Harmonizing Data collection in an Ontology for a Risk Management Platform

Désirée Hilbring[1], Jürgen Moßgraber[1], Philipp Hertweck[1], Tobias Hellmund[1], Hylke van der Schaaf[1], Anastasios Karakostas[2] and Efstratios Kontopoulos[2], Ilias Gialampoukidis[2], Stelios Andreadis[2], Stefanos Vrochidis[2], Ioannis Kompatsiaris[2]

[1] Fraunhofer IOSB, Fraunhoferstr.1, 76131 Karlsruhe, Germany
[2] CERTH-ITI, Information Technologies Institute, Thessaloniki, Greece

Abstract. In every disaster time is the enemy and getting accurate and helpful real time information for supporting descision support is critical. Data sources for Risk Management Platforms are heterogeneous. This includes data coming from several resources: sensors, social media, the general public and first responders. All this data needs to be analyzed, aggregated and fused and the semantics of the data needs to be understood. This paper discusses means for integrating and harmonizing data into an ICT platform for risk management and gives examples for semantic analysis.

Keywords: Risk Management, Internet of Things, Ontology, Knowledge Base, Decision Support, Mobile App

1.   Introduction

Recent research from the World Meteorological Organisation showed that the global temperature increase caused by climate change could reach 3°C by the end of the century. Consequently, the probability for the occurrence of natural catastrophes like heat waves, floods and forest fires is increasing. The United Nations have long been calling for the intensified worldwide development of early warning systems. Thus, the authorities of potentially affected regions need to prepare themselves. Early warning, risk and crisis management systems can provide these authorities with the necessary methods for decision support and therefore improve the safety during crisis events.

In this context, the Horizon 2020 project beAWARE develops a general framework to facilitate decision support by data integration via sensors and semantic data analysis. Furthermore, this framework provides means for the participation of normal citizens via Mobile Apps, thus enabling them to report efficiently on crisis incidents (crowd sourcing). One of the main issues in beAWARE is to harmonize and analyse incoming information for the detection of incidents. All information regarding a specific incident, which is bound timely and locally, needs to be collected and constantly upgraded. Information about these incidents then serves as a basis for decision support. This paper concentrates on means to solve this issue.

The following sections of this paper will explain how some of the beAWARE components solves the data integration issues and harmonizes the resulting information in an ontology. The section "Related Work" discusses data integration and semantic technologies relevant for risk management platforms. The focus in the main parts of the paper lies firstly on the harmonized data collection (Section "Harmonized Data Integration") and secondly on the semantic modelling of the incidents during a crisis which can be

applied to multiple use case domains like flood, forest fires or heat waves (Section "Semantic Data Fusion"). Finally, the section "Conclusion and Future Work" gives an outlook for the ongoing project.

2.   Related Work

A Decision Support System combining current technical possibilities is introduced by Fang et al. in "An Integrated System for Regional Environmental Monitoring and Management Based on the Internet of Things" [1]. The system combines IoT, Cloud computing and GIS technologies into an information system for environmental monitoring. Fang points out that research for data acquisition and fusion is important. However, in difference to our approach semantic integration is not considered.

Sensor data to be integrated into an ICT platform can be accessed via Internet of Things (IoT) or Machine-2-Machine (M2M) standards. For integrating and harmonizing the data and for creating a common basis for further analysis, the Open Geospatial Consortium (OGC) developed a data model and API called the SensorThings API [2]. The data model is based on the OGC/ISO Observations and Measurements model but simplified to make it more suitable for the use in the Internet of Things domain. The REST interface allows Create, Read, Update and Delete actions on all entities. Additionally, it offers powerful search capabilities, including geospatial and temporal filtering. Besides the REST interface, the OGC SensorThings API offers a Message Queue Telemetry Transport (MQTT) extension allowing push notifications when entities change. We use this standard for accessing data and writing into our proposed platform.

Semantic integration is discussed in other recent work. Wanner et al. present an "ontology-structured knowledge base from which then information relevant to the specific user is deduced and communicated in the language of their preference" [3]. Furthermore, a recent thorough review of the state of the art in crisis management ontologies is given in [4]. Besides the above, two of the most prominent approaches in crisis management and response are *MOAC* [5] and *SoKNOS* [6]. MOAC (Management of a Crisis Vocabulary), is a lightweight vocabulary. SoKNOS, on the other hand, is a set of ontologies ensuring that newly created information, as well as integrated sensor information, is semantically characterized.

Since the existing ontologies only cover a subset of climate-related crisis management, the proposed beAWARE ontology consists of modules for representing *all* aspects of crisis management adopting concepts from some of the existing ontologies [7].

Possibilities for public participation via Mobile Apps have been researched in the project WeSenseIt [8]. They developed a crowd sourcing Mobile App WaterDetective for uploading pictures or messages.

2.   Harmonized Data Integration

The beAWARE developments are tested in three use case pilots focussing on different climate hazards: floods, fires and heat waves. Various end users from each use case of the project provided information regarding the requirements for the Risk Management Platform. These requirements have been sampled and organized and are used as base requirements for the design of the platform. The following describes how information of different data sources is integrated into the platform. This corresponds to several user requirements: e.g. "Detect elements at risks from text by Mobile Apps or social media (UR_112)", detect

"River overtopping (UR_118)" or "Manage assignments" of first responders "based on river level overtopping" (UR_119). Thus, different kinds of data sources needs to be addressed: sensor data, data from Social Media, data from the general public. This data needs to be integrated and fused inside the platform, such that a combined analysis of all information is possible to support the stakeholders in finding suitable actions for improving the crisis situation.

Sensor Data: Depending on the use case different sorts of sensors are available and their data needs to be integrated into the platform. Meteorological data (temperature, humidity and rainfall) is importatnt to identify an ongoing crisis. In a flood crisis one needs access to water levels of rivers in the affected region to detect overtopping. The Open Geospatial Consortium (OGC) developed a data model and API called the SensorThings API capable of creating a common basis for further analysis [2]. beAWARE adopts this standard for integrating all provided sensor data. The OGC SensorThings API provides a REST interface and a data model for exchanging sensor- and metadata. One of the first implementations of the OGC SensorThings API is made by Fraunhofer IOSB [9]. This implementation, called FROST, is open-source. This implementation is used to include the meteorological data like temperature, rainfall and the water level data into the demonstrators of the beAWARE pilots. The sensor data from all available sensors in the flood pilot gets pushed hourly into a FROST server instance. Besides data from these sensors, AWAA (Alto Adriatico Water Authority) provides water level forecasts for every river section in the area. This integration matches the use case to identify river overtoppings. The sections are about 500m long, resulting in 304 sections. The water level prediction models generate hourly data, projecting about two days into the future. Depending on the crisis level, the models are executed at different frequencies. Having all sensor data as well as weather data inside the FROST instance, stakeholders can use an interactive graph component to query and display the data. This allows to easily analyse correlations between data series. For examples: Figure 1 shows every precipitation peak is followed by a water level peak.

Figure 1: Correlation between rainfall and water level

Another important data source in a crisis situation is the **Public** itself. Therefore Twitter data is included through a **Social Media** crawling component into the beAWARE framework using a Streaming API through a client that receives tweets the moment they are published. For the three pilots Twitter data is

crawled by keywords representing the use case topics flood, fire and heatwave. One goal for beAWARE is to find relevant tweets and successfully assign them to identified incidents matching requirement "Detect elements at risks from text by social media" (UR_112). An example for such a tweet is: "Matteotti Square is underwater! #flooding". Currently we are working on clustering the detected Tweets via DBScan for the identification of incidents. Clusters with an indentified location will then be matched to corresponding incidents in an ontology.

Another data source can be the people, which are affected by a crisis themselves. They might be willing to share observed information not only via Social Media but also directly to the decision makers of a crisis. Within the beAWARE framework, a mobile application is developed which supports sending information in an easy way. Multimodal input (pictures, videos, audio and text) is supported. This data, together with the geographic coordinates of the sender will be analysed and integrated into the ontology of the beAWARE framework. Additonally **First Responders** are considered as data sources as well. To meet UR_119, decision makers can sent first responders to specific places for action execution.

3.   Semantic Data Fusion

All collected data needs further analysis after the integration. One of the main issues to solve is the detection of critical incidents (UR112). For this purpose semantic technologies are used. As a first step, beAWARE defined an ontology based on the experience of the end users participating in the project [9]. The aim of the ontology is twofold: (a) to represent natural disasters (in this part the ontology's design is inspired by MOAC and SoKNOS) and (b) to represent the analysis results of input items (image, video, text) fed from sensors and the public to the various beAWARE analysis components. We concentrate in the description on goal (b) of the ontology.

The information from the heterogeneous data sources serves as input data for the beAWARE analysis components. The analysis results are then fed into the ontology. At this point the fusion takes place: important information snippets collected from all integrated data collected from different sources are stored into a specific place in the ontology: the incident. The analysed information is modelled in the ontology and bound locally. Referring the above example: the tweet (class "Media Item") is to be analysed via the Social Media Analysis (class "Task"). The analysis produces a report (class "Detection"), which can contain vulnerable objects at "Matteoti Square" (see Figure 2)
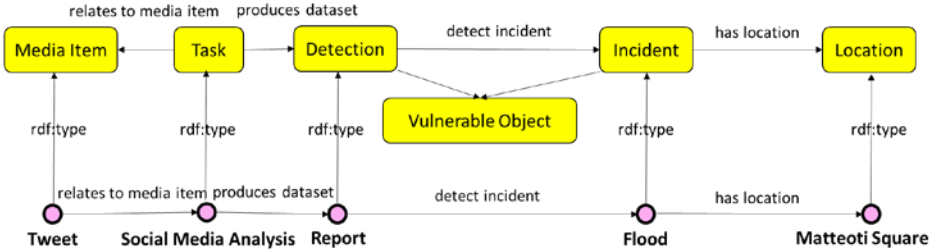

Figure 2: Representing analyzed data in the beAWARE ontology.

Representing **Unit Assignments:** Another component of the proposed ontology is responsible for semantically representing response unit assignments. The adopted representation is based on the approach

proposed by the *OASIS* project [10], mainly the part for representing mission assignments to units and associating missions to incidents.

Figure 3 displays the respective concepts in the proposed ontology. First responders (class "*Responder*") are assigned to one or more missions (class "*Mission*"), which in turn relate to incidents that involve participating entities (class "*Vulnerable Object*"). A mission is also characterized by start and end time, status and mission priority.
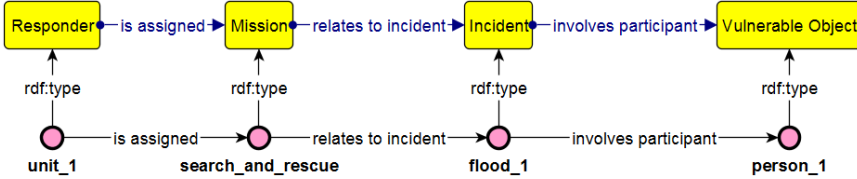

Figure 3: Representation of mission assignments to first responder units in the ontology.

The ontology is used to provide a contextualized overview of the crisis situation. The key aim is to support end users with combined information assisting them in decision making.

A corresponding component called *KB Service* (*KBS*), serving as the interface to the knowledge base/ontology, has been developed, which is responsible for the semantic fusion of the incoming analysis results. Each analysis component provides its results in JSON[1] format, and KBS's role is to parse these results, and to generate and add respective instances to the ontology. Independently from their source, all results will lead to creating instances of media items, tasks, datasets, incidents and vulnerable objects.

A set of SPARQL[2] rules running "on top" of the ontology are responsible to perform further inferences. Representative examples include the following:

- Grouping incidents taking place in neighbouring locations together.

- Calculating incident priority levels, e.g. an incident where people are in danger has higher priority than an incident where traffic is reported;

Figure 4 shows an example query retrieving all incidents with a certain confidence affecting humans.

```
1   SELECT ?incident
2   WHERE
3 ▼ { ?incident a <#Incident> .
4   ?impact <#isCausedByIncident> ?incident .
5   ?human <#suffersImpact> ?impact .
6   ?human a <#Human> .
7   ?human <#participantIsDetectedBy> ?detection .
8   ?detection <#hasDetectionConfidence> ?detectionConfidence .
9   FILTER ( ?detectionConfidence > 80)
10  }
11  GROUP BY (?incident)|
```

Figure 4: SPARQL query that retrieves incidents affecting humans.

The results of the above inferences are added to the ontology. Then the ontology is able to respond to a wide variety of queries; typically referred to as Competency Questions (CQs). The list of CQs includes queries such as:

---

[1] JSON: http://json.org/

[2] SPARQL: http://www.w3.org/TR/2013/REC-sparql11-query-20130321/.

- All incidents in a specific time interval
- Number of humans involved in incidents
- Involved objects in incidents

Each CQ is formulated as a SPARQL query; CQs may either be submitted to the ontology at periodic intervals (e.g. every 1 minute), or, alternatively, the end-user may explicitly request that the CQs be submitted at a specific point in time.

5. Conclusion and Future Work

This paper presented means for integrating, harmonizing and fusing data for an ICT platform for risk management. The information retrieved from the heterogeneous data sources are collected in an ontology and bound to the identified incident to create a condensed overview of the situation. This information can be then used to identify necessary actions to improve the crisis situation. The project is still ongoing: further work will be done in the detection and localization of incidents of Twitter messages. The concepts will be tested in field demonstrations in the pilot regions of the beAWARE project.

References

[1] Fang, S., Xu, L.D., Zhu, Y., Ahati, J., Pei, H., Yan, J. and Liu, Z. (2014) An Integrated System for Regional Environmental Monitoring and Management Based on the Internet of Things, *IEEE Transactions on Industrial Informatics*, Vol. 10, No. 2, May 2014.

[2] Liang, S., Huang, C. and Khalafbeigi, T. (2016) OGC SensorThings API Part 1: Sensing, Version 1.0, 15-078r6, http://docs.opengeospatial.org/is/15-078r6/15-078r6.html, visited in January 2018.

[3] Wanner, L., et Al (2014) Getting the environmental information across: from the Web to the user, *Expert Systems: The Journal for Knowledge Engineering*, Volume 32 Issue 3, June 2015, Pages 405-432.

[4] Liu, S., Brewster, C., & Shaw, D. (2013). Ontologies for crisis management: a review of state of the art in ontology design and usability. In: *Proceedings of the 10th International ISCRAM Conference* – Baden-Baden, Germany, May 2013, pp. 349–359.

[5] Limbu, M. (2012). *Management of a Crisis (MOAC) Vocabulary Specification*. Available online: http://www.observedchange.com/moac/ns/, last accessed: Mar'18.

[6] Babitski, G., Probst, F., Hoffmann, J., & Oberle, D. (2009). Ontology Design for Information Integration in Disaster Management. *GI Jahrestagung*, 154, 3120-3134.

[7] Kontopoulos, E., et Al (2018). Ontology-based Representation of Crisis Management Procedures for Climate Events, ISCRAM 2018, Rochester USA.

[8] When, U, Rusca, M, Evers, J, Lanfranchi, V. (2015). Participation in flood risk management and the potential of citizen oberservatories: A governance analysis. In: *Environmntal Science and Policy*. Volume 48 pp.225-236, April 2015

[9] Van der Schaaf, H. (2016) SensorThingsServer, Fraunhofer IOSB, https://github.com/FraunhoferIOSB/SensorThingsServer/issues, visited in January 2018.

[10] Couturier, M., & Wilkinson, E. (2005, April). Open advanced system for improved crisis management (OASIS). In *Proceedings of the 2nd International Information Systems for Crisis Response and Management (ISCRAM) Conference*, Brussels, Belgium.