IADIS International Journal on WWW/Internet
Vol. 14, No. 2, pp. 38-57
ISSN: 1645-7641

AN ONTOLOGY FOR THE USE OF QUALITY EVALUATED SOCIAL MEDIA DATA IN EMERGENCIES

Matthias Moi, Nikolai Rodehutskors and Rainer Koch . *University of Paderborn. 33095 Paderborn.*

ABSTRACT

As social media services are on the rise, people increasingly share information about emergencies on social media. Sometimes information about disasters even finds its way faster to social media websites than it reaches regular news companies and emergency services. Yet emergency services still have not found a way to put this potential to an effective use. Within our project "EmerGent", we are developing a system for emergency services to process and analyse information from social media. The goal is to collect the vast amount of data generated in social networks during a crisis, evaluate the quality of this information and to transform the high volume of noisy data into a low volume of rich content that can be presented to emergency personnel in a useful manner. In this paper, we describe the steps taken during the development process of an ontology that covers the fields of social media and emergency management (SMEM) and the structure of this ontology itself. Also a solution for the measurement and description of information quality in form of a graph is explained, where an overall information quality value is calculated based on several criteria and indicators.

KEYWORDS

Ontology; Semantic Data Model; Emergency Services; Social Media; Information Quality; Crisis Management

1. INTRODUCTION

Social media services grow exponentially each year. Facebook as the largest social network, receiving 85% of the clicks on social media websites (Statista 2016), shows ever growing numbers of monthly active users (Facebook 2016). And Twitter for instance had more than 100 million active users in 2011, posting 230 million tweets a day. One year later, the number of active users and the number of tweets per day had doubled (Etherington 2012). This means, that the general usage of social media is getting more and more important to share information. Additionally, Twitter has a history of breaking important stories before traditional news media (Petrovic et al. 2013). Hence, to handle the vast amount of data, information from social media streams has to be transformed into a low volume but rich information model. Because of the huge amount of potentially useful data created on a variety of social networks, direct usage of social media is impracticable to extract useful information. Hence, advanced filter techniques are required (Pirolli 2009). This need is amplified by the fact that even the main social network providers are using different concepts for communication, information sharing and the establishment of relationships. The concept of relations between users in social networks is also implemented in different ways. Beside ‘friendship’ relations, users are able to come together in groups, follow other users and institutes, or can share information publicly for everyone to see.

With requirements on data processing and information querying continuously growing, several approaches have emerged to represent machine-readable information. When file-based approaches reached their limits in structuring data, relational databases became popular (Martinez-Cruz et al. 2012); however, these are not well suited to the challenges of semantic processing. With the emergence of the Semantic Web, the role of semantic technologies became increasingly important. The Semantic Web is often discussed as the predecessor of novel concepts like the “Internet Of Things” and “Ubiquitous Computing”. It extends the network of hyperlinked human-readable web applications by machine-readable metadata about applications and how they are related to each other (World Wide Web Consortium (W3C) 2009). It standardises the way to build semantically enriched applications and can be used to build them on the top of vocabularies, taxonomies and ontologies. (World Wide Web Consortium (W3C) 2009) Ontologies, which define a set of representational primitives with which to model a domain of knowledge or discourse (Gruber 2009), provide a very effective way to structure and categorise knowledge. This means “an ontology is an explicit specification of a conceptualization” (Gruber 1993). In emergency response, ontologies can also unify and normalise data from different resources, e.g. social media, syntactically and semantically and associate it with emergency domain knowledge. Furthermore, on-going research has shown that the need for a common understanding of concepts within and across domains is important to avoid misunderstandings (Galton & Worboys 2011; Grolinger et al. 2011). They help to create meaningful relationships between information resources and to allow machines to process, infer, or combine the information from different sources automatically into a consistent body of knowledge.

These ontology features are needed for our project, as social media offers many information sources to enhance social media awareness. People in the affected area are able to report about the situation and emergency services can use social media information for decision support. Social media may also help in understanding the overall situation (Yin et al. 2012; Houston et al. 2015). Social media communication can also allow emergency services to get in direct contact with citizens and forward important information. While many emergency and healthcare services already try to use social media for information gathering and authority to citizen communication (Thackeray et al. 2012; Reuter, Ludwig, et al. 2016), the full potential of these networks is still not used. Emergency services fight to keep up with the huge and daily growing load of unfiltered social media messages and try to gather valuable information, while forwarding important information to citizens. Meanwhile citizens already have a very positive attitude towards social media and its use during emergencies. A large majority of citizens also thinks that emergency services should use social media for emergency information communication. (Spielhofer & Reuter 2015)

Often ontologies are built from scratch, which does not tap the existing potential of relevant, domain-related knowledge bases. Thus ontologies are often implicitly tailored to a specific need (Bontas et al. 2005). To facilitate information exchange with external systems, projects or domains it is necessary that new developments build upon existing standards. Therefore information models like FOAF (Brickley & Miller 2010), SIOC (Bojars & Breslin 2009) or MOAC (Limbu 2012) must be considered in order to build an ontology that associates information from social media with domain knowledge.

So far, there is no ontology in existence that unifies information from social media with the emergency domain and enhances it with additional information and emergency domain knowledge adequately. This forces us to design an ontology that serves exactly this purpose by extending existing ontologies. This paper describes a subset of the requirements placed in the ontology we developed, the evaluation process of existing ontologies in this area, as well as the development and structure of the ontology. This ontology handles social media in emergency management (SMEM) and is therefore abbreviated SMEM ontology.

The research done for this work is part of the European Union's Seventh Framework Programme funded project "Emergency Management in social media Generation", short "EmerGent". The project aims to analyse the impact that social media has during emergencies. The project regards positive impacts, as well as negative ones. In the end, the goal is to enhance the objective and perceived safety and security of citizens before, during and after emergencies. In EmerGent, we work on several sub objectives to realize this goal. We regard the gathering and mining of information, especially from social media, the quality assessment of emergency related information, the analysis of social media use in emergencies today and potentials for the future, assessment of social media impact and the development of guidelines for emergency management services and their implementation in emergency management processes.

By completing these objectives, EmerGent will strengthen the role of European companies that provide services and products related to the project's results. During the project, the means to gather, mine and evaluate vast amounts of data, effectively interlink it with other, already gathered information and to associate it with emergency domain knowledge, are being developed. The linking of information allows us to effectively query the data using the semantic querying language SPARQL and to make sophisticated conclusions.

The following figure (Figure *1*) describes the concept that is pursued in EmerGent. Especially it shows the current state of the relations between stakeholders, methods and tools and the objectives and tasks during which the current state will be analysed and changed into the final state that is set as the goal for the future.

As described before, one of the main goals of EmerGent is to assess the impact of social media in emergencies for citizens and emergency services (Figure *1*: O1). This assessment is based on the results of an analysis of social media in emergencies (Figure *1*: O2), where citizens and emergency services participate in workshops, case studies and questionnaires, and where as well past emergencies are analysed regarding social media usage. To handle the vast amount of data that is gathered as part of EmerGent, methods will be developed for information mining and for assessing information quality (Figure *1*: O3). Finally, based on all research done before, we develop guidelines that help stakeholders to understand the benefits of social media in emergencies (Figure *1*: O4), and how to integrate social media into their process on different levels (Figure *1*: O5).

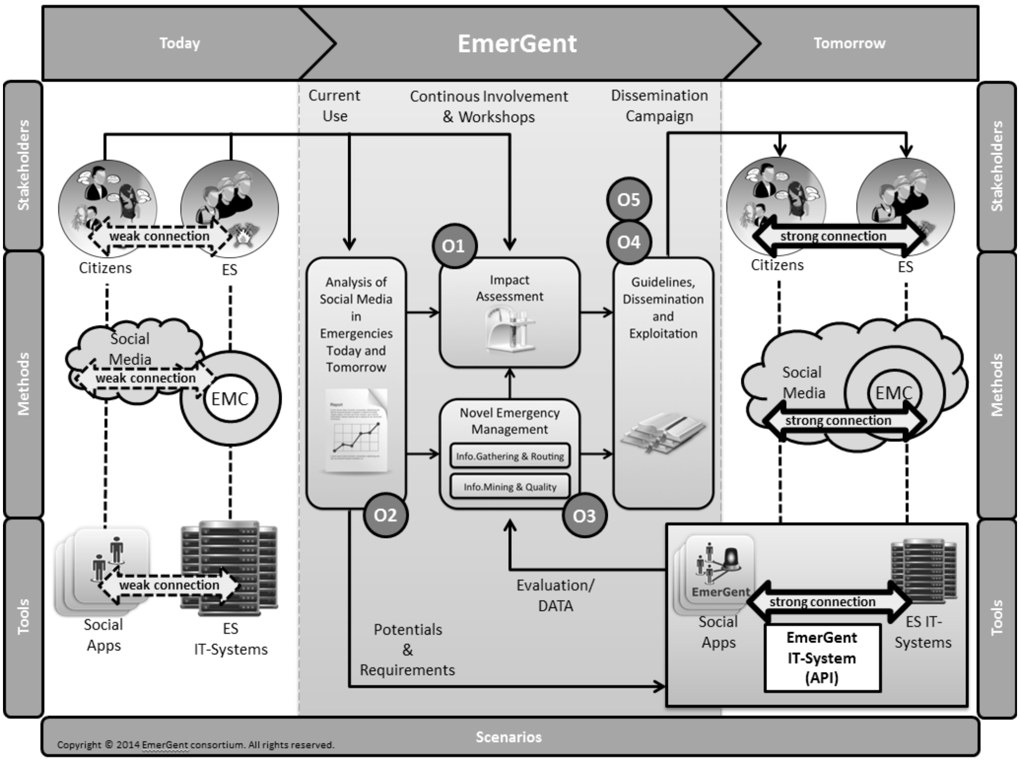

Figure 1. General EmerGent concept

The research in this paper was done as part of the “Novel Emergency Management” Task (O3) and the overall EmerGent IT-System. As described above, the goal of objective 3 is to develop and deploy methods for information gathering, mining and quality assessment. In this paper, a special focus is placed on the SMEM ontology, its creation and design, as well as important parts and methods that were considered or developed in the process, e.g. a graph that is included in the ontology and enables EmerGent to assess information quality, which is crucial for the use of information obtained from social media.

1.1 Related Work

There have been and there are several research projects underway, that regard the use of social media or social media data in emergencies. Project Pronto (Löffler 2009) especially researched event recognition and resource management in general, but also in emergencies. Project Slándáil researches natural crises (e.g. floods, hurricanes) and to analyses how social media can be better used to spread messages about the worst affected areas during these crises (Slándáil 2015). SecInCoRe , an FP7 project, aims to build a common information base for emergency services (SecInCoRe 2016).The RESCUE project researches the use of social media and reactions to social media messages by communicators during emergencies (RESCUE 2016). The SUPER project aims to understand citizens’ reactions against emergencies in social media, while at the same time empowering security forces and civil protection agencies to fully leverage social media in their operations (SUPER 2016). The DISASTER project researched emergency management, but regarded emergency information exchange as well, aiming to reduce misunderstandings between emergency responders. In addition, an ontology has been developed that describes common knowledge and concepts related to emergencies and the stakeholders involved in a crisis (DISASTER 2015; DISASTER 2012). Ontologies have already been thoroughly researched and many domains and applications, such as the semantic web, personal assistant programs and artificial intelligence, profit from the use of ontologies. For the social media domain as well as the emergency domain, several ontologies are in existence. These ontologies are often used to represent the knowledge domain. From the social media domain, ontologies like the Friend-Of-A-Friend ontology (Brickley & Miller 2010) and the Semantically-Interlinked Online Communities ontology (Bojars & Breslin 2009) are common examples for this domain and have been reviewed in the literature research. Both ontologies describe the structure and functions of social networks. Regarding the emergency domain, ontologies like the Management of a crisis vocabulary (Limbu 2012), EMERGEL (DISASTER 2012) and EDXL-RESCUER (Barros et al. 2015) have been reviewed. The ontologies from the emergency domain treat small parts of the domain, such as information exchange or the description of situations. However, these ontologies only handle the corresponding domain, or just a part of it, while the SMEM ontology aims to combine both domains and cover a larger part of these domains.

2. METHODOLOGICAL APPROACH

The SMEM ontology has to be understood as an ontology that connects the information generated by citizens on social media platforms and emergency services on an information level. It is part of the EmerGent IT system and offers a solution for emergency services to handle social media, combine social media information with emergency domain knowledge. This solution allows for the gathering and processing of emergency relevant social media information for use in emergency management, as well as two-way communication between citizens and authorities (Moi et al. 2015). Several ontologies describe either social media or emergencies. A unified ontology that covers both domains adequately does not exist and is direly needed to effectively put information from social media to use in emergency management and help the responders during an operation. Therefore, the emergency and the social media domains have to be considered simultaneously. Before developing the ontology, the main stakeholders for this topic were identified, then the development of the ontology took place in three steps.

First, in cooperation with the main stakeholders, the emergency services and the citizens, we developed a set of requirements, scenarios and use-cases, which needed to be regarded during the development of the ontology. Secondly, we used these requirements to evaluate already existing ontologies and standards that describe either social media or emergencies that could be usefully included in our final SMEM ontology and that enable the linking of both domains. E.g. Tweets and other social media messages can be linked to emergencies, such as a fire or a flooding. Meanwhile, we also researched potential methods to measure information quality and ways to implement it into the SMEM ontology. The evaluation of information quality is important, as a lot of noise and false information is created during catastrophes

(Shrivastav et al. 2012; Hiltz & Plotnick 2013). Therefore, it is important to filter such noise and enable emergency services to make decisions based on trustworthy information. Finally, after the evaluation of these ontologies, we built a mapping between domain-related information and information from social media and designed a first draft of the SMEM ontology, which also contains the possibility to represent information quality as a graph. To map this information, semantical relations are used, which are represented by data and object properties in the ontology. These properties function as predicates and link the different entities in form of triplets with a subject, a predicate and an object. E.g. "*Fire alert*" (subject) has a "*source*" (predicate) "*Facebook Post*" (object).

3. DEVELOPING OF THE SMEM ONTOLOGY WITH IQ ASPECTS

This section deals with the development of the SMEM ontology and the information quality component. First, the requirements and use-cases that were developed for this ontology in cooperation with a group of experts will be explained. Next, the process of evaluating existing ontologies that fit in the emergency or social media domain, as well as the assessment of common emergency data exchange formats is described. Finally, the information quality part of the SMEM ontology is covered, which helps assessing the quality and content of user-generated content from social networks and is crucial for the use of this information in the emergency context.

4. REQUIREMENTS & USE-CASES

As emergency services are the most relevant stakeholders of the IT system, they had an active part in the definition of requirements. For this purpose, we selected a group of 19 experts from ten European countries. This group of experts is called the End-user Advisory Board (EAB) and serves as a platform by which the experts in the domain of social media and emergency management can share their vision, expertise and domain knowledge. They can give feedback and their advice is captured. Citizens were also regarded as the requirements were defined. Citizens are not the end-users of the IT system, but are seen as stakeholders of this project and therefore their opinion is important for the design of the SMEM ontology.

The final IT system is supposed to show emergency services alert related posts, which can be seen in the following mock up (Figure 2). The ontology has to help in classifying and categorizing the gathered data in order to make it properly usable for this system.

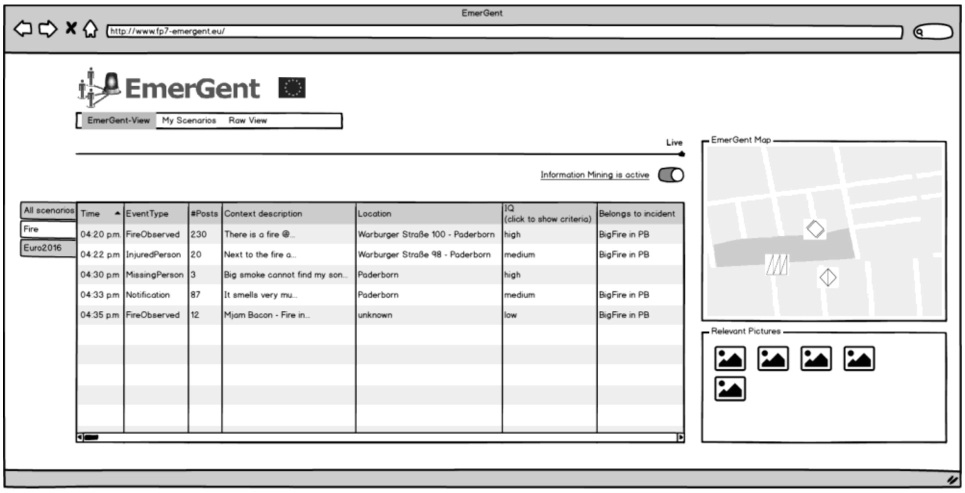

Figure 2. Mock-up of the IT system's frontend

The result of workshops held with the EAB and citizens was a set of two main use-cases for the ontology and a list of requirements that were placed in the SMEM ontology. For the ontology, two main use cases were identified. The first use case regards the communication from citizens to authorities (C2A). During emergencies, considering this use case, citizens can share information within social media as done in the past but with the possibility and with the awareness that their contribution may help emergency services during emergency handling. This means, that citizens have two possibilities to interact with emergency services. The first possibility is to share relevant information in the same way as before, by posting it publicly on a social media platform. ES can then gather this public information and use it. To draw the attention of an ES to a post, the citizens can use hashtags and mentions. Using a mention, citizens link the social media page of an ES, which then automatically gets notified. The second way is to send the information directly to ES, e.g. via a direct message. The second use case describes authorities to citizens (A2C) communication. Here the authorities can speak to the citizens directly and can share information about incidents or guidelines. This also can take place directly or indirectly. ES have the possibility to get into contact with certain individuals by directly messaging them. Here they can directly spread or ask for information. ES also can publish information on a social media platform, which represents the indirect way to share information.

The two most important requirements that became known were the possibility to add multimedia files to created alarms and to evaluate the information quality of social media information. The addition of multimedia files, like photographs, videos and audio files represents a huge advantage that results from the connection of social media and emergency services. Citizens get the ability to share information with the emergency services, and emergency services can gain the ability to gain an impression of an incident, even before the situation has been reconnoitred. At the same time, the consideration of information quality, especially the trustworthiness of an author and a post in social media is regarded as important. The information about the quality of a post gives the emergency services the ability to assess, whether a post depicts the truth about a situation and if a post should be taken seriously (Akerkar et al. 2016). Figure 3 depicts the exemplary logical structure of a post and appended information, such as media files, connected alerts and information quality.

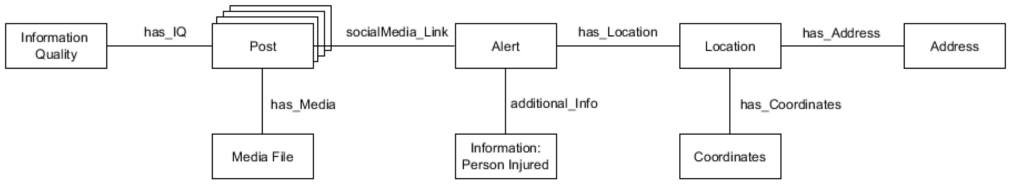

Figure 3. Connection of social media information with an exemplary alert

5. EMERGENCY DATA EXCHANGE FORMATS AND EXISTING ONTOLOGIES

During the literature research, we realized that both, the emergency and the social media domain were already well researched and thoroughly described with ontologies and standards. In the emergency domain, ontologies, like the Management of a Crisis vocabulary (MOAC) (Limbu 2012), and the standard HXL, the Humanitarian Exchange Language (Kessler & Hendrix 2015), already describe many different aspects of emergencies and emergency management. In the domain of social media, commonly ontologies used to describe social networks are the Friend-of-a-friend vocabulary (FOAF) (Brickley & Miller 2010) and SIOC, the vocabulary for Semantically-Interlinked Online Communities (Bojars & Breslin 2009).

Besides these well-known and common ontologies, many other ontologies exist in the emergency service and social media domain, and many more can be used to describe other aspects of these domains, e.g. the BasicGeo ontology, which describes the World Geodetic System in its 1984 revision (Brickley 2003). To evaluate all these already existing ontologies and to assist in the decision, whether the ontologies could or should be used in our SMEM ontology, we needed to regard and potential use-cases for our ontology and define requirements for the ontology and the ontologies that should be used in it.

Additionally, many countries already implement standards to communicate information about disasters and alerts. To exchange emergency related information between different and international senders and receivers, standards have to be chosen and adhered or, if non-existent, created. So far, several different standards already have been implemented and some of these standards are already used frequently. It is important to regard these standards and build the EmerGent IT system in accordance to them. A completely new development would not be preferable, as the EmerGent IT system aims to be compatible with existing structures and standards.

The Specific Area Message Encoding (SAME) is a very well-known emergency message format used in Mexico, Canada and the United States, where it is part of the Emergency Alert System. SAME contains very fundamental information about emergency, which is needed to protect life and property of citizens.

Another and newer, very common exchange standard is the Common Alerting Protocol (CAP). CAP has been approved by the Organization for the Advancement of Structured Information Standards (OASIS) and aims to provide a single and standardized input for alerting and warning systems. It is compatible with SAME and offers additional data exchange capabilities. While SAME only contains location, time and the type of event in form of event codes, CAP additionally offers capabilities such as, updates, multi-lingual messages, more detailed geographic location description and some other capabilities important to describe emergencies (Jones et al. 2010).

Finally, EDXL has to be regarded. The Emergency Data Exchange Language has also been approved by OASIS and is not only a format for exchanging emergency related information, but consists of a complete set of tools for different emergency related usages. It functions as a successor or extension to the CAP standard. The EDXL "Distribution Element" is used to transmit the general emergency information and can also contain CAP messages. Other modules allow the management of hospital capacities, emergency patient tracking, resource requests and situation reporting. (Raymond et al. 2006)

The Humanitarian Exchange Language (HXL) does not represent a format to describe single emergencies. HXL aims provide a standard to more effectively describe larger scale emergencies and humanitarian disasters, with special regard to the affected population (HXL Working Group 2016).

To decide whether a standard or ontology should be used as part of the SMEM ontology, each had to be assessed and evaluated. According to the proposal for ontology evaluation published by Zelewski et al., additionally to the mentioned requirements and use-cases, we also took several other indicators (Zelewski et al. 2001) into consideration. For the evaluation of the ontologies, that have already been designed for the domains social media and emergency services, we used an adaption of the MoSCoW prioritization method developed by Clegg and Barker (Clegg & Barker 2004). This prioritization technique usually categorizes the requirements into must-have, should-have, could-have and won't-have. Following this approach, a set of should-have and must-have criteria were developed in consideration of the findings of (Zelewski et al. 2001) and a set of 14 requirements that resulted from our work with citizens and the EAB. This set of requirements especially regarded scenarios for which the EmerGent IT system is being developed and which concern the communication of emergency services with citizens. These given requirements can be summarized in several categories. The most demanded requirement was the possibility to work with different types of multimedia files. Other requirements described the need to tackle information quality and to describe events and event types. In addition to this information, a proper integration of posts from different social networks and the use of metadata, e.g. location, was demanded (Moi & Rodehutskors 2016a; Akerkar et al. 2016).

To pass the must-have criteria, the ontologies needed to be helpful in fulfilling the defined requirements for the SMEM ontology, had to be publicly available for use and download and had to be documented. The outcome was an ordered list of ontologies that passed the must-have criteria, which helped us selecting ontologies that were suitable for the SMEM ontology. An overview over all criteria used for the evaluation can be seen in Table 1. Out of 40 evaluated ontologies, the following 15 ontologies and standards were integrated into the SMEM ontology listed in Table 2.

Table 1. Criteria for the ontology evaluation (Moi & Rodehutskors 2016a)

Must-have criterion	Description
Concordance with EmerGent scenarios and use-cases	Any ontology or external resource used has to help fulfil the predefin requirements, or enable EmerGent to use the ontology in certain scenarios ar use-cases.
Documentation available	Each used resource or ontology must have a documentation that at least contai definitions of each component. It would be desirable if the documentatic contained further descriptions as well.
Availability in OWL/RDF	Each ontology that will be imported into the EmerGent ontology has to l downloadable as OWL/RDF File. This is irrelevant for other external resources, such as certain standards ar databases.
Should-have criterion	**Description**
Perspicuity	The perspicuity of a resource is defined as good, if the resource is easi understood by the user. Therefore, definitions and descriptions have to l complete and have to be available in natural language.
Extendibility	The extendibility of a resource is defined by the possibility to add oth components to the original resource, without the need to change the existir constructs. Databases and standards will be checked for their compatibility with oth standards.
Functional/Local completeness	Functional or local completeness is defined as the ability to describe all concep and relations of a field, which the resource describes.
Efficiency	Describes the ontologies ability to give explanations for the represent knowledge. This is irrelevant for other external resources.
Minimality	The resource is minimal, if it uses the minimum number of objects to represe knowledge in comparison to other resources. This is irrelevant for other external resources.
Generality	Generality describes the possibility to use a resource in many different domain
Dependencies	Dependencies describes the dependence of a resource on other resources. This is irrelevant for other external resources.
Precision	Precision demands unequivocalness and selectivity of a resource. This is irrelevant for other external resources.

Table 2. List of included ontologies (Moi & Rodehutskors 2016b)

Ontology Abbreviation	Name	Description
CAP	Common Alerting Protocol	The CAP standard is used to describe alerts in a standardized way (Jones et al. 2010).
CiTO	The Citation Typing Ontology	CiTO allows us to describe referencing of different posts in social media, e.g. one author cites the post of another (Shotton & Peroni 2015).
DC-Terms	DCMI Metadata Terms	DCMI Terms offers a great range of standardized metadata. It is used to describe metadata of posts and media files (DCMI Usage Board 2012).
EM-DAT	Emergency Events Database	EM-DAT is an online database of disasters, reaching back into the 20^{th} century. It has a very precise categorization of events, which we used to describe incidents (Université Catholique de Louvain 2009).

FOAF	Friend of a Friend	FOAF describes relations between persons, such as which people know each other. In combination with SIOC it is used to describe social media networks (Brickley & Miller 2010).
BasicGeo	W3C Basic Geo	Basic Geo describes the most fundamental information about locations, coordinates (Brickley 2003).
HXL	Humanitarian Exchange Language	HXL describes additional information for disasters, e.g. number of people affected or injured (Kessler & Hendrix 2015).
iContact	International Contact Ontology	iContact offers many different classes to describe international addresses. It is used in the description of locations (Fox 2015).
LODE	Linking Open Descriptions of Events	LODE allows us to link information to events (Shaw et al. 2010).
MA-Ont	Ontology for Media Resources	MA-Ont provides a set of descriptive properties for media files and a mapping of different metadata formats (Michel 2012).
RDFG	Named RDF Graphs	RDFG is used to describe graph structures in ontologies. We used it to describe a graph for information quality (Carroll et al. 2005).
SCOT	Social Semantic Cloud of Tags	SCOT describes tagging in the Semantic Web (Scerri et al. 2012).
SEM	Simple Event Model Ontology	SEM defines the context of events and describes involved actors and other properties (van Hage & Ceolin 2011).
SIOC	Semantically-Interlinked Online Communities	SIOC describes the structure of social networks and is used in combination with FOAF to describe social networks and social media (Bojars & Breslin 2009).
SKOS	Simple Knowledge Organization System	SKOS defines a data model to describe knowledge organization systems and is used to describe our ontology (Miles & Bechhofer 2009).

Figure 4 depicts the included ontologies and their general category (rectangular boxes) which describes with which purpose we use the in the SMEM ontology. Most ontologies that we included were built to describe metadata information. Metadata in the social web describes locations, such as coordinates, addresses or countries, time, and post and file specific information, such as the author, or the file type. Furthermore, metadata is needed to describe events like single emergencies. In case of an emergency event, the information needs to be available in common standards, which have to be readable for machines and humans. Here existing standards for emergency data exchange were regarded and used, such as the Common Alerting Protocol. Regarding social media, several ontologies that help in describing social networks in general and user relations were included. Finally, for the description of information quality, we decided to rely on graphs, which allows us to represent information quality as a tree with nodes, branches and leaves. This graph-tree structure enables a detailed description of indicators for information quality and a simple change of indicators and their weight for the overall information quality value.

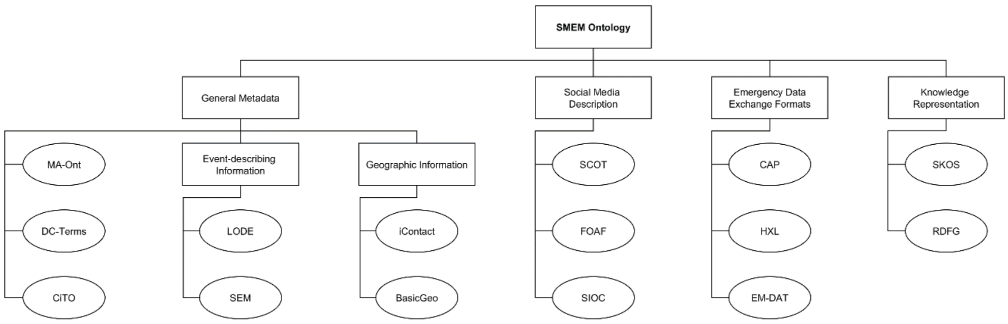

Figure 4. Included ontologies and standards by category

5.1 Incident Classification

The emergency services, the end-users of the EmerGent IT system need to classify information to be able to evaluate threats. During workshops held with the project's end-user advisory board, it became apparent that a classification by threat, consisting of certainty, urgency and severity, as well as a classification by incident type is a good method for such classification. Possible disaster types range from natural catastrophes to human made disasters and each disaster brings about its own properties and available information. We therefore decided to allow the classification of different types of incidents. These different types of incidents will allow adding event specific information that can be crucial for emergency service. An example for this use can be seen when regarding different types of fires. E.g. a forest fire and an industrial fire are both fires with a certain, equal base of information needed, such as location, size and citizens affected. Additionally, the industrial fire needs information about the materials and chemicals involved, as these could influence the methods needed to extinguish the fire. In addition, the classification makes it possible to use interdependences between incidents and other entities. This can be used to automatically recognize dangerous situations in critical infrastructure, such as a fire in a nuclear power plant.

To classify disaster information in a common, standardized manner, we decided to categorize the incidents according to the EM-DAT international disaster database from the Centre for Research on the Epidemiology of Disasters (Université Catholique de Louvain 2009), as can be seen in Figure *4*. Following the EM-DAT classification, we classify incidents in natural and technological disasters. Natural disasters cover geophysical, meteorological, hydrological, climatological, biological and extra-terrestrial disasters. The extra-terrestrial category describes meteorite, asteroid and comet incidents, as well as space weather, like solar flares. Technological disasters cover industrial, transport and other incidents that are man or technology influenced.

6. INFORMATION QUALITY

Although the vast amount of information from social media sites could be extremely valuable for emergency responders, the services are reluctant to rely their operations on them (Hughes & Palen 2012). This is due to misinformation that can be found on social media during crisis, which so far causes most of the work for emergency services that are represented on social media. Emergency managers, which already adopted social media as part of their processes report, that the identification and counteracting of misinformation is one of the most important factors of their work so far (Latonero & Shklovski 2011; Hughes & Palen 2012). Regarding the EmerGent project, the question was raised how uncertain, false or unrelated information could be recognized and handled.

When considering these different kinds of information, three categories of information quality become apparent: misinformation, unrelated information and other, which is everything else. Unrelated information describes every information, that is collected by the IT system, but that is irrelevant for the situation at hand or does have nothing to do with emergencies. Misinformation has to be divided into two different subcategories itself: false positives and false negatives (Goodchild & Glennon 2010). False positives represent real false information, such as wrong rumours, ambiguous statements and lies. The system may detect an emergency here, when there is none. False negatives represent real emergencies, with no information to be found concerning these emergencies. E.g. a wildfire in a remote area, about which no information has been shared. As this emergency is existent, but no information has been published, the system will not detect an emergency, a false negative.

To tackle this problem in EmerGent, an information quality system was designed as part of the overall IT system. We developed an approach that regards information quality as a graph, or tree. The root node describes the overall information quality score, which is calculated from different criteria and indicators. Each criterion is calculated based on several sub criteria and indicators, which represent the leaves of the tree. The end-users are able to customize the tree and the weighting of each criterion and indicator. Next to several indicators within the Tweets themselves, also other factors, such as the reputation of the source, the closeness to the situation and the timeliness after the incident are used. In addition, the Tweets are analysed for semantic information that allows us to draw conclusions regarding the quality of the tweet. This can tackle the problem of false rumours, but false positives, which cannot be regarded in relation to a previously known incident and false negatives, where no information is available about a real incident, are still problems that need to be tackled. (Markham et al. 2015)

The structure of the information quality graph can be seen in Figure 5. As mentioned before, a tree structure was chosen for the graph, which is made up of criteria, sub-criteria and indicators. The graph consists of five criteria, understandability, relevance, completeness, timeliness and believability. Understandability examines how easy a post is to understand for the reader in an emergency service. This includes the language of the post, as well as a readability score. Relevancy calculates the probability that a post is interesting for an emergency service, which is based on the count of keywords and relevant entities, e.g. victims. Completeness regards the existence of certain indicators to determine whether it contains the information needed for the use by an emergency service. Timeliness of a post is used to determine, whether it is up-to-date and therefore could be useful. The believability measurement not only evaluates if a post itself is potentially true, but also takes the reputation of its author into account. Believability of a post can be significantly influenced by its location, the interaction of other users with it and by the existence of media files and URLs that could support the information in the post.

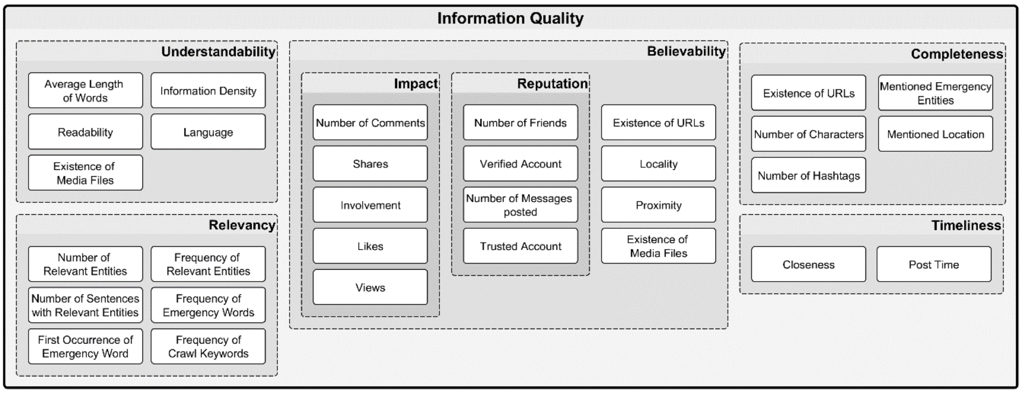

Figure 5: Structure of the IQ graph

7. SMEM ONTOLOGY STRUCTURE

Based on the evaluation results and the requirements placed in the ontology as described before, we designed our SMEM ontology. It unifies several different and already existing ontologies and extends them to perfectly fulfil its objective, as described in section 4. The ontology is designed to be easily extendable, which will be useful for possible functionality extensions in the future. An exemplary use of these relations can be found in Figure 3. It does not contain all used information and is only supposed to describe the general concept of our design.

Here a set of posts that was gathered from a social network, containing media files, e.g. images, gets enhanced with additional information about the quality of the posts. These posts are linked to a specific emergency alert that contains information about an operation to save an injured person. The alert also contains emergency related information, like the location, or the type of an event. With the usage of media files, like pictures, the emergency responders gain insight into the situation, before even arriving on site. This saves valuable moments that can be used to save the injured person's life.

The SMEM ontology consists of six main parts, describing alerts, incidents and events, social media, information quality, agents and metadata, like locations and time (Table 3). To describe the relations between these parts, we use semantic relations and properties. These connections and relations enable us to link information with a lot of other information and to draw conclusions from these relations.

Table 3. Parts of the SMEM ontology

Ontology Part	Description
Alerts	The alert part defines the usage of the Common Alert Protocol (CAP) alert and an enhanced alert. Both containing crucial information about a certain emergency, e.g. location, time, source. The enhanced alert also contains references to social media sources and media files.
Events	The event part further describes and defines different event and incident types. It allows a classification according to the specifications of the International Disaster Database (EM-DAT), therefore making our results widely compatible and able to handle a huge variety of different incident types. Later, this ontology will be expanded to contain further information for specific types of incidents, e.g. special emergency responses.
Information quality	The information quality part defines a RDF Graph used to evaluate information quality (IQ Graph). Here, different criteria, like trustworthiness, are used to calculate a value that represents the information quality. Using this value, it is possible to assess each information and each informant.
Social media & the web	The web part describes social media platforms, online information and user interaction on the web. Here external ontologies like SIOC and FOAF are used to describe social networks.
Agents	The agent part describes agents and their respective roles on the web. Persons, organizations, online accounts, software, etc. can be agents.
Metadata	The metadata part describes standardized metadata that is used to define events. This part includes but is not limited to the usage and reference of DCMI-terms and HXL.

To ensure interoperability of the IT system, for which the SMEM ontology was created, with existing command and control software, and to simplify data exchange it was important to regard existing exchange formats for alert information that can be used in real emergencies. The Common Alerting Protocol (CAP) standard is a format for exchanging all hazard emergency alerts and public warnings over all kinds of networks and is already successfully implemented in several countries (Jones et al. 2010). A CAP Alert has been defined, which is derived from OASIS' CAP standard and contains all information demanded by the standard, to ensure compatibility with real world incidents and to allow for standardized emergency communication.

Realizing that the CAP standard alone cannot contain much of the information, collected from social media, e.g. images, and needed for the use of social media data, e.g. information quality, we decided to extend it. We called this extended alert the "SMEM Alert" and it allows us to additionally transmit information about social media information and posts that reference the incident at hand. Also this alert allows us to transmit information that is demanded in other emergency related data exchange formats, such as EDXL (Raymond et al. 2006). These SMEM Alerts are also used to generate notifications in the EmerGent IT system. (Reuter, Amelunxen, et al. 2016)

In addition to the SMEM Alerts, we defined a graph that allows us to represent the quality of gathered information. Such assessment of information quality for the information gathered in social networks was one of the major concerns of the EAB. The graph serves as a general representation of the information quality of a posting and the trustworthiness of a content creator. Information quality refers to the content of the information received, capturing a wide range of variables, which are divided into indicators and criteria. Criteria summarize all of its corresponding indicators. The value of each criterion consists of the value of its sub-criteria and indicators. The user can decide the weighting of each criteria and indicator individually.

The following graph (Figure 6) depicts the main classes used in the SMEM ontology. Many classes have several subclasses, e.g. “Incident” has 80 subclasses, describing different types of incidents. To allow for better comprehensibility we decided to not depict all classes and relations in the following figure, but to only show a high-level view of our ontology, which only contains the main classes and relations. These classes just depict a very abstract view of the ontology, which is far larger and more complex.

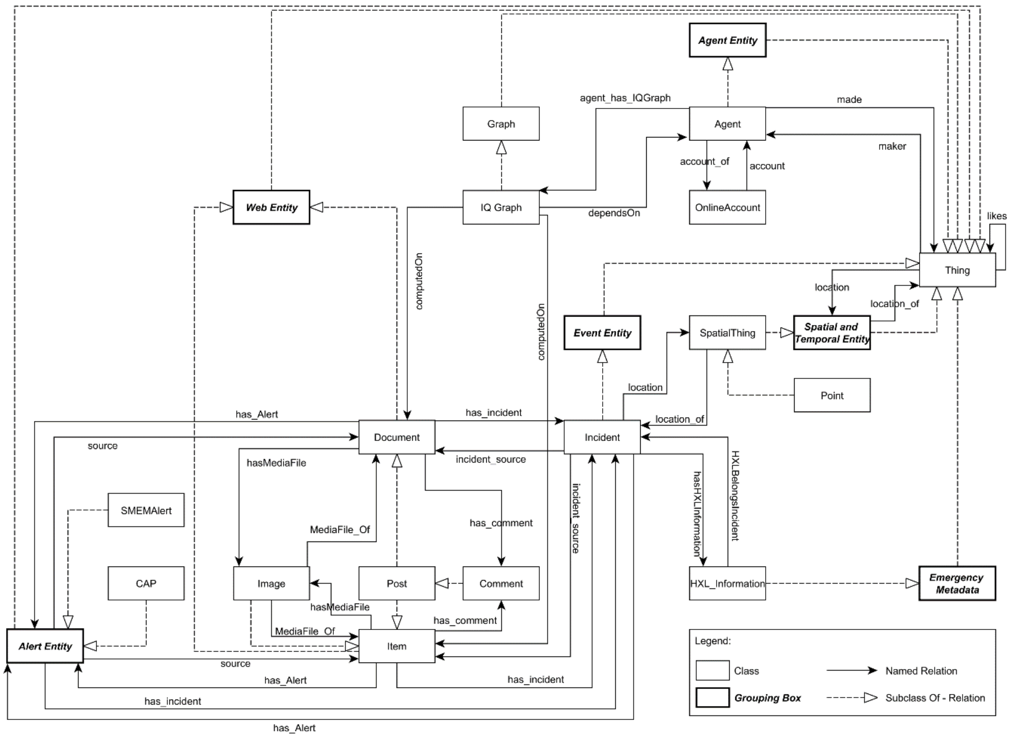

Figure 6. Main classes and relations of the SMEM ontology

In the context of the overall IT system, which is being developed in EmerGent, the ontology is used to structure the gathered data in a condensed way, which can be used effectively by the IT system and emergency services. The structuring according to alerts or situations makes it possible to link mined social media posts, to already existing information and domain knowledge, thus reducing the overall noise generated during situations and making relevant information easily displayable. The mock up presented in Figure 2 shows the IT system’s frontend, which displays the gathered, mined and stored data in our final system. Emergency services can use this simple interface to view posts from social media and their corresponding alerts.

8. CONCLUSION & FUTURE WORK

Ontologies, as a way to categorize and structure information, provide a good basis to build an IT system for emergency services to for linking social media data with emergency related information. This paper described the development process of our SMEM ontology, which will, as part of an IT system, help emergency services to achieve this goal. We intended to assist emergency services with the efficient use of social media. By designing this ontology for social media in emergency management (SMEM ontology), we developed the means to categorize data from social media, associate it with emergency domain knowledge and interlink it with other information to allow effective semantic querying using SPARQL. To increase interoperability, we evaluated existing ontologies in the social media area and the emergency domain. Following this evaluation, we built our SMEM ontology, which provides the grounds to categorise and link the data and information from social media with specific emergency domain knowledge and the additional information that is needed to properly describe an incident and aid the emergency services before, during and after emergencies.

Further, it is planned to include additional standards and to regard research conducted on this topic in the US. In addition, we plan to extend the ontology in order to include emergency response and criminal acts, such as terroristic attacks, and to describe incident specific information about different kinds of incidents. This breakdown with event specific information would enable us to store information about how to react to certain incidents, which might be a useful extension in the future. Furthermore, it would be helpful for planning an appropriate or even automated response. Regarding information quality, further research has to be done in the field of detecting and handling false negatives in general and false positives, if the incident was previously unknown.

As our IT system needs to handle high volume data sets in near-real time, we already started with the technical analysis of RDF storage solutions in terms of scalability. It is commonly known that dealing with high volume data sets in RDF storage solutions may cause scalability issues in terms of storage and processing time (Khadilkar et al. 2012). We need to tackle this issue to ensure good overall performance for our IT system.

ACKNOWLEDGEMENT

This research has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608352.

REFERENCES

Akerkar, R., Friberg, T. & Amelunxen, C., 2016. *Deliverable 3.5: User Requirements, Version 2*, Paderborn. Available at: http://www.fp7-emergent.eu/wp-content/uploads/2016/03/20160229_D3.5_RequirementsVersion2_EmerGent_final2.pdf.

Barros, R. et al., 2015. Edxl-rescuer ontology: Conceptual model for semantic integration. In *Proceedings of the 12th International ISCRAM Conference*.

Bojars, U. & Breslin, J.G., 2009. SIOC Core Ontology Specification. Namespace document, DERI, NUI Galway.

Bontas, E.P., Mochol, M. & Tolksdorf, R., 2005. Case Studies on Ontology Reuse. In *Proceedings of I-KNOW'05*. pp. 345–353.

Brickley, D., 2003. Basic Geo (WGS84 lat/long) Vocabulary. *W3C Semantic Web Interest Group*. Available at: http://www.w3.org/2003/01/geo/.

Brickley, D. & Miller, L., 2010. FOAF Vocabulary Specification. *Namespace Document*, 3, p.http://xmlns.com/foaf/spec/. Available at: http://xmlns.com/foaf/spec/ [Accessed March 26, 2016].

Carroll, J.J. et al., 2005. Named graphs, provenance and trust. In *Proceedings of the 14th international conference on World Wide Web*. pp. 613–622.

Clegg, D. & Barker, R., 2004. *Case Method Fast-Track: A Rad Approach*, Addison-Wesley.

DCMI Usage Board, 2012. DCMI Metadata Terms. Available at: http://dublincore.org/documents/dcmi-terms/ [Accessed July 16, 2015].

DISASTER, 2015. DISASTER - Data Interoperability Solution At STakeholders Emergency Reaction.

DISASTER, 2012. emergel.

Etherington, D., 2012. Twitter Passes 200M Monthly Active Users, A 42% Increase Over 9 Months. *TechCrunch*. Available at: http://techcrunch.com/2012/12/18/twitter-passes-200m-monthly-active-users-a-42-increase-over-9-months/.

Facebook, 2016. Facebook - Monatlich aktive Nutzer weltweit 2016 | Statistik. *Statista*.

Fox, M.S., 2015. International Contact Ontology: Addresses, phone numbers and emails. Available at: http://ontology.eil.utoronto.ca/icontact.html [Accessed July 16, 2015].

Galton, A. & Worboys, M., 2011. An Ontology of Information for Emergency Management. *International Conference for Crisis Response and Management (ISCRAM)*, 8(May), pp.1–10.

Goodchild, M.F. & Glennon, J.A., 2010. Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3(3), pp.231–241.

Grolinger, K., Brown, K. & Capretz, M., 2011. From Glossaries to Ontologies: Disaster Management Domain.

Gruber, T., 2009. Ontology. In L. Liu & M. T. Özsu, eds. *Encyclopedia of database systems*. Springer-Verlag, pp. 1963–1965.

Gruber, T.R., 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), pp.199–220. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.101.7493.

van Hage, W.R. & Ceolin, D., 2011. The Simple Event Model. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(2), pp.128–136. Available at: http://www.researchgate.net/publication/229066776_The_Simple_Event_Model/file/d912f4ffe7e452324b.pdf [Accessed October 14, 2015].

Hiltz, S.R. & Plotnick, L., 2013. Dealing with information overload when using social media for emergency management: emerging solutions. In T. Comes et al., eds. *Proceedings of the 10th international ISCRAM conference*. pp. 823–827.

Houston, J.B. et al., 2015. Social media and disasters: a functional framework for social media use in disaster planning, response, and research. *Disasters*, 39(1), pp.1–22.

Hughes, A.L. & Palen, L., 2012. The evolving role of the public information officer: An examination of social media in emergency management. *Journal of Homeland Security and Emergency Management*, 9(1), pp.1–20.

HXL Working Group, 2016. Humanitarian Exchange Language (HXL). Available at: http://hxlstandard.org/ [Accessed February 16, 2016].

Jones, E., Systems, W. & Westfall, J., 2010. Common Alerting Protocol Version 1.2. , (July), pp.1–47. Available at: http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2.pdf [Accessed October 14, 2015].

Kessler, C. & Hendrix, C., 2015. The Humanitarian eXchange Language: Coordinating disaster response with semantic web technologies. *Semantic Web*, 6, pp.5–21.

Khadilkar, V. et al., 2012. Jena-HBase: A distributed, scalable and efficient RDF triple store. *CEUR Workshop Proceedings*, 914(ii), pp.85–88.

Latonero, M. & Shklovski, I., 2011. Emergency management, Twitter, and social media evangelism. *Latonero, M. & Shklovski, I.(2011). Emergency management, Twitter, & Social Media Evangelism. International Journal of Information Systems for Crisis Response and Management*, 3(4), pp.67–86.

Limbu, M., 2012. Management of a Crisis (MOAC) Vocabulary Specification. *Online at http://observedchange. com/moac/ns*. Available at: http://observedchange.com/moac/ns/.

Löffler, D.J., 2009. The FP7 Project PRONTO.

Markham, D. et al., 2015. *Deliverable 4.6: Specification of Information Mining and Information Quality Components*, Oxford.

Martinez-Cruz, C., Blanco, I.J. & Vila, M.A., 2012. Ontologies versus relational databases: Are they so different? A comparison. *Artificial Intelligence Review*, 38(4), pp.271–290.

Michel, T., 2012. Ontology for Media Resources 1.0 Namespace Document. Available at: https://www.w3.org/ns/ma-ont [Accessed March 20, 2016].

Miles, A. & Bechhofer, S., 2009. SKOS simple knowledge organization system reference. *W3C recommendation*, 18, p.W3C.

Moi, M. et al., 2015. Strategy for Processing and Analyzing Social Media Data Streams in Emergencies. In *Proceedings of the International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*. Brest, France.

Moi, M. & Rodehutskors, N., 2016a. *Deliverable 4.2: Semantic Data Model – Version 2*, Paderborn. Available at: http://www.fp7-emergent.eu/wp-content/uploads/2016/04/20160330_D4.2_Semantic-Data-Model_final_pub.pdf.

Moi, M. & Rodehutskors, N., 2016b. Design of an ontology for the use of social media in emergency management. In P. Kommers, A. P. Abraham, & J. Roth, eds. *International Conferences ICT, WBC, BIGDACI and TPMC 2016*. Madeira, Portugal: IADIS Press, pp. 129–136.

Petrovic, S. et al., 2013. Can twitter replace newswire for breaking news? *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, pp.713–716.

Pirolli, P., 2009. Powers of 10: Modeling Complex Information-Seeking Systems at Multiple Scales. *Computer*, 42(3), pp.33–40.

Raymond, M., Webb, S. & Aymond, P.I., 2006. Emergency data exchange language (edxl) distribution element, v. 1.0. *OASIS Standard EDXL-DE v1. 0*, 1, p.354.

RESCUE, 2016. RESCUE.

Reuter, C., Ludwig, T., et al., 2016. Emergency services' attitudes towards social media: A quantitative and qualitative survey across Europe. *International Journal of Human-Computer Studies*, p. Available at: http://www.sciencedirect.com/science/article/pii/S1071581916000379.

Reuter, C., Amelunxen, C. & Moi, M., 2016. Semi-Automatic Alerts and Notifications for Emergency Services based on Cross-Platform Social Media Data –Evaluation of a Prototype. In H. C. Mayr & M. Pinzger, eds. *Informatik 2016: von Menschen für Menschen*. Klagenfurt: GI-Edition-Lecture Notes in Informatics (LNI).

Scerri, S. et al., 2012. SCOT Ontology Specification. Available at: http://rdfs.org/scot/spec/ [Accessed July 16, 2015].

SecInCoRe, 2016. SecInCoRe.

Shaw, R., Troncy, R. & Hardman, L., 2010. LODE: An ontology for Linking Open Descriptions of Events. Available at: http://linkedevents.org/ontology/ [Accessed July 16, 2015].

Shotton, D. & Peroni, S., 2015. CiTO, the Citation Typing Ontology. Available at: http://www.sparontologies.net/ontologies/cito/source.html [Accessed July 16, 2015].

Shrivastav, H. et al., 2012. Facebook News Feed: relevance or noise?

Slándáil, 2015. Slándáil.

Spielhofer, T. & Reuter, C., 2015. *EUROPEAN SURVEY OF CITIZENS' ATTITUDES TOWARDS SOCIAL MEDIA*, Available at: http://www.fp7-emergent.eu/wp-content/uploads/2014/05/CITIZEN-SURVEY-SUMMARY-REPORT.pdf.

Statista, 2016. Marktanteile von Social Media Seiten nach Seitenabrufen weltweit im August 2016. *Statista*. Available at: http://de.statista.com/statistik/daten/studie/241601/umfrage/marktanteile-fuehrender-social-media-seiten-weltweit/ [Accessed September 3, 2016].

SUPER, 2016. SUPER.

Thackeray, R. et al., 2012. Adoption and use of social media among public health departments. *BMC Public Health*, 12(1), pp.1–6. Available at: http://dx.doi.org/10.1186/1471-2458-12-242.

Université Catholique de Louvain, 2009. EM-DAT: The OFDA/CRED International Disaster Database. Available at: http://www.emdat.be [Accessed July 16, 2015].

World Wide Web Consortium (W3C), 2009. W3C EIIF Incubator Group Report. *W3C*. Available at: http://www.w3.org/2005/Incubator/eiif/XGR-Framework-20090428/.

Yin, J. et al., 2012. Using Social Media to Enhance Emergency Situation Awareness. *IEEE Intelligent Systems*, 27(6), pp.52–59. Available at: http://sfx.hbz-nrw.de/sfx_pad?sid=google&auinit=J&aulast=Yin&atitle=Using social media to enhance emergency situation awareness&id=doi:10.1109/MIS.2012.6&title=IEEE Intelligent Systems and their Applications&volume=27&issue=6&date=2012&spage=52&issn=1541-.

Zelewski, S., Schütte, R. & Siedentopf, J., 2001. Ontologien zur Repräsentation von Domänen. In *Wissen in Unternehmen. Konzepte, Massnahmen, Methoden*. Berlin, pp. 183–221. Available at: https://www.pim.wiwi.uni-due.de/uploads/tx_itochairt3/publications/Zelewski_Schuette_Siedentopf_-_Ontologien_zur.pdf.