Linked Open Data and the Design of Information Infrastructure for Emergency Management Systems

Marcos R. S. Borges, Kelli de Faria Cordeiro, Maria Luiza M. Campos, Tiago Marino
Graduate Program in Informatics
Universidade Federal do Rio de Janeiro, Brasil
{kelli, tiago.marino, mluiza}@ufrj.br, mborges@nce.ufrj.br

Abstract— Correct information is a vital resource during disasters. Without adequate information, response actions may be ineffective. What is the source of information of emergency management systems? Besides that originated from sensors, the emergency team and the public in general, a very relevant source is government data, such as demographic and geographic data, road maps, etc. The heterogeneity of information formats is a well-known problem that affects organizations and communities that want to access public data. Today, most public agencies provide access to their data, but the great majority is unreadable by automated mechanisms. Besides, most of them do not provide a dictionary meaning for the published content. A solution to this problem is of particular importance to emergency response organizations that need access to all information available to better respond to disasters and crisis. The linked open data (LOD) initiative allows the interconnection of data, using standards in the context of the semantic web approach. In ideal conditions, government agencies publish their public data, thus allowing the use of automated data concerned consumers, whether they are other government agencies or citizens. Efforts aimed to link government data are growing in several countries around the world. This talk presents the LOD concepts and describes an architecture that uses LOD in the design of an Emergency Management System. It describes a scheme for collecting available data from government agencies, such as departments of health, transport, works, that can supply information needs during an emergency response operation.

Keywords-Emergency Management, Linked Open Data, Collaborative Decision Making

I. Introduction

Information is a valuable resource during emergencies or disasters. It is the basis for coordination and decision making during the emergency response. It is an essential aspect of an organization's ability for gaining (or losing) visibility and credibility. It has a powerful impact on how national and international resources are mobilized. Above all, information is necessary for providing rapid and effective assistance to those affected by a disaster. It is also essential for after-action analysis, evaluation, and lessons learned.

Where does information come from during an emergency situation? Many valuable sources are available and can be useful to response operations if they can be treated by information systems. However, one of the main challenges is the design of communication and information management that contributes to more effective and timely emergency response. Another challenge is the adequate organization of command and operational activities to reduce the bad impact of disasters and emergencies.

This presentation discuss a scheme for collecting available data from government agencies, such as departments of health, transport, works, that can supply information needs during an emergency response operation. The scheme also includes the treatment and integration of voluntary collaborative data, i.e., data coming from citizens and public media, involved in the emergency settings. The scheme is based on the Linked Open Data (LOD) approach that provides an appropriate solution to manage heterogeneous data from different public sources. We focus on the integration and analysis of both static (previously available) and dynamic (current) information. The scheme also combines official (coming from official sources) and unofficial (coming from voluntary sources) information.

In case of government information, coming directly from its own public channels, for example, websites, the challenge is the creation of a common protocol to be applied to all government sectors. A solution to this problem would allow changing information without generating inconsistencies. The development of e-government programs is rapidly growing the volume of data available for citizens, although, data and access heterogeneity creates difficulty for their integration.

The linked open data (LOD) initiative, incorporated in this scenario, allows the interconnection of data, using standards in the context of the semantic web approach. In ideal conditions, government agencies publish their public data, thus allowing the use of automated data concerned consumers, whether they are other government agencies or citizens. Today, most public agencies provide access to their data, but the great majority is unreadable by automated mechanisms., even when they are in digital formats (doc, pdf, xls, images, etc.). Besides, most of them do not provide a dictionary meaning for the published content. These raw data sets and documentation structure should be reliable, otherwise its utility is null.

Efforts aimed to link government data are growing in some countries around the world. Interesting examples are given by Britain, the United States and Spain. In 2009, the British government began to adopt Linked Data as official standard for publication of public domain data[1]. During the elections of May 2010, British authorities were encouraged to publish the results in the form of Linked Data[2]. In the future, it is expected that data on education, justice, finance, among others are progressively made available and linked following this pattern.

This presentation is divided as follows. In Part 2 we emphasize the importance of information management in emergencies. Part 3 presents an overview of Linked Data. Part 4 is dedicated to describe an architecture for applying previous

[1] http://data.gov.uk/
[2] http://openelectiondata.org/

and current data connected through Linked Data. The last part presents an application scenario to illustrate the concepts.

II. PREVIOUS AND CURRENT CONTEXTUAL KNOWLEDGE IN EMERGENCY RESPONSE

Information Management plays an important role in all phases of an emergency. Some information is more static but a high flow of information can dynamically emerge depending on the situation. The need for a combined usage of different types and sources of information makes the information management a complex task, particularly during the response phase when time pressure and information overload are common.

Diniz et al. introduced a conceptual map (Figure 1) that organizes three types of knowledge that the command team, designated to manage an emergency, uses for decision making and for directing field team operations. These decisions are based on combinations of three types of knowledge: the previous (Personal and Formal) knowledge and the current contextual knowledge [1].

Previous Personal Knowledge: It is the knowledge that each member of this team develops over time and experience from previous operations, including training. Generally, people who integrate this team are experienced, having worked in many previous missions, and who understand and have control of the entire emergency process.

Previous Formal Knowledge: General information originated from government agencies not necessarily designed for emergency use. It usually refers to information about an affected area that is considered useful to command decision-making. Data such as road maps, number of hospital beds, construction plants and water supplies can be relevant to assess the emergency situation. A critical aspect is the reliability of this information, especially regarding the question of time stamping. In many cases, the information is outdated and therefore useless. Another related aspect is the lack of standardization. Different formats create difficulties in information sharing and integration, as well as updating. If govern departments do not adopt a common standard, each piece of information has to be processed differently creating barriers to its universal use.

Current Contextual Knowledge: This is information generated during the emergency evolution process. Situation assessment done by field agents, including information about victims and damages, orders issued by the command and their effects are examples of current contextual knowledge. When this information is combined with previous knowledge it builds what is called "Combined Knowledge" in the diagram of Figure 1, which helps the Command decision-making. This knowledge is very dynamic. It changes all the time and therefore it has to be constantly updated. Most updates come from field agents, but it can also be enriched by voluntary workers, directly involved in the emergency or not.

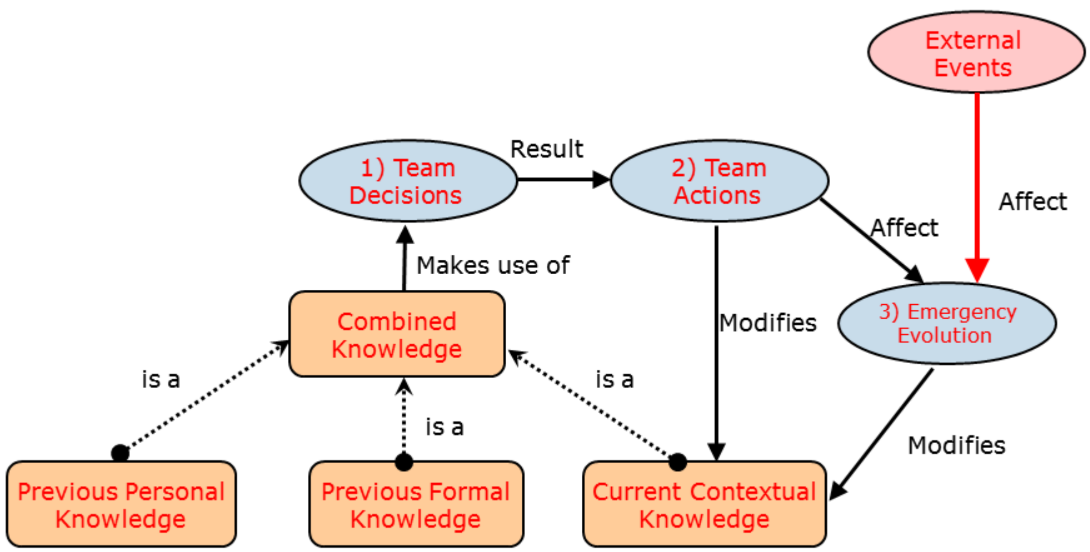

Figure 1 - Conceptual map of knowledge support during an emergency response phase [1]

III. LINKED DATA: AN APPROACH FOR INTERRELATING HETEROGENOUS DATA

The web emerged as a distributed hypermedia that allows interlink between documents, designed for human consumption, with low degree of structures in its objects and implicit semantic of content and links. The result is a "Web of Documents", a huge amount of publish data available as silos of information [2].

In 2006, Tim Berners-Lee posted, on his W3C´s design issues site, some statements on the realization of the semantic web vision, in what is called Linked Data [3]. He argued that "The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data." The proposal of a "Web of Data", considers connecting related data, not originally associated, across the web, using URIs (Uniform Resource Identifier), HTTP and RDF.

RDF (Resource Description Framework) is at the heart of this proposal, as a well-established specification consisting of a suite of W3C Recommendations, published in 2004. It consists basically of a triple-based representation formalism, supporting connections between resources and as well as their description. Among its advantages, it facilitates interoperability, allowing the use of multiple vocabularies and mappings between them.

A. *Basic Linked Data Publishing Process*

To take advantage of the Linked Data paradigm, it is necessary to follow some recommended best practices for exposing, sharing, and connecting pieces of data, information, and knowledge on the web, as a combined effort from both producers and consumers of data. The producers have to expose their raw data on the web to feed a process of publishing that filters, cleans and converts data to RDF triples (building the datasets).

As the exposed data usually comes from multiple sources, an important issue must be handled: provenance tracking, since it helps to determine the quality and trust of the data [4]. In collaborative environments that support the decision-making process, both ontologies and provenance metadata are critical issues, especially in Emergence Response scenarios, due to the various sources of data and domains involved.

Finally, the consumers have to use appropriate engines to explore the Linked Data. As data are exposed, many links can be pointed to them, and many applications to explore these data can be built to take advantage of the interlinked data [5].

B. *Linked Open Data*

The appeal for the government to open and expose data to the public as Linked Data became known as Linked Open Data (LOD). With the growth of e-government programs, the available data to citizens is growing in volume every day, but to make it a useful source of information, to be referenced and integrated more easily by different applications, it should to be published according to the best practices of Linked Data.

Efforts aimed to interlink government data are growing in some countries around the world. Interesting examples are given by Britain, the United States and Spain. In 2009, the British Government began to adopt Linked Data as the official standard for publication of public domain data [6]; it is expected that data on education, finance, among others, are progressively made available and linked following this standard.

C. *Linked Data Perspectives*

The current status of the Linked Open Data cloud includes 203 datasets, approximately 27 billion triples and 400 millions of RDF outgoing links. Outgoing links refer to the links that are set from data sources within a domain to other data sources [2]. Another recent fact is the approval of the European project LOD Around The Clock (LATC) Support Action [3] to support institutions and people in publishing and consumption of Linked Open Data.

Due to the growth of the Linked Data initiative, in September 2010, the European Union started the LOD2 project [7] to handle some challenges of the LOD paradigm associated to intelligent information management: the exploitation of the web as a platform for data and information integration in addition to document search. Some of their main concerns are related to: coherence and quality of data published on the web, establish trust on the Linked Data Web, methodologies for exposing, high-quality multi-domain ontologies, automatically interlinking and fusing data, standards and methods for reliably tracking provenance, ensuring privacy and data security as well as for assessing the quality of information.

D. *Collaboration with Linked Data*

An aspect about Linked Data that is of special interest in the context of this work is the interface for collaboration. Davies et al. [10] report a controlled experiment in which novices attempted to use a prototype Linked Data interface. They suggest several specific design approaches for Linked Data authoring environments. Luczak-Rosch and Heese [9] developed an application, called Loomp that allows Linked Data Authoring by non-experts, enabling them to produce and publish semantically annotated content as easy as formatting text in word processors. Passant and Laublet [8] introduced MOAT, a lightweight Semantic Web framework that provides a collaborative way to let Web 2.0 content producers to give meanings to their tags in a machine readable way.

Although several studies have been made to address issues associated with linked data, there is still lack of work on how to support collaboration in this scenario: how to facilitate the publication and linking of user´s contributions.

IV. CAPTURING, INTEGRATING AND DISPLAYING CURRENT CONTEXTUAL KNOWLEDGE AS LINKED DATA

Collaborative Knowledge Management Systems in Emergency Response face many challenges, especially those related to the information infrastructure, which must support information dynamics, trustee and integration. The Linked Open Data paradigm can address some of these challenges in order to make available current contextual information, originated from different sources, such as citizens, emergency teams and journalists, combined with previous knowledge.

To do so, we describe an architecture, illustrated on Figure 2, comprising mechanisms to deliver relevant and reliable information to support decision making, avoiding, as much as possible, information starvation or overload during the emergency response phase. The main components of each layer and their functionality are detailed in the following sub-sections, organized according to the Linked Data cycle (expose, publish and consume).

A. *Expose Data by a Collaborative Interface*

The current contextual data is captured through an interface designed for collaboration which is prepared to receive information independent of its structure. This interface can be used by the emergency team and other people like citizens and journalists interacting through their own mobile devices. The emergency team can feed the system with contextual data combined with previous personal knowledge through the Reliable Network Client interface, specially designed to receive these information with high level of trustee (a). Citizens and journalists can collaborate with information about the scenario using the Social Network Client interface (b) that is more flexible to allow an easy input of facts and special situations, as

[3] http://latc-project.eu

current contextual data, such as the one usually posted on social networks.

B. Publishing Current Contextual Data

The current contextual information captured through the collaborative interface is stored in the Data Layer. The structure of this new information might have been planned in the schema design of the knowledge base or not. If it has been expected, the new information will be *triplified* as an individual in an existing RDF schema. If not, the new information will be *triplified*, with its own structure, and interlinked with other triples using the RDF URI reference.

After that, the data is processed by the "Raw to Linked Data Processing Layer" where filtering, conversion, cleaning, provenance and many other tasks might be executed. All these steps are supported by ontologies and vocabularies.

The proposed architecture components consists of six layers. Four of them are already present in a typical information system: data, integration, application and user layer, and two others were specially created to support the recommendations of Linked Data: Raw to Linked Data Processing Layer and Semantic Layer.

C. Consuming and Exploring Combined Knowledge as Linked Data

Linked data can be consumed and explored by the management system using appropriate interface applications, directly or through a combined knowledge view (e). By analyzing the data, the managers can feed the system with new information (f), which will be linked to the data already stored.

D. Semantic Support

The implementation of the Linked Data Cycle´s steps are support by a set of ontologies and vocabularies of the Emergency Domain, which can be re-used or developed, and are available on a layer that interacts with all others layers. These terminology resources are used to semantically enrich the triplification, interlink and consume process, caring about the meaning of the information that flows thought the environment.

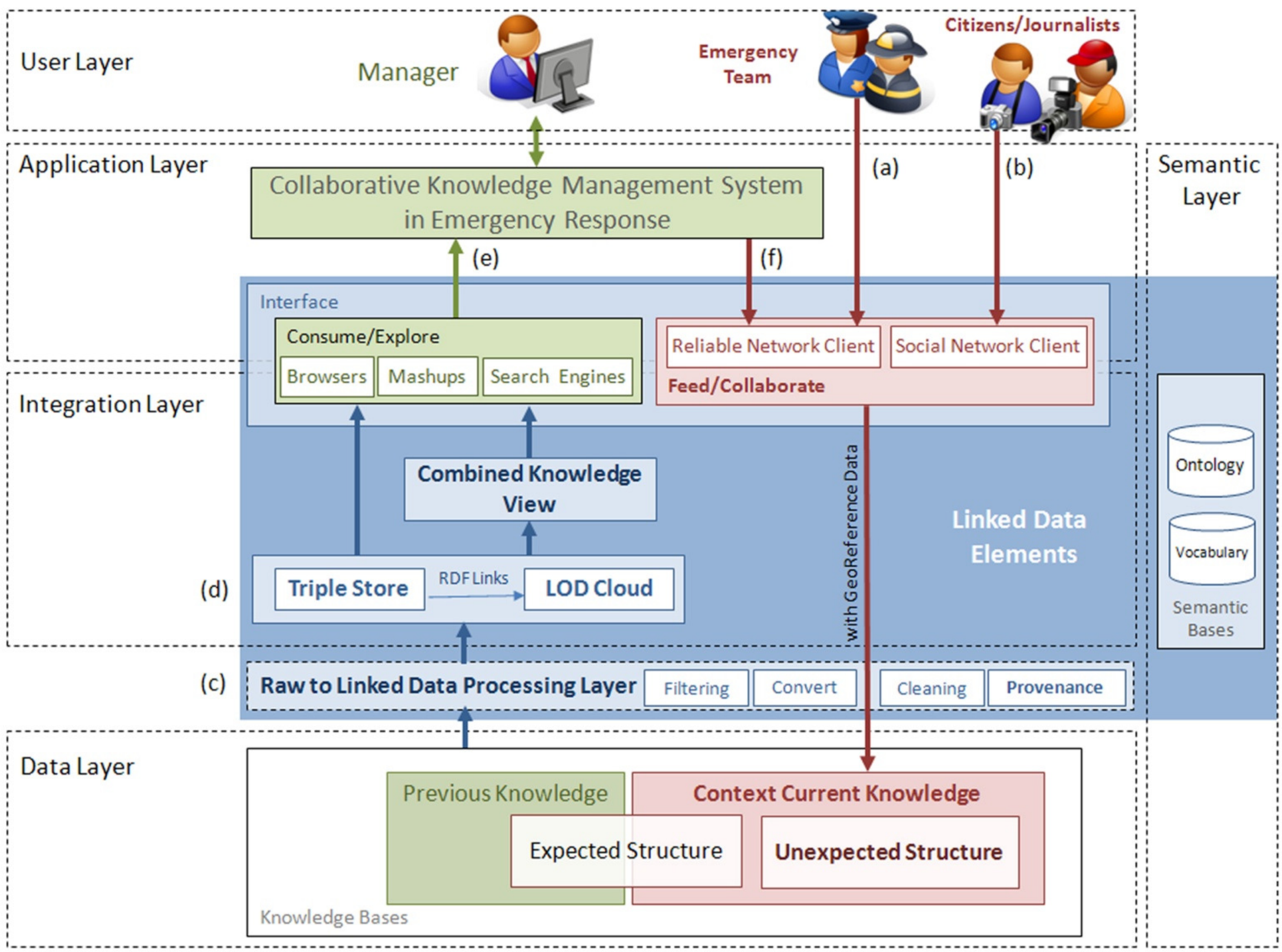

Figure 2. Linked Data for Collaborative Knowledge Management Architecture

V. EXAMPLE CASE

The example case of an emergency situation takes place on a flood and landslide scenario caused by a long period of heavy rain. Facing this chaotic situation, different teams work in emergency management and rescue operation, dealing with safety, health, firefight, media, army and non-governmental organizations providing different kinds of support. Acting directly and officially in the field is the "Emergency Team", agents of the organizations cited before, each one playing a specific role. All sorts of information gathered by these teams should be available to all others involved. The media, for example, trying to get information about the number of victims, (coming from civil defense and rescue teams), roads blocked (originated from roads department), weather forecast, and many other types of information, provided by these Emergency Teams.

Citizens and "Emergency Teams" compose the field agents, who transmit their perceptions in real time directly through a communication channel in order to feed the Team Manager´s "Knowledge Base". They also consume information, logically organized and certified, from all other teams. The "Team Manager" is composed by highly experienced professionals, commanders and public administrators, working in emergency management offices. They direct the Emergency Teams and support them with useful information in order to reduce the response emergence time, making their work more effective. These three groups compose the "User Layer". At the level of the "Application Layer", we have the system interfaces for data feeding via safe and easy front-ends. This layer also interfaces with social networks as a channel for information input from ordinary citizens.

At the "Integration Layer", field collected information is treated according to its provenance. In the field, time available for non emergency related tasks is very scarce. Thus, they need direct and user-friendly interfaces to send and receive information, so to ensure that teams spend as little time as possible in this task. In our scheme this interface is named "Reliable Network Client". Common citizens feed the system through social networks, uploading text and multimedia information. This is a channel with low complexity and very familiar to the community.

The main difference between the official and the "non-official" interfaces is the information reliability. Information uploaded from Emergency Team does not need to pass by the filtering and testing process. This is not always the case with information coming from social networks. All this information is classified as "Context Current Knowledge", because it´s related to information collected on the fly, nearly in real time. Joining the current with the previous contextual information, from government agencies, we structure the "Data Layer". We should remember that information provided by the collaborative environment comes embedded with geotagging information. Today, social networks already offer methods of providing the geographic location of users, by capturing it directly and transparently from GPS receivers embedded devices. As shown by Hickson [11] even mobile devices that do not have a GPS receiver are capable of capturing their location using routing and antennas triangulation methods.

In the next step, data goes back from the "Data Layer" to the "Integration Layer" passing through standardization, cleansing, conversion and testing processes, rewarding providers of reliable information and penalizing providers of false information. Unreliable data do not go to the next step because they have been discarded in the filtering step (c). After this filtering, the standard RDF *triplifying* process occurs, by making data available through a Linked Data unified standard. In the next step, the *triplified* data pass through the "Application Layer". At this point, all available collected data is standardized and able to be classified by precedence and theme, presented in simple mashup front-end.

Finally, the "User Layer", now represented by the Manager, is ready to consume data that will support the decision-making process. The decisions and any other knowledge and information acquired directly by managers may also be inserted through the same interface as the "Emergency Team" in order to enrich the "Knowledge Base" for next decisions and missions.

VI. CONCLUSIONS

In this presentation we described an architecture based on the knowledge framework proposed by Diniz et al. [2, 34] that makes use of the Linked Data approach for data integration. We described a scheme built to deal with emergency response scenarios, particularly the knowledge provided to decision-making and operational teams. We showed how a collaborative supply of information can be integrated into a comprehensive scheme aimed to support the decision-making process.

REFERENCES

[1] V. B. Diniz, M. R. S. Borges, J.O. Gomes, J.H. Canós, "Knowledge management support for collaborative emergency response" In Proceedings of the 9th International Conference on Computer Supported Cooperative Work in Design (Vol. 2, pp. 1188-1193). 2005, IEEE.

[2] C. Bizer, A. Jentzsch, R. Cyganiak, "State of the LOD Cloud", Freie Universität Berlin, Version 0.1, 10/19/2010, Preliminary Release.

[3] T. Berners-Lee, "Linked Data", 2006, http://www.w3.org/DesignIssues/LinkedData.htm.

[4] L. Moreau, "The Foundations for Provenance on the Web", Foundations and Trends in Web Science, Vol. 2, Nos. 2-3 (2010) 99-241, 2010.

[5] DBpedia Mobile http://wiki.dbpedia.org/DBpediaMobile.

[6] UK Government Data , "Opening up government", http://data.gov.uk/.

[7] "LOD2 - Creating Knowledge out of Interlinked Data", Funded under 7th FWP (Seventh Framework Programme), Project start date: [2010-09-01].

[8] A. Passant, P. Laublet, "Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data", Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China, Apr (2008).

[9] M. Luczak-Rosch, R. Heese, "Linked Data Authoring for Non-Experts", Proceedings of the 2009 Linked Data on the Web Workshop (LDOW09), 2009.

[10] S. Davies, J. Hatfield, C. Donaher, J. Zeitz, "User Interface Design Considerations for Linked Data Authoring Environments", LDOW2010, April 27, 2010, Raleigh, USA.

[11] I. Hickson, "HTML5: A vocabulary and associated APIs for HTML and XHTML", W3C, 2010.

[12] V. B. Diniz, M. R. S. Borges, J.O. Gomes, J.H. Canós, "Decision Making Support in Emergency Response" In: Encyclopedia of Decision Making and Decision Support Technologies.1 ed.London : Information Science Reference (an imprint of IGI Global), v.1, 2008, p. 184-191.