

# The EmerGent project: Emergency Management in Social Media Generation

## Dealing with Big Data from Social Media Data Streams

Reynold Greenlaw, Andrew Muddiman  
Oxford Computer Consultants Ltd  
Oxford, UK  
www.oxfordcc.co.uk

Therese Friberg, Matthias Moi  
University of Paderborn  
Paderborn, Germany  
<https://www.cik.uni-paderborn.de>

Massimo Cristaldi  
IES Solutions Srl  
Roma, Italy  
www.i4es.it

Thomas Ludwig, Christian Reuter  
-University of Siegen  
Siegen, Germany  
www.cscw.uni-siegen.de

**Abstract**— EmerGent will use social media to support the management of large scale emergencies. The project includes the construction of a big online store of data which will be continuously mined to provide emergency information and alerts. The overall objective is a stronger connection between citizens and emergency management authorities through social media.

**Keywords**-component; emergencies; data mining; social media; information mining; information quality

### I. INTRODUCTION

This paper describes EmerGent which is a recent EU FP7 project currently engaging with users, gathering requirements and writing initial technical specifications and which deals with the impact of social media in emergency management. This paper describes the overall objectives of the project and the plans to create an online big data semantic store of social media.

### II. SOCIAL MEDIA AND EMERGENCIES

Citizen-based crisis communities are currently only weakly connected through social media to Emergency Services (ES) during crises. The emergency management cycle (EMC) does not capture social media with its heterogeneous but valuable information. Currently existing apps (e. g. for mobile devices) are used by citizens to share observations and feelings but are weakly connected to existing ES systems.

### III. THE OBJECTIVES OF EMERGENT

The EmerGent project has five main objectives:

- (1) to analyse the impact of social media for citizens and for ES in the whole EMC,
- (2) to show the positive impact of information mining, information quality, information gathering, and information routing of social media in emergencies,

- (3) to identify the requirements for implementation, and the methods and tools for evaluation of novel emergency management in social media generation,
- (4) to provide officials and the public with guidelines for the use of social media in emergencies and finally
- (5) to clarify the potential for exploitation of social media in emergencies

In order to achieve these objectives an EmerGent IT-system will be developed, which will act as a tool-set for further analysis.

### IV. INFORMATION TECHNOLOGY IN EMERGENT

With the help of the proposed EmerGent IT-system the connection between citizens and ES will be enhanced, because ES will be able to find effective channels within social media to manage emergencies and citizens will be encouraged to report information.

Considering the citizen-to-authority (C2A) interaction (see Reuter 2012 [1] for a classification), citizens can share information through social media as normal but with the knowledge that their contribution may help ES. All information sent by citizens on social networks selected by the project (Facebook, Twitter, YouTube, Google+, InstaGram) will be coarsely filtered and gathered by EmerGent. The processing and analysis system of EmerGent (based especially on Information Quality (IQ) and Information Mining (IM) processes) will skim the list of messages in order to retrieve concise and accurate information to be communicated to ES.

At the end of the process, from a very large quantity of social media data, EmerGent will produce alerting and information to be sent to ES in all phases of the EMC. At this point EmerGent will adapt the data to specific formats that are used by ES: protocols like CAP and EDXL. Alerting messages will be completed using routing rules (provided by a routing system) and then sent to ES using mechanisms developed and used in other EU projects (REACT FP6 033607 and IDIRA FP7 261726).

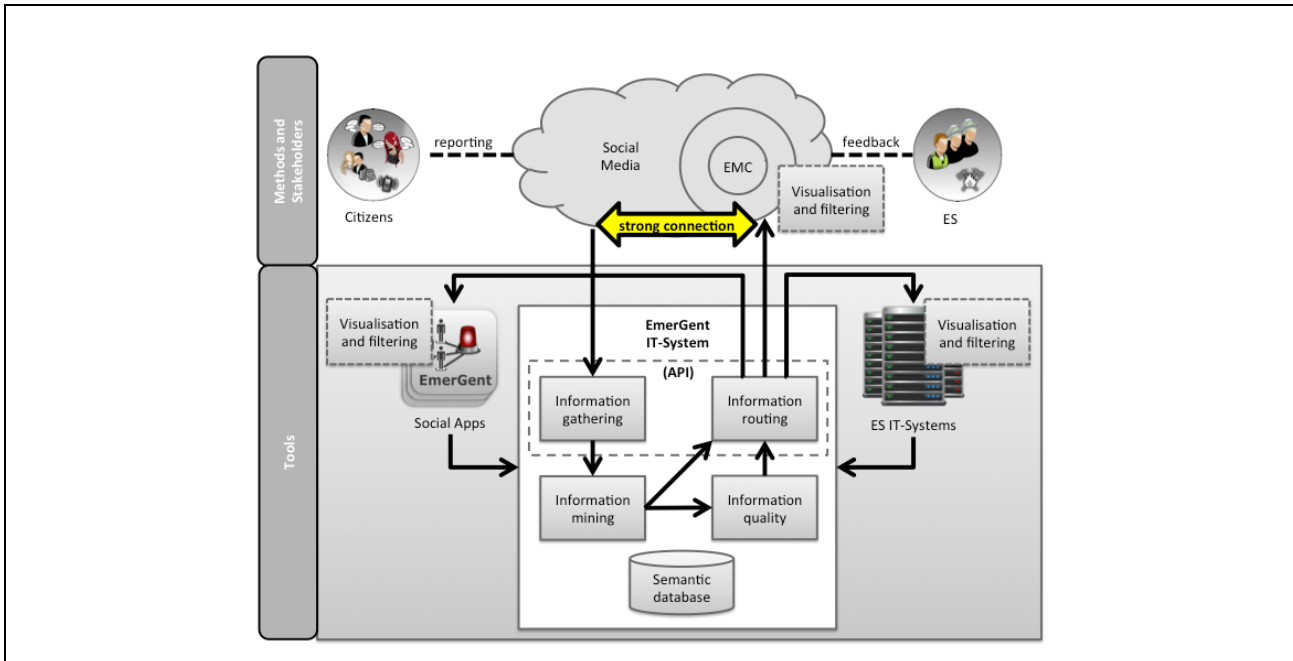


Figure 1: Schematic of the EmerGent IT-system

At this point Authority to Citizen (A2C) communication will be enabled. Although ES systems are not equipped for social media, they use interoperability alerting protocols through which they can delegate the A2C communication to EmerGent. In this way, ES systems can be both uncoupled from social media and yet able to use it for the broadcast of messages. The social media information to be shared by EmerGent will be adapted CAP-format messages. EmerGent will extract information from alerting messages and route them toward social media using available APIs. Citizens therefore both read and feed social media information through an emergency.

## V. EMERGENT AND ONTOLOGY

The information from social media has to be structured in a useful way as a prerequisite to its use. The most promising approach is to structure and enrich information semantically, because it simplifies the way to create context-aware queries. Bontcheva & Rout [2] summarized it as follows:

“Semantic technologies have the potential to help people cope better with social media-induced information overload. Automatic semantic-based methods that adapt to the individual’s information seeking goals and summarize briefly the relevant social media could ultimately support information interpretation and decision making over large-scale, dynamic media streams.”

With continuously growing requirements on data processing and information querying several models have emerged to render machine-readable information. In particular relational databases became popular when file-based approaches reached their limits in structuring data [3]. Now, with the

growing emergence of the Semantic Web, semantic technologies have become more important. To represent information in the Semantic Web a structural framework is required and ontologies are now a very common method of representing a hierarchy of concepts within a domain. To denote the types, properties and interrelationships, a shared vocabulary is needed [4]. Based on Vergara [5], the ontology in EmerGent is characterised as follows:

- It is explicit, because it defines concepts, properties, relationships, functions, axioms and constraints that compose it
- It is formal, because it is machine readable and interpreted
- It is a conceptualization because it is an abstract model, resp. a simplified view of the domain it represents
- It is shared, because there is a consensus about the information and it’s accepted by a group of experts

EmerGent intends to firstly develop an ontology to structure emergency-related data from social media sources (based on standards like SIOC, MOAC or FOAF) and then develop a scalable Semantic Data Store that implements that ontology.

From the modelling perspective the main steps are:

- The development of a model of social media data which captures its relevant features, properties and constraints. In particular meta-data such as replies, mentions, references, keywords, timestamps and geo-location, have to share a common structure for information mining and information quality processing. It is necessary to understand the

semantics of social media data to exploit its full potential.

- The development of a model to describe domain-related information for emergency services. The purpose is to model concepts like “incident descriptions”, “alerts” or “requests”.
- The creation of a mapping between the two: domain-related information to information from social media. This captures the associations between emergency-related information and processed social media data.

Thus the ontology will enable semantic analysis and data mining for the detection of patterns, incidents, unusual events and to discover correlations.

#### VI. THE PROPOSED EMERGENT SEMANTIC DATA STORE

EmerGent will build a large elastic data store in the cloud. It will include semantic data stored as RDF for OWL. In order to handle data of this size we expect a requirement for parallel computation, subdividing information and execution between different machines that work (generally) in the same network. Although some relational databases are horizontally scalable by sharding rows between machines (but with loss of ACID characteristic properties), NoSQL solutions are most suitable for these purposes.

The main challenge to storing RDF objects in NoSQL databases is to find the right way to represent graph inside them. Different studies have tried to use HBase (based on Hadoop) as a NoSQL database coupled with a semantic data framework like Apache Jena (e.g. Khadilkar [6]).

To use a Jena+HBase solution it is necessary to adapt the Jena Framework so that it can execute SPARQL queries; this type of approach should be carried out for each solution that does not use a Jena standard database. Considering data organisation inside HBase, for EmerGent we suspect the Hybrid Layout will provide a good compromise between performance and storage space efficiency – but this needs to be tested.

In EmerGent we will explore a combination of NoSQL plus a purely Semantic database. For data gathering, a NoSQL solution possibly based on MongoDB may be able to fulfill requirements in terms of performance and scalability. For the semantic store, scalability to big data sizes is challenging. For standard databases with a semantic framework like Jena, scalability is difficult to apply and EmerGent may need to rely on NoSQL and a scalable engine with tools to link their native language with SPARQL. There are also stand-alone solutions (e.g. Jena TDB) that will be analysed for whether they can be used for EmerGent (using previous crisis events to estimate the amounts of data that EmerGent could receive).

The strategy EmerGent will explore consists of maintaining in the semantic storage only the data that will be analysed (performing queries and apply data mining techniques), and creating a parallel storage for data that is already processed. The amount of data is unpredictable and if

big enough to require scalability, hybrid solutions may provide a good compromise. For example we can load data on both stand-alone and scalable data stores and use the first when the amount of data does not exceed its performance threshold, otherwise switching to the scalable version.

#### VII. DATA PROCESSING IN EMERGENT

Data from social media are characterized by complexity and heterogeneity. The data contains complex relationships and dependencies within itself and this, combined with its heterogeneous nature, imposes strong limitations on the data models that can be used and the scope of information that can be discovered. Data mining in the context of social media must continuously transform raw social media data into a processable form by selectively using specific characteristics needed in the analysis process. Three key challenges of data mining need to be overcome if EmerGent is to succeed. These are

1. To reduce high-volume, low-quality social media reports into low-volume, high-quality data appropriate for further mining,
2. To identify networks, especially social networks, within the data to understand the context of an emergency and the individual users and
3. To identify, categorise and match requests for, and offers of, help and supplies during and after an emergency.

These will be the primary objectives of the data mining process that EmerGent will develop. EmerGent is due to complete in 2017.

#### ACKNOWLEDGMENT

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

#### REFERENCES

- [1] Reuter, C., Marx, A., & Pipek, V. (2012). Crisis Management 2.0: Towards a Systematization of Social Software Use in Crisis Situations. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, 4(1), 1–16
- [2] Bontcheva, K., & Rout, D. (2014). Making sense of social media streams through semantics: A survey. *Semantic Web*, 5, 373–403
- [3] Martinez-Cruz, C., Blanco, I., & Vila, M. (2011). Ontologies versus relational databases: are they so different? A comparison. *Artificial Intelligence Review*, 271–290
- [4] Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 199–220.
- [5] Vergara, J., Villagrà, V., & Berrocal, J. (2002). Semantic Management: advantages of using an ontology-based management information meta-model. *Proceedings of the HP Openview University Association Ninth Plenary Workshop (HPOVUA'2002)*, Böblingen, Germany, (pp. 11–13).
- [6] Khadilkar, V., Kantarcioglu, M., Thuraisingham, B., & Castagna, P. (2012). Jena-HBase: A Distributed, Scalable and Efficient RDF Triple Store. *11th International Semantic Web Conference*. Boston MA.