Architecture for gathering and integrating collaborative information for decision support in emergency situations

Tiago Marino*,
Maria Luiza Machado de Campos and
Marcos R.S. Borges

Graduate Program in Informatics (PPGI),
Universidade Federal do Rio de Janeiro, Brazil
Email: tiagomarino@ufrj.br
Email: mluiza@ufrj.br
Email: mborges@nce.ufrj.br
*Corresponding author

John G. Breslin and Maciej Dabrowski

Insight Centre for Data Analytics,
National University of Ireland Galway, Ireland
Email: john.breslin@insight-centre.org
Email: maciej.dabrowski@insight-centre.org

Abstract: The involvement of citizens in supporting crisis situations is no longer a new phenomenon. In the past, the lack of data was one of the main barriers faced by public managers in the decision-making process. Today the situation has been reversed, such that the challenge faced is managing an excessive mass of data, which is totally dynamic and originating from different sources such as remote environmental sensors, social networks, response teams in the field. During an emergency response, the concern is no longer being able to collect data for a better understanding of the affected environment, but in knowing how to organise, aggregate and separate what is actually useful for crisis managers. This research proposes a collaborative information architecture that considers aspects from environmental complexity in the context of emergency scenarios, in order to support response teams with decision making, through gathering and integrating information that has originated from different media resources and, hence, enriching a decision team's 'current contextual knowledge' base.

Keywords: emergency; heterogeneous information sources; complex systems; information integration; crisis and disaster management; social media; collaboration; domain vocabulary; decision support; data integration.

Reference to this paper should be made as follows: Marino, T., de Campos, M.L.M., Borges, M.R.S., Breslin, J.G. and Dabrowski, M. (xxxx) 'Architecture for gathering and integrating collaborative information for decision support in emergency situations', *Int. J. Information and Communication Technology*, Vol. X, No. Y, pp.000–000.

Biographical notes: Tiago Marino is researcher and Professor at the Geosciences Department of Rural Federal University of Rio de Janeiro (UFRRJ). His research focus in geographic information systems (GIS) field. For over 15 years, he supports the Centre for Scientific Support in Disasters (CENACID), developing and applying GIS solutions, and is also working on field. He experienced dozens of missions and training activities, including Haiti's Earthquake (2010). He holds a DSc in Informatics from Federal University of Rio de Janeiro (PPGI/UFRJ), Rio de Janeiro, Brazil (2015): Internship at Social Software Unit (SoSo), Digital Enterprise Research Institute (DERI), Republic of Ireland (2013).

Maria Luiza Machado de Campos is researcher and Professor at the Computer Science Department of the Federal University of Rio de Janeiro. She received her PhD in Information Systems from the University of East Anglia, Norwich, England, in 1993 and for the period of 2014–2015 she engaged as a one-year Visiting Research Scholar at the Laboratory for Applied Ontology (ISTC/CNR), in Italy. Her research interests include metadata management, ontology engineering and conceptual modelling, applied to decision support systems and information integration in general. She has led many projects in the domains of bioinformatics, energy, emergencies and government data, funded by Brazilian agencies.

Marcos R.S. Borges is a Full Professor of Computer Science at the Federal University of Rio de Janeiro, Brazil. He earned his Doctorate in Computer Science from the University of East Anglia, UK in 1986. From 1994 to 1996, he was a Visiting Research Scholar and a member of the Object Technology Laboratory at Santa Clara University, California, USA. He has also served as Visiting Professor at University of Paris VI (2001) and at Polytechnic University of Valencia, Spain (2004–2005). He has published over a hundred research papers in international conferences and journals related to decision support systems. His research interests include CSCW, group decision support systems, resilience engineering and collective knowledge.

John G. Breslin is a Senior Lecturer in Electronic Engineering at NUI Galway. He is also a Funded Investigator at the Insight Centre for Data Analytics there. He has been PI/budget holder for funding of €1.75M, was co-PI on the €5M DERI CSET, and led the Eurapp EU app economy study for the EC. He created the SIOC framework, implemented in hundreds of applications on over 25,000 websites. He has 150 publications and co-authored the books *The Social Semantic Web* and *Social Semantic Web Mining*. He is Vice Chair of IFIP's Working Group 12.7 (Social Networking Semantics and Collective Intelligence).

Maciej Dabrowski is a Postdoctoral Researcher at DERI, NUI Galway. His multi-disciplinary work combines innovative recommender system techniques, data mining and analytics, in particular in enterprise platforms. His background includes degrees in computer science, business, and information systems. His research interests focus on recommender systems, decision support systems and data analytics, as well as social networking tools and analysis, both in enterprise and web context. He has authored over 20 peer-reviewed publications, served on several programme committees of international conferences and led a number of applied research projects on these topics in close collaboration with industrial partners, such as ADVANSSE in collaboration with Cisco systems.

1 Introduction

During emergency situations, information becomes an important ally, as the basis for coordination and decision making during a response. It is an essential aspect of an organisation's ability to earn (or lose) visibility and credibility. It exerts a strong influence on how national and international resources are mobilised. In particular, this information is needed to provide quick and effective assistance to citizens involved in the crisis. It is also essential for a post-event analysis, evaluation and reflection on lessons learned and achieved during the response relief.

Information management plays an important role during all phases of an emergency. While some types of information are more static (such as plant buildings and geographical maps), others can dynamically emerge, spread and, depending on the circumstances, quickly become obsolete. The need for a combined usage of information generated from heterogeneous data structures, collaborator's profiles and communication sources makes its management a complex task, particularly during the response phase when time pressure and information overload are frequent.

An emergency response activity usually involves several teams from distinct organisations, working cooperatively with the purpose of saving lives or properties. Commanders are frequently tasked with making many decisions under constant environmental pressures and during a short period of time. The team's decision is frequently supported by the 'current contextual knowledge' (Diniz et al., 2005) that is available, i.e., one that is obtained from new information acquired by teams directly involved in the emergency action, throughout the response evolution. This knowledge is very dynamic. It changes all the time and, therefore, it has to be constantly updated. Most updates may have originated from field agents, but can also be enriched by voluntary workers involved in the emergency.

The evolution of mobile communication devices and services and the expansion of social media has turned common citizens into 'field agents'. Using real-time messaging, citizens are able to report their perceptions regarding a current emergency scenario via text and multimedia, on the fly, through their own mobile devices (Kamel Boulos et al., 2011).

The microblogging platform Twitter[1] is frequently used among citizens as a collaborative tool to share information regarding traffic conditions, risks of violence, and weather conditions. Thus, during emergency situations, it can also be used to report on victims or hazards (Hughes and Palen, 2009), helping to build a more accurate and complete 'current contextual knowledge' base.

Therefore, this paper proposes a flexible and adaptable architecture which considers aspects related to the complexity of emergency environments (imprecision, uncertainty, dynamism, heterogeneity) (Coskun and Ozceylan, 2011), and supports response teams by enriching the current contextual knowledge base. The architecture is based on procedures related to information gathering and integration, posted by distinct generator's profiles

(automatic, official agents, collaborators), and submitted from heterogeneous communication resources (such as social networks, SMS, e-mail, environmental sensors, etc.).

This paper is divided as follows. Section 2 presents the motivation for this work. Section 3 addresses some requirements and issues related to this kind of proposal. Section 4 is dedicated to detailing the architecture and its involved agents, processes and roles. Section 5 presents results from the experiments conducted, using datasets generated during a real emergency situation. The last two sections are dedicated towards the conclusions and to approaches for enhancing the proposed architecture and, hence, improving the effectiveness of its results.

2   Motivation

Social collaboration during crisis response is no longer new. The advent of new communication technologies coupled with the popularisation of mobile devices has created various opportunities for increased contextual information generation, originating in online social participation, especially when citizens observe imminent hazardous situations among their own community.

2.1   Collaboration through social networks

Social networks, powered by user-friendly interfaces, may facilitate collaboration between citizens and emergency response teams by making use of the advantages provided by resources integrated into mobile devices (camera, voice recorder, GPS, GPRS).

Figure 1 depicts a real case, whereby a common citizen shared an emergency event with the world via social media just ten minutes after its occurrence, simply by taking advantage of the integrated features on his mobile device (camera, GPS and 3G). Social media can quickly and easily facilitate collaboration between individuals and teams involved in the response to emergency events, and bring this response to the awareness of the command team from a different perspective, i.e. the angle of the citizens themselves (Alexander, 2013; R. Sabra Jafarzadeh, 2011; Spring, 2013).

Figure 1   U.S. Airways aircraft crash into Hudson River in January 2009. Information posted just 10 minutes after the aircraft emergency landing (see online version for colours)

Source: Janis Krums Twitter [online] http://twitpic.com/135xa

2.2 Social collaboration in disaster response

Heinzelman and Waters (2010) reported on the Ushahidi[2] platform, an open-source project based on collaborative mapping. It originated in Kenya in 2007, when a local group of developers created a website for collecting and mapping population reports related to violent acts stemming from the contested national election results.

In 2010, having already been improved and supported by a bigger community of collaborators, the platform was used to support disaster relief for the 7.0-magnitude earthquake that struck Haiti in January of that year, collapsing 90% of the buildings in Port-au-Prince (the Haitian capital), killing 222,000 people and injuring over 300,000.

Submissions were sent by e-mail and short message service (SMS) messages, free of cost to the public, thanks to agreements set with local carriers. At the time, the project also incorporated submissions from social networks.

As long as information was submitted by local citizens, a content management team worked in data triage tasks, such as relevance verification, validation, classification and, finally, its mapping and publishing.

In this case, all information screening and organisational processes were not automated, consequently requiring a longer processing time. In some cases, translation support was also required, which could be collaboratively performed by local citizens. In the Haitian disaster relief, where the native languages are Creole and French, local citizens joined the Ushahidi team to translate messages from the population's native languages to English (a standard language adopted by the United Nations response teams).

Figure 2 Ushahidi/Haiti crisis map: hotspots shown for relevant incidents submitted by Haitian citizens after the earthquake in Haiti (2010) (see online version for colours)
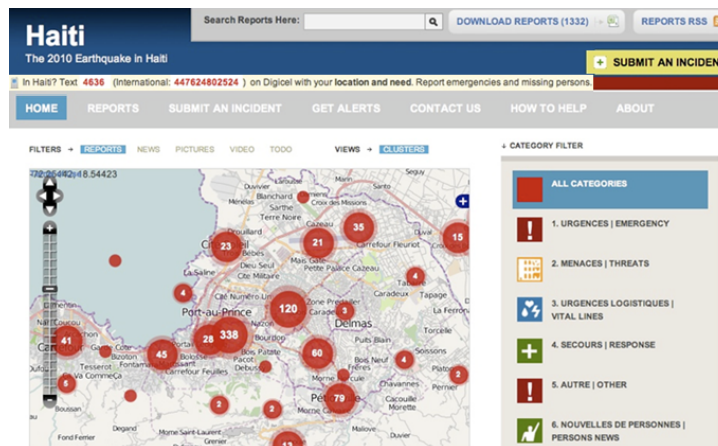
Source: Heinzelman and Waters (2010)

According to Thomas McKenzie (Heinzelman and Waters, 2010) – a team member of the crisis mapping project Ushahidi/Haiti – by the end of the first phase of disaster response, the platform had helped 4,636 people with a total of 25,186 (peaking at 2,000 posts in a day) SMS messages, e-mails and communications through social networks, resulting in 3,596 records considered sufficiently relevant to be mapped in the

platform (Figure 2). Notifications related to people under the rubble, medical emergencies and needs, such as food, water and shelter, were received, processed and mapped in almost real time by an international group of volunteers. Four days after the earthquake, response teams that operated in the country began to consume this information, focusing on determining how, when and where to direct their technical and human resources.

The concrete examples presented above provide evidence that collaboration in emergency situations yields savings in time, material and human resources, extends the data acquisition coverage area and promotes a current contextual knowledge base's evolution.

3   Architectural requirements

Evidently, collaboration can bring numerous benefits such as resource savings, an increased capacity of monitoring, and enriching the crisis response knowledge base. However, a solution which considers social collaboration and which must process the information generated through heterogeneous sources faces some requirements, such as: uncertain frequency of information generation/updates; distinct attribute sets; incompatible data representation formats; geolocation checking; content translation; detecting and forwarding incomplete and inconsistent information; identifying and establishing associations among correlated information; modelling a flexible data storage structure, and so on.

Therefore, it is essential to control and organise this new vast and heterogeneous volume of information. For each considered information source (i.e., SMS, e-mail, social networking, environmental sensors, data originating from field agents, etc.), it is necessary to inspect its available dataset structure. This is needed to identify, for example, who, where, when, and under what circumstances some information has been originated, in order to present it as clearly and in as organised a manner as possible for consumers, so that they can determine whether the collaborative information is (or is not) useful and effective for decision making.

Regarding the origin of collaborative information, even if not judging its quality or utility, it is necessary to flag whether the information is official (originating from agents officially involved in the response) or unofficial (generated by common citizens). Under certain circumstances, a citizen (collaborator) can promote or enhance their own personal interests and feelings in order to be served first, even if his or her situation is not actually a priority among many others that may exist.

The unpredictability related to human behaviour is a typical complex systems feature (Mitchell, 2009) which can compromise the effectiveness of crisis response. Thus, it must be considered and modelled on a solution that aims to support decisions based on collaborative information.

From the technological aspect, different information sources result in a high degree of heterogeneity involving different sets of data and transmission protocols (Benbya and McKelvey, 2006). It demands the creation of a common protocol for collecting and processing data, as well as being able to adapt (complexity) to the likely future aggregation of new sources.

For each piece of gathered information, it is necessary to evaluate if the available data complies with the minimum requirements for its processing. Otherwise, in the case of insufficient or conflicting data, it is necessary to promote solving mechanisms, in order to keep their screening alive inside the processing chain and, hence, prevent its premature rejection.

The following situations address some critical requirements regarding collaborative information processing:

1   How to quantify qualitative information?
     e.g. 'many injuries' is numerically equivalent to what?

2   How to deal with a format mismatch of two pieces of correlated data originating from different sources? e.g. *Facebook.create_date = 2010-01-10 16:45 ≡ Sensor.Sysmic.date = 10/01/2010-16:40:10*

3   How to handle incomplete information?

4   How to detect and handle conflicting information? e.g.
     '*10 injured people*' versus '*few injuries*' versus '*20 injured*'

5   Is this actually a conflict or just a scenario evolution?

6   How to identify and group correlated information?

7   How to prioritise the processing of incoming information?

8   How to decide whether or not to discard incomplete information?

9   How to identify and measure how updated some information is?

10   How to evaluate the quality of information?

The requirements presented above are directly related to information quality attributes, such as completeness, timeliness, consistency and clarity [as listed by Friberg et al. (2011)]. Therefore, it can be assumed that an architecture proposal that addresses methods to deal with these issues can increase the quality of the current contextual knowledge base. Otherwise, what would be able to substantially contribute to emergency response, with the proper treatment, can just become a threatening element within a 'flood' of useless information.

4   Architecture and workflow proposal

Current contextual knowledge formation and evolution during emergency situations is characterised by high levels of dynamism and uncertainty, as the result of intense environmental transformations caused by the disturbing event. In this scenario, information plays a key role and becomes the main working tool to guide the situation awareness and support decisions for crisis response activities.

This section presents and describes an architecture that considers aspects related to complexity, and implements procedures for collecting and integrating information generated from heterogeneous sources during emergency situations.

The diagram depicted in Figure 3 identifies the main elements involved in the proposed architecture, grouped into user, application and data storage layers, and also segmented according to the environment in which they operate: internal or external to the system. The layers interact with each other and with end users through interfaces that enable communication among all architecture elements.

Figure 3   Architecture for gathering and integrating collaborative information for decision support in emergency situations (see online version for colours)
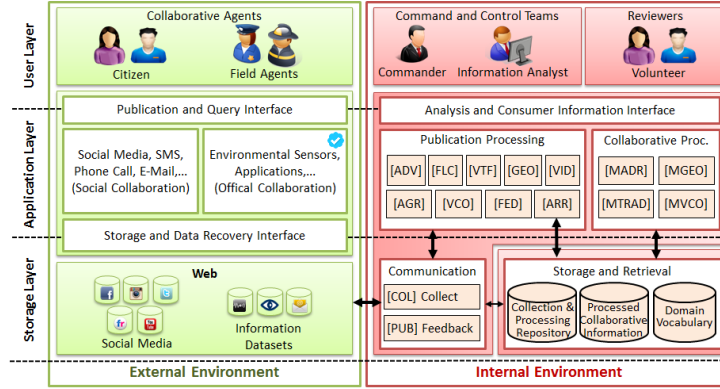

The user layer provides the information generator agents (external environment), collaborators and information consumers (internal environment). The application layer presents its external environment data input interface elements, such as social networks, SMS, e-mail, phone calls (channels for collaborative content submission) and specific applications, in which field teams may officially submit information through exclusive and reliable channels. In the internal environment, the distinct elements involved in the collaborative information processing are presented and will be detailed in the following sections.

The storage layer consists of external data repositories, such as social media datasets and others repositories from which the architecture must extract collaborative information in its original format. The internal environment of this layer consists of three repositories:

1   the 'repository for collection and processing' registers all information processing histories, since its capture in an external environment, through the stages of each partial transformation during processing inside the internal environment

2   the 'processed collaborative information' repository stores the relevant information already processed and semantically annotated, based on an organised domain vocabulary

3   the vocabulary domain brings together a set of terms related to some emergency information. A vocabulary term will be linked to each information source's corresponding attribute, in order to promote their integration.

Figure 4 presents a workflow that illustrates the screening sequence performed by collaborative information and depicts the interactions among agents during the generation, collection and processing steps.

Figure 4 Workflow: agents, processes and interactions involved in the collaborative information processing triage (see online version for colours)
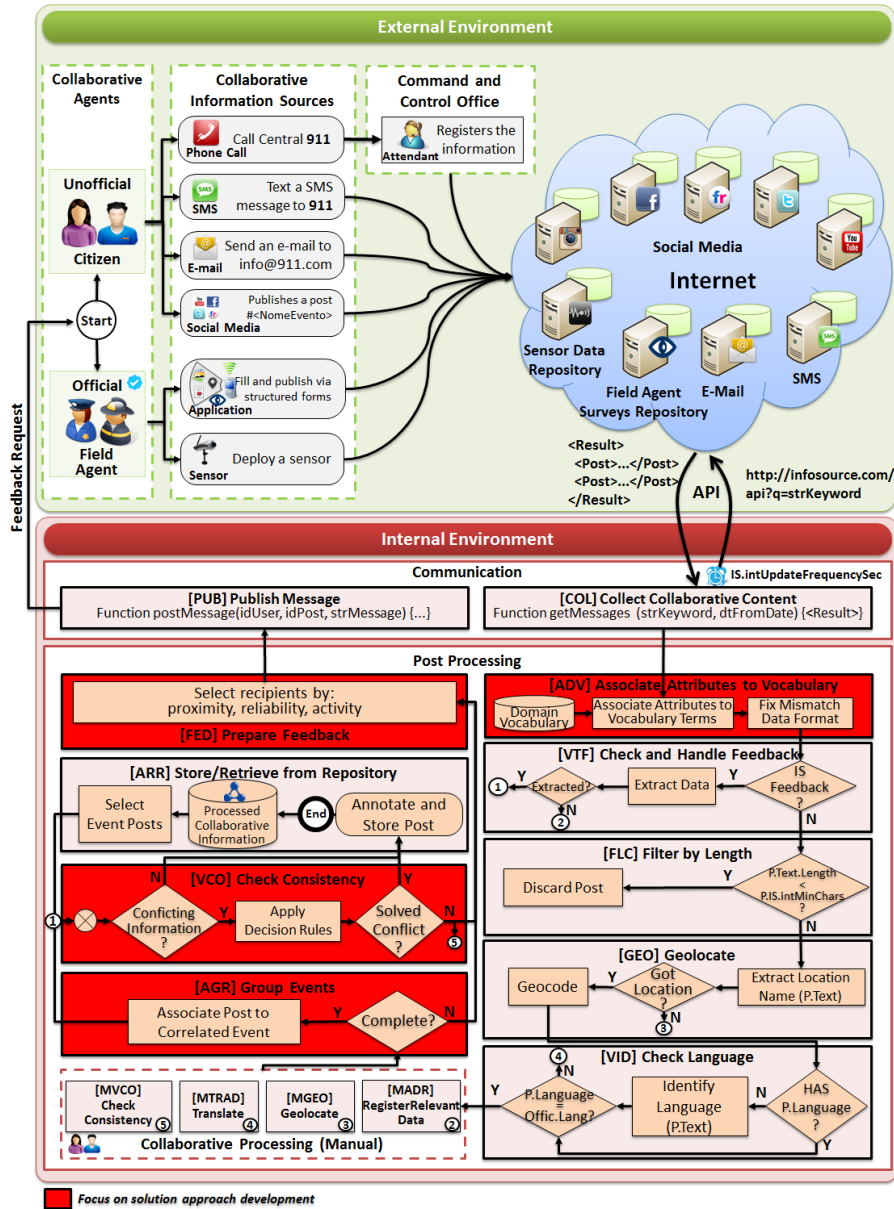
4.1   External environment

4.1.1   Collaborating agents

Collaborative information can be generated under the following circumstances:

- Through the free initiative of citizens (social cooperation), who witness an unusual situation (from their perspectives) related to an emergency event, and decide to report their testimonies to the crisis response managers through a communication channel.

- Through an information analyst's request, seeking new information, such as confirmations of or additions to a certain topic or place of interest, demanded from the field teams and also from the citizens. The following examples illustrate some of these request situations:
  a "Office command asks volunteers to confirm the number of collapsed houses downtown."
  b "Commander requires photos from downtown since yesterday."

Each information analyst's request must be immediately forwarded to field agents, through trusted communication channels and, as needed, also to the citizens, through their own social networks and other available contact channels, such as e-mails and mobile calls. Collaborators who are closer to the scene may also contribute by publishing their findings directly from the site. In some situations, collaboration may still be enriched by multimedia content. All generated collaborative content will then be subjected to a new screening cycle within the system and, eventually, available to support the coordinators' decisions.

4.1.2   Information resources

For contributions incoming from phone calls, a call centre has attendants trained to filter and register all relevant information gathered through the dialogue, and store it into forms previously structured by the command (SOS, 2008). An exclusive phone number can also concentrate incoming SMS collaborations, such as the initiative performed by 'Mission 4636' during Haiti's earthquake recovery in 2010 (Munro, 2013).

From field surveys, agents feed the information database through exclusive mobile device applications. By filling structured forms, agents can standardize communications among their entire team. Marino et al. (2012) presented an operating platform designed for this purpose.

The crisis response command can consume any available environmental monitoring sensor dataset[3] (which has been granted public access by its owner) or deploy the installation of sensing devices in specific locations, according to their monitoring interests. Thus, instant and continuous environmental readings may be performed, for example, to monitor the level of a river, the rainfall volume in certain periods, or retrieve live images from a critical place. Once connected to the internet via integrated general packet radio service (GPRS) modules, all gathered data can be transmitted in real time to data repositories on the web and therefore, be available for consumption through web services.

4.2   Internal environment

4.2.1   Communication module

To attach a new information source, the platform's communication module requires a library presenting the following communication procedures:

- *Function getMessages (strKeyword, dtFromDate) {return <results>}* is performed in the process '[COL] collect collaborative content' to capture messages from the collaborative information source. The parameter *strKeyword* constrains the search results to specific keywords. The *dtFromDate* parameter determines that the set of publications returned must contain only records created after the date specified in this parameter;

- *Function postMessage (idUser, idPost, strMessage)* is a procedure called by the process '[PUB] publish message' to establish the return contact with a given user in a collaborative information source. Considering that some information sources do not provide mechanisms for establishing return contacts, this is not a mandatory function in the library. However, feedback requests are possible for those sources that enable return communications through this function.

Figure 5 illustrates the interface for information source management and parameter configuration in the experimental platform. After uploading the library file, a diagnosis is automatically executed in order to check if collecting and publishing procedures are present and operational for further executions.

Figure 5   Information source configuration interface: parameter settings and results of diagnostics on an uploaded library file (see online version for colours)

The available configuration parameters defined by the administrator through that interface are:

- *Reliability level:* integer value representing the source reliability level. For instance, higher confidence levels must be assigned to field agents (official), rather than to social media (unofficial). This hierarchy will automatically allow the platform to:
  1   make decisions based on the information source's reliability in cases of conflict detection [VCO]
  2   suggest whom to request feedback from, based on the source's reliability [FED].

- *Gather frequency (seconds):* frequency, in seconds, that the platform will execute the procedure [COL].

4.2.1.1 [COL] Collect collaborative content

As an initial step for the screening cycle, collaborators generate event-re-related information, which are stored in different structures of repositories distributed over the internet.

Today, the most popular social networks already offer application program interface (APIs) that allow developers to build automated queries to their data repositories through requests based on the simplicity of a REST architecture proposed by Fielding (2000).

Standardised protocols for information extraction provided by the APIs reduce the complexity for integrating new information sources to the system. The column '*Get|Post*' in Figure 6 indicates the presence and operation of procedures for collecting (down-arrow) and publishing (up-arrow), respectively, for each information source attached to the platform.

Figure 6   Experimental architecture prototype: information sources management interface (see online version for colours)

In order to acquire data almost in real time, the process '[COL] collecting collaborative content' is triggered according to each information source's update frequency (defined by the parameter *gather frequency* shown in Figure 5). On each data gathering iteration, the process queries the information source data repository and extracts only the most recent items, i.e., those created after the last collection. Content filtering by keywords related to the emergency event can be applied to queries in order to reduce and refine the result set.

4.2.1.2 [PUB] Publish message

The process '[PUB] post message' provides communication from the internal environment to the external environment. For each information source, it is possible to post messages just by requiring as input the contact's unique identifier and the message

text to be published. As mentioned it is important to note that some information sources do not provide mechanisms for establishing a return contact.

The informer's unique identifier is the key that uniquely identifies the publisher in a determined information source. For example, in social networks every user has a unique identifier, usually represented by an extensive numerical sequence[4]. For phone calls and SMS messages this identifier is the telephone number. Thereby, the command response can establish contact with collaborators to request feedback.

4.2.2 Post processing module

4.2.2.1 [ADV] Associate attributes to vocabulary

Each information source has a specific data structure as well as its own terminology to describe its set of attributes. Manuals and reference documentation provide details concerning the metadata terminology explanation and available functions for coders.

Figure 7   Process '[ADV] associate attributes to vocabulary': data collected from different sources are interrelated through an automatic associative mapping step between the domain vocabulary terms and information source's correspondent attributes (see online version for colours)
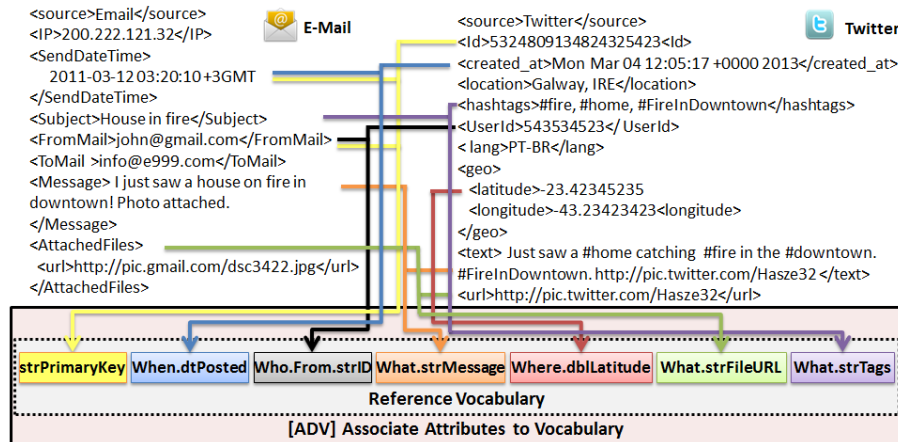
For our approach, we have established a domain vocabulary containing terms related to the most relevant and essential attributes for the architecture's information processing stages. Once the platform's administrator establishes the links between domain vocabulary terms and their correspondent information source attributes, the process [ADV] will automatically identify and establish associations between each collected information attribute and its respective equivalent term in the vocabulary domain (as illustrated in Figure 7).

This action results in a standardised and homogeneous data structure, thereby simplifying the identification of these data items in subsequent workflow steps.

4.2.2.2 [VTF] Check and handle feedback

Before starting a message processing cycle, it is necessary to check if it is related to a feedback request. In case it refers to an information update or inconsistency feedback return, automated processes will try to extract the returned value and submit it to be considered in a new consistency check [VCO] round, and possibly help to solve the pending conflict [the forwarding is represented by the connector (1) in the workflow].

While feedback return messages are generated by humans, in some cases, automated procedures (such as natural language procedures) may fail to extract values within expected ranges. In these cases, the message must be forwarded to the collaborative processing module for human-assisted data extraction and curation procedures (Heinzelman and Waters, 2010; Howard, 2013; Rogstadius et al., 2013).

4.2.2.3 [FLC] Filter by length

This initial verification aims to prevent any waste of time in processing messages without useful information, something that can often occur in social networks (Rogstadius et al., 2013). Logically, this restriction does not apply to messages flagged as feedback returns, characterised by short and compact contents that are extremely relevant for the knowledge base.

4.2.2.4 [GEO] Geolocate

Geographical location information is essential for coordinators during the crisis response. For instance, field agents can be directed to activities based on their proximity to a task's site location. Also, feedback responses may be more effective when requests are broadcasted to nearby citizens and agents.

Therefore, this step attempts to identify, flag, and, if possible, repair some casual conflicts caused due to the difference between the informer's current location (at the moment that the information is posted) and the place which is referred to in his or her post (the location which really matters for the response). This means that even messages that already contain geolocation information should be analysed by this process.

This geographic location information can also support other architecture processes, such as duplicated entries detection and message grouping. For example, when comparing two distinct messages and identifying similarities regarding date and location, it would provide more accurate evidence that both messages may refer to the same event and, hence, are probably correlated. On the other hand, inaccurate georeferencing can cause a cascading effect on subsequent stages of information processing and thus compromise the entire architecture efficiency.

4.2.2.5 [VID] Check language

Some data sources (e.g. social networks, e-mail, phone calls) may provide information regarding the content language in its original attributes set. For other data sources, such as SMS messages, this information is missing. Thus, content must be submitted to an automated procedure with the purpose of identifying its original language.

Natural language processing (NLP)-based tools – such as Google Translate[5], Microsoft Translator[6], AlchemyAPI[7] – present satisfactory results for detecting the native

language of the text content. On the other hand, unfortunately, these and other services currently available on the web do not present satisfactory results for text translation.

Therefore, in order to avoid prematurely discarding potentially useful information published by local citizens [when their language is different from the response team's one, as in the Haitian case (Munro, 2013)], the remaining alternative is to forward them to a user-assisted translation service (usually local collaborators), represented by the process [MTRAD].

4.2.2.6 [AGR] Group events

This step focuses on searching and establishing associations between the currently processed message and the set of previously processed messages (already stored in the database). Automation is essential since it is impossible to compare thousands of messages manually. This pre-processing step is also required so as to decrease processing costs during information retrieval and clustered visualisations.

For instance, grouping by geographic location restricts the dataset to a refined subset referring just to a specific area of interest for response commanders, discarding those that are not directly related to that place. Similarly, information can be categorised by source types, informant's reliability levels, date windows, etc.

Initially, a completeness test checks if the published item presents the minimum information required for executing grouping methods. In case of any essential information that is missing, a specific request for information will be forwarded to the [FED] process as an attempt to request more information, which could possibly help to solve the stated issue. A feedback request can be forwarded to the original informant or even a group, based on the proximity, informant types (e.g. field agents, citizens, media), depending on its level of importance and urgency.

4.2.2.7 [VCO] Check consistency

This workflow stage consists of identifying and solving conflicting data within an information event grouping (this explains why it occurs after the process [AGR]).

Reasoning techniques will be applied based on predefined rules, as an attempt to make decisions without human intervention. Decisions are weighted by the source's reliability level, the information's timeliness and the informer's reputation.

As an incremental and evolutional approach, whenever a conflict cannot be automatically solved through decision rules, it must then be forwarded to a feedback request, in an attempt to consider an 'external opinion'.

As long as new contributions arrive, this new related information can be aggregated to the knowledge base and perhaps it could decisively contribute towards solving the pending conflict. The pending conflict in question can also be flagged and forwarded to a supervised decision maker through collaborative processing [MVCO] [connector (5) in the workflow].

4.2.2.8 [FED] Prepare feedback

Posts related to feedback returns must be processed with high priority in order to keep the information updated as much as possible (timeliness is an attribute of information quality).

During feedback preparation, some parameters can be defined to achieve a better return efficiency:

a   For who? A specific person or a set of people, according to their current location, reliability and/or kind of information source.

b   Processing priority level.

c   Timeout period.

d   How to proceed once a timeout occurs? Notify the sender, request it to a new group of people and/or discard the information.

Collaborators must provide information more clearly and as objectively as possible in order to achieve the best performance from automated NLP methods and avoid forwarding it to manual processing, which increases the treatment time of the message, and, consequently, makes it obsolete. The following cases illustrate types of feedback request messages:

• "How many people are homeless in 46th Street? Type an INTEGER number."

• "Is it raining in your neighbourhood? Type YES or NO."

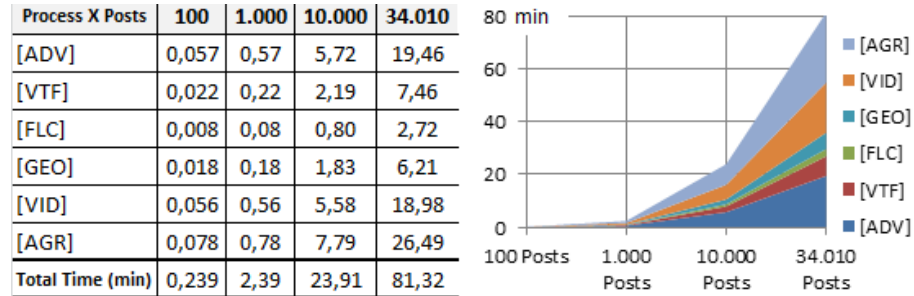## 5   Architecture experimentation

5.1   Experimental datasets

The experimental simulation has processed heterogeneous datasets in a developed prototype, according to the specifications required by the proposed architecture. The acquired datasets refer to the disaster resulting from the magnitude 7.0 earthquake that hit the city of Port-au-Prince, Hai-ti's capital, on the afternoon of January 12th, 2010. Help requests from local citizens, assessment reports produced by field agents, and seismic sensor data generated during the disaster response phase constitute the seven distinct sources (listed in Figure 6) acquired through internet searches, e-mail contacts, and also obtained from personal participation in supporting the response.

These datasets are satisfactory for experimental purposes, allowing observations of the systemic behaviour and the complexities particular to real emergency situations. Another positive factor of the acquired set is the high degree of heterogeneity among sources, encompassing different levels of reliability, structure and frequency in terms of content generation.

5.2   Results

In terms of system scalability, Figure 8 presents the time distribution to process each step involved in the architecture's information processing chain for distinct amounts of posts. For the currently conducted experiment, the prototype has processed all the 34,010 acquired posts in approximately 1 hour and 21 minutes.

Figure 8 Time distribution for distinct amounts of posts and architecture's information processing steps (see online version for colours)

Process X Posts	100	1.000	10.000	34.010
[ADV]	0,057	0,57	5,72	19,46
[VTF]	0,022	0,22	2,19	7,46
[FLC]	0,008	0,08	0,80	2,72
[GEO]	0,018	0,18	1,83	6,21
[VID]	0,056	0,56	5,58	18,98
[AGR]	0,078	0,78	7,79	26,49
Total Time (min)	0,239	2,39	23,91	81,32

Source: Platform SocialCol [online] http://greco.ppgi.ufrj.br/socialcol

Figure 9 shows two published items grouped into the same 'story'[8], as a result of the experiment. Regardless of their having originated from distinct information sources, and hence, presenting different attribute sets, both items refer exactly to the same event. The correlation index between these publications reached its maximum value (1.0 out of 1.0), which testifies the effectiveness of the heuristic designed to associate similar posts.

Figure 9 Similarity between two items grouped in the same story (see online version for colours)

ID:	36477 - Ushahidi Platform - SMS	ID:	153287 - Mission 4636 - SMS
Date:	2010-02-28 16:23:00	Date:	2010-02-28 18:05:00
Classes:	[3] Request for Need.Food	Classes:	[3] Request for Need.Food
	[13] Request for Need.Shelter		[13] Request for Need.Shelter
Coordinates:	18.616315:-72.268066	Coordinates:	18.61631473:-72.26806641
DESCRIPTION:	english: Hi; my name is Jislene; i live Lathan MBC #25. Please; we need [13] tent; rice; bean; touth paste; soap; spaghetti. Thanks. -- Category: --4d. Distribution daliments	[3] food distribution creole: Bonjou.mwen se jislne;mwen abite lathan MBC No.25.silvoupl nou bezwen tant diri pwa lwil pat savon espageth.Msi	translation:	Hi, my name is Jislene, i live Lathar MBC #25. Please, we need [13] tent, rice, bean, touth paste, soap, spaghetti. Thanks. NOTES: Its so important, This is a crowd at Lathar who needs [3] food, [13] tent,Rice,Bean, Touth paste, soap, Spaghatti.

Source: Platform SocialCol

6 Conclusions and next steps

Motivated by the continuous growth of environmental data acquisition resources, which also encourages and creates favourable conditions for social collaboration practices, the main contribution of this work was to demonstrate that, even facing heterogeneous data structures, origins and generation circumstances, correlations among information can be measured and highlighted in order to increase the understanding of the emergency context in which they were generated and, consequently, support decisions by emergency response commanders.

Exploratory research from the state of the art revealed that initiatives and existing approaches only operate over homogeneous sets of information. However, unlike our solution, no other approach proposes flexible and adaptable solutions to integrate,

correlate and detect conflicts among similar information, regardless of their origins and data structures.

As an alternative to increasing social collaboration through feedback requests, some user profile information may help to identify potential collaborators and increase the effectiveness of the information request campaign, such as: reputation, daily activity averages, reply times, their last publication place and date. The promotion of a 'top collaborator ranking' can also support information analysts in selecting the most active and reliable collaborators for specific tasks (Tang et al., 2011).

Concerning procedures related to conflict detection and solving, resolution rules previously defined can be combined with machine learning algorithms, creating an autonomous (or human-supervised) decision-making environment, and resulting in processing time and backlog decreases.

Currently, this work is still under development as PhD research. Issues and results obtained from new experimentation rounds and developments shall be presented and shared in future publications.

References

Alexander, D.E. (2013) 'Social media in disaster risk reduction and crisis management', *Science and Engineering Ethics*, doi:10.1007/s11948-013-9502-z [online] http://www.ncbi.nlm.nih.gov/pubmed/24306994.

Benbya, H. and McKelvey, B. (2006) 'Toward a complexity theory of information systems development', *Information Technology & People*, Vol. 19, No. 1, pp.12–34, doi:10.1108/09593840610649952.

Coskun, E. and Ozceylan, D. (2011) 'Complexity in emergency management and disaster response information systems (EMDRIS)', *International Conference at Information System for Crisis Response and Management*, Lisboa, Portugal, p.5.

Diniz, V.B., Borges, M.R.S., Gomes, J.O. and Canós, J.H. (2005) 'Knowledge management support for collaborative emergency response', *Proceedings of the 9th International Conference on Computer Supported Cooperative Work in Design*, pp.1188–1193, Coventry, England: IEEE.

Fielding, R.T. (2000) *Architectural Styles and the Design of Network-based Software Architectures*, University of California [online] http://www.ics.uci.edu/~fielding/pubs/dissertation/ (accessed 3 March 2015).

Friberg, T., Prödel, S. and Koch, R. (2011) 'Information quality criteria and their importance for experts in crisis situations', *Proceedings of the 8th International ISCRAM Conference*, Lisbon, Portugal, p.10.

Heinzelman, J. and Waters, C. (2010) *Crowdsourcing Crisis Information in Disaster-Affected Haiti*, p.16, Washington, DC, USA [online] http://www.usip.org/files/resources/SR252-Crowdsourcing Crisis Information in Disaster-Affected Haiti.pdf (accessed 17 November 2013).

Howard, B.C. (2013) *Scanning Social Media to Improve Typhoon Haiyan Relief Efforts*, pp.11–14 [online] http://news.nationalgeographic.com/news/2013/11/131108-typhoon-haiyan-philippines-crisis-mapping/ (accessed 22 June 2014).

Hughes, A.L. and Palen, L. (2009) 'Twitter adoption and use in mass convergence and emergency events', *International Journal of Emergency Management*, Vol. 6, Nos. 3/4, p.248, doi:10.1504/IJEM.2009.031564.

Kamel Boulos, M.N., Resch, B., Crowley, D.N., Breslin, J.G., Sohn, G., Burtner, R. and Chuang, K-Y.S. (2011) 'Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples', *International Journal of Health Geographics*, Vol. 10, p.67, doi:10.1186/1476-072X-10-67, ISSN: 1476-072X [online] http://www.pubmedcentral.nih. gov/articlerender.fcgi?artid=3271966&tool=pmcentrez&rendertype=abstract.

Marino, T.B., Nascimento, B.S. do and Borges, M.R.S. (2012) 'GIS supporting data gathering and fast decision making in emergencies situations', *Proceedings of the 9th International ISCRAM Conference*, pp.1–6, Vancouver, Canada.

Mitchell, M. (2009) *Complexity: A Guided Tour*, Oxford University Press, USA.

Munro, R. (2013) 'Crowdsourcing and the crisis-affected community', *Information Retrieval*, Vol. 16, No. 2, pp.210–266, doi:10.1007/s10791-012-9203-2.

R. Sabra Jafarzadeh, J. (2011) 'Emergency management 2.0: Integrating social media in emergency communications', *The New England Journal of Medicine*, Vol. 9, No. 1, pp.13–18, doi:10.5055/jem.2011.0063.

Rogstadius, J., Vukovic, M., Teixeira, C.A., Kostakos, V., Karapanos, E. and Laredo, J.A. (2013) 'CrisisTracker: crowdsourced social media curation for disaster awareness', *IBM Journal of Research and Development*, Vol. 57, No. 5, pp.4:1–4:13, doi:10.1147/JRD.2013.2260692.

SOS (2008) *112 in Sweden: The 2008 Annual Activity Report*, p.28, Stockholm, Sweden.

Spring, A., Rogstadius, J., Vukovic, M., Teixeira, C.A., Kostakos, V., Karapanos, E. and Laredo, J.A. (2013) 'CrisisTracker: crowdsourced social media curation for disaster awareness', *IBM Journal of Research and Development*, Vol. 57, No. 5, pp.4:1–4:13.

Tang, J.C., Cebrian, M., Giacobe, N.a., Kim, H-W., Kim, T. and Wickert, D.B. (2011) 'Reflecting on the DARPA red balloon challenge', *Communications of the ACM*, Vol. 54, No. 4, p.78, doi:10.1145/1924421.1924441.

Notes

1   http://www.twitter.com.

2   http://www.ushahidi.com/.

3   Xively [online] https://xively.com/.

4   Facebook Account ID Help Page [online] https://www.facebook.com/help/329684710497463.

5   Google Translate API [online] https://developers.google.com/translate/.

6   Microsoft Translator API [online] http://msdn.microsoft.com/en-us/library/dd576287.aspx.

7   AlchemyAPI [online] http://www.alchemyapi.com/.

8   As a terminological convention for this work, it has been assumed that a 'story' consists of a grouping of posts presenting a high correlation degree among them. The correlation among posts is determined by comparing their temporal (when), spatial (where) and taxonomic (what) dimensions.