open data in fire fighting. Publishing dispatching messages in RDF, enriching and linking these message with public data (e.g. information about road works), which is as well published as linked open data, enables the generation of a mashup and a map that shows important information on how to reach the scene on the fastest way. This can help the fire fighters who leave for a fire scene enormously to get an overview of the situation.

System described in [5] (Schulz, Paulheim, & Probst, 2012) - The system described in [5] belongs to category 3. It is a disaster Management Information System based on RDF and covering the whole linked data lifecycle - collecting, processing, classifying, enriching, and search. The processed and pre-filtered based on input from response personal. They authors suggest the use od DBpedia spotlight for text annotation, and DBpedia and LinkedGeoData for enriching.

System described in [6] (Borges, de Faria Cordeiro, Campos, & Marino, 2011) - The system described in [6] belongs as well to category 3. The authors suggest the use of public (open) data (e.g. road maps, construction plants, water supplies, number of hospital beds) and the combination of previous formal and current contextual knowledge for higher situational awareness.

Twitcident system - The Twitcident system , described in [7] belongs to category 4. Through the analysis of tweets in a given geographical area a list of incidents is identified. Further tweets a collected by incident and iteratively refined. Tweet analysis (content, time, type, pattern analysis) helps to further classify the incident.

The area of LOD based EM systems is an emerging research area, as all found articles about the topic were published between 2010 and 2012.

According to the examples we analyzed we agree that the use of linked data in this kind of systems brings advantages and precedes the state-of-the-art. When gathered and processed in real-time and at the same time presented in a good way to responding personnel, linked data has the potential to generate increased situational awareness.

Moreover we think that the use of real-time Twitter data combined with linked data paradigms (as introduced in [7]) is an area which should further be explored.

4. LOD2 STACK COMPONENTS AND EM

One area which was not covered by the previous approaches is text annotation. While not so important during the emergencies, it can affect the learning curve of new personnel. Many terms can be unknown to a newcomer making annotation and links to dbpedia entries very useful. Here, we will give the results obtained when using DBPediaSportlight, an LOD2 stack component, to annotate texts related to emergency management. Then, we will examine how situational awareness can be enhanced by using another LOD2 stack component called LinkedGeoData, which is an effort to add a spatial dimension to the Web of Data

4.1 THE USE OF DBPEDIASPOTLIGHT

DBpedia Spotlight[8] is a tool for automatically annotating mentions of DBpedia resources in text. If it is reliable enough, such a feature can improve the learning curve of new personnel by linking terms in EM texts used for learning to entries in DBpedia. Therefore, in order to assess its usefulness, we tested it to evaluate the annotation accuracy of the tool with different kind of EM texts, namely:

- Text 1: a formal document (an airport emergency plan - AEP),
- Text 2: a table of content of ICAO document (Convention on International Civil Aviation), and
- Text 3: a report/ newspaper article style text (retrieved from National Fire Protection

Association, Fire Investigation Summary of the Düsseldorf Airport Terminal Fire).

The Table bellow gives an overview of the test configuration. Column 1 defines the test number and column 2-4 define the values of the DBPedia Spotlight parameters (C – Confidence, CS – Contextual Score, P – Prominence). The other parameters were always set to “No common words”, “Default Disambiguation” and “Show best candidate”. Column 5-7 give an overview which tests were performed with which text.

Test #	C	CS	P	Text 1 (AEP)	Text 2 (ICAO)	Text 3 (NFPA)
1	0	0	0	X	X	X
2	0.5	0	0	X	X	X
3	1	0	0	X	X	-
4	0.5	0.5	0	X	X	-
5	0	1	0	X	X	X
6	0.3	0	0	-	-	X

Table 1 Test parameters overview

The tests were performed by putting each text into the online application, setting the parameters and then annotating the text. The measures correct, borderline, wrong and completely wrong were defined/ interpreted by the tester. Annotated words are counted as correct if the correct DBPedia resource with the correct semantic meaning of the annotation was selected for annotation. Annotated words are counted as borderline if a correct DBPedia resource was selected but the semantic meaning in this context is not correct. Annotated words are counted as wrong if a similar DBPedia resource was selected (a resource that contains for example the annotated word, but together with other words it leads to a completely distinct semantic meaning). Annotated words are counted as completely wrong if the DBPedia resource doesn’t have any relation to the annotated word. Annotations were only counted once per word, thus the percentage values refer to each annotated word not to each annotation. The results of the testing can be summarized as follows.

Text 1 had an overall length of 422 words of which a maximum number of 55 words was annotated (in test 1, including several annotations of the same word). Test results varied from 55.8 % correctly annotated words (test 5) to 67.9 % correctly annotated words. The configuration as performed in test 2 (confidence = 0.5, contextual score = 0, prominence = 0) lead to the best solution. When setting the confidence to 1 no annotations were made.

Test #	Correct	Borderline	Wrong	Completely wrong
1	56.8 %	29.7 %	11.8 %	2.6 %
2	67.9 %	21.4 %	10.7 %	0 %
3	66.7 %	26.7 %	6.6 %	0 %
4	57.1 %	32.2 %	10.7 %	0 %
5	55.8 %	26.5 %	11.8 %	5.9 %

Table 2 Testing results for text 1 (AEP)

Text 2 had an overall length of 227 words of which a maximum number of 33 words was annotated (in test 1, including several annotations of the same word). Test results varied from 66.7 % correctly annotated words (test 1 and 5) to 75 % correctly annotated words. The configuration as performed in test 2 (confidence = 0.5, contextual score = 0, prominence = 0) lead to the best solution, even though the number of correctly annotated words there is lower as in test 4, but the number of wrong annotated words is much lower than in test 4. When setting the confidence to 1 no annotations were made.

Test #	Correct	Borderline	Wrong	Completely wrong
1	66,7 %	25,9 %	7,4 %	0 %
2	66.8 %	16.6 %	16.6 %	0 %
3	0 %	0 %	0 %	0 %
4	75 %	0 %	25 %	0 %
5	66,7 %	25,9 %	7,4 %	0 %

Table 3 Testing results for text 2 (ICAO)

Text 3 had an overall length of 530 words of which a maximum number of 83 words was annotated (in test 1, including several annotations of the same word). Test results varied from 59.2 % correctly annotated words (test 1 and 6) to 100 % correctly annotated words. The configuration as performed in test 5 (confidence = 0, contextual score = 1, prominence = 0) lead to the best solution. When setting the confidence to 0.5 no annotations were made.

Test #	Correct	Borderline	Wrong	Completely wrong
1	59.2 %	10.2 %	28.6 %	2 %
2	100 %	0 %	0 %	0 %
5	100 %	0 %	0 %	0 %
6	59.2 %	10.2 %	28.6 %	2 %

Table 4 Testing results for text 3 (NFPA)

When comparing all texts it seems that medium confidence leads to the best results. Text 3, which is a general text with not too many specific words from the airport and emergency management domain, seems to produce the worst results.

4.2 THE USE OF LINKEDGEODATA

LinkedGeoData is an effort to add a spatial dimension to the Semantic Web. LinkedGeoData uses the information collected by the OpenStreetMap project and makes it available as an RDF knowledge base according to the Linked Data principles. It interlinks this data with other knowledge bases in the Linking Open Data initiative.

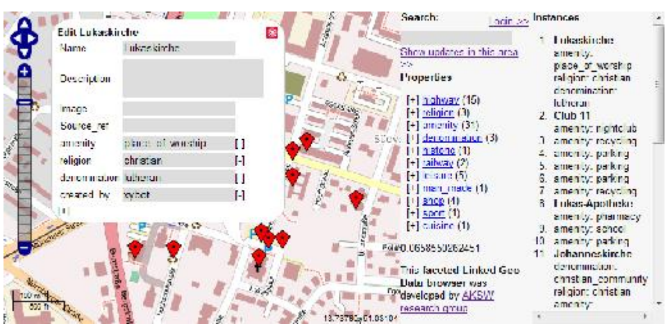

Figure 1 LGD browser demo

During emergencies operators need to be aware of the situation at hand, including:

- location of the incident(s)
- location of nearby points of interest, like hospitals, fire stations, shelters, etc.
- location of responders
- location of resources that are needed to effectively manage the emergency

Ideally, all this information should be shown on a single map. Just by looking at the LGD browser demo (see Figure 1) at http://browser.linkedgeodata.org/# it can be seen that it is quite good at managing information about various points of interest. Furthermore, it provides faceted browsing, allowing the operator to filter only some types of information, which can come in handy in situations where a quick reaction is needed. For example, in case of an emergency, like an earthquake, in order to better plan a response, the person making decisions and suggesting solutions might want to know about nearby hospitals, emergency rooms, fire stations, etc. The approach taken by LinkedGeoData where information is collected from community driven sources is good in a sense that more information is acquired this way, but it raises a question about the correctness of the data. If the data were to be used in emergency management, there would have to exist a strict authorization process.

In addition to rather static information, operators would get great benefit from being able to track dynamic information. Incidents like floods and earthquakes can result in a lot of additional incidents, e.g. collapsing of buildings. Operators need to be aware of all these problems, so a map with faceted browsing as is the case in LDG browser would be ideal. Also, the position of rescue services and any resources that are needed for dealing with the emergency at hand should be presented to the operator on a single map with an easy way to filter different kinds of information. For example, the operator might want to see the location of all incidents and the location of rescue services in order to determine which incidents will need more attention. Originally, the LGD browser was not meant to be used in this way, i.e. for tracking various resources, so an analysis will be needed in order to determine if it would be appropriate for these tasks, and what kinds of modifications or improvements would be needed.

4.3 THE USE OF OTHER COMPONENTS

For other LOD2 components such as OntoWiki, Poolparty, CKAN, or SILK it seems that they are rather meant for editing, enhancing and maintaining existing and/or new linked data than for processing linked data on-the-fly. Consequently, they are not very suitable for responding to emergencies, but they can enhance training and awareness. CKAN could be used for publishing emergency related datasets, OntoWiki for browsing emergency related information available in RDF, while Silk could be used to link data from different sources. In this way emergency experts would be up-to-date with the latest information from different sources, enabling them to prepare better training scenarios. At this time it is not clear to which extent we can leverage the LOD2 Stack in this respect, since it depends on the amount and format of currently available information. Since this is an area which is covered by almost any Linked

Data use case, at first we will focus on specifics of the emergency management use case which is using data produced in real-time to help stakeholders get a clear picture of the current situation.

5. CONCLUSIONS

At the moment there are some components in the LOD2 stack which might be used for the EM use case. DBpedia spotlight could be used to annotate words from messages and to link them to DBpedia. However, in general, the linking to DBpedia can include very general links ("Common words" problem), thus it should be used carefully or only with prior training. Using LinkedGeoData could be considered to create a spatial mash up of relevant data; however as the data in LinkedGeoData is based on OSM it has to be ensured that the needed instances are available yet and that the synchronization with OSM is sufficient. On the other hand, if using OSM data is not satisfying, one could use the possibility to manually set the graph that will be used to retrieve and display the data.

Other components were built for editing, enhancing and maintaining existing and/or new linked data, not for processing linked data on-the-fly. Therefore, they should be used to enhance training and awareness, but the applicability of these tools should be further investigated as their potential depends on the amount and format of currently available information.

The analysis presented in this paper shows that LOD2 stack components and Linked Data in general can be used to improve emergency management in different ways, from increasing the learning curve to enhancing situational awareness, and should therefore be further investigated.

6. ACKNOWLEDGMENTS

This work was supported by a grant from the European Union's 7th Framework Programme (2007-2013) provided for the project LOD2 (GA no. 257943), as well as by the Ministry of Education and Science of the Republic of Serbia under grant agreement SOFIA, no. TR-32010.

7. REFERENCES

[1] Kraus, L., Janev, V., & Vraneš, S. *"Different approaches to ICT-enhanced emergency management"*, Proc. of the 56th Conference for Electronics, Telecommunications, Computers, Automation, and Nuclear engineering (ETRAN, Zlatibor, Serbia). The ETRAN Society, 2012.

[2] Ortman, J., Limbu, M., Wang, D., Kauppinen T. *"Crowdsourcing Linked Open Data for Disaster Management"*, Proc. of the 10th International Semantic Web Conference. Bonn, Germany, 2011.

[3] http://www.ushahidi.com/, September 2012.

[4] Van Leeuven, B., *"Open Government Data for fighting fires: when no traffic rules apply"*, In: Nodalities – The magazine of the semantic web, Issue 12, Birmingham, Talis, UK, 2010.

[5] Schulz, A., Paulheim, H., Probst, F. *"Crisis Information Management in the Web 3.0 age"*, Proceedings of the 9th International ISCRAM Conference. Vancouver, Canada, April 2012.

[6] Borges, M. R. S., de Faria Cordeiro, K., Campos, M. L. M., Marino, T. *"Linked Open Data and the Design of Information Infrastructure for Emergency Management Systems"*, Proceedings of the 10th International ISCRAM Conference, Lisbon, Portugal, 2011.

[7] Terpstra, T., de Vries, A., Stronkman, R., Paradies, G.L. *"Towards a real-time Twitter analysis during crises for operational crisis management"*, Proceedings of the 9th International ISCRAM Conference, Vancouver, Canada, April 2012.

[8] https://github.com/dbpedia-spotlight/dbpedia-spotlight/, October 2012.