20th International Conference on Information Fusion
Xi'an, China - July 10-13, 2017

# Using Social Media in Crisis Management

## SOTERIA Fusion Center for Managing Information Gaps

Claire LAUDY
Thales Research & Technology
Palaiseau, France
claire.laudy@thalesgroup.com

Fabio RUINI, Alessandro ZANASI
Zanasi & Partners
Italy
fabio.ruini@zanasi-alessandro.eu;
alessandro.zanasi@zanasi-alessandro.eu

Marcin PRZYBYSZEWSKI, Anna STACHOWICZ
ITTI
Poland
mprzybysz@itti.com.pl; anna.stachowicz@itti.com.pl

***Abstract*— The development of mobile devices as well as social media platforms recently lead to the necessity of monitoring the latter during crisis and emergency situations. Paradoxically, the huge amount of information available through these new sources may lead to information gaps, within the Public Safety Organization operators' awareness. We describe some specific types of information gaps due first to imprecise or unreliable information and second to information overload. We then propose a set of tools aiming at reducing these information gaps and supporting the human operators in the social media generated information during crisis and emergency management. The first tool aims at geolocalising tweets relying on the content of the messages. The second tool provides sentiment analysis and clustering of multi-lingual messages and the third tool provides means for semantic information fusion and hypothesis evaluation relying on the contents and metadata of the tweets reporting about an event.**

***Keywords—social media; emergency & crisis management; information gaps; information fusion, text based geolocalisation; multilingual text analysis***

## I. Introduction

The role played by social media in crisis and emergency events takes action from the situation assessment phase, with new sources of information about the ongoing (potentially remote) situation, to the dispatch of response efforts. Therefore, the use and management of social media during crisis is an emerging and rapidly growing trend in the crisis management and IT research communities ([1][2], [3], [4], [5], [6], [7], [8], [9], [10], [11]).

As information is now provided and shared by anyone to anyone, precision, quality and credibility of information provided to the Public Safety Organizations (PSO) is a major issue. Furthermore, the development of social media platform as well as portable devices enables any citizen to widely provide, forward and share information on any emergency or crisis event. The availability of all these information sources may be an opportunity for PSO, in order to get information from the inside as soon as possible, and possibly answer to specific call for helps, however, this also results in a huge amount of information available and that has to be analyzed.

In this paper, we focus on the specific problem of information gaps at the level of the PSO officers themselves. This situation is paradoxically caused both by the diversity of information sources, as well as the huge amount of information and the inability to manage such a huge mass of information. We propose to overcome some of the gaps using specific tools laid out within a so called *Fusion Center*.

In the second section of this paper, we describe the context of our work, namely the SOTERIA project, the specific information gaps we focused on and the architecture of the Fusion Center. Each subsequent section is then dedicated to each tool: section 3 describes the Text analysis tool tweet locator, section 4 describes the multi lingual social stream analysis and section 5 describes the Semantic information Fusion and Evaluation tool. Section 6 presents some of the field experimentations that we conducted. We then conclude describing an experimentation the tools were involved in.

## II. Context

### A. *Use of social media during crisis: the SOTERIA Project*

Social Media monitoring has become a major issue in crisis and emergencies management. Emergencies throughout the World recently prompted new attention to the role new mobile technologies and social media platforms play in emergency situations and response efforts [12]. By studying the dynamics between PSOs (Public Safety Organizations) and citizens, the SOTERIA Project aims to research and develop recommendations and tools for leveraging the potential of social and mobile media in emergencies.

During the discussions held with the PSOs throughout the SOTERIA project, it appeared that the huge mass of information available through the social networks, may itself, paradoxically, lead to situations where PSOs face information gaps at their level. This may be caused, among others, to the inability to manage a huge mass of information. Therefore, tools are needed to support the PSOs, and summarize information, while highlighting important aspects. The multiplication of information sources may also cause these information gaps, as it is impossible manage them all ”manually”.

### B. Information gaps

This section describes shortly the problem of information gaps in the information received. Generally the detection of a data-gap is a result of the analysis performed at the Operations Centre (OC):

- Operations center receives incoming data;
- Operations center looks at the incoming data and makes an assessment;
- Operations center may conclude after the assessment that they are missing certain data the make an informed decision
- Operations center may require additional information (demand side driven).

At first, the definition of an information gap seems straightforward. However, analysis of collected data from Social Media channels during emergency is a very complex task as SOTERIA tools receive data from multiple sources and also can be created by both - citizens and public safety organizations, taking advantage from nowadays mobile and online communications.

Within this work, we focus on three types of data gaps, differentiated by their causes:

- Unprecise and/or unreliable information,
- Missing of information due to information overload

#### 1) Unprecise/unreliable information

The first one occurs when one or more sources of information are not precise enough to provide all the necessary information. Within SOTERIA and crisis management supported by social media, this is the case, for instance, when a citizen provides information to PSO though a tweet message, but the geolocalisation is off. The PSO will receive the information contained in the textual message, but cannot relate the information to a specific place. As tweets messages are very short messages, this may be an important limitation for PSOs, in order to take the information into account. A non-geolocalised request for help, for instance will be difficult to process. A witness report of an incident without information on the localisation of the incident will be useless.

Information gleaned from Social Media channels are often too generic to be processed, e.g. tweet consists of keywords like e.g. fire, accident, etc. without any emergency context. Therefore, the SOTERIA toolbox aims at synthesis of semantic information that comes from different sources. One interesting opportunity, when dealing with multiple sources of data, would certainly be to extract correlations between information coming from different sources, in order to validate and reinforce the relevance of the analysis carried out by the SOTERIA platform tools. The Fusion Centre also helps to make connections between different messages that may be providing information on the same event i.e. through clustering or comparison of the messages, and using semantic events models.

In many occasions, the authors of tweet messages provide information on their localisation, either in the text message or using hash tags, but this is difficult to process automatically and thus result in missing information.

#### 2) Missing information due to overload

The second type of data gap is missing either important information or link between several information items, due to the vast amount of data that cannot be processed by a single individual. This can be related to situation in which relevant information is lost in ocean of irrelevant information.

The data gap resulting from not being able to make connections between messages usually is also due to the vast amount of data that cannot be processed manually. This can be related to situation when PSO is faced with multiple messages, each important, but each containing only a small piece of information about the overall situation thus making it hard to draw the full picture because of too many small pieces.

Since during emergency there can be a huge amount of data collected and not all incoming information may be relevant, thus mechanism for processing information before getting to the PSO in the Fusion Centre is needed, thereby offloading the job by reducing the amount of information that user has to directly process. At the same time, since in emergency events it is not desirable to obscure any information from PSO, SOTERIA gives PSO the ability to view the source data on demand.

In order to identify most relevant information reducing the amount of information that user has to directly process SOTERIA tools parse data and search for relevant information which can contribute to the emergency situational awareness. The approach that we propose enables to process and analyse data and thus contributes to provide more relevant information and filter out information not relevant for emergency purposes.

### C. SOTERIA Fusion Center

To mitigate data-gaps, the SOTERIA project provides a Fusion Center connected to the main Platform named OZONO. This section describes how Fusion Centre helps to mitigate the data-gaps resulting from a vast number of messages that may come through Social Media channels. Each component of the Fusion Center is described in more details in the next sections.

**Error! Reference source not found.** presents the simplified view on the Fusion Centre. The numbers in the circles in the figure correspond to the numbers in the description of the Fusion Centre operation. This description provides a general view, without going into details of specific tools or listing their full functionality.

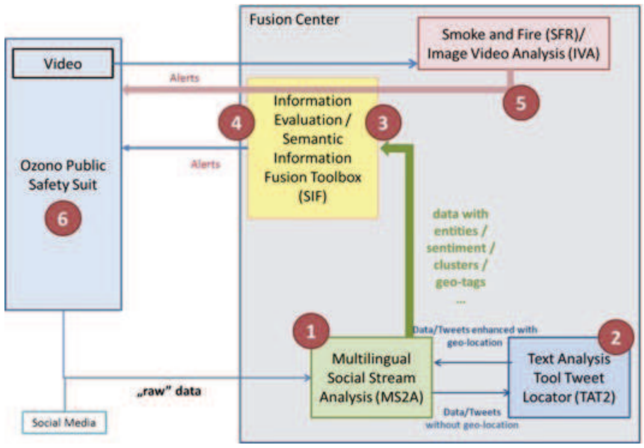


Fig. 1. SOTERIA Fusion Center

(1) In Fusion Centre the input data coming from OZONO/Social Media is first processed by the Multilingual Social Stream Analysis (MS2A) service. The analysis performed in there, which relies on natural language processing and text mining techniques, aims to automatically classify and prioritize the messages gathered from the platform according to how serious appears to be the situation they portrait.

(2) During the analysis, if any Tweeter data (or other textual data) is identified to be lacking geo-tags, the Tweet / text is sent for processing in the TAT2 service. TAT2 service then attempts to perform place name extraction and the toponym resolution (place name disambiguation) based on the context of the message and message metadata. As a result, if a toponym resolution is successful, TAT2 returns the message with additional metadata on potential geo-locations, each with confidence value assigned.

(3) After completing the analysis and providing additional metadata, the MS2A service sends the processed data to Information Evaluation and Semantic Information Fusion Toolbox (IE-SIF). The IE-SIF then performs the in-depth analysis of the available data including the additional metadata provided within MS2E and TAT2. IE-SIF uses semantic models of certain events such as car accident or fire. These models are defined by the PSO officers as generic descriptions of the events they happen to manage. The models may be defined as domain ontologies, where instances of situations are not included.

In general, the IE-SIF compares the data against these models using multiple data mining techniques. As a result, the IE subsystem can answer a question about the certainty of occurrence of a given event, whereas the SIF subsystem provides a synthetic summary of a set of information items related to the same real-life event.

(4) Finally, if certain confidence threshold is crossed, then the FC system (and IE-SIF in particular) generates the alert about the possibility of event occurrence. The alert is then presented to the OZONO PSO user.

(5) Moreover, the Smoke and Fire / Image Video Analysis (SFR-IVA) is processing video data for detection of smoke or e.g. man-down behaviours. Since detecting smoke/fire/mandown in video is an important event, thus SFR-IVA immediately informs the OZONO PSO user with an alert about the detection. Still information from SFR-IVA could be fed as additional piece of information into the Information Evaluation.

(6) Finally, the PSO user, having to make decision based on the information received from Fusion Centre (i.e. forward /confirm / discard alert) can request the source data, on which the decision was based upon.

In this paper we focus on the tools that enable us managing information gaps within crisis management. Regarding the three different types of gaps mentioned above, we rely on the following tools:

- **TAT2**, the **text analysis tool locator**, will provide estimated location information on missing geolocalisation information.

  TAT2 provides means to overcome the first type of information gaps, namely **imprecise information**.

- **MS2A**, the **multilingual social stream analysis**, will provide sentiment analysis and clustering of social media messages, in order to mitigate the risk of incurring in information overload by the PSO in the Operations Center.

  MS2A provides means to support the end-users with the second type of gaps, namely the **unnoticed information due to information overload**.

- **IE-SIF,** the **semantic information fusion and analysis** tool, provides means to filter out irrelevant information find links between short messages related to a same event and evaluate the truthfulness of a reported event within a complex situation. These functions support the PSO in its social media analysis task.

  IE-SIF provides means to overcome the first type of information gap, namely unreliable information. It also supports the user in reducing information overload by the **removal of redundant information** and the **discovery of links between information**.

## III. TAT2: TEXT ANALYSIS TOOL TWEET LOCATOR

Location of the emergency event is one of the crucial in the response actions, however only small percentage of Twitter messages is geo-tagged. Thus, Text Analysis Tool TWeet lOcator (acronym TAT2) has been developed to provide estimated localisation of given tweet origin based on analysis of message body, text of the tweet. The tool aims at identifying location, from which tweet has been sent, not the user location indicated by his/her profile. In order to differentiate location names mentioned in the text disambiguation process is incorporated. In TAT2 existing open source tool AIDA[13] (Accurate Online Disambiguation of Named Entities in Text and Tables) is used for entity detection and disambiguation. Tweet message consisting of content and metadata is the input for TAT2 tool. First tweets are processed in preparation module, where mentions of location are detected by using two built-in methods for information extraction. Next, detected mentions are resolved by identifying the meaning by the use of one of the six disambiguation algorithms in AIDA. Returned results have links to Wikipedia URL, which is used by TAT2 to extract geographical coordinates if they are available in Wikipedia article. To extract coordinates Google Geocoding web service is also used in TAT2. To narrow results time zone and area filters can be added. The final output from the TAT2 web service contains Twitter message together with a list of possible locations associated with that tweet, locations names have geographical coordinates and confidence value for each of the locations.

TAT2 architecture has been presented in [14] in more details, together with test scenarios and their results.

## IV. MS2A

The Multilingual Social Stream Analysis (MS2A) module has the purpose of analysing textual multilingual social media messages gathered by the SOTERIA platform. MS2A, by relying on natural language processing and text mining techniques [25] is able to perform two different operations: sentiment analysis and clustering.

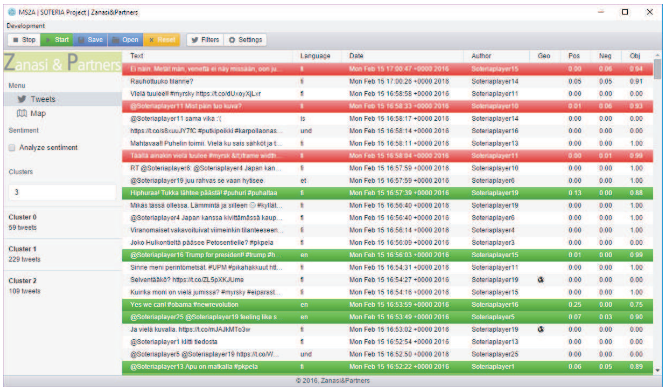

Fig. 2. Screenshot of a GUI for MS2A specifically developed for the campaign of experimentation that took place in Kuopio, Finland. Listed are the different social media messages, gathered in real time and coloured in red (negative) or green (positive), depending on the outcome of the sentiment evaluation.

### A. *Sentiment analysis*

Sentiment analysis (SA, also known as *opinion mining*) is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes [18].

The sentiment analysis process, as applied by MS2A on an individual message, is articulated in the following steps:

1. Identify language;
2. Tokenise message;
3. Remove stop words;
4. Compute sentiment.

MS2A, at first, needs to identify the language of the message, as this information is required to perform some language-specific operations (steps 3 and 4). To achieve this goal, the module uses *langid*, a dedicated Python library [19].

The module subsequently performs two *pre-processing* operations in order to prepare the original message for the computation of the sentiment: tokenisation and stop word removal. Tokenisation is the process of breaking a stream of text up into words, symbols, or other minimal elements called *tokens*. This operation is performed regardless of the language of the message under analysis. The removal of *stop words*, which is a language-dependent operation instead, consists in eliminating from the message the tokens corresponding to the most common words in the used language (grammar articles, preposition, etc.) under the assumption that they convey little or no indication about the sentiment of the whole message.

To compute the sentiment, MS2A relies on *WordNet* [20], a lexical database in which words are linked together by means of their semantic relationships, with the main relationship amongst words being synonymy. Wordnet groups synonyms into unordered collections called *synsets* (*synonym set*), each expressing a distinct concept. The same synonym may be part of one or multiple synsets. MS2A processes all the tokens left into the message and retrieves the list of all the WordNet synsets each token is part of. The module then assigns a sentiment score to each token by means of *SentiWordNet* [21], a lexical resource which associates to each WordNet synset three numerical scores, respectively indicating to what extent positive, negative, and objective (*i.e.*, neutral) the terms contained in the synset are. For every token, MS2A averages the sentiment scores of all the synsets the token is part of, thus deriving a single three-fold sentiment for the token. Once the sentiment has been computed for every token, MS2A calculates the overall message sentiment score by computing the mean value of the sentiments across all the tokens composing the message.

Table I reports an example of sentiment computation for the sentence "It is a sunny day. I am enjoying the warm weather". It is worth noting that the words "it", "is", "a", "I", "am" and "the" were removed during step 3, so they do not contribute to sentiment's computation.

TABLE I. EXAMPLE OF SENTIMENT COMPUTATION

| | **Negative score** | **Objective score** | **Positive score** |
|---|---|---|---|
| sunny | 0.0 | 0.5 | 0.5 |
| day | 0.0 | 0.95 | 0.05 |
| enjoying | 0.05 | 0.475 | 0.475 |
| warm | 0.154 | 0.625 | 0.221 |
| weather | 0.104 | 0.896 | 0.0 |
| **Message avg.** | **0.062** | **0.689** | **0.249** |

### B. *Clustering*

In the literature, the clustering problem is defined as finding groups of similar objects in the data [22]. When applied to the text domain, clustering consists in grouping a set of texts in such a way that texts in the same cluster are more similar to each other than to those in other clusters.

Clustering operates on sets of elements, rather than on single ones. To cluster messages, MS2A follows those steps:

1. Tokenise message;
2. Remove stop words;
3. Stem the tokens (Snowball);
4. Vectorise message (bag of words);
5. Calculate TF-IDF weights;
6. Apply the clustering algorithm.

While the first two steps are implemented in the same way as for the sentiment analysis, an additional pre-processing step introduced by the clustering procedure is *stemming*. Stemming, whose aim is to reduce words to their *stem* (*i.e.*, its morphological root), is a language pre-processing operation frequently adopted [23] as it simplifies analyses by treating different variations of a word in a common fashion (e.g., “fish”, “fishing”, and “fisher” are all reduced to the same stem, which is “fish”).

After pre-processing, the messages are transformed into numerical feature vectors (*vectorisation*). MS2A does so by using the *bags of words* representation. According to this model, a text is represented as the “bag of its words”: the model counts the number of times each word appears in each message (*occurrences*), without considering grammar and word order. Once the bag of word has been created, a *term frequency-inverse document frequency* (tf-idf) weighting is applied on it, with the purpose of scaling down the impact of terms that occur very frequently across the whole message set.

An example of bag of words creation is shown in Table II, referring to a collection composed by two messages: “John likes to watch movies. Paul likes movies too” and “John also likes to watch football games”.

Once the bag of words has been created, MS2A can proceed with the actual clustering. Clustering can be achieved by using various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. MS2A relies on the *Mean Shift* algorithm [24]. Given a set of data points, the algorithm iteratively shifts each data point towards the closest cluster centroid. The direction to the closest cluster centroid is determined by where most of the points nearby are. Unlike other common clustering algorithms (e.g., *K-Means*), the Mean Shift automatically determines an “ideal” number of clusters. The algorithm is fed with the list of pre-processed and vectorised messages, returning the number and composition of identified clusters. Thus, at the end of this procedure each message is assigned a single cluster identifier (*hard-clustering*).

TABLE II. EXAMPLE OF BAG-OF-WORDS REPRESENTATION AND TF-IDF WEIGHTING

| | **John** | **likes** | **to** | **watch** | **movies** | **Paul** | **too** | **also** | **football** | **games** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Message 1 – word occurrences** | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 0 |
| **Message 1 – tf-idf weighted** | 0.11 | 0.22 | 0.11 | 0.11 | 0.29 | 0.14 | 0.14 | 0 | 0 | 0 |
| **Message 2 – word occurrences** | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| **Message 2 – tf-idf weighted** | 0.14 | 0.14 | 0.14 | 0.14 | 0 | 0 | 0 | 0.19 | 0.19 | 0.19 |

## V. INFORMATION EVALUATION AND SEMANTIC INFORMATION FUSION

When an emergency call arrives, the PSO pile up a large amount of information in order to have a better understanding of the ongoing situation and evaluate the level of emergency and severity.

The different information items are testimonies and reports provided either by PSO, tools or citizens. They are linked together through a big information network that represents the on-going situation, as information sources express part of their reasoning and testimony through links between different pieces of information. One of the issues for using such an amount of information is to be able to access relevant parts of it efficiently. For example, this enables highlighting schemes of emergency events for instance. The Semantic Information Fusion module provides means to manage and analyse networks of information and support PSO in gathering a synthetic and accurate vision of an on-going emergency.

The Information Evaluation & Semantic Information Fusion module (IE/SIF) relies on the use of semantic graph structures to store information and uses a graph algorithm to carry out the fusion process [15]. It enables three different operations on networks of information:

- Report synthesis
- Information query
- Information evaluation

### *A. Event Typing*

In order to manage emergency messages, IE/SIF relies on the deep analysis of the messages received by the PSOs. The first step of this deep analysis is to type the emergency descriptions (i.e. the messages), according to the type of emergency they describe. The possible emergency types are defined in the domain ontology provided to the fusion module.

TABLE III. EXAMPLE OF EVENT TYPES/KEYWORDS DATASET PROVIDED TO THE IMS AS PARAMETERIZATION

| | |
|---|---|
| Accident | Accident, crash, omnettomuus |
| CarEvent | Car, truck, road |
| TrainEvent | Train |
| Fire | Fire, palokunta, incendie |
| Flood | Flood, innondation |
| BusEvent | bus |

To type the emergency messages, we use a data set of keywords associated to one or more emergency types (see TABLE III. for an example). As keywords are detected in a message, all possible emergency types are selected. The type finally associated to the emergency is the most general common subtype, according to the domain ontology (see Fig. 2 for an example).

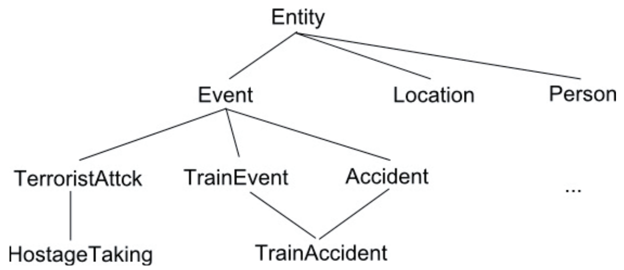


Fig. 2. Example of a type hierarchy

For example, the following messages will be typed as follows:

- Sentence: There are lots of cars stuck in a tunnel due to snow. There might be a crash inside the tunnel
- Potential event types : [CarEvent, Crash, Accident]
- Selected event type : CarAccident

### B. Reports Synthesis

The emergency messages related to the same emergency situation are processed through the SIF module, so to provide the PSOs with a synthetic and non-redundant description of the situation.

The event descriptions are fused by giving the most precise description that contains all the elements of the initial messages. The type of event reported is the most generic common subtype of all the initial reports, and the content of the synthetic report contains all the information contained in the initial reports.

Whenever a piece of information is redundant from one report to another, the fusion function eliminates this redundancy. See [16] for more details on the approach.

To illustrate this function, let us consider the following messages that could be collected in an emergency center.

- T1: There are lots of cars stuck in a tunnel due to snow.
- T2: There might be a crash inside the tunnel.

T1 will be detected by typing service of type CarEvent and T2 of type Accident. Therefore, the synthesis of T1 and T2 will generate an event report of type CarAccident. Furthermore, the synthetic report will contain both sentences of T1 and T2.

### C. C. Information Query

All the instances of information corresponding to a specified graph pattern may be found within a network of information, through the information query function. For example, PSO will look for all instances of events of type fire in a habitation building during a week-end that are listed in their data bases.

The graph patterns are semantic models defined by PSOs, such as domain ontologies for instance.

### D. D. Information Evaluation

The function provided by the Information Evaluation service is the evaluation of the degree of certainty one may grant an event, given all the information available on this event. This evaluation relies on the use of a network or database of event and/or testimony descriptions.

The testimonies and descriptions are scored according to the degree of confidence the PSO have in this information. It is either done manually according to PSO expertise of the types of events, types of testimonies and source of the testimony or information. Or the evaluation may be supported by tools. The main function provided by the Information Evaluation module, is the query for degree of certainty that a specific situation or event is on-going. As for Information Query, the hypothesis to be evaluated is given as a –partially– instantiated graphs. Testimonies related to this hypothesis are sub-graphs of the overall information network. These testimonies may confirm or infirm the hypothesis.

When the PSO want to evaluate the level of certainty that an event occurred or is on-going, they query the available information items on this event. Given the degree of certainty associated with each information item, a global evaluation of the veracity of the event is processed. The approach used for processing the likelihood that an event is on-going is inspired by works on the Transferable Belief Models ([26], [27]), widely used in the fusion community in order to manage uncertain information (see [17]).

## VI. EPERIMENTATIONS

The work presented here was applied within two European funded projects the iSAR+ project [28] and the SOTERIA project [29]. Several experimentations were performed. The different tools developed for the projects were tested, and among others, the PSOs interacted with the Fusion Center tools.

### A. iSAR+ Experimentations

iSAR+ project gathered together 16 partners from 8 European countries. One of the objectives was to develop a platform dedicated to ease communication between citizens and PSO during crisis. Several experimentations were conducted during the project. Detailed information about these experiments can be found in iSAR+ document D7.731 (see [28]).

Among others, the second experimentation took place in France, in near to real conditions. The scenario took place in a big train station and included several emergency events: an unattended luggage found and the start of a fire in technical premises near to the metro. The citizens were played by Red Cross volunteers and students. Their role was to broadcast information on social network and behave as asked by authorities through social media platforms and other communication means.

Regarding the IE/SIF module, we tested the detection of new event reports that have to be checked as being true by the PSO. Numerous tweets were sent notifying the abandoned luggage and bringing attention to a potential terrorist attack. An alarm was triggered and information sent to citizens on their smart-phones. The luggage owner was found and the rumor blocked through information of actual situation to the citizens.

In the second part of the scenario, numerous citizens reported the smoke, thus the PSO were advised of the situation both from regular calls to the 112 and through their twitter account. Instructions were given to citizens by PSO in order to evacuate the metro station. The report synthesis function of the IE/SIF enabled to gather tweets and re-tweets about the event. It then provided the PSO with a single event report, made of an aggregation of all the initial tweets.

The experiment was a success and tools were assessed as useful and easy to use. We collected feedback from several PSO, among others the followings.

• A colonel for Gendarmerie (78 department) assessed his interest for “intelligent tool able to qualify information collected on social networks”.

• A PSO from SDI 78 assessed his interest for social networks, but states that “the information should be qualified and reliable”.

### *B. SOTERIA Experimentations*

The SOTERIA project took place as the continuation of iSAR+. The hypothesis evaluation function of IE/SIF module was tested during the last experimentation of the SOTERIA project. The experimentation was held in Portugal, and officers from Guarda Nacional Republicana (GNR) tested the tools. The scenario of the use case comprised an earthquake resulting in building collapsing, chemical and biological contamination of population and road accident close to the river.

As numerous events occurred in the same area, the Report synthesis and hypothesis evaluation of the IE/SIF module capabilities were very welcome by PSO officers. Thanks to hypothesis evaluation, they could distinguish between new event reports about an actual emergency or ongoing rescue operation (reported both by PSO and citizens with lower level of trust) and rumors of terrorist attacks when GNR drones flew over the experimentation zone. Indeed, the potential terrorist attack was only reported by citizens and later contradicted by a GNR message.

## VII. CONCLUSION

We presented here a series of tools aiming at supporting the management of social media within crisis and emergency situations. The development of social media and mobile platforms make it possible, for the PSO, to have a huge amount of information provided directly by citizens.

Counter-intuitively, this may lead to gaps in information at the level of PSO awareness. Indeed, the huge amount of information makes it impossible, for a single operator, to analyze all the information by himself. Furthermore, as citizens get involve in information spreading about the situation, the quality of that information might decrease, due to the lack of training of citizens, regarding crisis communication.

The tools we developed to support crisis and emergency management target three axis:

- TAT2 enables giving precision concerning the geolocation of reports,
- MS2A provides a mean of selecting important information given the underlying sentiment of messages,
- IE-SIF enables finding missing connections between information items provided through different messages.

The SOTERIA toolbox, and among other, the Fusion Center, was presented to end-users during various campaigns of experimentations. PSO officers from several countries and several organizations tested the tools, which were very welcome.

## ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community’s Seventh Framework Program (FP7/2007-2013) under grant agreement no.606796 (SOTERIA Project).

## REFERENCES

[1] Mukherjee, S. (2014). **The Use of Twitter, Facebook, LinkedIn etc. as Strategic Tools for Crisis Communication**. International Journal of Management and International Business Studies, 4(2), 175–180.

[2] Mergel, I. (2014). **Social Media Practices in Local Emergency Management. Results from Central New York**. Retrieved from http://sotechem.syr.edu/wp-content/uploads/2014/11/SoTechEM_Executive_Report-mwedit111914.pdf

[3] Yin, J., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2012). **Using Social Media to Enhance Emergency Situation Awareness**. IEEE Intelligent Systems, 27(6), 52–59.

[4] Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Hentenryck, P. Van, Fowler, J., & Cebrian, M. (2016). **Rapid assessment of disaster damage using social media activity**. Science Advances, 2(3).

[5] Jin, Y., Liu, B. F., & Austin, L. L. (2014). **Examining the Role of Social Media in Effective Crisis Management: The Effects of Crisis**

**Origin, Information Form, and Source on Publics' Crisis Responses**. Communication Research, 41(1), 74–94.

[6] Abbasi, M.-A., Kumar, S., Filho, J. A. A., & Liu, H. (2012). **Lessons Learned in Using Social Media for Disaster Relief - ASU Crisis Response Game**. In Proceedings of SBP12, the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (pp. 282–289).

[7] Cresci, S., Cimino, A., Dell'Orletta, F., & Tesconi, M. (2015). **Crisis Mapping During Natural Disasters via Text Analysis of Social Media Messages**. In Proceedings of WISE 2015, the International Conference on Web Information Systems Engineering. http://doi.org/10.1007/978-3-319-26190-4

[8] McClendon, S., & Robinson, A. C. (2012). **Leveraging Geospatially-Oriented Social Media Communications in Disaster Response**. In Proceedings of the 9th International ISCRAM Conference (pp. 2–11).

[9] Mukherjee, S. (2014). **The Use of Twitter, Facebook, LinkedIn etc. as Strategic Tools for Crisis Communication**. International Journal of Management and International Business Studies, 4(2), 175–180.

[10] Chen, R., & Sakamoto, Y. (2013). **Perspective Matters: Sharing of Crisis Information in Social Media**. In Proceedings of the 2003 46th Hawaii International Conference on System Sciences (pp. 2033–2041). Ieee. http://doi.org/10.1109/HICSS.2013.447

[11] Olteanu, A., Vieweg, S., & Castillo, C. (2015). **What to Expect When the Unexpected Happens: Social Media Communications Across Crises**. In Proceedings of CSCW '15, the 18th ACM Conference on Computer-Supported Cooperative Work and Social Computing.

[12] M. Manso and B. Manso, "**Role of Social Media in Crisis: A European Holistic Approach to the Adoption of Online and Mobile Communications in Crisis Response and Search and Rescue Efforts,**" in Proceedings of the 17th International Command and Control Research and Technology Symposium (ICCRTS), Fairfax, VA, 2012.

[13] http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/aida/

[14] Flizikowski, A., Przybyszewski, M., Stachowicz, A., Olejniczak, T. and Renk, R. (2015). **Text Analysis Tool TWeet lOcator - TAT2**. *Proceedings of the ISCRAM 2015 Conference*

[15] Laudy, C. Hidden **Relationships Discovery through High-Level Information Fusion**. *Proceedings of International conference on information fusion*, 2015

[16] Laudy, C. **Semantic Knowledge Representations for Soft Data Fusion**—Efficient Decision Support Systems - Practice and Challenges From Current to Future. Chiang Jao Publisher, 2011.

[17] Simon Fossier, Claire Laudy, Frédéric Pichon.**Managing uncertainty in conceptual graph-based soft information fusion**.*Proceedings of the 16th International Conference on Information Fusion, 2013*.

[18] B. Liu and L. Zhang, "**A survey of opinion mining and sentiment analysis**". In "Mining Text Data" (pp. 415-463). Springer US.

[19] M. Lui and T. Baldwin, "**langid.py: An Off-the-shelf Language Identification Tool**," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012), Jeju, Republic of Korea, 2012.

[20] G. A. Miller, "**WordNet: A Lexical Database for English**," Communications of the ACM, vol. 38, n. 11, pp. 39-41, 1995.

[21] A. Esuli and F. Sebastiani, "**SENTIWORDNET: A publicly available lexical resource for opinion mining**," in Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06), Genova, Italy, 2006.

[22] C. C. Aggarwal and C. X. Zhai (eds.), "**A survey of text clustering algorithms**," in Mining Text Data, Springer Science+Business Media, LLC, 2012.

[23] C. Moral, A. de Antonio, R. Imbert and J. Ramírez, "**A survey of stemming algorithms in information retrieval**," Information Research, vol. 19, n. 1, 2014.

[24] K. Fukunaga and L. Hostetler, "**The estimation of the gradient of a density function, with applications in pattern recognition**," IEEE Transactions on Information Theory, vol. 21, n. 1, 1975.

[25] A. Zanasi (ed.), "**Text Mining and its Applications to Intelligence, CRM and Knowledge Management**", Southampton, UK, WIT Press, 2007.

[26] F. Pichon, D. Dubois, and T. Denoeux, "**Relevance and truthfulness in information correction and fusion**," International Journal of **Approximate Reasoning, vol. 53, no. 2, 2012.**

[27] P. Smets, **"Belief functions: The disjunctive rule of combination and the generalized bayesian theorem,"** International Journal of Approximate Reasoning, vol. 9, no. 1, 1993.

[28] iSAR+ Project. url: http://isar.i112.eu.

[29] SOTERIA Project. url: http://soteria.i112.eu/publications/index.html.