

# Masters Program in **Geospatial Technologies**



## **INTEGRATION OF CROWDSOURCED INFORMATION WITH TRADITIONAL CRISIS AND DISASTER MANAGEMENT INFORMATION USING *LINKED DATA***

**Minu Kumar LIMBU**

Dissertation submitted in partial fulfilment of the requirements  
for the Degree of *Master of Science in Geospatial Technologies*



ifgi  
Institute for Geoinformatics  
University of Münster

Supported by:



Education and Culture

**Erasmus Mundus**

# **Integration of Crowdsourced Information with Traditional Crisis and Disaster Management Information using Linked Data**

**by**

**Minu Kumar Limbu**

minulimbu@gmail.com  
*Institute for Geoinformatics,  
University of Muenster,  
Muenster, Germany*

Dissertation supervised by:

**Dr. Tomi Kauppinen, PhD**

*Institute for Geoinformatics,  
University of Muenster,  
Muenster, Germany*

Dissertation co-supervised by:

**Professor Werner Kuhn, PhD**

*Institute for Geoinformatics,  
University of Muenster,  
Muenster, Germany*

**Professor Marco Painho, PhD**

*Instituto Superior de Estatística e Gestão de Informação  
Universidade NOVA de Lisboa,  
Lisbon, Portugal*

**Dr. Edward Pultar, PhD**

*Institute of New Imaging Technologies  
Universitat Jaume I,  
Castellón, Spain*

**28 February 2012**

*To my family*

## ACKNOWLEDGMENTS

I would like to take this opportunity to express my sincere gratitude to my thesis supervisor Dr. Tomi Kauppinen (PhD) and co-supervisor Professor Werner Kuhn (PhD) from the Institute for Geoinformatics, University of Munster in Germany. Their regular and timely inputs during my research tenure have been instrumental towards the completion of this study. A sentence is not enough to summarize how grateful I am to my supervisor and Linked Open Data course coordinator - Dr. Kauppinen for showing me a completely new approach to better facilitate humanitarian information management processes. I have indescribable respect for Prof. Kuhn for his motivational speeches, his inspiring ideas and especially for his support even after being in a sabbatical.

Please allow me to thank my co-supervisor Professor Marco Painho (PhD) from Universidade Nova de Lisboa in Portugal. The report greatly benefited from Prof. Painho's invaluable guidance and academic insights. I will forever remain indebted and thankful to my co-supervisor Dr. Edward Pultar (PhD) from Universitat Jaume I in Spain for his constructive comments on my thesis drafts. I would like to extend my profound appreciation to Dr. Pultar and Prof. Painho for believing in my vision and for trusting my commitment to explore new innovative tools for crisis management. For without their guidance and persistent help, this research could not have been completed.

My special thanks to PhD candidates Mr. Jens Ortmann and Mr. Dong Wang from the University of Munster in Germany as this research greatly benefited from a jointly co-authored scientific paper on a relevant topic. Their willingness to help and give their best suggestions throughout my research period have been extremely helpful. Let me also thank Dr. Carsten Keßler (PhD) from the University of Munster for allowing me to be a small part of ongoing UN project on "Humanitarian Exchange Language" which directly contributed and highlighted the practicalities and possibilities of future implementation of ideas proposed in this research.

This report was prepared with generous contributions of many individuals and organizations that form the humanitarian and global crowdsourcing/volunteering community. In addition to thanking field and headquarter based colleagues from UNOCHA and the Emergency Shelter Cluster, I wish to acknowledge the support and cooperation from Ushahidi Inc, Standby taskforce Committee and Non Governmental Organisations. Sincere acknowledgements are expressed to: Andrej Verity, Andrew Alspach, CJ Hendrix, Cecilia Utas, Craig Williams, Gregg Grimsich, John Marinos, Paul Handley, Wendy Cue from UNOCHA; Neil Bauman, Catherine Lefebvre, Gregg McDonald, Miguel Urquia from the Shelter Cluster; Patrick Meier, Rob Baker from Ushahidi Inc.; Christoph Dennenmoser from Humanity Road, Inc. The completion of this report is the result of various surveys followed by peer interviews for evaluation of the research. A page is not enough to list the names of those priceless colleagues or to express how grateful I am for their support.

This International Masters of Science in Geospatial Technologies is supported through the Erasmus Mundus Master Program (contract no. 2007-0064/001/FRAME MUN123). It is a pleasure for me to thank Dr. Christoph Brox (PhD) and course secretariat for their unwavering support during the entire Masters Programme. I would like to thank the Erasmus Munds course coordinators in advance for allowing me to submit this report as a part of short term research goals for the UN lead global Community of Interest (COI) group on "Decision Makers Needs".

Above all, I would like to dedicate my thesis to my beloved family and my wife for their cheer leading and unwavering love and support through the good times and bad.

# **Integration of Crowdsourced Information with Traditional Crisis and Disaster Management Information using Linked Data**

## **ABSTRACT**

This paper reviews to what extent Linked Data can support information management in the humanitarian decision-making processes during the early stages of a disaster response. On the immediate onset of a disaster there is little or no reliable information, especially about the affected population and the presence of relief agencies including their services. At the cost of time and people in need of help, decision-makers are faced with uncomfortable choices where they must act on less reliable information or wait longer for additional verifications. This study proposes to link data using specific semantic vocabularies for crowd generated content as well as information extracted from humanitarian agency specific assessment reports. The research explains the basic technology and knowledge about how disaster information managers can identify needs of the people affected by a crisis and the 3W that is which organizations (Who) are carrying out what activities (What) in which locations (Where) using Linked Data. The study also discuss the underlying concepts of crowdsourcing 3W linked data, including quality issues and gives some concrete examples using emergency shelter cluster coordination during the Haiti earthquake in 2010. In the end the research explores the future of disaster information management and crowdsourcing Linked Data along with the usefulness of management using a crisis vocabulary.

## **KEYWORDS**

Linked Data, Management of a Crisis, Vocabulary, Disaster Management, Crisis management, Crowdsourcing, Information Management

## ACRONYMS

3W	Who What Where?
4W	Who What Where and When?
CSV	Comma Separate Values
IASC	Inter Agency Standing Committee
HHI	Harvard Humanitarian Initiatives
IFRC	International Federation of Red Cross and Red Crescent Societies
MOAC	Management of a Crisis Vocabulary
NFI	Non Food Items
PDNA	Post Disaster Needs Assessment
RDF	Resource Description Framework
SBTF	Standby Task Force Committee
URIs	Uniform Resource Identifiers
UN	United Nations
UNOCHA	UN Office for the Coordination of Humanitarian Affairs
VCT	Volunteer Technical Committees

# Table of Contents

Acknowledgements.....	iv
Abstracts .....	v
Keywords .....	v
Acronyms.....	vi
Index of Tables .....	ix
Index of Figures .....	x
1. Introduction.....	1
1.1 Background.....	1
1.2 Rationale and Motivation.....	2
1.3 Aims/Objectives.....	3
1.4 Research approach and Contribution .....	4
2. Literature Review.....	6
2.1 Principles of humanitarian information management and exchange in complex emergencies .....	6
2.2 Information gap impact on Decision Makers Needs .....	9
2.3 User Generated Content and crowdsourcing in disaster management .....	10
2.4 Humanitarian Information: Who is doing what, Where and When?.....	11
2.5 Linked Data: An Approach for interrelating information in disaster management.....	13
3. Methodology.....	14
3.1 Humanitarian responders expert survey on information management and crowdsourcing.....	14
3.2 Management of a Crisis Vocabulary.....	14
3.3 Implementation of linked data .....	15
3.4 Evaluation .....	15
4. Expert Survey: Crisis Information Management, Crowdsourcing, Volunteered Technical Committee - Opportunities and Bottlenecks .....	16
4.1 Introduction.....	16
4.2 Expert Survey Methodology .....	16
4.3 Results.....	16
4.4 Discussions .....	18
5. Management of a Crisis Vocabulary (MOAC) .....	21
5.1 The story so far .....	21
5.2 Vocabulary Development Process .....	21
5.2.1 Specification .....	21
5.2.2 Knowledge Acquisition .....	23
5.2.3 Conceptualization .....	24
5.3 Who What Where when? Vocabulary.....	24
5.4 Traditional Humanitarian Coordination Agency Vocabulary: Shelter Cluster .....	27
5.5 Crowdsourcing and Volunteer Technical Committee - Ushahidi Haiti Platform Vocabulary.....	28
5.6 Integration of crowdsourcing and crisis management vocabulary .....	29
5.7 Discussion: Completeness, Consistency and maintenance .....	30

6.	Implementation of Linked Data for crowdsourced and traditional disaster information.....	31
6.1	Linked Open data: MOAC semantic annotation.....	31
6.2	Classes and subclasses: Who what where? Humanitarian Needs and Gaps Analysis.....	33
6.3	Linked Data Principle: An Approach for connecting data for disaster management .....	36
6.4	Haiti Earthquake Response 2010: Linking Data sources .....	36
6.5	Triplification of Raw Data: .....	37
6.5.1	Shelter Cluster traditional reports .....	38
6.5.2	Ushahidi Haiti Project reports .....	40
6.6	Queries and Visual Interpretation of integrated data .....	41
7.	Evaluation .....	42
7.1	What is Management of a Crisis Vocabulary Evaluation?.....	42
7.2	Foundation for Evaluation .....	42
7.3	Requirements for Evaluation .....	43
7.4	Evaluation Results .....	44
7.4.1	Humanitarian "Needs / Response" and "Who What Where (3W)/ When (4W)" queries.....	44
7.4.2	Domain Expert Evaluation .....	47
8.	Discussions, Limitations, Contributions and Conclusion .....	51
8.1	Discussions .....	51
8.2	Conclusion .....	53
8.3	Research Contributions.....	54
8.4	Limitations and Future Work.....	55
9.	Bibliography .....	57
10.	Appendices: .....	65
10.1	Appendix 1: Triplification Sample .....	65
10.2	Appendix 2: Vocabulary Evaluation Survey Questionnaires.....	66
10.2.1	Survey 1:Sample of Linked Data Enthusiasts Evaluation Questionnaires.....	66
10.2.2	Survey 2:Domain Expert Interview Survey and Questionnaires .....	67
10.3	Appendix 3:Profiles of Research Results Evaluators.....	70
10.3.1	Linked Data Enthusiasts .....	70
10.3.2	Domain Experts .....	71
10.4	Appendix 4: Expert Survey and Linked Data Enthusiasts Samples.....	72
10.4.1	Summary of Expert Survey for Section 4.....	72
10.4.2	Sample of Domain Expert Evaluation for Section 7.....	72
10.4.3	Sample of Linked Data Enthusiasts Evaluation for Section 7 .....	72
11.	Declaration of Originality .....	73

## INDEX OF TABLES

Table 1 Management of A Crisis (MOAC) Vocabulary Requirement Specification Document	22
Table 2 Who What Where and When concept Description. For details description of other "things", please visit <a href="http://observedchange.com/moac/ns/">http://observedchange.com/moac/ns/</a> (one of the output of this research)	26
Table 3 MOAC and other URIs used in a single Care International message .....	39
Table 4 Ushahidi Haiti Incident Reports Raw Data, incident report number 2307.....	40
Table 5 Evaluation Time Distribution across different types of surveys and peer interviews ....	48
Table 6 MOAC Vocabulary Evaluation Processes in relation to thesis objectives.....	53

## INDEX OF FIGURES

Figure 2-1 Information Management Cycle (Credit: UNOCHA)	6
Figure 2-2 Information Management at the core of Humanitarian Decision Making Process for all the Clusters throughout Crisis/Disaster Management phases (Credit: UNOCHA)	9
Figure 2-3 Panofsky-Shatford facet matrix (source Shatford 1986)	12
Figure 3-1 Management of a Crisis Vocabulary (MOAC)	14
Figure 4-1 Expert Survey "Knows" vs "Uses" in Disaster Management	17
Figure 4-2 Expert Survey - How would crowdsourcing/volunteer platform help?	17
Figure 4-3 Expert Survey - Why did the crowdsourcing/volunteer platform not help?	18
Figure 5-1 Example of a Humanitarian Contact Card showing how various terms are conceptualized across sectors like Shelter, Food, Health. (Credit: UNOCHA)	24
Figure 5-2 Who what where and When (4W) Concept Derivation in Humanitarian Response. Detailed definitions in <a href="http://observedchange.com/moac/ns/">http://observedchange.com/moac/ns/</a>	25
Figure 5-3 Inter Agency Standing Committee Core Terms (source IASC). For this research Emergency Shelter Cluster convened by IFRC will be reviewed.	27
Figure 5-4 IASC Shelter Cluster (Natural Disasters) Information Collection Graph (Source IFRC, Cluster Coordination/Information Management Focal Point)	28
Figure 5-5 Ushahidi Haiti Platform Concept Terms derived from online mapping platform (source Ushahidi Haiti Project)	29
Figure 5-6 Integration of core concept to form a Management of a Crisis Vocabulary	29
Figure 5-7 A Section of Management of a Crisis Vocabulary Version 0.4 (source:ObservedChange)	30
Figure 6-1 RDF Graph component visualized using Inspector	33
Figure 6-2 Who What Where Class with RDF Framework	33
Figure 6-3 Who What Where and When (4W) view of a single Shelter Response report	38
Figure 6-4 Demonstration of Haiti Earthquake ShelterNeeds (Affected Population Plea) and ShelterOffered (Humanitarian Agency Response) using SIMILE Exhibit	41
7-a Linked Data Enthusiasts Evolution Participants	44
7-1 Evaluators response for survey question "When (date) was the first Emergency shelter Needs reported and where?"	45
7-2 Evaluators response for survey question "In the given reporting period, please indicate the number of locations where Oxfam (Agency) provided shelter assistance."	45

7-3 Evaluators response for survey question " Based on the given data, please indicate which was the first agency to provide shelter, where (Location) and when (Time)? "	46
7-4 Response for survey question "From timeline graph at the bottom of the map, can you indicate how many days there were NO reported shelter offered on the first week of Haiti Disaster i.e. from 12 Jan 2010 - 19 Jan 2010?"	46
Figure 8-1 Example of Google search to locate the name of the agency who might have the first agency to provide emergency Shelter Response during Haiti 2010 Relief Efforts	52
Figure 8-2 Inclusion of Linked Data and MOAC in Disaster Relief 2.0 Flows of Data to Crisis Responders (Adapted from Disaster Relief 2.0, page 19)	55



# 1. Introduction

## 1.1 Background

In immediate response to many humanitarian emergencies the primary focus is on saving lives and life saving relief assistances as well as assured protection of the vulnerable population [4, 68, 110]. Ensuring a reliable, timely search and rescue mission in parallel to providing adequate relief efforts are highly dependent upon decisions made according to life saving interventions. Such informed decisions are influenced by whether relevant information was provided to the right person at the right time and in a usable format that facilitated situational understanding [41, 114]. Information in this sense is the foundation on which decision-making for a coordinated and effective response is based. However, in the early stages of a disaster there is no or little data available about the needs of the affected population and their location: what kind of humanitarian assistance is needed where and which agency might be closest to that area, also commonly known as "Who What Where?" information. Hence having a large overall effect on humanitarian response.

The UN Emergency Relief Coordinator, Sir John Holmes, says that “information is very directly about saving lives. If we take the wrong decisions, make the wrong choices about where we put our money and our effort because our knowledge is poor, we are condemning some of the most deserving to death or destitution” ([44, 53], page 2) . In a major emergency, there is always a massive combined effort to collect and analyze large volume of data, convert it to information so that it could be used to form the knowledge needed to target humanitarian aid most efficiently. But the humanitarian response to Haiti Earthquake in 2010 was different. When the traditional humanitarian agencies were conducting rapid assessments to estimate the needs of the *people affected by a disaster, for the first time in Haiti*, the very same members of the community divested by the earthquake were already issuing pleas for help using social media and widely available mobile technologies [48]. At the same time, thousands of *ordinary citizen through the world* were voluntarily mobilized to aggregate, translate, and plot these pleas for help into a near real time interactive maps in support of the ongoing humanitarian response. Information and knowledge managers in a disaster rely heavily on common operational baseline datasets [4, 114] like political/administrative boundaries, populated places, transportation network, transportation infrastructure, hydrology and city maps. In Haiti, details of road network and infrastructures crucial for planning timely and coordinated response were not available. To help with this missing information, geospatial *volunteers* throughout the world used fresh satellite imagery and literally rebuild missing maps of Haiti and plot a picture of the current, changed reality on the ground [50, 94, 123]. This particular work done through OpenStreetMap<sup>1</sup> became an essential component of the response mapping, providing much of the street-level data that was used for humanitarian logistics and camp management planning [86].

The author of this research was one of the humanitarian information mangers deployed to support Haiti earthquake response to support IASC<sup>2</sup> Shelter Cluster's<sup>3</sup> needs and response activities. It became evident in Haiti that the broader international humanitarian agency systems were not tooled to handle two new information sources: one from the disaster-affected community and one from a mobilized swarm of global volunteers [47]. Amidst trying to cope

---

<sup>1</sup>See [http://wiki.openstreetmap.org/wiki/WikiProject\\_Haiti](http://wiki.openstreetmap.org/wiki/WikiProject_Haiti)

<sup>2</sup> <http://www.humanitarianinfo.org/iasc/>

<sup>3</sup> <https://www.sheltercluster.org/Pages/default.aspx>

with information overflow from within and amongst humanitarian community, the relief agency communities were not ready and at times even reluctant to accept this new introduction of information management [108]. Using social networking website and mobile phones, people are already telling us everywhere in real-time the very information the humanitarian agencies collect via rapid assessments and household surveys. The challenge is to turn it into something actionable by policy makers. If somehow we can link data about specific needs and problems collected from people in real time to information collated by traditional humanitarian agencies on its activities, it will provide the relief agencies with the agility they need to protect affected populations in a world of fast moving crises. Humanitarian coordination in a complex emergency is not possible without information sharing. With the rise of social networking in 21st century, the challenge is to find a systematic way to integrate *crowdsourced* information from both the affected communities and from global volunteered geographic information sources with that of traditional humanitarian communities reports of needs and response activities.

### 1.2 Rationale and Motivation

The UN Secretary General Mr. Ban Ki-Moon<sup>4</sup> clearly indicated that "(o)ur traditional 20th Century tools (...) simply cannot keep up. Too often, by the time we have hard evidence of what is happening at the house hold level, the harm has already been done. Our inability to understand the impact of a crisis while there is still time to adjust our policies and programmes threatens to reverse years of hard earned development gains. The irony is we are actually swimming in the ocean of real time information." The spread of information networks with public participation is forming a new nervous system for our planet. The US Secretary of State shared that "(w)hen something happens in Haiti or Hunan, the rest of us learn about it in real time – from real people. And we can respond in real time as well." This presents a window of opportunity for both humanitarian relief agencies and the communities of social media to support disaster response before, during and after a crisis. The idea is not new but the challenge is to find a way to integrate heterogeneous crowdsourced and volunteered information with well assessed humanitarian traditional information management systems.

Enormous amounts of data and information are generated in every disaster [47, 103] . Due to the agency specific heterogeneous information collection and sharing procedures, humanitarian community usually face a mammoth tasks in trying to collate data regarding affected population's needs and humanitarian response. This problem is amplified in the early stages of a disaster response as datasets are generally not standardized across different actors and the process for sharing the data are often ad hoc. A recent UN lead humanitarian information exchange standardization initiative documented that "(i)n the best case, this (heterogeneous information) results in a significant delay between the collection of data and the formulation of that data into a common operational picture. In the worst case, information is simply not shared at all, leaving gaps in the understanding of the field situation." ([52], page 1). Relief agencies have spent millions of dollars in developing tools for collecting data in a more agreeable and standardize manner. However, most of the attempts were not universally accepted as it required relief agencies to change their internal information management processes. When humanitarian information managers of different agencies are already over burdened with their own internal workload, such concept did not materialized during disaster response as it required additional workload either by filling forms or creating extra layer of information upon existing ones.

---

<sup>4</sup> UN Secretary General Speech on the importance of bringing the UN's work into the digital age, and a summary of Global Pulse's work in 2011. <http://www.unglobalpulse.org/ResearchAndBigDataVideos>

With the rise of social networking platforms, crowd information sharing across disaster affected zones is undeniable [41, 51, 123]. After the Haiti earthquake in 2010, there are also increasing trends where social media network continuously generated overwhelming heterogeneous crowdsourced contents for almost real time live crisis mapping. Independent review of Ushahidi Haiti Project 2010, a volunteer-driven effort to produce post disaster crisis mapping, documented that “lives were saved” [83, 84, 108] with the help of volunteered crowd information. The report also suggests that the platform was not widely used across bigger agencies like UN. One of the primary reasons could be issues related to liability, or accuracy, and reliability, or authority of crowdsourced and volunteered information. One year later, despite its limitations, there are promising examples where United Nations and other international humanitarian agencies are exploiting crowdsourced social media reports for live crisis mapping in Libya and Japan [117]. From my own experiences as a disaster manager, this is indeed a much needed positive cultural change to accept the potential utility of crowdsourced information in a humanitarian environment.

On the immediate onset of a disaster, there is little or no reliable information especially about the affected population and the presence of the relief agencies including its services. At the cost of time and people in need of help, decision makers are faced with uncomfortable choices between acting on less reliable information or waiting longer for additional verifications. One solution could be the integration of data between crowdsourced and volunteered information (e.g. Ushahidi, OpenStreetMap) and traditional humanitarian agencies. My research argues that such information integration can be the key to new ways of managing information during disaster response. As Goodchild explains “(i)nfomation obtained from a crowd of many observers are likely to be closer to the truth than information obtained from one observer.” [40] “(S)ome information is ultimately better than no information since some information can at least be verified.” [81] This is true in the earlier stages of a disaster when humanitarian agencies usually struggle to acquire timely information. Hence, this study argues that crisis mapping of “who what where (3W)” through crowdsourced volunteered geographic information (VGI) using *Linked Data*<sup>5</sup> [19] can fill the information *black hole* in the early stages of a disaster.

### 1.3 Aims/Objectives

The idea to use of Semantic Web<sup>6</sup> to put individual data in the web and to link it to other existing data not new. However, humanitarian agencies have yet to test this option for better information management in a crisis [47]. Hence the main research question of this study is to discover if within the context of a disaster response the concept of *Linked Data* can help data integration between traditional humanitarian agencies and crowdsourced volunteered information. Based upon this a primary goal is to publish a vocabulary for crisis management and test its usability with real disaster related data to convert it into standard 5-starred Linked Open Data (LOD) [14].

---

<sup>5</sup> Linked Data “ (...) is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods”. (Source: <http://linkeddata.org/>)

<sup>6</sup> The Semantic Web “provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.” (source:W3C)

In doing so, the study will reflect on the following:

1. Linked Data - Management of a Crisis Vocabulary:
  1. Are there existing, ready-to-use linked data versions of crisis management vocabulary to interlink data across traditional humanitarian agencies, crowdsourced volunteered technical committees and affected populations?
    - If yes, what does this research contributing to the existing ones?
    - If no, how does one build, test and propose a Crisis Management Vocabulary?
  2. Can *Linked Data* and crowdsourcing support informed decision making processes? For example, which organizations(Who) are carrying out what activities (What) in which locations (Where)?
2. Heterogeneous Data and Linked Data : Can Linked Data with proposed management of a disaster vocabulary help convert and link internal contents or data (request for help) within crowd generated messages? Is it possible to take data (Food response, Shelter response) from a document (UN report) and convert and connect data within a document?
3. 5 star Linked Open Data:
  1. Can Linked Data help in the conversion of traditional humanitarian organizations reports, excel, PDF and word documents into machine-readable structured data?
  2. Can Linked Data help in the transformation of unstructured crowdsourced and volunteered information into machine-readable structured data?

### **1.4 Research approach and Contribution**

Instead of relying only on the traditional humanitarian agency specific information systems, this research explores how a humanitarian agency can interlink data with contents generated from social media and publish it in a machine-readable format using *Linked Data*. The research also argues that with the help of *Linked Data*, a relief agency can not only link information across totally different domains, but it can also help connect data within a document to other existing data in the web. Which brings us to Sir Tim Berners-Lee's concept of better data integration by allowing everyone who puts individual items of data on the Web to link them with other pieces of data. This is done using standard formats turning the web into a semantic web and thereby producing a wider network of knowledge.

In order to propose crowdsourcing Linked Open Data as the next step towards a full exploitation of crowdsourced information in disaster management, I along with other semantic and linked data experts wrote a workshop paper that was published at the International Semantic Web Conference<sup>7</sup> in 2011. The paper suggests crowdsourcing as a promising candidate for disaster management where *Linked Open Data* can serve as a common exchange format [94]. As a part of my Master's thesis, I plan to expand such a concept of data integration across relief agencies, the affected population and the global communities of volunteers.

My research is structured as follows. The introduction provides a brief snapshot of what the study is all about. In the literature review, I document already existing scientific researches and humanitarian documentation that cut across the domains of disaster management, information management, and crowdsourcing. The methodology chapter discusses the process which leads to the construction, testing and evaluation of a management of a crisis (MOAC) semantic

---

<sup>7</sup>See <http://iswc2011.semanticweb.org/workshops/terra-cognita-foundations-technologies-and-applications-of-the-geospatial-web/>

vocabulary. To draw the foundation for my thesis, I elaborate the findings from the focused disaster expert survey. To address the issue of data processing and information integration collected from the survey assessment I develop a vocabulary called MOAC: Management of a Crisis Vocabulary. The output of this chapter is a well-structured and standard semantic vocabulary that can be used to link data between crowds and traditional humanitarian organisations. Based on expert inputs, I revise the contents again and test the relevance and consistency of MOAC with official disaster data from the Haiti earthquake 2010 and the Arab Spring of 2011. For the evaluation of the thesis work, I will conduct a user survey on the usefulness of the MOAC vocabulary in a disaster scenario with the help of recent disaster simulation exercise in Samoa Island. Further, to propose that *Linked Data* is one of the research frontiers for interlinking and integrating information, I conduct a final expert user survey on the usefulness of the MOAC vocabulary and present its findings. The research also explores the future of disaster information management and crowdsourcing using Linked Data and the use of crisis management vocabulary.

It is worth mentioning that *Linked Data* in support of real time disaster information management support is arguably a new approach. At no point does the research claim to prove that the given set of MOAC vocabulary will solve all the problems of information integration. On the contrary, it highlights that the vocabulary can be continuously improved with the help of crowd participation and expert reviews. There is no denying the rise of affected population participation in sharing information about what they need. Additionally, there are increases in global social media volunteers that help generate near real-time disaster maps. There are over 800<sup>8</sup> million active Facebook users and Twitter is one of the fastest growing social networking sites. Over 5 billion mobile phones exist out of which 75% are used in developing countries. This research presents an idea on how to integrate data between traditional humanitarian agencies, the affected population and global crowdsourced volunteers. This report also seeks to understand and make recommendations on ways Linked Data can help us adapt to this new reality where collective action can lead to collective intelligence.

---

<sup>8</sup> Facebook Statistics: <https://www.facebook.com/press/info.php?statistics>

## 2. Literature Review

The research brings together literature from three different fields of study. Section 2.1 reviews literature pertaining to information management and its principles in complex emergencies like Haiti earthquake response. Section 2.2 reviews literatures related to crowdsourcing and volunteered geographic information in relation to crisis management. Section 2.3 reviews literature from the field of human-computer interaction, semantic web and Linked Data pertaining to data integration methods. Later it discusses the values embedded in the design of the web, Web 2.0, and social media services. Both the academic and practical aspects of "who what where" in relation to this research are explained in section 2.4. The last section weaves together interdisciplinary perspectives from these three fields as the basis for understanding the application of *Linked Data* to integrate data between traditional humanitarian agencies, social media and the global network of volunteers in the crisis domain.

### 2.1 Principles of humanitarian information management and exchange in complex emergencies

UNOCHA<sup>9</sup>, the United Nations Humanitarian Agency's official definition of a complex emergency [27, 27, 67, 110] is "a humanitarian crisis in a country, region or society where there is total or considerable breakdown of authority resulting from internal or external conflict and which requires an international response that goes beyond the mandate or capacity of any single agency and/ or the ongoing United Nations country program."([116],page 6). Complex emergencies are "situations of disrupted livelihoods and threats to life produced by warfare, civil disturbance and large-scale movements of people, in which any emergency response has to be conducted in a difficult political and security environment."[120]. Manmade or natural disasters<sup>10</sup> are types of emergencies. Depending on the nature and scale and the capacity of the local/regional capacity to respond to the situation, international relief agencies are requested for global support - for example with the Pakistan Floods in 2011[109], Haiti Earthquake in 2010, Philippines Typhoons in 2009.

*Information management*<sup>11</sup> covers 'the various stages of information processing from production to storage and retrieval to dissemination towards the better working of an organization; information can be from internal and external sources and in any format.' This



Figure 2-1 Information Management Cycle (Credit: UNOCHA)

<sup>9</sup>See <http://www.unocha.org/>

<sup>10</sup> See <http://www.unisdr.org/we/inform/terminology#letter-d>

<sup>11</sup> AIM, Association for Information Management website (accessed on 22 Aug,2010), <http://www.aslib.co.uk>

chain of data collection, processing, analysis and dissemination is called the *information management cycle* as shown in figure 2-1. Timely and accurate information is recognized as integral to humanitarian action in both natural disasters and complex emergencies [44, 72, 77, 115, 118]. Based on the same concept, OCHA's Information Managers created guides and articles explaining the details of what kind of activities should take place in order to better support an informed decision-making process [80, 113].

To streamline the global approach towards coordinated information management in a crisis, various high profile events have taken place. In 2002, the first ever global symposium on information management was organised at Palais des Nations in Geneva, Switzerland, to discuss best practices in humanitarian information exchange. The Symposium brought together information management professionals from organizations throughout the sector including UN agencies, NGOs, and governmental agencies, as well as representatives from donors, the private sector and academia. The symposium highlighted that the international humanitarian community's ability to collect, analyze, disseminate and act on key information is fundamental for effective response. Most importantly, to assist people at risk and to promote effective humanitarian action, participants of the symposium identified and agreed on a set of humanitarian principles. It was agreed that the following operational principles should be used to guide IM and information exchange activities in emergencies, namely, *Accessibility, Inclusiveness, Inter-operability, Accountability, Verifiability, Relevance, Objectivity, Humanity, Timeliness, Sustainability*.

As explained in ([115], page 1-2), the principles are explained below:

- **Inclusiveness.** Information management and exchange should be based on a system of collaboration, partnership and sharing with a high degree of participation and ownership by multiple stakeholders, especially representatives of the affected population.
- **Inter-operability.** All sharable data and information should be made available in formats that can be easily retrieved, shared and used by humanitarian organizations.
- **Accountability.** Users must be able to evaluate the reliability and credibility of data and information by knowing its source. Information providers should be responsible to their partners and stakeholders for the content they publish and disseminate.
- **Verifiability.** Information should be accurate, consistent and based on sound methodologies, validated by external sources, and analyzed within the proper contextual framework.
- **Relevance.** Information should be practical, flexible, responsive, and driven by operational needs in support of decision-making throughout all phases of a crisis.
- **Objectivity.** Information managers should consult a variety of sources when collecting and analyzing information so as to provide varied and balanced perspectives for addressing problems and recommending solutions.
- **Humanity.** Information should never be used to distort, to mislead or to cause harm to affected or at-risk populations and should respect the dignity of victims.
- **Timeliness.** Humanitarian information should be collected, analyzed and disseminated efficiently, and must be kept current.
- **Sustainability.** Humanitarian information and data should be preserved, cataloged and archived, so that it can be retrieved for future use, such as for preparedness, analysis, lessons learned and evaluation.

A follow up Global Symposium +5 on "Information for Humanitarian Action" was organized in 2007 [44, 77]. Three additional humanitarian principles were added - *Reliability, Reciprocity and Confidentiality*. As explained in ([44], page 25), the principles are explained below:

- **Reliability.** Users must be able to evaluate the reliability and credibility of data and information by knowing its source and method of collection. Collection methods should adhere to global standards where they exist to support and reinforce credibility. Reliability is a prerequisite for ensuring validity and verifiability.
- **Reciprocity.** Information exchange should be a beneficial two-way process between the affected communities and the humanitarian community, including affected governments.
- **Confidentiality.** The processing of any personal data shall not be done without the prior explicit description of its purpose and will only be done for that purpose, and after prior informed consent of the individual concerned. Sufficient safeguards must be put in place to protect personal data against loss, unauthorized processing and other misuse. If sensitive information is publicly disclosed, the sources of such information will not be released when there is a reasonable risk that doing so will affect the security or integrity of these sources.

The importance of these two symposiums and its impact on humanitarian information management systems are also highlighted by [4, 44, 80, 118] . Further, these principles been formalized under the "Operational Guidance on Responsibilities of Sector Cluster Leads and OCHA in Information Management" section as the functional and foundational principles of humanitarian information management and exchange [114].

The documented principles are also the foundation for the protection of humanitarian space and the people in need of humanitarian support in emergencies. The term “espace humanitaire” was first coined by former Médecins Sans Frontières (MSF) President Rony Brauman, who described it in 1990 as: “A space of freedom in which we are free to evaluate needs, free to monitor the distribution and use of relief goods and have a dialogue with the people”([46], page 1-2). As a part of monitoring access to humanitarian space, such principles are also used as the key indicators in measuring decreasing safety and security of beneficiaries, and providing access to them [56]. Furthermore, these principles have also become the bread and butter for humanitarian practitioners which set a common vision of the central role of information in support of effective decision making through humanitarian preparedness, response and recovery.

Haiti earthquake independent Humanitarian Emergency Response Review indicated that new technological innovation like Ushahidi Haiti Project was able to produce almost real-time crisis mapping by tapping into information received from the affected community. The platform was a good example that could help improve *accountability* towards the affected population by allowing them to participate in ongoing humanitarian response (*Inclusiveness*) [7]. The new emergence of unexplored strength of social media and crowdsourcing in collating real-time information directly from the affected population and global volunteers could be one of the turning points in the disaster management information cycle [47].

On the onset of a disaster, when humanitarian agencies are striving for information on affected population and their needs, the Haiti earthquake proved that people were already sharing information using available technology. **With my research I argue that, as a part of principles of humanitarian information management exchange, humanitarian agencies must find a coordinated way to use crowdsourced information.** The research strives to show that Linked Data can support a connection of data from the emerging mecca of crowdsourcing and volunteer technical committees [117] with already existing humanitarian information systems.

## 2.2 Information gap impact on Decision Makers Needs

At a global level, humanitarian agencies like the International Federation for Red Cross and Red Crescent Movement, have already warned that 2.57 billion urban dwellers living in low and middle income nations are exposed to unacceptable levels of risk [62, 68]. If Haiti Earthquake in 2010 is one of the prime examples of major urban catastrophes, let us all note that this is not going to be the last one. Hence, to save additional lives there is an even more pressing need to find a reasonable solution to address the humanitarian gap in the early stages of a disaster. In December 2005 the IASC<sup>12</sup> Principals generally welcomed the “cluster approach” as a mechanism that can help to address identified gaps in response and enhance the quality of humanitarian action [55].

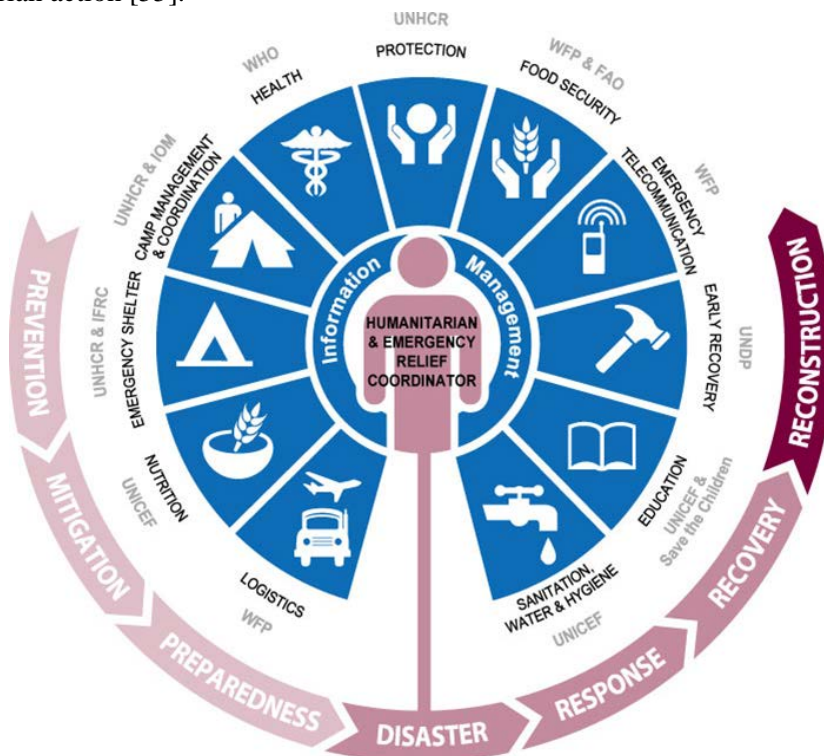


Figure 2-2 Information Management at the core of Humanitarian Decision Making Process for all the Clusters throughout Crisis/Disaster Management phases (Credit: UNOCHA)

The cluster approach is one component of a broader humanitarian reform agenda designed to contribute to enhanced humanitarian response capacity, predictability, accountability, and partnership. First, the approach aims to develop and maintain adequate global capacity in key areas to ensure that the responses to new crises are timely and effective. Second, the approach seeks to provide predictable leadership in areas of response in which there are gaps in capacity or resources, in an initiative called the Humanitarian Reform Agenda. A number of agencies have agreed to take overall responsibility (as shown in figure 2-2) for certain aspects of relief operations, called ‘clusters’: IFRC with UNHCR convenes Emergency Shelters [24, 25, 61] for example, and the UNICEF leads the water sanitation and hygiene. Targeting better information management practices within and amongst such "clusters", IASC agreed on operational

<sup>12</sup> The Inter-Agency Standing Committee (IASC) was established in 1992 by General Assembly Resolution 46/182 on strengthening the coordination of humanitarian assistance. The IASC provides a broad representation of today’s international humanitarian organisations covering 90% plus of total international humanitarian spending, consisting of three “families”: The United Nations, IOM, the Red Cross Red Crescent Movement and NGOs, represented by three consortia (source: IASC)

guidance for responsibilities of sector cluster leads and OCHA in information management [114]. This is one of the single most important documents for disaster information managers which stresses not only the need of information as a crucial part of decision support but also highlights that without information sharing across partners there will be "no coordination".

In order to provide much needed humanitarian relief support to the affected population, information coordination becomes a core part of decision-making process [67]. On the immediate onset of a disaster, everyone wants to know what the humanitarian needs are and their relevant properties [110]. However, information managers are faced with critical information gaps. When it is of utmost importance to provide evidence-based information to support timely humanitarian intervention, at times, it can take weeks and even months to collect, process and analyse such information [22, 94, 119].

Some of the first responders to be deployed in the disaster zone are the search and rescue teams and rapid assessment experts. The latter focus is solely on finding the urgent humanitarian needs by accessing the affected zone and affected population [71, 112]. Based on the nature of a crisis, the assessments of affected population could take days if not weeks due to either geographical accessibility or access to affected victims. At the cost of time and people in need of help, decision makers are faced with uncomfortable choices between acting on less reliable information or waiting longer for additional verifications [4, 58].

**Therefore, this research intends to fill this information gap by investigating the possible use of linked data connected with crowdsourced, volunteered information and comparing it to the efforts of traditional humanitarian relief agencies.**

### *2.3 User Generated Content and crowdsourcing in disaster management*

Throughout history a crowd has been defined based on physical human proximity. With the development of the internet there are less boundaries and collections of people are able create virtual crowds. This is a fundamentally new development in the course of human history that communities are able to form simply out of shared interest or passion for a hobby, craft or art.

*Crowdsourcing*, as a term, was first used by Jeff Howe in a 2006 Wired article [54] and described it as a distributed problem-solving model where the task of solving a challenge or developing an idea gets "outsourced" to a crowd. It is crowdsourcing when a work that is supposed to be performed by an employee is outsourced as an open call to a large undefined group of people using the internet. "Crowdsourcing is Wikipedia with everything"<sup>13</sup>. When we use a large group of people to share, report and communicate a story it is called crowdsourcing information [92]. Crowdsourcing geospatial data refers to generating a map using volunteered geographic information [39] from informal social networks and web 2.0 technology [51, 99].

Good communication is central to effective coordination in all major disasters. Whether the medium is email, phone, two-way high frequency radio or face-to-face, information sharing is undeniably one of the most important aspects for better managing crises. As a result, many institutions made significant investments in information and communication technologies (ICTs). With the growth of the mobile phone users and an unimaginable rise of social networking websites including Facebook and Twitter, there has been a significant change from "one-to-many forms of communication, such as television and radio, to many-to-many forms of communication using social networking and crowdsourcing websites" ([23], page 7). Examples of this include the Haiti earthquake in 2010 and Arab Spring in the Middle East during 2011 [5,

---

<sup>13</sup> <http://www.crowdsourcing.org/editorial/towards-a-new-taxonomy-by-jeff-howe/5458>

81]. This has significantly changed the way in which information is delivered and exchanged [23].

There is a multitude of scientific research suggesting the importance of information generated using social networks in order to meet the challenges of different disaster management phases such as preparedness, response, recovery and rebuilding [23, 95, 96] and how such platforms were used during the Haiti earthquake [41, 50, 76, 84, 86, 108, 123]. The opportunities presented by social networking with web systems for humanitarian information management are also explained by [122]. However, there is also evidence as to how humanitarian communities failed to utilize such information resources to better support decision-making processes. For example, the emergency response system in Haiti was not organized to integrate intelligence from affected communities [48] but rather was designed for collaboration and information sharing between international teams of responders [47, 50].

The joint report from the Harvard Humanitarian Initiative, the United Nations Foundation, the UN Office for the Coordination of Humanitarian Affairs (OCHA), and the Vodafone Foundation documented two major observations on how people communicated after 2010 Haiti earthquake. According to the report, for the first time people directly affected by disaster issued pleas for help using available mobile technologies. At the same time, thousands of ordinary citizens helped to aggregate, translate, and plot these pleas on maps and to organize technical efforts to support the disaster response. "The international humanitarian system was not tooled to handle these two new information fire hoses -one from the disaster-affected community and one from a mobilized swarm of global volunteers." ([47], page 9). **This research intends to highlight that linked data could provide a stable solution to tap into crowdsourced volunteered information and messages received directly from the affected population using Ushahidi platform.**

### *2.4 Humanitarian Information: Who is doing what, Where and When?*

According to UNOCHA, the UN humanitarian coordination and advocacy office, the key information that is important to assess and ensure that humanitarian needs are met in any emergency/disaster are: 1) to know which organizations (Who), 2) are carrying out what activities (What), 3) in which locations (Where), also universally referred to as 3W (Who does What Where). The Who does What Where information (3W) is one of the key elements and core products necessary during a disaster response [111, 113].

The earliest version of 3W can be traced back as far as the Kosovo crisis in 1999 where the first Humanitarian Information Center (HIC) was established [1]. The tool is of such importance that it is also used for key indicators that help in identifying humanitarian needs and possible gaps based on the location and capacity of relief agencies as well as the affected population. Most recently the same tools with temporal components of "When?" were also used as a humanitarian agencies performance indicator to measure if the needs of the Pakistan floods beneficiaries were covered or not [109].

Information Scientists' use of "Who What Where When?" as a tool of information analysis was first discussed by Panofsky [97] and explicated by Markey [78]. The method was further refined by Shatford [104] and is now commonly referred to as the Panofsky-Shatford facet matrix. The matrix has been used in various fields of information analysis like image archiving [6] [30]. It is also used as a means of support for semantic descriptions of place [28] and for ontological relationship analysis between entities [13].

On the onset of a disaster, the first days of a response can shape the type and scale of humanitarian assistance [119]. When resources are scarce, making the right decisions is critical

and it is very much driven by humanitarian needs. Additionally, this knowledge has to be based on well-founded evidence and observation. Hence information managers on the ground or at the headquarters are under pressure to figure out what those needs are and, based on the scale of the disaster, it should also be compared across countries and across continents, and across crises of very different kinds and origins [53]. Who Does What Where (3W) information may be a useful starting point for anticipating capacities for response by sector or geographic area [113, 114].

Facets	Specific Of	Generic Of	About
<b>Who?</b>	Individually named persons, animals, things	Kinds of persons, animals, things	Mythical beings, abstraction manifested or symbolised by objects or beings
<b>What?</b>	Individually named events	Actions, conditions	Emotions, Abstractions manifested by actions
<b>Where?</b>	Individually named geographic locations	Kind of place geographic or architectural	Places symbolised, abstractions manifest by locale
<b>When?</b>	Linear time; dates or periods	Cyclical time; seasons, time of day	Emotions or abstraction symbolised by or manifest by

Figure 2-3 Panofsky-Shatford facet matrix (source Shatford 1986)

Despite its usefulness and even with the commitments of all the humanitarian actors to share information about their activities, 3W information collection can take days to weeks. Such delays are unacceptable when lives of a vulnerable population are at stake. However, during the Libya 2011 crisis, UNOCHA with the newly formed Standby Task Force Team<sup>14</sup> (SBTF) was able to map 3W information within 48 hours. This was an unprecedented achievement in two ways. First, traditionally 3W mapping involved a lot more than just collecting information. Therefore, it usually took at least a week to get a better picture [117]. Second, one needs to bear in mind that there was no or very little humanitarian access to the country. V&TC not only broke the information gathering record, it also showed that crowdsourcing humanitarian work can be done without geographical boundaries and from different parts of the world.

Even Tim Berners-Lee [15] used the concept of who?, where? and why? to explain how semantic links can be created across totally different, unrelated but identifiable objects using the Resource Description Framework<sup>15</sup> (RDF), a standard model for data interchange on the web [9, 15, 18, 79].

In emergency situations, every need is paramount and timely collection and analysis of who is doing (or in need of) what and where activities is second to none. **During the information collection step of the information cycle, I propose to use Panofsky-Shatford facet matrix (figure 2-3) to semantically categorize and link the facets like who, what, where and when to link data between crowds and traditional humanitarian agencies.**

<sup>14</sup> Standby Task Force online volunteer community for live mapping

<http://blog.standbytaskforce.com/about/introducing-the-standby-task-force/>

<sup>15</sup> <http://www.w3.org/RDF/>

### *2.5 Linked Data: An Approach for interrelating information in disaster management*

The term Linked Data was coined by Tim Berners-Lee [14] in his Linked Data Web architecture note. It refers to a style of publishing and interlinking structured data on the Web. One of the main idea behind Linked Data is that the value and usefulness of data increases as more data are available in machine readable format with possibilities to link with in the web of data. In summary, Linked Data is a principle that uses the Web to create typed links between "things" like data from different sources. Linked Data [3, 9, 14, 18, 19, 49, 63] is about using web techniques to share and interconnect pieces of data by following a few simple principles. Linked Open Data builds on Semantic Web technologies, where data is encoded in the form of subject, predicate, object also know as RDF triples. The main goal of Linked Data is to break the "data hugging" practices and provide "raw data now"<sup>16</sup> to enable people to share structured data on the Web as easily as they can share documents today. There are different documents talking about the modules or the infrastructure [21, 31, 32, 94] to include linked data for managing disaster information but not much work has been done in terms of developing and publishing a disaster management vocabulary. **With this research, I plan to collect vocabulary for crisis management and use the concept to demonstrate how Linked Data principles can be used to integrate information across crowdsourcing and traditional humanitarian agencies.**

---

<sup>16</sup> See "Raw Data Now" - One year on in the UK. <http://www.guardian.co.uk/open-platform/blog/raw-data-now-one-year-on-in-the-uk>

### 3. Methodology

This section comprises the conceptual and analytical framework that directs the interpretation, analysis and discussion to respond to the research goal and to meet the objectives of the study.

#### 3.1 Humanitarian responders expert survey on information management and crowdsourcing

Experienced humanitarian responders and information managers from the United Nations, the Red Cross, International/Non Governmental Organizations have firsthand experience regarding information coordination in the field (e.g., in a disaster affected area). To understand the ground reality of crowdsourced information and to document challenges associated with heterogeneity of data, a targeted expert survey is conducted. This section will support all the objectives mentioned under section 1.3.

#### 3.2 Management of a Crisis Vocabulary

One of the primary goals of the research is to link crisis information from three different sources: a) traditional humanitarian agencies (Shelter Cluster), b) Volunteer and technical committees (Ushahidi platform) and c) disaster affected communities. From the literature review in section 2.5 we can summarize that there are existing linked data infrastructures for designing emergency management [21, 31, 103]. However, during the course of research period, the author did not come across vocabularies that could be used to gist of this research.

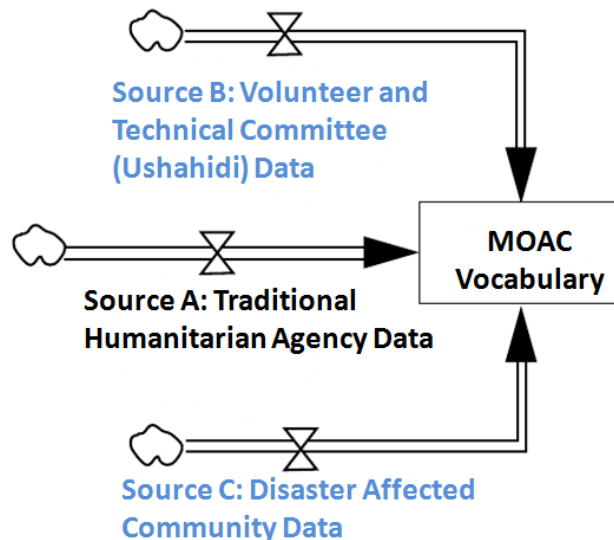


Figure 3-1 Management of a Crisis Vocabulary (MOAC)

Therefore, the vocabulary also known as management of a crisis vocabulary (MOAC) is planned based on the combination of concepts and methodologies explained in:

1. Terms of traditional data analysis [26],
2. Werner Kuhn's conceptualization of interoperability [69, 70],
3. Tomi Kauppinen et al. approach of Linked Science[64, 66],
4. Methontology cycle and linked data design [31, 33, 69, 70, 103], and
5. A co-authored ISWC<sup>17</sup> 2011 workshop paper Crowdsourcing Linked Open Data for Disaster Management [94]

<sup>17</sup>See <http://iswc2011.semanticweb.org/>

This section meets objectives 1 and 2 under section 1.3 by documenting the process of creating a semantic vocabulary for crisis management also known as MOAC.

### **3.3 Implementation of linked data**

The real time disaster documents or messages from the volunteer technical committee (Ushahidi), the affected population and the humanitarian community are converted into data and then to Tim Berners Lee's 5 star linked data standards [14]. This section demonstrates how the research meets objectives 2 and 3. For this, actual disaster data from emergency Shelter Cluster and Ushahidi Haiti Project are tested.

### **3.4 Evaluation**

Even though, there is no set of concrete guidelines on how to evaluate vocabulary that support Knowledge Sharing Technology like Management of a Crisis Vocabulary (MOAC), evaluation in this research is based on the framework proposed by Gómez-Pérez [36], Gómez-Pérez al. [34] and Elhadad et al.[29] in 1994, 1995, 2011 respectively. With the identified method of evaluation, the proposed MOAC vocabulary is tested and then verified with the help of domain experts and linked data enthusiasts. This section also weaves together all the three objectives on how the research meets the goal of *integrating crowdsourced information with traditional Crisis and Disaster Management Information using Linked Data*.

## 4. Expert Survey: Crisis Information Management, Crowdsourcing, Volunteered Technical Committee - Opportunities and Bottlenecks

### 4.1 Introduction

A survey was conducted to gather the perspective of humanitarian responders and relief workers concerning the use of information contributed by non-experts (also known as volunteered sources or crowdsourced information) on the immediate onset of a disaster. The survey intends to highlight a disaster information manager's view on usefulness, problems and possible improvements for the use of volunteered or crowd-generated information using social media or similar networking technologies.

### 4.2 Expert Survey Methodology

The survey was comprised of seven multiple-choice questions and one free text field for comments or personal observations. The first two questions assessed participants' information management competencies in a crisis situation and the other five questions targeted their experiences with crowdsourcing tools and services.

To narrow down the tools used in a crisis only a handful of social networking and crisis management platforms were chosen for the survey. The collection was based on the research of technologies used in the Haiti earthquake [47, 48, 76, 123] as well as on my own first-hand experience as a humanitarian information manager.

Since the idea was to gather expert opinion on crowd tools and volunteer technical committee (VTC) technologies in a disaster setup, the survey was intentionally targeted towards selected individuals known as emergency first responders [74, 101]. The survey questionnaire was generated using Google spreadsheet<sup>18</sup> and the responses were collected online for about one month.

### 4.3 Results

Out of 25-30 email requests sent out, 17 unique survey were gathered from experts, including but not limited to, agencies like the United Nations, Red Cross, non-governmental organizations and donor communities. The introductory questions revealed that over 75% of those who contributed to the survey were humanitarian information managers with a wide range of experiences in complex emergencies (73%).

---

<sup>18</sup> <http://support.google.com/docs/bin/answer.py?hl=en&answer=87809>

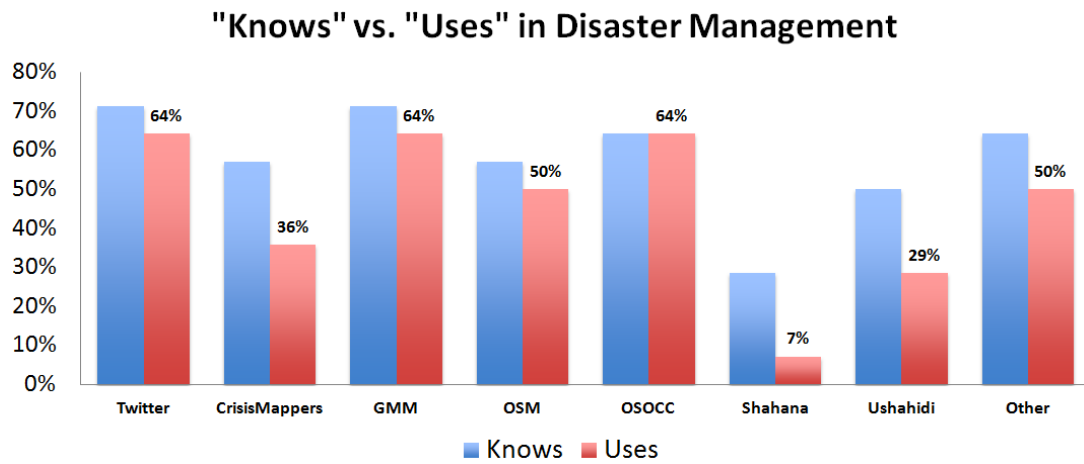


Figure 4-1 Expert Survey "Knows" vs "Uses" in Disaster Management

The survey started with question asking participants about their awareness of certain platforms that could allow the crowd to contribute or process information. The experts were more familiar with services like Twitter<sup>19</sup>, Google Map Maker<sup>20</sup> (GMM), and Open Street Map<sup>21</sup> (OSM) followed by crisis and message broadcasting platforms like Ushahidi<sup>22</sup>, Sahana<sup>23</sup> and On site Operations Coordination Center<sup>24</sup> (OSOCC). When responders were asked if they or their team members use the tools for managing disasters there were mixed reactions as shown in figure 4-1.

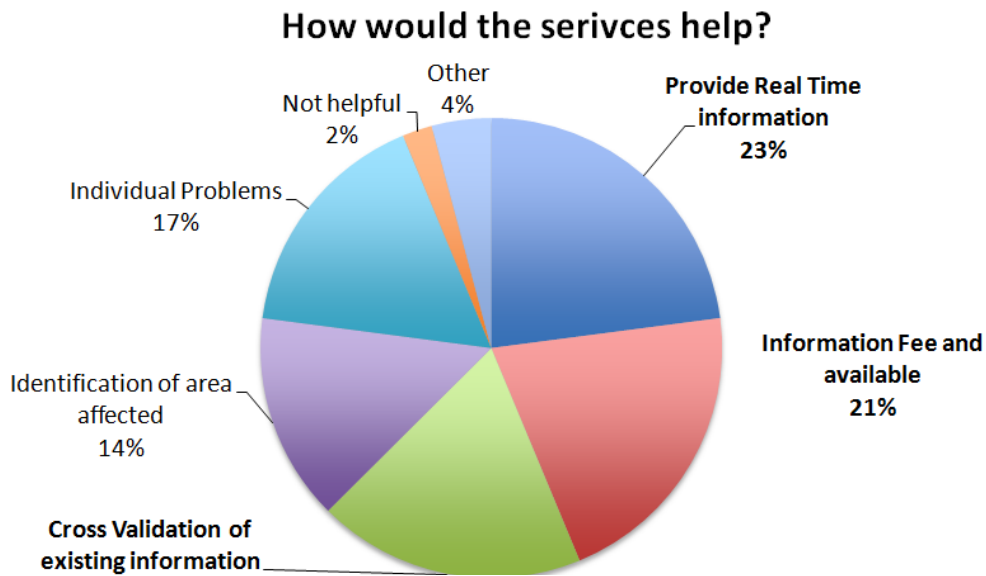


Figure 4-2 Expert Survey - How would crowdsourcing/volunteer platform help?

Experts were also specifically asked if the given crowdsourced volunteer committees, platforms and services can help improve their information management work and hence support decision-making processes. Except for one, everyone responded positively as shown in Figure 4-2.

<sup>19</sup> <http://www.twitter.com>

<sup>20</sup> <http://www.google.com/mapmaker>

<sup>21</sup> <http://www.openstreetmap.org/>

<sup>22</sup> <http://ushahidi.com/>

<sup>23</sup> <http://www.opensource.org/node/501>

<sup>24</sup> <http://vosocc.unocha.org/>

#### 4.4 Discussions

Since disaster information managers are on the front line to collect, process, analyze and provide refined, timely information products to support the decision-making process, this survey was particularly insightful in summarizing their personal views on the usability and constraints of crowdsourcing / volunteering platforms.

##### *Expert Survey - How would crowdsourcing/volunteer platform help?*

From figure 4-1: "Knows vs. Uses", we can argue that disaster information managers are aware that social networking and crowdsourcing platforms have become an integral part of collecting and sharing information in a disaster. Still they are not able to integrate the information into their system. The data administrator in a disaster is already over burdened with work and pressure to produce outputs [80, 119]. Therefore even if he or she wanted to use the crowdsourced information, they would have neither the time or the necessary tools to integrate such information.

In the Haiti earthquake, a different crowdsourced and volunteered information platform was launched. During the independent evaluation of such efforts in Haiti, it was made evident that the platform helped in integrating messages from the affected community as well as from social media [51, 75]. Also, the humanitarian agencies information system was neither equipped nor ready to integrate information from the volunteer technical committee [47, 50]. When you are working 98% of your time and sleeping for 2%, asking the humanitarian information managers to find a way to integrate such untested yet valuable information during a disaster response would not be the priority of the relief agencies. This message has been acknowledged through the survey free text section.

Like many other ideas, such new innovations don not come without limitations (figure 4-3). In this survey, experts also shared their own thoughts about the possible improvements required by VCT systems before integrating it into the traditional humanitarian information management cycle. Some of the extracts from the survey free text are grouped below:

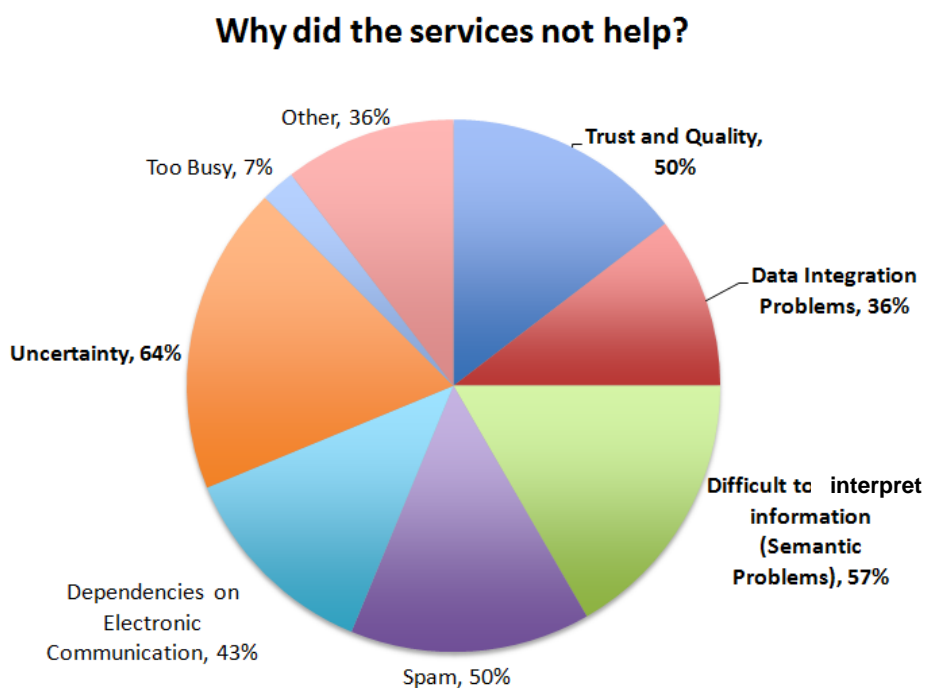


Figure 4-3 Expert Survey - Why did the crowdsourcing/volunteer platform not help?

- *Uncertainty-information as a noise*: One expert said that " (v)olunteer technical committees is exciting and represents an opportunity, (but) I have yet to see an example of any operationally interesting or useful information being produced. It all seems to be a lot of "noise" " .
- *Standardization and data integration problem*: The Shelter information management expert said " (...) the short of the problem is that socially collected data is unverifiable and unstandardized".
- *Information overload*: ".. the multiplicity of assessments, information sources, platforms make it difficult/ wasting time to look for the information you are looking for and capture the overall picture. The willing to share information is sometime a problem too!"
- *Heterogeneous system and incompatible data formats*: "Information managers most often have to deal with heterogeneous system and incompatible data formats. To make these systems operational from very beginning of disaster response, these systems should be localized beforehand and introduced to the humanitarian communities. In addition, if the functionality of these systems extended to cater the need of disaster preparedness activities and establish disaster decision support system before had, these could be more useful in the response time."
- *Semantic problem*: One of the expert shared that we need "(..) to (somehow) reduce duplication of similar / same information products (..)" and " there are still difficulties to manipulate the (information management) devices".

Volunteering and crowdsourcing services have been proven to be useful and, at times, exemplary in gathering real time information [50, 76, 101]. It is a strong statement that over 75% (survey question 3) of the expert survey responders indicated it is useful to have knowledge and experience from "other" forms of social media and crowdsourcing platforms during a crisis. The survey also clearly reflects that the humanitarian responders did use VCTs for confirming or cross-referencing isolated incidents collecting from their own traditional sources. Since the concept of volunteerism and crowdsourcing costs no money by definition and the information was already available in the web, a number of experts also suggested that the platform was useful to see the trends of affected population requests collected directly through social networks and mobile phones [92].

Mr. Andrej Verity, who is a disaster first responder and also a veteran information management officer with UNOCHA (also contributing author of [47, 117]) shared through the survey that "(f)rom the OCHA side, I coordinated the deployment of the Libya Crisis Map (in 2011) by the Standby Task Force on the Ushahidi platform. They used public data sources such as Twitter, regular news media sites, etc to compile data. Challenge is to provide an *appropriate categorization* (with sufficient explanation) so that volunteers can classify correctly from the beginning. Although the tool mapped the various reports, we also found that the data was very rich and could be used for various other products that would aid in the response. That will be addressed in the future."

Based on all these observation, the survey summarizes that there is a need for a reliable yet simple technique to address the issue of information categorization, explanation of the terms used and semantic problems of unstructured data. The finding of this survey answers part of the research questions<sup>1,2</sup> and 3 and also provides the foundation for construction of a common linked data vocabulary for managing a disaster. One of the first steps one can take is to support both the humanitarian partners as well as the crowdsourcing and volunteer technical committees by establishing different vocabularies. As a result, I will use this survey as the disaster expert

proof to propose *Linked Data* as a tool to convert documents into data and create linkages in the form of web of data also known as semantic web [15].

To link data about crisis management a vocabulary of terms is needed, which is discussed in the following chapter.

## 5. Management of a Crisis Vocabulary (MOAC)

This chapter is associated with the second objective of my research that is to establish a Linked Data vocabulary for the *Management of a Crisis*, also known as MOAC.

### 5.1 *The story so far*

Today, most data from humanitarian agencies are in the form of spreadsheets, PDFs and word processor documents hidden somewhere in a desktop. A number of agencies are making gradual efforts to publish these data as an open source [1]. However, based on the literature review (section 2) and the expert survey (section 4), there are no suitable disaster management vocabularies with well-defined URI<sup>25</sup>s and RDFs<sup>26</sup>[11]. A solution to this problem is of particular importance to emergency response organizations that need access to all of the information available to better respond to disasters and crises.

This chapter explains a series of steps taken to build the management of a crisis vocabulary, also known as MOAC. As explained in the methodology section 3, the combination of "Methontology", semantic and vocabulary concepts [20, 33] are used to define MOAC. For generalization, this research views the term "ontology" synonymously as defined by Prof. Werner Kuhn [70] that "(..) ontology defines the vocabulary of an information community in computer-readable form." So, the author refers methontology as a method to structure vocabularies where ontology can be referred as kind of information object or computational artefact [42]. Since there is no linked data based disaster vocabulary (explained in section 2.5 and 4), this research aims utilize methontology lifecycle to construct and propose a common MOAC vocabulary to easily "read-write in the web"<sup>27</sup>.

MOAC will be used as a part of the terminologies and controlled vocabularies while building a linked dataset using RDF triples as demonstrated in section 6 and evaluated in section 7.

### 5.2 *Vocabulary Development Process*

For the development of a structured vocabulary, the methontology life cycle [33] explains series of sequential states and activities. It consists of states like specification, conceptualization, formalization, integration, implementation and maintenance. Activities like knowledge acquisition, evaluation and documentation are carried out throughout the whole life of the vocabulary.

#### 5.2.1 Specification

The first stage of the MOAC is to define the purpose and scope of the vocabulary including its intended uses, scenarios of uses and end users as shown in table 1. It focuses on why the MOAC vocabulary is being built. The vocabulary presented in this study is primarily based on the information gap and the need to utilize crowdsourced/volunteered technical committees as discussed in Sections 2.3 and 4. Since each disaster is different from the next [98], there is no way to prove the total completeness of the vocabulary. Hence, MOAC is proposed as an evolving, informal and partially complete vocabulary set.

---

<sup>25</sup> URI or uniform resource identifier is a string of characters used to identify a name or a resource. (Wikipedia)

<sup>26</sup> RDF or Resource Description Framework is a standard model for data interchange on the Web. (source: W3C)

<sup>27</sup> Tim Berners Lee, <http://news.bbc.co.uk/2/hi/technology/4132752.stm>

<b>Management of A Crisis (MOAC) Vocabulary Requirement Specification Document</b>		
<b>Item</b>		<b>Specification</b>
Domain	Related field	<ul style="list-style-type: none"> <li>Humanitarian Information Management in Complex Emergencies</li> </ul>
	Knowledge covered	<ul style="list-style-type: none"> <li>Humanitarian needs and response</li> <li>New innovative use of Crowdsourcing and Volunteering committees for disaster response using the concept of Linked Data</li> </ul>
Purpose	Main purpose	<ul style="list-style-type: none"> <li>Conceptualization and formal representation of humanitarian semantic vocabulary called management of a crisis (MOAC)</li> <li>Explore the use of Linked Data to publish and reuse disaster/crisis information from heterogeneous sources</li> <li>Integration of data hidden within VGI, crowdsourcing and non-expert communities</li> <li>Integration of data in the form of 5 star linked data standard [14]</li> </ul>
	Intended users	<ul style="list-style-type: none"> <li>International Humanitarian Communities</li> <li>Crowdsourcing communities</li> <li>Volunteer Technical Team - Standby Task Force for crisis mapping</li> <li>Linked Data enthusiasts</li> <li>UN Global Pulse, UNOCHA, Inter Agency Standing Committee Information Management Task force</li> </ul>
Level of Formality		<ul style="list-style-type: none"> <li>Semiformal and partially complete vocabulary</li> </ul>
Scope	Key concepts	<ul style="list-style-type: none"> <li>Complex emergencies, humanitarian needs, who is doing what activities and where?, emergency response, humanitarian needs, IASC cluster information, affected population</li> </ul>
	Properties	<ul style="list-style-type: none"> <li>Emergency Related: Shelter Cluster</li> <li>Humanitarian needs: Shelter Needed</li> <li>Humanitarian response: Shelter Offered</li> <li>3W:FoodResponse, 3W:WASHResponse</li> </ul>
Source of Knowledge	Humanitarian Information management	<ul style="list-style-type: none"> <li>Community of Humanitarian Partitions like UNOCHA, UN Global Pulse, Inter Agency Standing Committee, UNISDR</li> </ul>
	Crowdsourcing Volunteer Technical Committee (VCT <sup>28</sup> )	<ul style="list-style-type: none"> <li>Ushahidi Platform, CrisisMappers, Standby Task Force</li> </ul>

Table 1 Management of A Crisis (MOAC) Vocabulary Requirement Specification Document

<sup>28</sup>Standby Task Force - Online Volunteer Committee for Live mapping. <http://blog.standbytaskforce.com/>

### 5.2.2 *Knowledge Acquisition*

Once answering why, for whom and for what purpose, the MOAC vocabulary is being built the next step is to gather vocabulary and taxonomy contents from as many reliable sources as possible. Fernández-López et.al. described knowledge acquisition as an independent activity in the vocabulary development process which is also conducted in conjunction with the requirement specification. For this research following steps were taken:

1. Domain expertise: The author is himself is a humanitarian worker with over 8 years of experience in managing information during complex emergencies. Based on the experiences gathered from various disasters, the author chose to use his personal reflections from disasters in Nepal 2005 - 2008, the Philippines emergency in 2009 and the Haiti earthquake in 2010.
2. Non structured interviews with other domain experts:  
The author had the privilege of working with domain experts from the United Nations, the volunteering committee as well as with the affected population. In a period of 8 months, over 35-40 experts were consulted formally and informally to gather following information on:
  - What are the main concepts that describe the humanitarian information management processes in a disaster?
  - What are the possible scenarios to use Linked Data between the traditional humanitarian agencies, the crowd communities as well as people affected by a disaster?
  - If vocabulary were to be used for linking data, what are the core conceptual terms necessary for the three different information contributors?
3. Informal text analysis:  
To further populate and extend the core concepts of the vocabulary, additional key terms were extracted from humanitarian reports, crowdsourcing information forms and the affected communities' bulletins. Reports were mostly in the form of PDF, word documents, excel and CSVs existing on the internet. For cross validation and cross fertilization of concepts, the author explored his previous contacts within disaster management agencies like International Federation for Red Cross and Red Crescent Movement (IFRC), the United Nations, Government officials in various parts of the world.
4. Formal text analysis  
After steps 1 to 3, a considerable amount of time was spent in finding the right structures and the kind of knowledge contributed by each of the concepts, attributes, and relationships. As predicted, most of the concepts are centered on the humanitarian work and disaster management phase summarized in figures 5-2, 5-3 and 5-4. The detailed breakdown of glossary of terms is explained in the following section.

### 5.2.3 Conceptualization

CONTACT CARD		OCHA
First Name:		
Last Name:		
Job Title:		
Organization (Acronym):		
Phone Primary:	<input type="checkbox"/> Cell	<input type="checkbox"/> Land <input type="checkbox"/> Sat
Phone Other:	<input type="checkbox"/> Cell	<input type="checkbox"/> Land <input type="checkbox"/> Sat
Email:		
Location:	(Country \ Admin 1 \ Admin 2 \ Admin 3)	
Mission End Date (DD/MM/YY):		
Your Contact Status :	<input type="checkbox"/> Public <input type="checkbox"/> Private	
Sector(s) :	<input type="checkbox"/> Agriculture <input type="checkbox"/> Camps <input type="checkbox"/> Coordination <input type="checkbox"/> Early Recovery <input type="checkbox"/> Education <input type="checkbox"/> Environment <input type="checkbox"/> Food <input type="checkbox"/> Health <input type="checkbox"/> Information <input type="checkbox"/> Nutrition <input type="checkbox"/> Logistics <input type="checkbox"/> Protection <input type="checkbox"/> Shelter <input type="checkbox"/> Telecoms <input type="checkbox"/> WASH *Others: _____	
	<input type="checkbox"/> Modification \ Correction?	<a href="http://3w.unocha.org/">http://3w.unocha.org/</a>



Figure 5-1 Example of a Humanitarian Contact Card showing how various terms are conceptualized across sectors like Shelter, Food, Health. (Credit: UNOCHA)

Gómez-Pérez et. al [37] presented principles, design criteria and a verification process used in the knowledge conceptualization process of vocabulary construction. Similarly after the knowledge acquisition, methontology [33] proposes to 1) build a glossary of terms and 2) group terms as concepts and verbs. Concepts can be: abstract or concrete, elementary or composite, real or fictitious, description of a task, function, action, strategy, reasoning process, among others [38].

Depending on the intended meaning of the term, the main concepts are classified as core concepts and secondary concepts. Core concepts are those terms that are central to describing the domain. Secondary concepts are required to fully describe the core concepts and satisfy the purpose<sup>29</sup>. For example, humanitarian information management are core concepts whereas information collection, analysis, processing, and dissemination are secondary concepts in relation to the core concept. I use this notion of core concepts and secondary concepts, also known as classes and subclasses [90], in the formulating MOAC vocabulary.

Based on the knowledge acquisition steps carried out in Section 5.2.2 and based on the targeted audience of the vocabulary, the core concepts that are essential to define the 'Management of Crisis Vocabulary' are grouped into the following concepts:

1. Traditional Coordination Humanitarian Agency: Shelter Cluster (Humanitarian Needs of affected population and response of relief agencies)
2. Who, What, Where (3W)
3. Crowdsourcing and Volunteer Technical Committee

### 5.3 Who What Where when? Vocabulary

Referring back to the literature review in section 2.4, for humanitarian response, "who what where and when?" data is at the centre of the humanitarian information management processes.

<sup>29</sup> <http://www.ordnancesurvey.co.uk/oswebsite/ontology/>

Based on the humanitarian core concept of 3W [111, 113] and the Panofsky-Shatford mode/facet matrix [104], the who what where and when has the following core and secondary concepts (figure 5-2). Only a section of the concepts are described here. The complete descriptions and its relationships across concepts are given in details online at <http://observedchange.com/moac/ns/>.

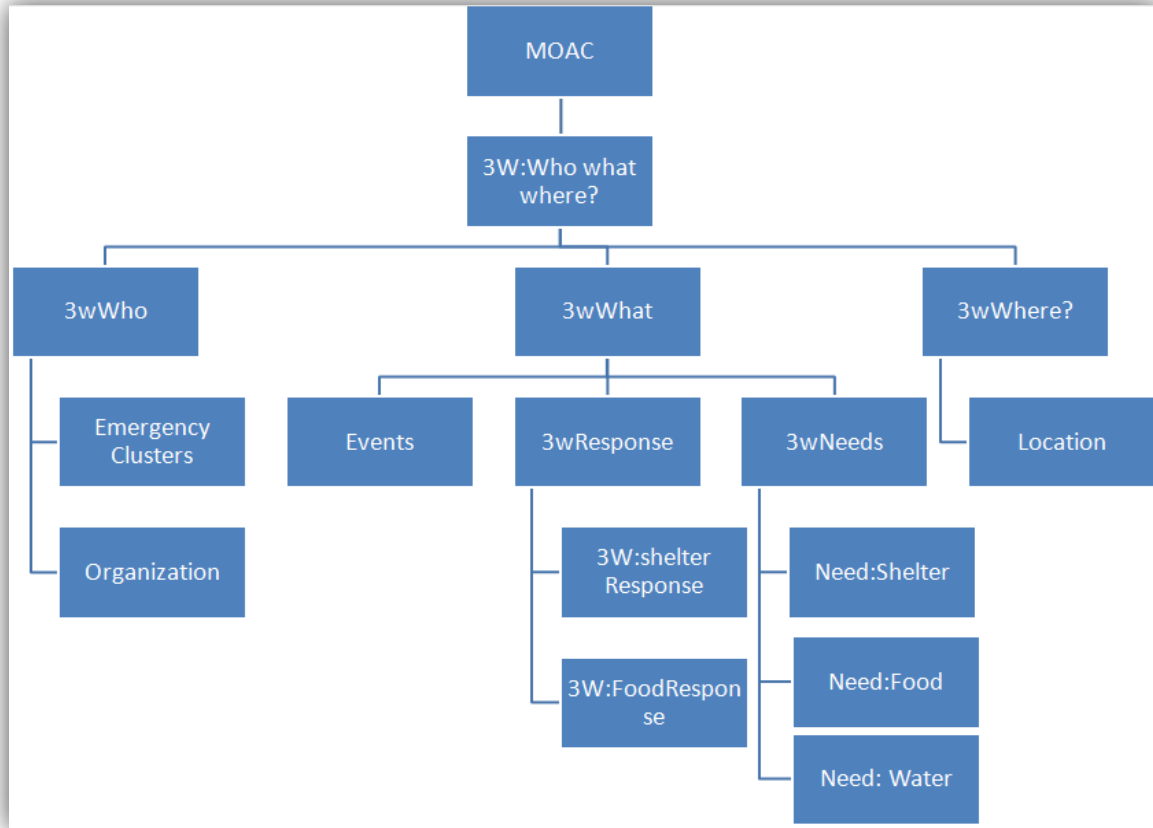


Figure 5-2 Who what where and When (4W) Concept Derivation in Humanitarian Response. Detailed definitions in <http://observedchange.com/moac/ns/>

Here, "When" is also the temporal component and comes from the informal and formal text analysis [20, 33] as well as from visual comparison with the traditional humanitarian agency mapping. Such as a breakdown of classes can be very helpful while connecting data across different domain using following URIs:

Knowing which organizations (Who) are carrying out which activities (What) in which locations (Where) is essential if those organizations and activities are to be coordinated in a way that ensures that humanitarian needs are met<sup>30</sup>. The 3W not only helps the humanitarian agency but it also helps the people affected by disaster to know where to go for help or who to ask for assistance. Besides, 3W information also helps decision makers access information by agency, by location and by activities and find ways to meet the gaps in relief activities. On the immediate on set of a disaster it is of paramount importance to make sure that there is as little duplication of coordination as possible.

Based on the Panofsky-Shatford mode/facet matrix [104] explanation of these conceptual (also called as classes [90]) terms, almost all concepts used for the vocabulary can be secondary concepts (or subsidiary nodes) to the four parent key terms, i.e., who what where and when. For example, the location terms discussed either by the crowd community or affected population or

<sup>30</sup>See <http://ochaonline.un.org/ethiopia/WhoWhatWhere/tabid/2944/language/en-US/Default.aspx>

relief agencies can be branched under the "where". Similarly, the names of the organisations, people, and groups can also be branched under "who". The location and time related vocabularies become the component of where and when respectively. This concept of associations, linkages and relationships make Linked Data one of the better options to interconnect data.

### Who What Where and When URI

Terms	URIs	Description [104]
moac: WhoWhatWhere	<a href="http://observedchange.com/moac/ns#WhoWhatWhere">http://observedchange.com/moac/ns#WhoWhatWhere</a>	<ul style="list-style-type: none"> <li>• Specific of who is doing what kind of activity and where</li> <li>• Generic of humanitarian need and response analysis</li> <li>• About finding humanitarian gap</li> </ul>
moac:Who3W	<a href="http://observedchange.com/moac/ns#Who3W">http://observedchange.com/moac/ns#Who3W</a>	<ul style="list-style-type: none"> <li>• Specific of individually named persons, animals, things</li> <li>• Kinds of persons, animals, things</li> <li>• Mythical beings, abstraction manifested or symbolised by objects or beings</li> </ul>
moac:What3W	<a href="http://observedchange.com/moac/ns#What3W">http://observedchange.com/moac/ns#What3W</a>	<ul style="list-style-type: none"> <li>• Specific of individually named events</li> <li>• Generic of Actions, conditions</li> <li>• About Emotions, Abstractions manifested by actions</li> </ul>
moac: Where3W	<a href="http://observedchange.com/moac/ns#Where3W">http://observedchange.com/moac/ns#Where3W</a>	<ul style="list-style-type: none"> <li>• Specific of individually named geographic locations</li> <li>• Generic of kind of place geographic or architectural</li> <li>• About places symbolised, abstractions manifest by locale</li> </ul>
moac:When3W	<a href="http://observedchange.com/moac/ns#When3W">http://observedchange.com/moac/ns#When3W</a>	<ul style="list-style-type: none"> <li>• Specific of linear time; dates or periods</li> <li>• Generic of cyclical time; seasons, time of day</li> <li>• About Emotions or abstraction symbolized by or manifest by</li> </ul>

Table 2 Who What Where and When concept Description. For details description of other "things", please visit <http://observedchange.com/moac/ns/> (one of the output of this research)

#### 5.4 Traditional Humanitarian Coordination Agency Vocabulary: Shelter Cluster

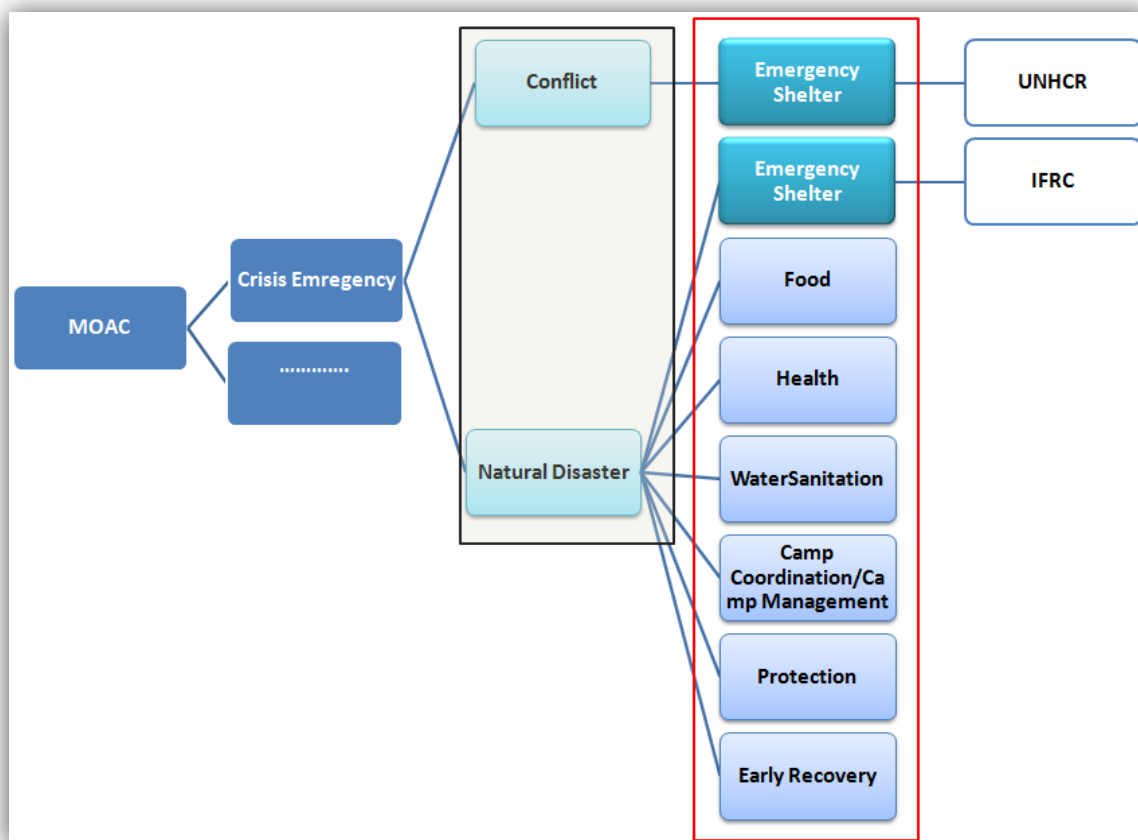


Figure 5-3 Inter Agency Standing Committee Core Terms (source IASC). For this research Emergency Shelter Cluster convened by IFRC will be reviewed.

The first set of core term for this core concepts are the organisations and the type of humanitarian work those organisations are responsible for during specific emergency.

The Inter-Agency Standing Committee (IASC) was established in 1992 by General Assembly Resolution 46/182 on strengthening the coordination of humanitarian assistance [57]. The IASC agreed to designate global “cluster leads” with defined terms of reference<sup>31</sup> for humanitarian emergencies.

Initially nine<sup>32</sup> sectors or areas of activity were identified as shown in the red box in figure 5-3. Agencies such as the World Food Program leads Food Cluster, UNICEF leads Water/Sanitation, Protection, Nutrition, and so on and so forth. For emergency shelter activities, (one of the core terms), IFRC agreed to convene as the cluster lead during natural disaster and coordinate activities to avoid unpredictable humanitarian responses and also to meet inevitable capacity and response gaps in some areas[55]. However, if the type of emergency is conflict then UNHCR with other partners will lead the Emergency shelter cluster coordination.

IFRC convened information management work during Haiti earthquake was regarded as one of the best practices for crisis management [7, 25]. The author was also a part one of the information managers supporting over 80 different cluster partners and also managing information to meet the needs of over 1.5 million affected people. The following figure 5-4

<sup>31</sup> See <http://onerresponse.info/Coordination/ClusterApproach/publicdocuments/Forms/DispForm.aspx?ID=24>

<sup>32</sup> See <http://www.humanitarianinfo.org/iasc/downloadDoc.aspx?docID=5712>

documents a portion of its information collection key terms. When examined closely, there is a relationship between the given set of terms and the proposed 3W terms:

- Who: is related to Organization
- What: is related to Shelter items, Non Food Items and activities
- Where: Location
- Affected population: covered under coverage
- Also the shelter items and NFI distribution reports can be included as the 3W:ShelterResponse in "who what where" matrix. Such linkages can be further exploited using RDFs, a technology for developing semantic web [15, 16, 66, 102], as shown in the upcoming chapter.

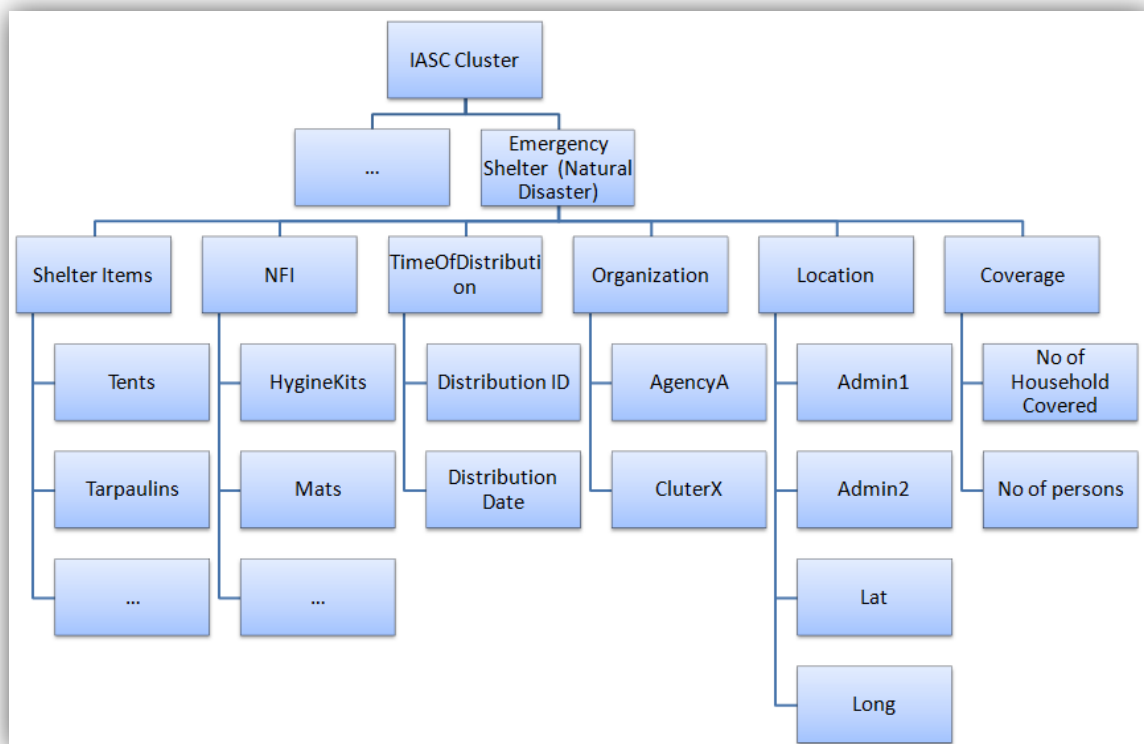


Figure 5-4 IASC Shelter Cluster (Natural Disasters) Information Collection Graph (Source IFRC, Cluster Coordination/Information Management Focal Point)

Shelter Cluster information collection form distributed to over 90 immediate cluster partner agencies and an unknown number of information sources via the internet. (source: Shelter Cluster Information Management Lead, IFRC). Detailed breakdown of definitions, its relationships with participating concepts and relevant referrals are explained in <http://observedchange.com/moac/ns/>. Selection and descriptions of each concept terms are made with reference to the

### 5.5 Crowdsourcing and Volunteer Technical Committee - Ushahidi Haiti Platform Vocabulary

Section 5.3 and 5.4 constituted humanitarian relief agencies related disaster management vocabularies. To link data between the traditional humanitarian information with crowdsourcing mechanisms, we need to understand the kind of data that has been collected and coordinated via social network or using crowdsourcing techniques. Ushahidi<sup>33</sup> platform is chosen for this task.

<sup>33</sup> See <http://ushahidi.com/>

Ushahidi is a free and open-source platform that allows interested individuals and groups to create live, interactive maps[89]. The workings of the tool has been explained in detail in [92]. During the Haiti earthquake response, Ushahidi Haiti Project (UHP) was launched as a volunteer-driven effort after the January 12, 2010 earthquake in Haiti [83]. The project was regarded as one of better examples of an application of crisis mapping and crowdsourcing to large scale catastrophes and a novel approach to the rapidly evolving field of crisis informatics [47].

All the categories used in the Online crisis mapping<sup>34</sup> was converted into uniquely identifiable resources and given a unique identifier as shown below. The entire list of vocabulary can be accessed from ObservedChange [73]. Based on this vocabulary, the author along with ontology and linked data experts wrote a paper suggesting that Linked Data can help ease the challenge of information triage in early stages of disaster management [94].

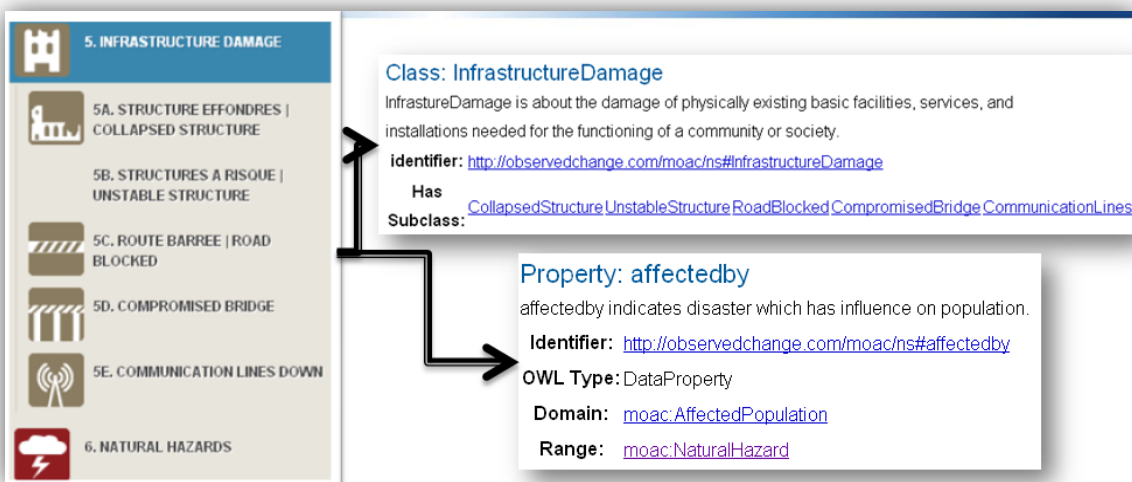


Figure 5-5 Ushahidi Haiti Platform Concept Terms derived from online mapping platform (source Ushahidi Haiti Project)

### 5.6 Integration of crowdsourcing and crisis management vocabulary

Here is discussed the vocabulary in support of linking data across traditional humanitarian communities and the crowdsourced/volunteer technical committees. The concepts generated in section 5.3,5.4 and 5.5 are integrated under the revised version of the Management of a Crisis Vocabulary version 0.4<sup>35</sup>.

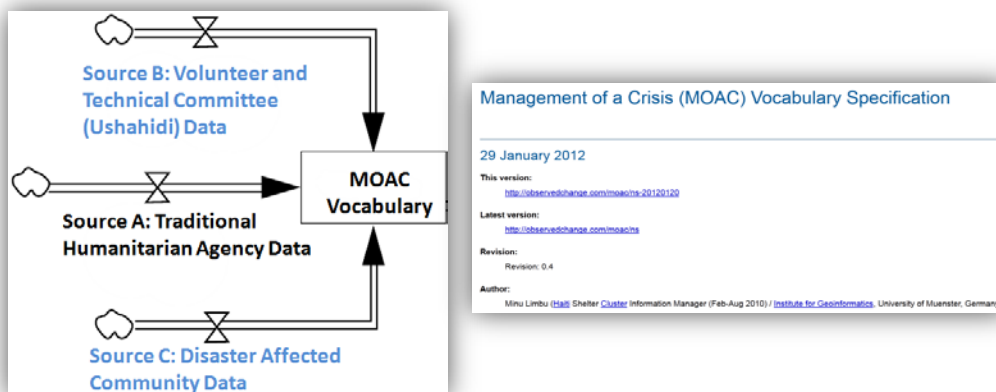


Figure 5-6 Integration of core concept to form a Management of a Crisis Vocabulary

<sup>34</sup>See <http://haiti.ushahidi.com/>

<sup>35</sup>See <http://observedchange.com/moac/ns/>

## 5.7 Discussion: Completeness, Consistency and maintenance

At no point does this research claim that the given set of vocabulary is complete. The study does not portray MOAC to be "the" solution towards semantic interoperability. Having said that, human understandable and computer readable vocabulary is one step closer towards that reality [31, 69]. The value of MOAC is based on the two vocabulary mantra's suggested by Halpin [45]:

- The MOAC's machine (URIs) readable and human readable specification document will remain reliably accessible on ObservedChange.com.
- The URIs will not be sold or repurposed.

The humanitarian communities see the importance [47, 117] of integrating collective efforts from volunteering and crowdsourcing technologies for collaborative intelligence [93]. At the same time, disaster information managers also show positive affinity towards using crowd contents provided the volunteers are trained in managing data for noise reduction. Just as no two businesses are the same, no two disasters are the same. Also, some 20,000 Ushahidi maps have been launched in more than 130 countries since 2008 [81]. This leaves the possibility for the vocabulary as a continuously evolving process. With each version, experts from the humanitarian field and crowd technologists are providing additional inputs for the needed or missing concepts. Based on disaster experts' use of MOAC, this leaves rooms for adjustments and improvements for the terms and concepts that this research might have missed.

## 2. The Management of a Crisis Vocabulary at a glance

An alphabetical index of terms from the Management of a Crisis Vocabulary, by class and by property, are given below. All the terms are hyperlinked to their detailed description for quick reference.

Classes: | [AffectedPopulation](#) | [Agency](#) | [AidManipulation](#) | [AnimalIllnessOrDeath](#) | [AskingToFowardAMessage](#) | [ChronicCareNeeds](#) | [CollapsedStructure](#) | [CommunicationLinesDown](#) | [CompromisedBridge](#) | [ContaminatedWater](#) | [ContaminatedWaterSupply](#) | [Deaths](#) | [EarthquakeAndAftershock](#) | [Emergency](#) | [FeedingCentersAvailable](#) | [FinancialServicesAvailable](#) | [Fire](#) | [Floods](#) | [FoodDistributionPoint](#) | [FoodShortage](#) | [FuelShortage](#) | [GroupViolence](#) | [HealthOfWomen](#) | [HighlyVulnerable](#) | [HospitalOperating](#) | [HumanRemainsManagement](#) | [IDPConcentration](#) | [Incident](#) | [InfectiousHumanDisease](#) | [InfrastructureDamage](#) | [InternetAccess](#) | [Landslides](#) | [Looting](#) | [MedicalEmergency](#) | [MedicalEquipmentAndSupplyNeeds](#) | [Menaces](#) | [MissingPersons](#) | [NaturalHazards](#) | [NonfoodAidDistributionPoint](#) | [Others](#) | [PeopleTrapped](#) | [PersonsNews](#) | [PortOpen](#) | [PowerOutage](#) | [PriceGouging](#) | [PsychiatricNeed](#) | [PublicHealth](#) | [Riot](#) | [RoadBlocked](#) | [RubbleRemoval](#) | [SearchAndRescue](#) | [SecurityConcern](#) | [ServiceAvailable](#) | [ShelterNeeded](#) | [ShelterOffered](#) | [TheftOfAid](#) | [UnstableStructure](#) | [UshahidiReport](#) | [VitalLines](#) | [WaterSanitationAndHygienePromotion](#) | [WaterShortage](#) | [WaterDistributionPoint](#) | [WhoWhatWhere3W](#) | [Who3W](#) | [What3W](#) | [Where3W](#) | [When3W](#) | [NeedsResponse3W](#) | [Response3W](#) | [Needs3W](#) |

Properties: | [affectedby](#) | [approved](#) | [atLocation](#) | [atGeoLocationLatitude](#) | [atGeoLocationLongitude](#) | [belongsToCluster](#) | [category](#) | [countOfAffectedPopulation](#) | [countOfAffectedHH](#) | [countInStock](#) | [countInPipeline](#) | [clusterActivity](#) | [date](#) | [dateOfDistbn](#) | [distributingPartner](#) | [description](#) | [has3Wcomponent](#) | [hasOrganisationID](#) | [hasShelterItems](#) | [hasNFItems](#) | [involvedIn](#) | [impact](#) | [location](#) | [NFNeededAt](#) | [organizationName](#) | [supplyingAgency](#) | [shelterOfferedAt](#) | [shelterNeededAt](#) | [tentDistbn](#) | [title](#) | [verified](#) |

Figure 5-7 A Section of Management of a Crisis Vocabulary Version 0.4  
(source:ObservedChange.com/moac/ns)

## 6. Implementation of Linked Data for crowdsourced and traditional disaster information

This section checks the implementation of MOAC - management of a crisis vocabulary by linking traditional humanitarian agency data with crowdsourced volunteered technical committee data. As explained in Section 2.5, Linked Data is a technique and a set of principles that enables data sharing and reusing data on the web [49]. The web has made it possible for anyone to publish documents on the web and discover it automatically by search engines or by human with browsers. Based on the same principle, Linked Data goes a step further and provides technical solutions to link not only the document but also the content of the document, i.e., data [17–19, 105].

As a part of the third objective of this study, this section defines how Linked Data principles are implemented for formalizing, supporting, connecting between resources explained in MOAC.

### 6.1 *Linked Open data: MOAC semantic annotation*

Tim Berners-Lee in his Web architecture note [14] introduced Linked Data principles as best practices to publish linked data by applying the general architecture of world wide web

1. Use URIs (uniform resource identifier) as names for things.
2. Use HTTP URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

RDF or Resource Description Framework provides a structure to describe different "things". These things can be anything that one can identify as a unique entity, also commonly known as resources. Since the concept of linked data is to connect data within a document, every possible identifiable or separable object that is necessary for the given domain can be regarded as a resource. Hence in order to make meaningful statements in RDF, the thing we are talking about has to be identified in some unique way. For example the need for shelter can be identified uniquely as:

*<http://observedchange.com/moac/ns#ShelterNeeded>* which is defined in the MOAC vocabulary as

#### **Class: ShelterNeeded**

ShelterNeeded is a term used to indicate number of emergency life saving basic structures like tents and tarpaulins.

**identifier:** <http://observedchange.com/moac/ns#ShelterNeeded>

**Subclass of:** [VitalLines](#)

Such a naming convention is called URI (uniform resource identifiers). It looks like a URL but it may or may not represent an actual web page.

RDFs are composed of three basic elements or resources: 1) the things being described, properties, 2) the relationships between things and classes, and 3) the grouping of things or resources. The combination of the three elements make simple statements in the form of Triples.

The RDF triple contains three<sup>36</sup> components the **subject** which is an RDF URI reference or a blank node; the **predicate**, which is an RDF URI reference; the **object**, which is an RDF URI reference, a literal or a blank node. An RDF triple is conveniently written in the order subject, predicate, object. In MOAC, the message for shelter needs which loosely translates into "Certain People are affected by Earthquake and aftershock (in location X) " can be written in RDF in the following format.

```
<http://observedchange.com/moac/ns#AffectedPopulation>      SUBJECT
  <http://observedchange.com/moac/ns#affectedby>             PREDICATE
  <http://observedchange.com/moac/ns#EarthquakeAndAftershock> OBJECT
```

The predicates which are used to connect subjects and objects in MOAC are also known as RDF Properties. They are type, subclassOf, subPropertyOf, range, domain, label, comment. Type defines a resource that belongs to a certain class. SubClassOf is a class that belongs to a parent class. SubPropertyOf is a property that has a parent property. Range and domain are used with MOAC to express which of the indicated resources have used the given property. In this case, for moac:location (predicate), moac:AffectedPopulation is the domain and moac:EarthquakeAndAftershock is the range. In MOAC, the name of the resources are labelled using the property label and comments are used to provide additional information about the resource. The turtle notation for a section of RDF triples looks this way:

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix moac: <http://observedchange.com/moac/ns#> . //link to MOAC vocabulary

moac:AffectedPopulation moac:affectedby moac:EarthquakeAndAftershock. //Statement

# two classes AffectedPopulation and EarthquakeAndAftershock
moac:AffectedPopulation
  a rdfs:Class ;
  rdfs:label "People Affected by a crisis" ;
  rdfs:comment "AffectedPopulation is a crowd of people affected by disaster" .

moac:EarthquakeAndAftershock
  a rdfs:Class ;
  rdfs:label "Earthquake Aftershock" ;
  rdfs:comment "An earthquake is the result of a sudden release of energy in the Earth's crust that creates seismic waves, and an aftershock is an earthquake that occurs after a previous earthquake, the mainshock." .

#Property
moac:affectedby
  a rdf:Property;
  rdfs:label "people affected by" ;
  rdfs:comment "affectedby indicates disaster which has influence on population. " ;
  rdfs:domain moac:AffectedPopulation ;
  rdfs:range moac:NaturalHazard.
```

<sup>36</sup>See <http://www.w3.org/RDF/>

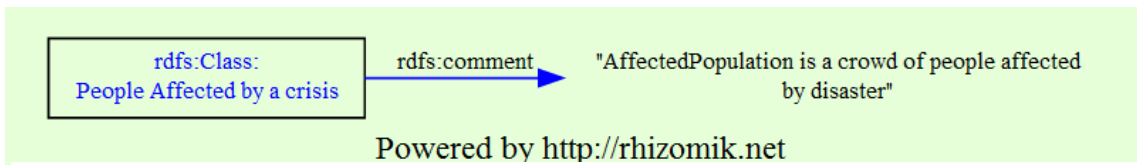


Figure 6-1 RDF Graph component visualized using Inspector

The detailed RDF Graph can be accessed from <http://xrl.us/bmqb8y> (temporary)

### 6.2 Classes and subclasses: Who what where? Humanitarian Needs and Gaps Analysis

Based on the 3W conceptualization, there are four core concepts which are turned into 4 separate classes using the RDF framework as shown in the red box in figure 6-2. As shown in the black rectangle the shelter distributing agency, which is also a subclass of Shelter Cluster is also the subclass of class:Who. This means that the shelter distribution agency belongs to a parent class called who where all the properties are also inherited and new properties specific to Shelter distribution agency can also be drawn. Similarly, the four classes in the red box are also subclasses of the Who What Where class known as 3W (as shown in who what where vocabulary conceptualization state).

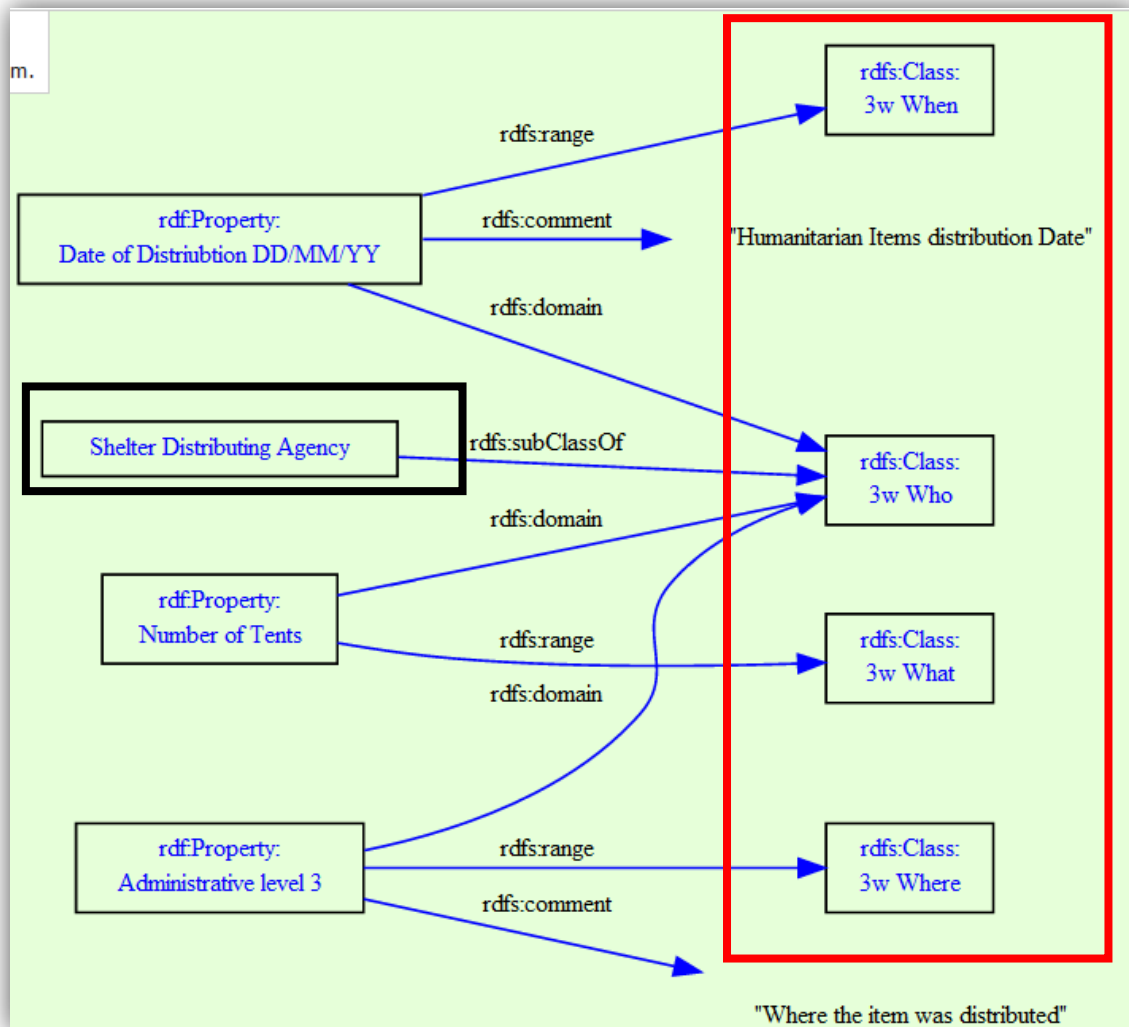


Figure 6-2 Who What Where Class with RDF Framework

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix moac: <http://observedchange.com/moac/ns#> . //link to MOAC vocabulary

moac:Who3w                                     //Class
  a rdfs:Class ;
  rdfs:label "3w Who" .
# rdfs:comment "Individually named Person, agency" .

moac:ShelterAgency
  rdfs:subClassOf moac:Who3w ;
  rdfs:label "Shelter Distributing Agency" .
# rdfs:comment "Refer to Shelter Cluster" .

moac:What3w
  a rdfs:Class ;
  rdfs:label "3w What" .
# rdfs:comment "Individually named events" .

moac:Where3w
  a rdfs:Class ;
  rdfs:label "3w Where" .
# rdfs:comment "Individually named geographic locations" .

##Property##
moac:tent                                     //Property from traditional humanitarian agency
  a rdf:Property;
  rdfs:label "Number of Tents" ;
# rdfs:comment "Tents provided on a given date by a given agency (refer to shelter guidelines)" ;
  rdfs:domain moac:Who3w ;
  rdfs:range moac:What3w .

moac:dateOfDistbn    // Refers to When?
  a rdf:Property;
  rdfs:label "Date of Distriubtion DD/MM/YY" ;
  rdfs:comment "Humanitarian Items distribution Date" ;
  rdfs:domain moac:Who3w ;
  rdfs:range moac:When3w.

moac:locationCommune //Refers to a Location
  a rdf:Property;
  rdfs:label "Administrative level 3" ;
  rdfs:comment "Where the item was distributed" ;
  rdfs:domain moac:Who3w ;
  rdfs:range moac:Where3w .
  rdfs:range moac:NaturalHazard.

Detailed turtle serialization for who what where specification can be accessed from:
http://xrl.us/bmqcey (temporary)

```

While trying to reach as many people as possible that have been affected by a disaster, it is extremely important to find the most correct estimation possible concerning humanitarian response, needs and underlying "gaps". Based on the vocabulary identification process, UNOCHA recommended the use of terms like moac:Needs3W which roughly translates into existing request for humanitarian assistance. Similarly the term moac:Response3W is a class that contains information about all the existing humanitarian responses by agency, location and time. For further analysis in latter sections ShelterResponse and ShelterNeeded URIs are used. Both these URIs subclasses of Needs and Response super classes.

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix moac: <http://observedchange.com/moac/ns#> .

moac:WhoWhatWhere a rdfs:Class . // Super Class Who What Where?

moac:Who3w a rdfs:Class; // Classes characterization
    rdfs:subClassOf moac:WhoWhatWhere.
moac:What3w a rdfs:Class;
    rdfs:subClassOf moac:WhoWhatWhere.
moac:Where3w a rdfs:Class;
    rdfs:subClassOf moac:WhoWhatWhere.
moac:When3w a rdfs:Class ;
    rdfs:subClassOf moac:WhoWhatWhere.

##Property explaining how Hunaitarian Needs is connected to the super class Who What Where and When

moac:Response3W
a rdf:Property;
    rdfs:domain moac:Who3w ;
    rdfs:range moac:When3w.

moac:Needs3W
a rdf:Property;
    rdfs:domain moac:Who3w ;
    rdfs:range moac:What3w

Detailed RDF Graph: http://tinyurl.com/7m8rco8
    
```

From the RDF graph and the turtle serialization given above, response and needs are properties that are attached to, classes like who, what, where and when. Such linkages across classes, properties and the creation of domain and range turned out to be helpful when trying to calculate the estimation of needs and gaps responses. When data from crowds or the VCT community is also inserted using the same set of vocabularies then such information integration across the

web of data or also within the humanitarian agency will open new data sources rather than relying on the specific agency system. When more classes are associated with the super class 3W, it links a broad range of humanitarian concerns that cut across various clusters like Shelter, Food, and health. This is a positive duplication to widen the reach of future information searches.

In MOAC, a number of similar conceptual associations are created as part of super class moac:WhoWhatWhere. Classes like emergencies, Cluster Agency, Incidents, and Locations belong to one of the parent classes like - "who, what, where and when" which are also sub classes of moac:WhoWhatWhere. From such associations, it can be concluded that the conceptual terms related to identity, location and time, i.e., "Who is doing what activities Where and when?" is one of the most mapped resources. As a result, it also fulfils the requirement of a commonly accepted URIs [49] that can be used across different disasters.

### 6.3 *Linked Data Principle: An Approach for connecting data for disaster management*

The primary reason for the development of the MOAC vocabulary is to show that various concepts and terms extracted from the disaster response agencies can be used to link data between traditional humanitarian agencies like the United Nations and the Red Cross with the data collected by crowdsourced volunteered technical committees (VCT).

Based on Tim Berners-Lee's best practices for publishing Linked Data, the research acknowledges that MOAC vocabulary fulfills all four of the fundamental Linked Data principles [14]. Therefore, Management of a Crisis Vocabulary is:

- online (version 0.2) and it can be accessed from ObservedChange<sup>37</sup> [65],
- The XML Namespace URI that must be used by implementations of this specification is: <http://observedchange.com/moac/ns#>, and
- The suggested prefix for the namespace is "moac" to be used by implementations of specification: @prefix [moac: http://observedchange.com/moac/ns#](http://observedchange.com/moac/ns#)

### 6.4 *Haiti Earthquake Response 2010: Linking Data sources*

This research tests integration of unstructured, crowdsourced VGI data and traditional humanitarian response data gathered from the Haiti earthquake in 2010. The two sources of data from the Haiti earthquake disaster chosen for the tests are:

#### 1. **Raw data source 1:Ushahidi Haiti Project<sup>38</sup> Data**

- File format: CSV<sup>39</sup>.
- File received from: Mr. Patrick Myer<sup>40</sup>- Ushahidi Director of Crisis Mapping & Partnerships and Rob Baker, Ushahidi Inc.
- Data contents and sources: It contains volunteered geographic information (VGI) or crowd information collected to facilitate (near real) time crisis mapping using Ushahidi. VGI in Haiti was collected from within Haiti via SMS, web, email, radio, phone, twitter, Facebook, television, list-serves, live streams, and situation reports.

*From this research interest point of view, the file contains vital information about pleas for help from people directly affected by the earthquake since 13 Jan 2010.*

---

<sup>37</sup> Management of a Crisis Vocabulary <http://observedchange.com/moac/ns/>

<sup>38</sup> Website <http://haiti.ushahidi.com/>

<sup>39</sup> See [http://en.wikipedia.org/wiki/Comma-separated\\_values](http://en.wikipedia.org/wiki/Comma-separated_values)

<sup>40</sup> Patrick Myer Bio <http://irevolution.net/bio/>

## 2. Raw data source 2: Emergency Shelter and Non Food Item Cluster humanitarian Response in Haiti

- File format: excel sheets.
- File received from: Mr. Neil Bauman - Global Focal Point for Shelter Coordination/ Information Management, International Federation of Red Cross and Red Crescent Societies
- Data Contents and sources: The files contain details about shelter and Non Food Items distribution (tents, tarpaulins, jerry cans, shelter kits) by agency related to emergency Shelter and Non Food Items Cluster.

*For this research point of view, the files contain information about "Where" a humanitarian agency (Who) provided shelter (What) and related relief support. It also has details about temporal distribution (When).*

Without overlooking human and economical costs of a disaster [59] added with humanitarian coordination chaos and data exhaust during Haiti earthquake response, the choice to use the Ushahidi Haiti Platform and Shelter Cluster traditional reports are based on following observations:

- For the first time [47], people directly affected by the crisis asked for help using social media and widely available mobile technologies.
- With the Haiti earthquake we saw the first wave of innovative technologies concerning how to aggregate, translate, and plot messages with requests for help from the affected people and create near real time maps with the help of volunteers around the world.
- From the humanitarian response perspective, the immediate priority after the Haiti earthquake was to provide emergency shelters [60] to over 1.5 million affected people.

### 6.5 Triplification of Raw Data:

Linked Data is about using web techniques to share and interconnect not only the documents but also the data within the documents by following simple 'rules' for publishing data on the Web outlined by Berners-Lee [14]. One of the first steps in the process is triplification.

*Triplification (noun) or triplify (verb) [8, 31, 32, 91] is the process of converting raw unstructured data into RDF statements using available vocabularies (in this case Management of a crisis Vocabulary) and publishing the data on the Web as Linked Data. Triplification in a way breaks down an entire document into data that can be related to a specific "thing" in the real world. Such a "thing" is defined as a unique URIs in MOAC vocabulary which is used to encode in the form of <subject> <predicate> <object> also known as RDF-triples[11].*

As explained in the literature review, considerable amount of information shared during disasters are either through hard copies, excel files or CSV files. In the case of Haiti Data, raw data sources 1 and 2 from section 6.4 are identified, accessed and linked through URIs as explained in section 6.1. As discussed and tested in [94], a Java program based on JENA<sup>41</sup> was used to convert the files into RDFs.

---

<sup>41</sup> See <http://jena.sourceforge.net/>

**6.5.1 Shelter Cluster traditional reports**

Emergency Shelter Cluster Information Managers during the Haiti earthquake were responsible to collect, process, analyse and disseminate (also known as information cycle, described section 2.2) shelter needs and agency response information during Haiti humanitarian response [61]. With the help of the coordination team within and outside the capital city of Port Au Prince, the information managers gathered humanitarian needs and response related information from over 80[24] shelter partners on a daily basis. After rigorous, at times manual, time-consuming data mining of qualitative and quantitative reports, (such as assessments, emails, situation reports, needs and response reporting forms) the data were entered into an excel based data entry form.

In order to meet the shelter and non food items needs of the people affected by Haiti earthquake, some of the primary information collected and entered into the excel sheet were shelter relief items distribution by different locations of agencies throughout Haiti, the GPS coordinates of the distribution points and the date of distribution. For the triplification processes, since there are no existing unique identifiers for such "things", these terms are described and coded into unique identifications as described in section 5.4.

Out of over 550<sup>42</sup> shelter response distribution points recorded in the raw data, the following triplification shows how a single report of shelter items distribution, Care International, is encoded into the building block of the data web using the linked data approach and the vocabulary this research has contributed.

The content of Care International can be as simple as:

*"Care International distributed 400 Tents to 400 households in Leogane at the given GPS location (18.5235,-72.5746) on 17 Feb 2010."* (Raw data source 2: Emergency Shelter and Non Food Item Cluster humanitarian Response in Haiti, section 6.4)

The message can be broken down into different components as shown below:

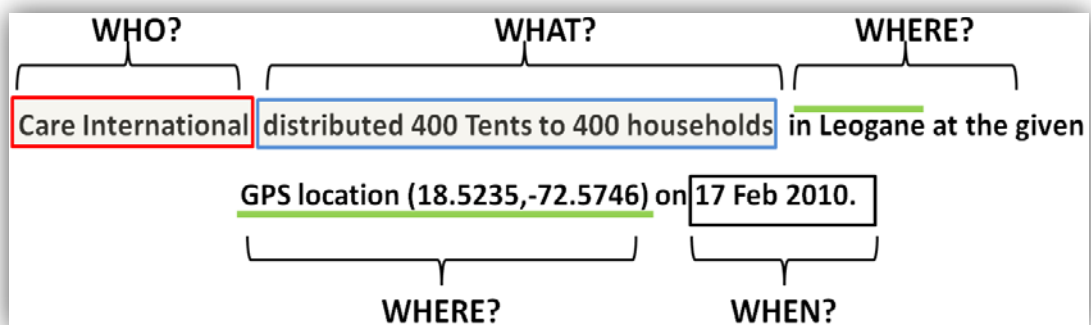


Figure 6-3 Who What Where and When (4W) view of a single Shelter Response report

<sup>42</sup> NOTE: The reported distribution points are based on the available raw data from January to April 2010.

A single report from Care International is triplified using the following URIs while creating RDF relations:

<i>Things or Concepts</i>	<i>MOAC Vocabulary and Other existing Vocabulary Unique URIs</i>	<i>WHO, WHAT WHERE and WHEN?</i>
Reporting Agency	<http://observedchange.com/moac/ns#Agency>	Who?
Report ID	<http://purl.org/dc/elements/1.1/identifier>	Who?
Emergency IASC Cluster	<http://observedchange.com/moac/ns#IASCCluster>	What?
<b>Shelter Offered</b>	< <b>http://observedchange.com/moac/ns#ShelterOffered</b> >	<b>What?</b>
Country	<http://www.loc.gov/marc/countries/Country>	Where?
Location Activity	<http://purl.org/dc/elements/1.1/coverage>	Where?
Latitude and Longitude	<http://observedchange.com/moac/ns#latlon>	Where?
Latitude	<http://www.w3.org/2003/01/geo/wgs84_pos#latitude>	Where?
Longitude	<http://www.w3.org/2003/01/geo/wgs84_pos#longitude>	Where?
Date of Distribution	<http://purl.org/dc/elements/1.1/date>	When?
Information details:	<http://rdfs.org/sioc/ns#content>	General
Title of a Report	<http://purl.org/dc/elements/1.1/title>	General
Information verification	<http://observedchange.com/moac/ns#verified>	General

Table 3 MOAC and other URIs used in a single Care International message

The complete RDF triples of the single message are given in Appendix 1. The reduced version of the same is given below:

```

<http://http://ifgibox.de/m_limb03/moac/L425> http://purl.org/dc/elements/1.1/identifier
"L425".
<http://http://ifgibox.de/m_limb03/moac/L425>
<http://observedchange.com/moac/ns#Agency>
"CARE"^^<http://www.w3.org/2001/XMLSchema#string> .
    <http://http://ifgibox.de/m_limb03/moac/L425>
    <http://purl.org/dc/elements/1.1/subject>
    <http://observedchange.com/moac/ns#ShelterOffered> .
<http://http://ifgibox.de/m_limb03/moac/L425> <http://purl.org/dc/elements/1.1/date>
"2010-02-17T00:00:00-05:00"^^<http://www.w3.org/2001/XMLSchema#dateTime> .
    <http://http://ifgibox.de/m_limb03/moac/L425>
    <http://observedchange.com/moac/ns#latlon> "18.5235,-
72.57463333"^^<http://www.w3.org/2001/XMLSchema#string>.
    
```

6.5.2 *Ushahidi Haiti Project reports*

The Ushahidi platform was used in Haiti as a crowdsourced (almost) real time mapping of pleas received from affected communities. Unlike traditional humanitarian agencies, a single Ushahidi incident report can contain more than one type of request or plea from people directly affected by the earthquake. For example, Ushahidi report number 2307 contains 5 different type of incidents which are categorised under the "category" column as shown in table 4.

#	INCIDENT	INCIDENT DATE	LOCATION	DESCRIPTION	CATEGORY	LATITUDE	LONGITUDE	APPROVED	VERIFIED
2307	30 people r	2/1/2010 11:55	59 rte de Dupuy,	we(30p) need food drink diapers milk medicine wather.. Adress: 59 rte de Dupuy, Thiotte. Sud-Est, HAITI	2b. Penurie d'eau   Water shortage, 2a. Penurie d'aliments   Food Shortage, 3c. Besoins en materiels et medicaments   Medical equipment and supply needs, 2d. Refuge   Shelter needed, 7c. Denrees non alimentaires	18.2282	-71.8343	YES	NO

Table 4 Ushahidi Haiti Incident Reports Raw Data, incident report number 2307

After going through the individual message the categories are defined by the Ushahidi Haiti team. In this case the message received is categorised as:

*2b. Penurie d'eau | Water shortage, 2a. Penurie d'aliments | Food Shortage, 3c. Besoins en materiels et medicaments | Medical equipment and supply needs, 2d. Refuge | Shelter needed, 7c. Denrees non alimentaires | Non-food aid distribution point*

All the categories identified for Ushahidi Haiti crisis mapping are given unique URIs in Management of Crisis Vocabulary (MOAC). Using RDF and MOAC, the message received directly from the affected population is then related to relevant categories as shown below.

**Portion of RDF Triples for Ushahidi Incident Report 2307**

```
<http://haiti.ushahidi.com/reports/view/2307> <http://purl.org/dc/elements/1.1/subject>
<http://observedchange.com/moac/ns#MedicalEquipmentAndSupplyNeeds> .

<http://haiti.ushahidi.com/reports/view/2307> <http://purl.org/dc/elements/1.1/subject>
<http://observedchange.com/moac/ns#ShelterNeeded> .

<http://haiti.ushahidi.com/reports/view/2307> <http://purl.org/dc/elements/1.1/subject>
<http://observedchange.com/moac/ns#FoodShortage> .

<http://haiti.ushahidi.com/reports/view/2307> <http://purl.org/dc/elements/1.1/subject>
<http://observedchange.com/moac/ns#NonfoodAidDistributionPoint> .

<http://haiti.ushahidi.com/reports/view/2307> <http://purl.org/dc/elements/1.1/subject>
<http://observedchange.com/moac/ns#WaterShortage> .
```

### 6.6 Queries and Visual Interpretation of integrated data

Rather than simply connecting the CSV and excel sheets between traditional humanitarian agency and the Ushahidi Haiti reports, Linked Data uses RDF to make typed statements that link relevant arbitrary "things" in the world. The data (not the document) about a humanitarian agency like Care International's response and the pleas (not only the statements but the data within statements) for help from affected people are linked using Linked Data standards. This creates what is known as Web of Data or web of things described in the real world by data on the web [19].

Converting documents into RDF triples is important (objectives 2 and 3) but it is only an intermediate step. The most vital aspect of the analysis is what it that the prospective users want out of the system (Objective 1). As a next step, this section attempts to visualize the integrated RDF triples of affected community pleas for "shelter needs" mapped by Ushahidi and reports of "shelter relief items offered" by traditional humanitarian agencies during Haiti response. A website<sup>43</sup> is set up that uses a SIMILE<sup>44</sup> Exhibit for visual demonstration of the Haiti earthquake shelter needs and response.

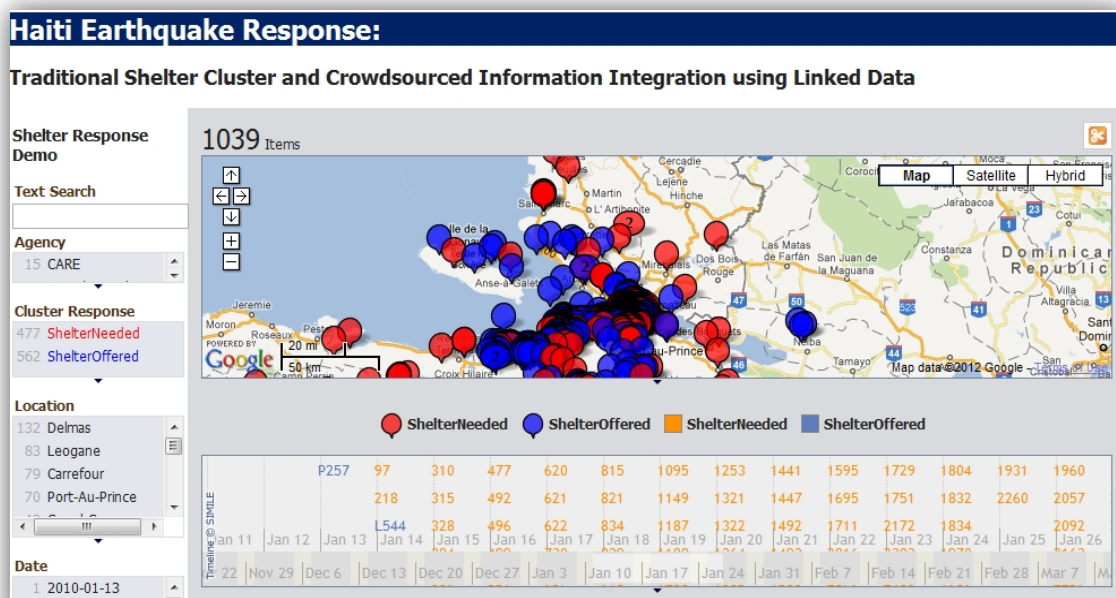


Figure 6-4 Demonstration of Haiti Earthquake ShelterNeeds (Affected Population Plea) and ShelterOffered (Humanitarian Agency Response) using SIMILE Exhibit

The website provides user-friendly interface to perform a web based search across URIs described in Management of a Crisis Vocabulary. On the left pane of the website, there are sections like Agency (Who?), Shelter Cluster Response (What?), Location (Where?) and Time (When?). The website also has a timeline at the bottom that illustrates temporal distribution of incidents reports from the first day after the Haiti disaster i.e. 13 January 2010. Using these parameters, the user can find answers to fundamental humanitarian queries like:

- Humanitarian Needs: What are their most urgent needs of the people affected by a crisis?
- Humanitarian Response: Which organizations (Who) are carrying out what activities (What) in which locations (Where) and when ?
- Humanitarian needs and gaps: What are the major humanitarian response gaps in terms of needs of the affected Population?

<sup>43</sup> See [http://ifgibox.de/m\\_limb03/minu-haiti-response/HTMLPage1.htm](http://ifgibox.de/m_limb03/minu-haiti-response/HTMLPage1.htm)

<sup>44</sup> see <http://www.simile-widgets.org/exhibit/>

## 7. Evaluation

The main accomplishment of this research is the compilation of 'Management of a Crisis Vocabulary (MOAC)' comprising unique URIs of terms used in a humanitarian crisis. This section discusses the evaluation of MOAC, its strengths and limitations as a part of the process *to integrate crowdsourced information with traditional Crisis and Disaster Management Information using Linked Data*.

### 7.1 What is Management of a Crisis Vocabulary Evaluation?

For clarity, the term "evaluation" means to carry out a technical judgment [20, 33, 34] of the MOAC vocabulary. As explained by Gómez-Pérez et al. [34], evaluation subsumes the terms Verification and Validation. "Verification"[36] is the technical process that guarantees the correctness of a set of vocabulary, its associated software environments, and documentation. Similarly, "validation"[33] guarantees that the MOAC vocabulary, the visualization software environment and documentation correspond to the system that they are supposed to represent.

### 7.2 Foundation for Evaluation

As explained throughout the research, MOAC vocabulary is created for the integration of crowdsourced information with traditional crisis and disaster management Information using Linked Data. Based on section 5.2.1, its intended users are traditional humanitarian agencies like United Nations, NGOs and network of crowdsourcing communities scattered through out the world. In that sense, MOAC is at the centre of Knowledge Sharing<sup>45</sup> practices and Knowledge Sharing Technology (KST)[85]. Even though there is no set of concrete guidelines on how to evaluate vocabulary that support KST [36], the MOAC evaluation in this research is based on the framework proposed by Gómez-Pérez [36], Gómez-Pérez et al. [34] and Elhadad et al.[29] in 1994, 1995, 2011 respectively. The authors propose two methods: extrinsic and intrinsic evaluation.

The *intrinsic evaluation* method measures the quality of MOAC vocabulary directly (e.g., structure of vocabulary) which can be evaluated as a standalone body of domain knowledge [87]. *Extrinsic evaluation* method measures how well the MOAC vocabulary helps in performing a particular task, which also defines the context of the evaluation (e.g., identification of initial humanitarian response, needs and gaps).

Even though intrinsic evaluation reflects structural property, the author selects the extrinsic method based on the aims and objectives of this research. The choice to select extrinsic method of evaluation is based on further additional scientific documentations listed below:

- Extrinsic method is a task-based evaluation [100]. This method is also referred to as the evaluation from the end users point of view [34, 82] or "as human task-performance evaluations" ([12], page 1).
- Gómez-Pérez suggested that "there doesn't exist an agreement among the methods to evaluate the results of the evaluation process"([36], page 6), and
- "Interpretation of (intrinsic) measures is eventually task-oriented"([29], page 6, experimental settings) which is again extrinsic.

---

<sup>45</sup> See [http://en.wikipedia.org/wiki/Knowledge\\_sharing](http://en.wikipedia.org/wiki/Knowledge_sharing)

MOAC is end user based and task oriented since it is compiled so that it could be used before, during and after a crisis. Hence, it needs to be evaluated by the prospective users of the vocabulary, i.e., disaster mangers, social media volunteers and Linked Data users.

### 7.3 Requirements for Evaluation

The usefulness of MOAC vocabulary lies in its ability to integrate unstructured data generated by the crowd to that of traditional humanitarian agencies. MOAC must also help in relating messages from two newly discussed sources of information [47]: 1) message from affected population 2) information gathered by global volunteers in almost real time. To determine the usability and continuity of MOAC, the vocabulary proposed by this research should support in relating available humanitarian data and retrieve answers for the following most basic, and yet, crucial information in support of informed decision-making processes:

- Question 1: What are the immediate needs of the people affected by a crisis? What humanitarian relief assistances are currently being provided to meet those needs? [also know as humanitarian needs and response information]
- Question 2: "Who What Where (3W) questions:
  - Which organizations (Who) are carrying out "What" activities in which locations (Where), and if possible
  - When are the humanitarian agencies providing humanitarian support? (4W)
- Question 3. In addition to these questions, MOAC should also help answer 3W and 4W questions related to people affected by a disaster.

To test these observations and questions, validation of MOAC vocabulary is carried out using both qualitative and quantitative research techniques[88]. It is based on carefully selected MOAC evaluators (details in Appendix 3) from three different user groups namely, the domain experts (Humanitarian Information mangers), Linked Data Enthusiasts (Master Degree and PhD researchers) and Social Media Humanitarian Volunteers. The following evaluations are conducted to document concrete feedback on the proposed vocabulary as a part of information integration process:

- Evaluation Scenario 1: Humanitarian "Needs / Response" and "Who What Where (3W)/ When (4W)" queries on integrated crowdsourced Ushahidi Haiti Platform and Shelter Cluster traditional reports from Haiti Earthquake 2010 humanitarian response.
- Evaluation Scenario 2: MOAC Vocabulary functionality and usability survey using questionnaires and peer interviews [107]
- Evaluation Scenario 3: MOAC Vocabulary URI testing with crowd (or in this case linked data experts with no or limited domain knowledge)

The evaluation process also depends on the way vocabulary is constructed. For this research, the evaluation scenarios and criteria are mapped to the following evaluation dimensions [29]:

1. *Functionality (task-based)*: evaluate how well MOAC serves its purpose to support informed decision making process in early stages of a disaster;
2. *Usability*: assess the pragmatic aspects of the vocabulary i.e., MOAC usefulness in support of Linked Data Principle as an approach for connecting data for disaster management and to integrate crowdsourced and traditional information
3. *Structural evaluation*: Consistency, Correctness, Completeness and Continuity of MOAC vocabulary structure.

## 7.4 Evaluation Results

### 7.4.1 Humanitarian "Needs / Response" and "Who What Where (3W)/ When (4W)" queries

The focus of this evaluation is to validate the functionality, usability and structural dimensions of MOAC vocabulary from the end user point of view. Professionals and academics from different backgrounds were requested to participate on query based exploratory search of the implementation of MOAC vocabulary. A set of nine questions<sup>46</sup> related to Emergency Shelter cluster humanitarian needs and response were given to five participants<sup>47</sup>. All of them were asked to perform the necessary queries using shelter response demonstration website discussed in section 6.6.



7-a Linked Data Enthusiasts Evolution Participants

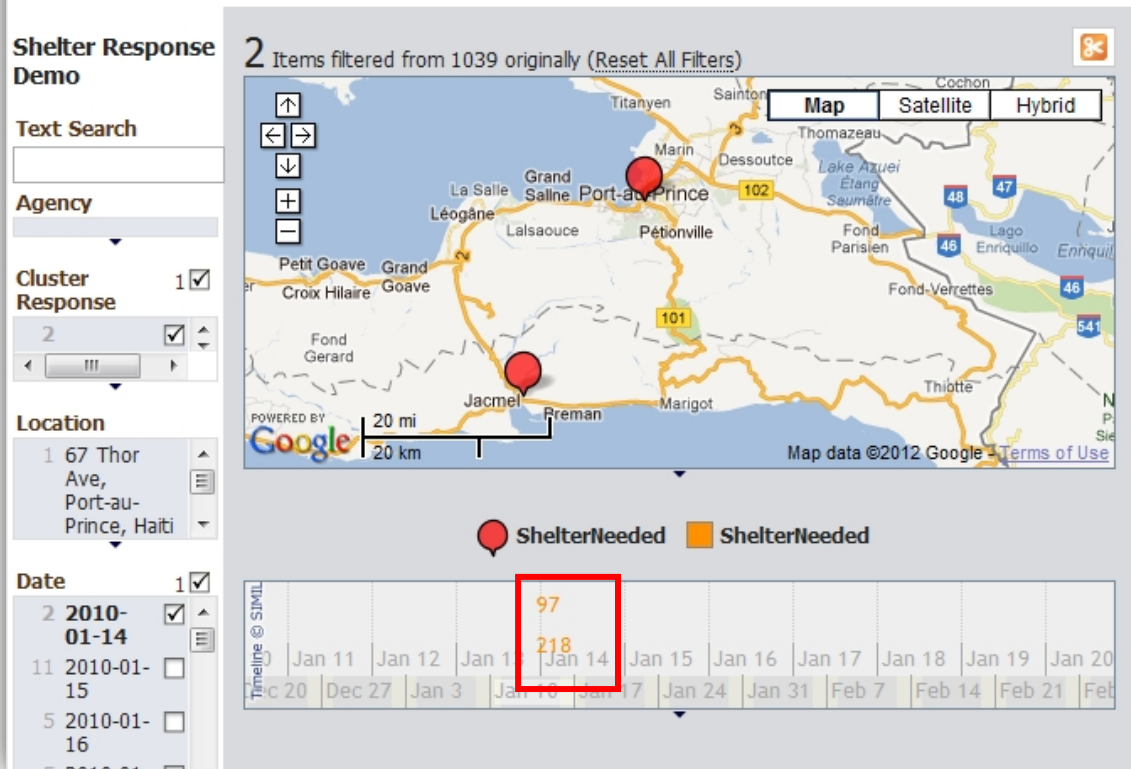
The results (figure 7-1 to 7-4) of the evaluation shows that, with the help of MOAC vocabulary, fundamental questions related to humanitarian needs and response stated in section 7.3 can be answered. For example, evaluators were able to locate immediate shelter needs (survey 1, questions 1, 4, 7, 9 ) that the communities affected by the earthquake were directly communicating as shown in figure 7.1. The figure clearly shows that people started sending pleas for help on the second day of the onset in location like Port Au Prince and Jacmel. Even survey participants with no or little disaster management domain knowledge were able to precisely locate *where* humanitarian relief agencies were providing much needed assistance and when (survey 1, questions 2,3, 6, 8).

As shown in figure 7-2 (survey 1, question 3), evaluators identified humanitarian agency specific shelter items distribution in a certain location. The figure shows the GPS based locations of emergency Shelter and NFI distribution points of a humanitarian agency. Humanitarian response search was further triangulated with time and precise result is shown in figure 7-3 (survey 1, question 2). Furthermore, using the web services, the evaluators also cross referenced to pleas recorded by Ushahidi directly from people in need of shelter items from the same location.

<sup>46</sup> For more details, please refer to Appendix 2: Sample of Linked Data Enthusiasts Survey Questionnaires.

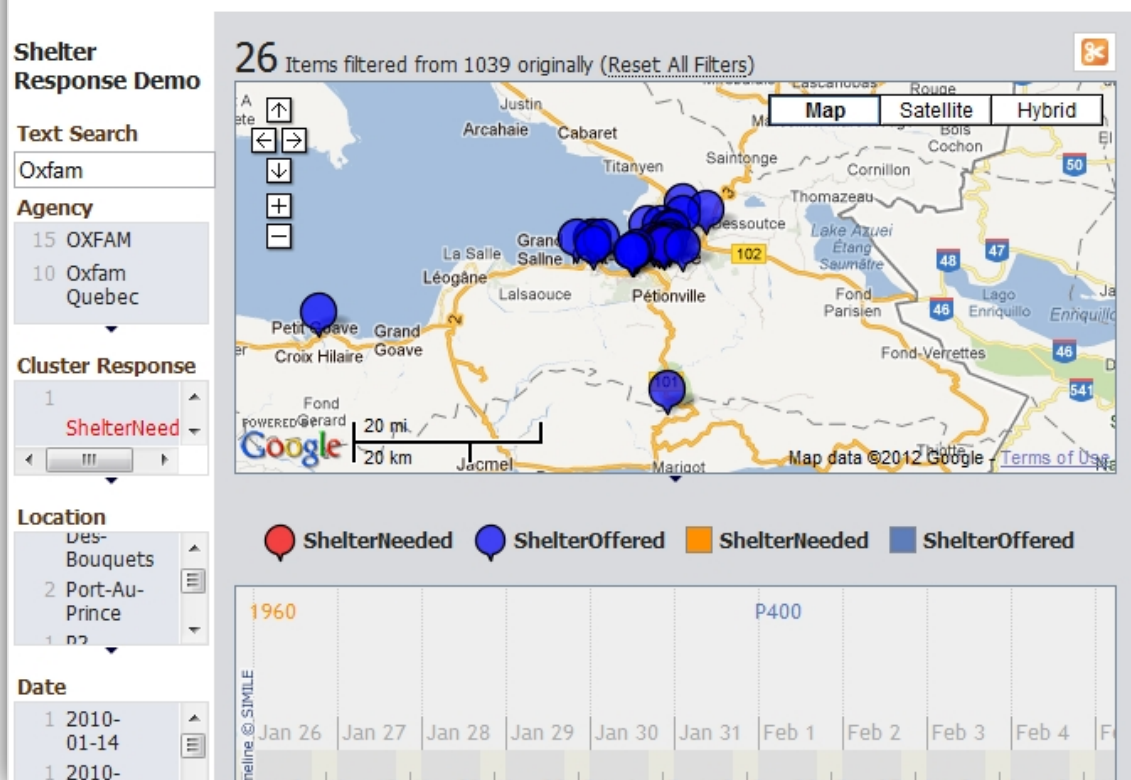
<sup>47</sup>For more details on evaluator profiles, please refer to Appendix 3: Profiles of Evaluators.

### Traditional Shelter Cluster and Crowdsourced Information Integration using Linked Data

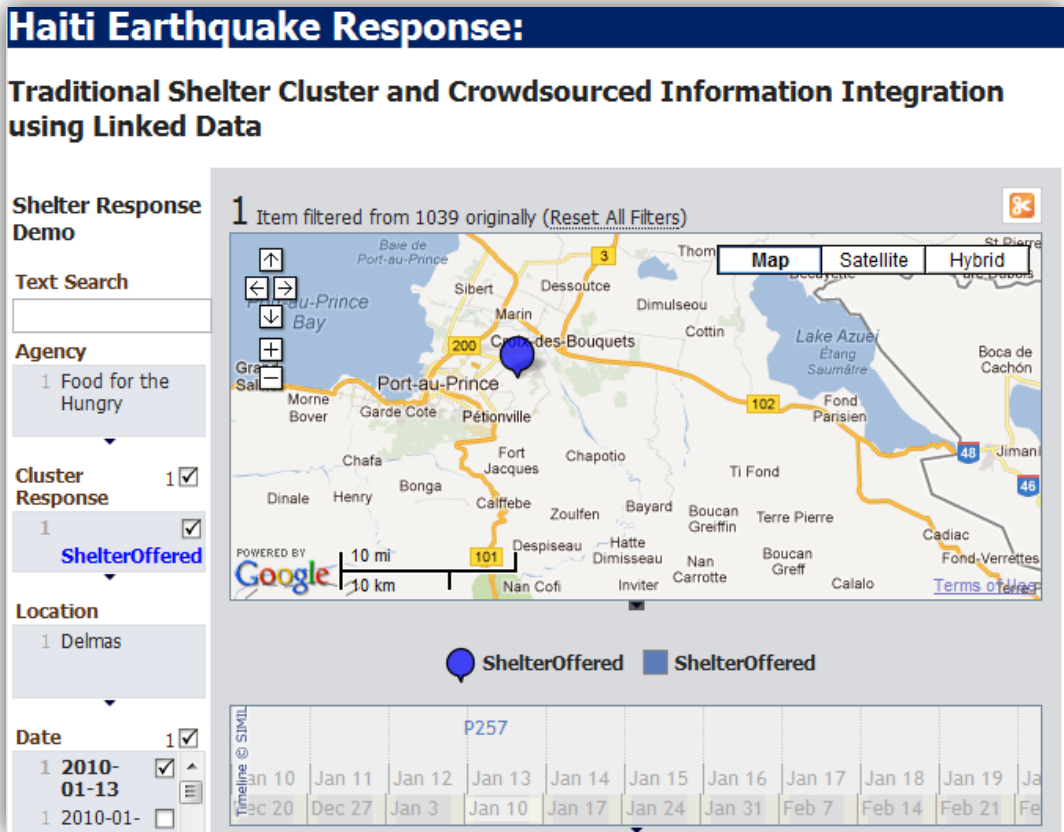


7-1 Evaluators response for survey question "When (date) was the first Emergency shelter Needs reported and where?"

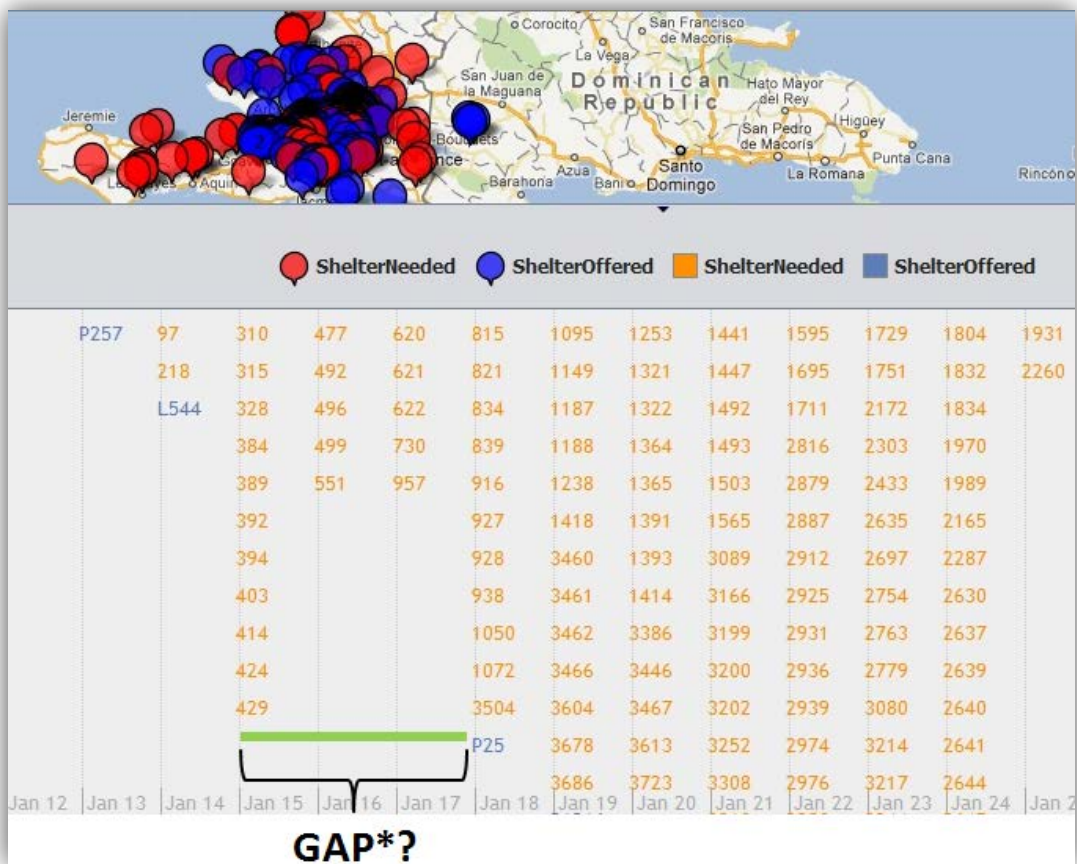
### Traditional Shelter Cluster and Crowdsourced Information Integration using Linked Data



7-1 Evaluators response for survey question "In the given reporting period, please indicate the number of locations where Oxfam (Agency) provided shelter assistance."



7-3 Evaluators response for survey question " Based on the given data, please indicate which was the first agency to provide shelter, where (Location) and when (Time)?"



7-2 Response for survey question "From timeline graph at the bottom of the map, can you indicate how many days there were NO reported shelter offered on the first week of Haiti Disaster i.e. from 12 Jan 2010 - 19 Jan 2010?"

All the survey participants found the visual timeline distribution of people's shelter needs and on ongoing humanitarian response very useful. One of the significant findings of the survey was the use of the timeline graph to determine what are the needs of affected population and where; what humanitarian aids have been provided and where; and most importantly what more needs to be done. The evaluators could visually identify humanitarian needs, response and gaps\* in early stages of a disaster as shown in figure 7-4 (survey 1, question no. 5). For example, the evaluators were able to identify that there were no "reported" shelter distributions for three straight days in Haiti. At the same time there were many pleas for shelter needs as visualized by evaluators in figure 7-4. The evaluators demonstrated that when official data between affected population pleas for shelter needs and humanitarian shelter response are integrated, it could clearly show where the cluster of requests were coming from and where the humanitarian response were concentrated and when. Assuming there were no credible assessment reports on the needs of the people in crisis, such mashups provides a basic, yet fundamental, starting point for the humanitarian rapid assessment team to check the authenticity of those pleas.

With the help of web based visual integration of crowd and traditional humanitarian linked data, the evaluators were able to assess the adequacy of MOAC vocabulary in support of exploratory searches related to humanitarian needs, responses and gaps. The evaluators asserted "Who (Agency), What (Needs/Response), Where (Coverage/Location) and When (Date)" queries using the webpage of *Shelter Response Demonstration* as the front end to access thousands of Ushahidi and Shelter cluster RDF triples linked with MOAC vocabulary on the backend.

### **7.4.2 Domain Expert Evaluation**

The main focus of the expert survey and peer questionnaire evaluation is to check whether MOAC can be used to suit the information integration needs of both the traditional humanitarian agencies and the volunteered crowdsourced platforms. This is based on a problem centred interview [121] with a combination of qualitative and quantitative methods[88]. As shown in Appendix 2 - domain expert survey<sup>49</sup> template, questionnaires are directed to meet the functional quality criteria defined by Gulla et al.[43]:

1. Concept familiarity to evaluate the terminology introduced by the vocabulary is strongly connected to commonly used terms by the user group,
2. Domain volatility to evaluate continuity, maintenance of vocabulary and robustness in case of updates.

The usability of MOAC vocabulary for information integration during, before and after a disaster is also accessed in this section.

Between January - February 2012, depending on the availability of the humanitarian information managers, Skype<sup>50</sup> peer meetings were (re)scheduled with four humanitarian affairs/emergency information managers<sup>51</sup>. One of the criteria was that experts needed to have more than 8 years of experience in complex emergency and were also part of Haiti earthquake response in 2010. The interview in an average took around 1 hour and 20 minutes. None of the experts had experiences with Linked Science or Linked Open Data. Therefore, extra care was taken not to rush the questions on the technicalities of vocabulary. To draw experts into the theme of the research, the interview started in the following manner:

---

\* Humanitarian Gaps =  $\pm$  [Humanitarian Response (Ongoing and Planned) - Humanitarian Needs]

<sup>49</sup> For more details, please refer to Appendix 2: Domain Expert Interview Survey and Questionnaires.

<sup>50</sup> See <http://www.skype.com/>

<sup>51</sup> For more details on evaluator's profiles, please refer to Appendix 3: Profiles of Evaluators: Domain Experts

<i>Expert Interview Questions Sections</i>	<i>Concept Covered in each Section (Humanitarian and non humanitarian)</i>	<i>Time spent</i>
1. Information Management General Practices in an emergency/ crisis	<ul style="list-style-type: none"> <li>Information in early stages of a disaster - what is needed? Sources? Trust of Information Sources?</li> <li>Data: Sources? Formats? Trusts?</li> <li>Information Collection and dissemination practices</li> <li>Raw Data</li> <li>Information Collection interoperability and heterogeneity problems</li> </ul>	15%
2. Two new Sources of information	<ul style="list-style-type: none"> <li>Information from Affected Population</li> <li>Crowd/VGI information, usefulness, trust</li> <li>Meaning of terms and its concepts</li> </ul>	25%
3. Integration of crowdsourced information with tradition crisis and Disaster management Information using Linked Data	<ul style="list-style-type: none"> <li>Importance of integration of Information from explained in (2)</li> <li>Use of crowdsourced contents in the absence of humanitarian traditional assessment in the early stages of a disaster</li> <li>Concept of web of data using Linked Data principles and MOAC</li> <li>Triplified RDF data visualization explained in section 6.5 (similar to Linked Data enthusiasts survey in 7.4.1)</li> <li>Finally - MOAC vocabulary consistency checking</li> </ul>	60%

Table 5 Evaluation Time Distribution across different types of surveys and peer interviews

In terms of *Information Management General Practices* in an emergency all experts were of the opinion that impact of the disaster to the affected population are primary. "(..)population needs and available capacity of the government and existing humanitarian NGOs will shape the humanitarian response for coming few weeks". As a part of information collection cycle<sup>52</sup>, humanitarian information managers collected both qualitative and quantitative data in all possible formats ranging from "assessment reports, emails, word of mouth, meetings" to "excel sheets, reporting formats". Their major concern was on three folds i.e. the unavailability credible data in the early stages of a disaster, interoperability issues and heterogeneity of available data. If the data were to be shared in the right format, the expert who recently completed her mission in Haiti said "(...), it will reduce 100% of my information collection problem and also reduce my workload by 50%".

The conversation was linked to the research when reviewers themselves started sharing problems associated with "meanings" of humanitarian activities. One expert said that "the difference in the understanding and interpretation of indicators( meaning) could change the outlook of the humanitarian response." In a separate interview, this was followed with an example of household and person data. "(..)so imagine one partner in the first week of a disaster reports 3000 people affected and the next day (agency removed) say it's not 3000 people but 3000 household. This is a huge problem as it makes a difference of 3000 - 15000 persons (...)". Another expert said "In a disaster a single line in an excel sheet is somebody or someone's lifeline".

The "Pressure cooker" feeling of need for credible information for making informed decisions was unanimously felt across all evaluators. When the author asked if they would consider using crowdsourced or VGI data in the absence of traditional assessment data, the following positive examples from Haiti were shared:

<sup>52</sup> Explained in section 3.1

- expert A said *"The humanitarian OpenStreetMap during Haiti disaster was instrumental in collecting baseline data required for day to day humanitarian mapping. This was done using global crowd (and VGI)"*
- expert B said *"There is no way out of it, technology is changing and so are the means to collect and share disaster information - crowdsourcing is something new but my office (name removed) is also looking into this (...)"*
- expert C said *"Based on where people were using mobile in Haiti, internally displaced people were mapped. This also gave a movement pattern of how people affected by earthquake were coping..( with the displacement)."*
- expert D said *"A UN agency (name removed) has used crowdsourcing technology to gather human right violations in Africa (particular country name removed) "*

The discussion followed with expert opinions on the two new sources of information: affected population and global volunteers discussed as in the joint UN Foundation, UN OCHA, Harvard Humanitarian Initiative report [47] . In this section, reviewers had to rank the importance of need of affected population (survey 2, question 6a.). *"If 5 is the best (most important), its 6"* clarifying that "need" of people affected by a crisis is the *"holy grail"* and one of the most important pieces of information that *"it cannot be ignored at any cost"*.

Based on the reviewers, best practices of crowdsourcing techniques and judging their sense of importance of information received directly from affected population, the interview was switched to garner their views and ideas about possible information integration between different sources. Unlike Haiti, when asked if they were now ready to integrate crowdsourced information into their traditional well practiced machine of information management system, all but one agreed that time has come to look into ways to make such *"things"* possible *"technically and as a part of behavioural change"*. Having said, cautions were raised on the accuracy, quality and reliability of data. On the other hand, one expert shared that *"every single data in the early stages of the disasters will be used to help make decisions to save people's lives. If I am asked to base by information analysis only on the crowd information, I would be hesitant."* On the contrary, the same person did agree that visualization of crowd information *"..does give you a sense of what is happening on the ground"* . Hence such information can be used to make decisions to humanitarian fact finding teams and follow up with appropriate humanitarian interventions if identified.

As the concluding portion of the evaluation process, the author showed the *Shelter Response visualization demonstration*<sup>53</sup> with integrated crowdsourced and traditional humanitarian Shelter Cluster data. The reviewers agreed that it is useful to see such platforms where the work of integration has already taken place. One reviewer immediately commented that *"#ShelterNeeded"* in the interactive map when analysed with the *"#ShelterOffered"* *"..you could see that shelter distribution in Loegane and Jacmel was much more later", "(...)as Port Au Prince was consuming all the resources (...)"*. One reviewer reported that based on the concentration of "shelter needs" message, you could draw two simple assumptions *"it shows the pattern of how people are moving around. On the other hand, it shows there is a need and a humanitarian assessment team can be sent to verify..(true or not)"*. One reviewer made an argument that it was *"a big mistake"* to concentrate most of the humanitarian response only in Port Au Prince. *"Such tools can certainly help"* in avoiding similar situations in future.

---

<sup>53</sup> Visualization webiste is explained in section 7.6 and tested as a part of evaluation earlier with Linked Data Enthusiasts in section 8.4.1.

After this, a brief explanation was given on the concept of Linked Data principles and how different "things" related to a crisis are used for the visualization process by first creating a unique identity called URIs. The reviewers were finally asked to review the Management of a Crisis Vocabulary and specify, whether MOAC vocabulary:

- Is strongly connected to commonly used terms by the user group.
- Is missing some key concepts related to managing a shelter humanitarian response in early stages

The reviewers shared that the "*listed terms would qualify to describe basic disaster response*" using Linked Data principles. In addition, the following additional suggestions were documented:

- Definitions of certain terms, like "#ShelterNeeded" to be redefined based on global standards.
- Classes like to describe things like "Shelter Items" to be integrated.
- "#ShelterPlanned" to be added as a part of Shelter Cluster humanitarian needs and response plans.
- "#InfrastructureDamage" and its classes to be defined on the basis of Post Disaster Needs Frameworks (PDNA)
- Addition of cluster activities other than shelter cluster
- Expansion of emergency structure

In summary, experts found it very useful to see how "things" described in MOAC vocabulary are being used with Linked Data Principles to integrate information between crowdsourced/VGI data and traditional humanitarian data. The conversion of triplified data into the form of interactive web based interactive visualization on a map gave a unique perspective for fact finding, exploration and comprehensive search. Interestingly, these searches are also the properties of evaluation for search ontology information needs [2, 35]. As humanitarian practitioners are often overburdened during crisis response, one of the first responders of an emergency said that it might be interesting if someone can "*link information*" in real time and "*demonstrate produced knowledge*" using such concepts. This does open avenues to test Linked Data principle and MOAC as a part of information collection process on the onset of a disaster.

Before concluding the interview, all the participants were asked to continue collaborating in the development and maintenance of MOAC.

## 8. Discussions, Limitations, Contributions and Conclusion

### 8.1 Discussions

The broad objective of this study is to examine to what extent Linked Data can support information management in the humanitarian decision-making processes during the early stages of a disaster response. It has been argued that in the early stages of a crisis or a disaster there is no or little reliable information. Without which, decision makers cannot take informed decisions on how best to support affected communities. This creates an awkward "information black hole" which could result in uncoordinated humanitarian response and at times death of people in need of help. At the same time, academic and humanitarian lessons learned evaluations from the Haiti earthquake in 2010 show that people affected by the earthquake sent pleas for help using mobile technologies and social media from the very first day of the onset. On the other hand, the global social media and crowdsourced/VGI communities collated these messages and converted them into real-time interactive crisis maps. It is now confirmed that most of the traditional humanitarian agencies during Haiti were either reluctant or not equipped to fully utilize these two new powerful information sources. Based on discussions with over 40 humanitarian information managers and crismappers who worked during the Haiti crisis, this study re-confirms that *trust and quality of crowd information, standardization and data integration problem, information overload, heterogeneous system and incompatible data formats and semantic problems* are some of the main concerns associated with crowdsourced/VGI information. This research proposes Linked Data Principles as one of the open source solutions to help mitigate some, if not all of these issues.

From the disaster management expert survey (discussed in section 4 and 7.4.2), it is evident that information received directly from the affected community is one of the gold mines in an emergency that is necessary to figure out the actual needs, response and existing humanitarian gaps. Which brings us to the question: why is it that there are hesitations to use or integrate crowdsourced or VGI data even when the information is from genuine disaster affected communities? This study therefore arises out of concerns to understand *how to integrate crowdsourced information with traditional Crisis and Disaster Management Information using Linked Data*. The specific objectives of this study were outlined as follows:

- 1.1 To determine if Linked Data and crowdsourcing can support informed decision-making processes
- 1.2 To investigate if there are existing, ready-to-use linked data versions of crisis management vocabulary to interlink data across traditional humanitarian agencies, crowdsourced volunteered technical committees and affected populations
2. To investigate if Linked Data principles with the help of Management of a Crisis Vocabulary can link data within the document of unstructured crowdsourced messages and heterogeneous traditional humanitarian agencies reports.
3. To assess if Linked Data can help convert crowdsourced and traditional humanitarian organizations information into 5 star Linked Open Data.

Linked Data is the principle used to publish data in the world wide web using relation between "things". To use it for data conversion or integration in the web, there is a need for vocabulary describing "things" like people, places or projects which has now emerged in the Linked Data community. To avoid heterogeneity, the study first investigates if "things" related to

crowdsourcing and traditional humanitarian crisis already exists in the form of a vocabulary. After determining that existing vocabularies do not provide sufficient terms that are needed to publish the complete contents related to this study (objective 1.2), the required terms are described as Management of a Crisis Vocabulary (MOAC). As discussed in section 6, the MOAC is the combination of unique URIs (a thing) of classes and properties related to terms used in Ushahidi Haiti Platform, Humanitarian Relief Response, Inter Agency Standing Committee and Emergency Shelter Cluster. This step is the first milestone of the research.

As a part of objective number 2 and 3, the vocabulary functionality and usability is tested while converting and connecting official data received from two authentic sources: 1) compilation of people's pleas for help compiled by Ushahidi Haiti Crisismappers, and 2) humanitarian shelter relief items distribution data maintained by Haiti Emergency Shelter cluster coordination. With the help of the triplification process explained in section 6.5, the data within the documents are triplified into thousands of machine-readable RDF triples of subject, predicate, object creating the web of data. URIs from the MOAC vocabulary (like #SheleterNeeded, #ShelterOffered) are used to link terms between two data sources. Based on Tim Berners-Lee's best practices to publish linked data, the research acknowledges that MOAC vocabulary helps to fulfill all the four fundamental link data principles[14] that are primary to publish Linked Data on the web. MOAC vocabulary provides a solid database and historical documentation of the URIs (principle number 1). All the URIs are in the form of HTTP URIs. Therefore, anyone who wants to check the underlying meaning of the concepts will have not only the meaning but also the sources of those information (principle number 2). This also makes MOAC machine readable using URIs and human readable using the links. MOAC is also based on the RDF framework (principle number 3). It also has considerable amount of links integrated into vocabulary using properties like OWL:SameAs and OWL:seeAlso. Since Linked Data advocates for the use of already existing vocabulary, the triples also use a number of well know vocabulary like Dublin Core Metadata Initiative (DCMI) Metadata Terms, Friend-of-a-Friend (FOAF), to name few. This also helps MOAC meet the fourth principle.

The web of data is used to create an interactive web application to relate people's needs and humanitarian relief response. The search could be triangulated by agency, by humanitarian activities, by location and by date (Who What Where and When). Among various available "things" described in the vocabulary, MOAC URIs like #ShelterNeeded is used to directly relate pleas from affected population to ongoing humanitarian intervention (#ShelterNeeded) by location and time. The precision of the MOAC based search can be compared with google search. The result of the question like "Which was the first agency to provide shelter relief items in Haiti?" in google will return over 1.3 million pages. However, with the help of RDF graph, one could not only precisely locate the agency as shown in figure 7-3, it will also provide other related information about Who, What, Where and When? Such humanitarian activity related queries and searches directly supports objective 1.1 i.e. support decision making process in a disaster.



Figure 8-1 Example of Google search to locate the name of the agency who might have the first agency to provide emergency Shelter Response during Haiti 2010 Relief Efforts

To further validate the notion that Linked Data and Crowdsourced information can help facilitate humanitarian emergency decision making process, the visualization of Shelter Response activities is used to test the authenticity functionality, usability and structural evaluation of MOAC vocabulary. Two different groups of evaluators namely Linked Data enthusiasts and domain experts were asked to help with the extrinsic evaluation of MOAC. Tasked based competency questions survey and peer interviews were conducted between December 2011 - February 2012. The process demonstrated that evaluators found the website more user-friendly to search and seek answers related to fundamental humanitarian questions using the web user interface. Based on the survey questions listed in Appendix 2, the surveyors could not only locate what humanitarian relief items the particular humanitarian agency is providing, they were also able to relate it to the pleas for help communicated from the same region. This is particularly important achievement because in the Haiti earthquake of 2010 "(..) the (humanitarian) cluster did not achieve any progress with respect to participation of the affected population in its activities. ([106], page 56-57)" . Therefore, as "the shelter cluster is an example of best practice in cluster information management, as well as in public information support to the cluster members.( [58], page 6, par. 22 )", based on research's evidences from shelter response demonstration, this research proposes to use Linked Data to integrate pleas for help from affected communities and further strengthen evidence based decision support.

Finally, following table summarizes the objectives of this research with the evaluation processes based on "who can evaluate, what can be evaluated, when and where the activity is carried out"([34], page 291 -292).

<i>Evaluation Dimensions</i>	<i>Evaluation Tests Qualitative and Quantitative)</i>	<i>Evaluators (Details in Appendix 2)</i>	<i>Meet Thesis Objectives</i>
<ul style="list-style-type: none"> <li>• Functionality (task based)</li> </ul>	<ul style="list-style-type: none"> <li>• Evaluation Scenario 1</li> <li>• Evaluation Scenario 2</li> </ul>	<ul style="list-style-type: none"> <li>• Domain experts</li> <li>• Linked Data enthusiasts</li> </ul>	1,2,3
<ul style="list-style-type: none"> <li>• Usability</li> </ul>	<ul style="list-style-type: none"> <li>• Evaluation Scenario 2</li> <li>• Evaluation Scenario 3</li> </ul>	<ul style="list-style-type: none"> <li>• Domain experts</li> <li>• Linked Data enthusiasts</li> <li>• Humanitarian Social Media Volunteer</li> </ul>	1,2,3
<ul style="list-style-type: none"> <li>• Structural</li> </ul>	<ul style="list-style-type: none"> <li>• Evaluation Scenario 1</li> <li>• Evaluation Scenario 2</li> <li>• Evaluation Scenario 3</li> </ul>	<ul style="list-style-type: none"> <li>• Domain experts</li> <li>• Linked Data enthusiasts</li> </ul>	2,3

Table 6 MOAC Vocabulary Evaluation Processes in relation to thesis objectives

## 8.2 Conclusion

The research concludes that the Principle of Linked Data, with the help of Management of a Crisis Vocabulary, does support the integration of crowdsourced information into traditional humanitarian relief agencies' crisis or disaster management information. This is more true in the early stages of a disaster when there is a scarcity of verifiable information on the needs of affected people. The disaster expert surveys verifies that the lack of information does impact the way decisions are made in emergencies and its consequent results are, at times regrettably, paid with the lives of people in need of help. Knowing that the relief items in the early stages are almost always going to fall short of the needs, humanitarian relief agencies are under tremendous pressure to make the right choice. However, during the evaluation of the information integration across different sources, this study also confirmed that such shortfalls of

information and relief items do not diminish either the urgency or the need to make informed decision on where to concentrate the humanitarian effort. Hence, decision makers are often faced with the dilemma to whether wait for more verified data or to force their hands and act based on available information. In either case, there are chances that one might still not capture the full dimensions of the needs of the affected population. To meet this challenge, a new breed of innovators has emerged during the Haiti earthquake where the volunteers and crowdsourced technologists used social media and mobile networks to compile information directly received from the people affected by the disaster. With the help of emergency shelter cluster and Ushahidi Haiti project official data, this study demonstrates that Linked Data with the help of the proposed MOAC can support informed decision making process by integrating information and creating mashup like a) pleas for help collected directly from the affected communities using crowdsourced mechanisms with b) traditional relief agencies humanitarian response data. Disaster Relief 2.0 summarizes that each major humanitarian disaster rips open a gap between the past and present, between what once was and what is now. The study also touches on the evidences that in every major disaster, there is a gap between which agency (who) is providing what kind of assistance and where" to what affected population are in need of, where (location) and when (time). This research is a small step to bridge that gap using the principle of Linked Data to integrate pleas for assistance from large numbers of an affected population collected through crowdsourcing concepts with traditional humanitarian management processes.

### **8.3 Research Contributions**

Implementation of Linked Data for the integration of crisis information on the onset of a disaster is a novel, emerging concept that is yet to reach its maturity. As Sir Tim Berners-Lee, the founder of world wide web said in a TED conference: the true potential of web of data can only be realized when everyone starts doing their share to publish Linked Data. This particular research summarized following foreseeable contributions; to both the academic and the professionals; on *integration of crowdsourced information with traditional crisis and disaster management Information*, such as:

- A joint post Haiti earthquake 2010 report from UNOCHA, Harvard Humanitarian Initiatives, UN Foundation titled "Disaster relief 2.0: The future of information sharing in humanitarian emergencies" [47] could consider the proposed concept of Linked Data (as shown in figure 8-2) while adding two new inflows of data - 1) from the people directly affected by a disaster and 2) information compiled by swarm of global volunteers also know as Volunteer Technical Committees (VCTs).
- Global consortium of "Decision Makers Need" *Community of Interest Group* managed by UNOCHA [117] could benefit from the research proposed idea of using Linked Data to integrate information from affected population and VCTs in the early stages of a disaster in support of informed decision making processes.
- The final report on Global Symposium 5+ [44, 77] and Inter Agency Standing Committee Operational Guidance on Responsibilities of Sector Cluster Leads and OCHA in Information Management [114], because the use of Linked Data in the research is also based on the principles of humanitarian information management and exchange in emergencies like accountability towards the affected population and humanitarian partner, inclusiveness of data received from multi sources, interoperability of data, and accessibility of information (discussed in section 2.1).
- OCHA Guidelines for Field Information Management in Humanitarian Context [113], because this study promotes the idea of using Linked Data to integrate information from the alternative sources (like the crowdsourced and VGI) in the early stages of a disaster when there is no or little reliable information about the needs of the people affected by a crisis.

- Cross-Cluster Information Management Framework within the Cluster Approach ([10], [72] as Annex 3), because Linked Open Data principle with the help of Management of a Crisis (MOAC) vocabulary could provide means to facilitate information integration within clusters and across the humanitarian response.
- United Nations Global Pulse program<sup>54</sup>, because the thesis proposes Linked Data as a part of knowledge sharing technique when it comes collecting, processing, analysis and disseminating *Big Data*<sup>55</sup> (or Data Exhaust).

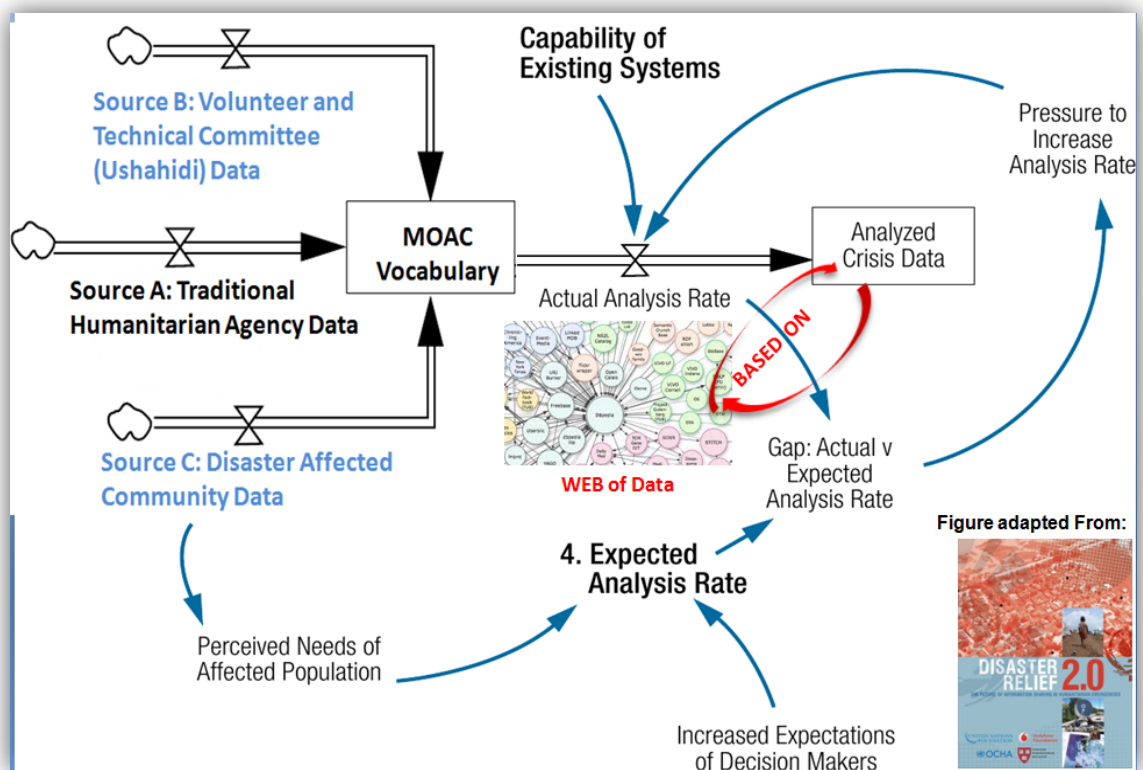


Figure 8-2 Inclusion of Linked Data and MOAC in Disaster Relief 2.0 Flows of Data to Crisis Responders (Adapted from Disaster Relief 2.0, page 19)

#### 8.4 Limitations and Future Work

At no point does this research claim to document a complete set of semantic vocabulary resources for the disaster information managers or the crowdsourced volunteer committees. Having said that, it is a missed opportunity that this study could not integrate vocabulary from all humanitarian clusters other than the emergency Shelter Cluster convened by IFRC. Most of the scientific documentation about Linked Data in relation to disaster management was limited to either disaster scenarios or proposed conceptual frameworks. This limited the ability to fully explore connections to existing disaster management vocabularies.

Based on the feedback received during evaluation of MOAC, the following future works are proposed:

- Test MOAC vocabulary in upcoming disasters and include missing classes and properties to guarantee maintenance and continuation of the vocabulary,
- Streamline MOAC from the ontological design concept discussed in [20],

<sup>54</sup> UN Global Pulse, see <http://www.unglobalpulse.org/>

<sup>55</sup> See [http://en.wikipedia.org/wiki/Big\\_data](http://en.wikipedia.org/wiki/Big_data)

- Test the MOAC integration into the Linked Data collaborative knowledge management architecture discussed in [31], and
- Evaluation of the MOAC vocabulary based on intrinsic approach.

## 9. Bibliography

1. Ahearn, S. et al.: Proof of concept and state of the art in FOSS Geospatial Technology. Report for Field Information Geospatial-database System (FIGS) for the United Nations Office for Coordination of Humanitarian Affairs (OCHA). (2006).
2. Alani, H., Brewster, C.: Ontology ranking based on the analysis of concept structures. Proceedings of the 3rd international conference on Knowledge capture. 51-58 (2005).
3. Alexander, K. et al.: Describing linked datasets. Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009). (2009).
4. Amin, S., Goldstein, M.P.: Data Against Natural Disasters: Establishing Effective Systems for Relief, Recovery, and Reconstruction. World Bank Publications (2008).
5. Anderson, L.: Demystifying the Arab Spring: Parsing the Differences Between Tunisia, Egypt, and Libya. Foreign Affairs. 90, 3, 2-7 (2011).
6. Armitage, L.H., Enser, P.G.B.: Analysis of user need in image archives. Journal of information science. 23, 4, 287-299 (1997).
7. Ashdown, P.: Humanitarian Emergency Response Review. <http://www.dfid.gov.uk/Documents/publications1/HERR.pdf> (2011).
8. Auer, S. et al.: Triplify: light-weight linked data publication from relational databases. Proceedings of the 18th international conference on World wide web. 621-630 (2009).
9. Bauer, F., Kaltenböck: Linked Open Data: The Essentials A Quick Start Guide for Decision Makers. The Semantic Web Company (2012).
10. Bauman, N.: Information Management Framework within the Cluster Approach v2.2. Online at [http://www.humanitarianinfo.org/hicstakeholder2007/CCIM/Information\\_Management\\_Framework\\_narrative\\_within\\_the\\_Cluster\\_Approach\\_v2.2.pdf](http://www.humanitarianinfo.org/hicstakeholder2007/CCIM/Information_Management_Framework_narrative_within_the_Cluster_Approach_v2.2.pdf). (2007).
11. Beckett, D., McBride, B.: RDF/XML syntax specification (revised). W3C recommendation. 10, (2004).
12. Belz, A., Gatt, A.: Intrinsic vs. extrinsic evaluation measures for referring expression generation. Proceedings of ACL-08: HLT, Short Papers. 197-200 (2008).
13. Benson, A.C.: Relationship Analysis of Image Descriptions: An ontological, content analytic approach. (2011).
14. Berners-Lee, T.: Design issues: Linked data. Online at <http://www.w3.org/DesignIssues/LinkedData.html>. (2006).
15. Berners-Lee, T.: Semantic web road map. Online at <http://www.w3.org/DesignIssues/Semantic.html>. (1998).
16. Berners-Lee, T. et al.: The semantic web. Scientific American. 284, 5, 34-43 (2001).

17. Bizer, C. et al.: DBpedia-A crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web*. 7, 3, 154-165 (2009).
18. Bizer, C. et al.: How to publish linked data on the web. Online at <http://www4.wiwiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>. 20 (October), (2007).
19. Bizer, C. et al.: Linked data-the story so far. *International Journal on Semantic Web and Information Systems*. 4, 2, 1-22 (2009).
20. Blázquez, M. et al.: Building Ontologies at the Knowledge Level using the Ontology Design Environment. (1998).
21. Borges, M.R.S. et al.: Linked Open Data and the Design of Information Infrastructure for Emergency Management Systems. *ISCRAM*. (2011).
22. Van Borkulo, E. et al.: Decision making in response and relief phases. *Geo-information for Disaster Management, First International Symposium on Geo-information for Disaster Management*. 47-53 (2005).
23. Coyle, D. et al.: New technologies in emergencies and conflicts: The role of information and social networks. *United Nations Foundation* (2009).
24. Davidson, S.: A Review of the IFRC-led Shelter Cluster Haiti 2010. (2011).
25. Davidson, S., Price, G.: Review of the International Federation's Shelter Cluster Commitment. (2011).
26. Dey, I.: *Qualitative data analysis: A user-friendly guide for social scientists*. Routledge (1993).
27. Dillon, M., Reid, J.: Global governance, liberal peace, and complex emergency. *Alternatives: Global, Local, Political*. 25, 1, 117-143 (2000).
28. Edwardes, A.J., Purves, R.S.: A theoretical grounding for semantic descriptions of place. *Proceedings of the 7th international conference on Web and wireless geographical information systems*. 106-120 (2007).
29. Elhadad, M. et al.: Automatic Evaluation of Search Ontologies in the Entertainment Domain using Text Classification. *Applied Semantic Technologies: Using Semantics in Intelligent Information Processing*. (2011).
30. Enser, P., Sandom, C.: Retrieval of archival moving imagery-CBIR outside the frame? *Image and Video Retrieval*. 85-106 (2002).
31. de Faria Cordeiro, K. et al.: Use of Linked Data in the Design of Information Infrastructure for Collaborative Emergency Management System. *Proceedings of the 2011 15th International Conference on Computer Supported Cooperative Work in Design*. (2011).
32. Feridun, M., Tanner, A.: Using linked data for systems management. *Network Operations and Management Symposium (NOMS), 2010 IEEE*. 926-929 (2010).
33. Fernández-López, M. et al.: Methontology: from ontological art towards ontological engineering. (1997).

34. Gómez-Pérez, A. et al.: Evaluation and Assessment of the Knowledge Sharing Technology. Towards Very Large Knowledge Bases. 289-296 (1995).
35. Gómez-Pérez, A.: Evaluation of Ontologies. *International Journal of intelligent systems*. 16, 3, 391-409 (2001).
36. Gómez-Pérez, A.: From Knowledge Based Systems to Knowledge Sharing Technology: Evaluation and Assessment. (1994).
37. Gómez-Pérez, A. et al.: Towards a method to conceptualize domain Ontologies. (1996).
38. Gómez-Pérez, A., Corcho, O.: Ontology languages for the semantic web. *Intelligent Systems, IEEE*. 17, 1, 54-60 (2002).
39. Goodchild, M.F.: Citizens as sensors: the world of volunteered geography. *GeoJournal*. 69, 4, 211-221 (2007).
40. Goodchild, M.F., Glennon, J.A.: Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*. 3, 3, 231-241 (2010).
41. Goolsby, R.: Social media as crisis platform: The future of community maps/crisis maps. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 1, 1, 7 (2010).
42. Guarino, N. et al.: What is an Ontology? *Handbook on Ontologies*. 1-17 (2009).
43. Gulla, J. et al.: Ontology learning for search applications. *On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS*. 1050-1062 (2007).
44. Haggarty, A., Naidoo, S.: Global Symposium+5 Final Report, Information for Humanitarian Action. UN OCHA, Geneva. ReliefWeb, UNOCHA, Global Symposium Secretariat ReliefWeb Palais des Nations 1211 Geneva (2008).
45. Halpin, H., Baker, T.: Vocabulary Hosting: A Modest Proposal. 2010 AAAI Spring Symposium Series. (2010).
46. Hansen, G.: Humanitarian action in Iraq—emerging constraints and challenges. *Humanitarianism and War Project*. 27, (2004).
47. Harvard Humanitarian Initiative et al.: Disaster relief 2.0: The future of information sharing in humanitarian emergencies. Washington, DC and Berkshire, UK: UN Foundation & Vodafone Foundation Technology Partnership, Washington, D.C. and Berkshire, UK (2011).
48. Hattotuwa, S., Stauffacher, D.: Haiti and beyond: Getting it right in Crisis Information Management. (2010).
49. Heath, T., Bizer, C.: Linked data: Evolving the web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology*. 1, 1, 1-136 (2011).
50. Heinzelman, J. et al.: Crowdsourcing crisis information in disaster-affected Haiti. *US Institute of Peace* (2010).
51. Heipke, C.: Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*. 65, 6, 550-557 (2010).

52. Hendrix, C.J.: Humanitarian Exchange Language (HXL). Online at <https://sites.google.com/site/hxlproject/home/project-overview>. (2012).
53. Holmes, J.: Keynote Speech by UN Under Secretary General Sir. John Homes, United Nations Global Symposium+5: Information for Humanitarian Action. Key Note Address. (2007).
54. Howe, J.: The rise of crowdsourcing. *Wired magazine*. 14, 14, 1-5 (2006).
55. IASC: Guidance note on using the cluster approach to strengthen humanitarian response. Online at <http://reliefweb.int/node/23017>. (2006).
56. IASC: Inter Agency Standing Committee 70th Working Group Meeting: Background Document: Preserving Humanitarian Space, Protection and Security. WO/0803/2583/7, Session 3.1 A review of concepts related to the issue of Preserving Humanitarian Space, Protection and Security. (2008).
57. IASC: Inter-Agency Standing Committee (IASC) Fact Sheet. Online at <http://www.humanitarianinfo.org>.
58. IASC: Preliminary Lessons Learned relating to the Humanitarian Response to the Haiti Earthquake. Inter Agency Standing Committee Principals Meeting. (2010).
59. IASC: Response to the humanitarian crisis in Haiti following the 12 January 2010 earthquake. Online at [www.humanitarianinfo.org/](http://www.humanitarianinfo.org/). (2010).
60. IASC, Partners: Haiti Earthquake Flash Appeal 2010. UNOCHA (2010).
61. IFRC: Case study - Coordinating shelter in Haiti: IFRC and the Shelter Cluster. IFRC. (2011).
62. IFRC: World Disaster Report 2010. Focus on Urban Risk. International Federation of Red Cross and Red Crescent Societies, 2010 ISBN 978-92-9139-156-1, International Federation of Red Cross and Red Crescent Societies, Geneva (2010).
63. Jain, P. et al.: Linked data is merely more data. *Linked Data Meets Artificial Intelligence*. 82-86 (2010).
64. Kauppinen, T. et al.: *Linked Science: Interconnecting Scientific Assets* (forthcoming 2012). *Data Intensive Science*. 1-17 (2012).
65. Kauppinen, T.: *ObservedChange.com: About Changes that should be Observed and Shared*. Online at <http://observedchange.com/>. (2011).
66. Kauppinen, T., de Espindola, G.: *Linked open science - Communicating, sharing and evaluating data, methods and results for executable papers*. *Procedia Computer Science: International Conference on Computational Science, ICCS 2011*. 4, 726-731 (2011).
67. King, D.: *Humanitarian Knowledge Management*. *Proceedings of the Second International ISCRAM Conference*. 1-6 (2005).
68. Knight, L.: *World Disaster Report 2009. Focus on early warning, early action*. International Federation of Red Cross and Red Crescent Societies, 210s. ISBN (2009).

69. Kuhn, W.: Geospatial semantics: why, of what, and how? *Journal on data semantics III*. 587-587 (2005).
70. Kuhn, W.: *Semantics for dummies*. Online at [http://www.geoconnexion.com/uploads/kuhn\\_intv5i9.pdf](http://www.geoconnexion.com/uploads/kuhn_intv5i9.pdf) . (2006).
71. Kurtz, j, Wheaton, W.: *Education Cluster: Joint Needs Assessment Toolkit for Education in Emergencies*. (2009).
72. Larsen, L.: *Strengthening Humanitarian Information Management: A Status Report*. Report prepared for OCHA/ECHO, Field Information Services Unit, United Nations Office for Coordination of Humanitarian Affairs, Geneva (2007).
73. Limbu, M. et al.: *Management of a Crisis (MOAC) Vocabulary Specification*. Online at <http://observedchange.com/moac/ns/>. (2011).
74. Liu, S.B.: *Grassroots heritage: A multi-method investigation of how social media sustain the living heritage of historic crises*. (2011).
75. Liu, S.B.: *The Rise of Curated Crisis Content*. *Proceedings of the 7th International ISCRAM Conference-Seattle*. 1, (2010).
76. Liu, S.B., Palen, L.: *The new cartographers: Crisis map mashups and the emergence of neogeographic practice*. *Cartography and Geographic Information Science*. 37, 1, 69-90 (2010).
77. Maitland, C., Tapia, A.: *Global Symposium+ 5 Information for Humanitarian Action Survey Report*. Prepared for United Nations Office for the Coordination of Humanitarian Affairs. (2007).
78. Markey, K.: *Computer-Assisted Construction of a Thematic Catalog of Primary and Secondary Subject Matter*. *Visual Resources*. 3, 1, 16-49 (1983).
79. Matthews, B.: *Semantic web technologies*. *E-learning*. 6, 6, 8 (2005).
80. McDonald, B., Gordon, P.: *United Nations' Efforts to Strengthen Information Management for Disaster Preparedness and Response*. *Data against natural disasters: establishing effective systems for relief, recovery, and reconstruction*. 59-82 (2008).
81. Meier, P.P.: *Verifying Crowdsourced Social Media Reports for Live Crisis Mapping: An Introduction to Information Forensics*. (2011).
82. Meurs, M.J. et al.: *Towards Evaluating the Impact of Semantic Support for Curating the Fungus Scientific Literature*. *WS2*. 34 (2011).
83. Morrow, N. et al.: *Independent Evaluation of the Ushahidi Haiti Project*. (2011).
84. Munro, R.: *Crowdsourced translation for emergency response in Haiti: the global collaboration of local knowledge*. *AMTA Workshop on Collaborative Crowdsourcing for Translation*. (2010).
85. Neches, R. et al.: *Enabling technology for knowledge sharing*. *AI magazine*. 12, 3, 36 (1991).
86. Neis, P. et al.: *Collaborative mapping and Emergency Routing for Disaster Logistics-Case studies from the Haiti earthquake and the UN portal for Afrika*. *Geospatial Crossroads@*

- GI\_Forum. 10, (2010).
87. Netzer, Y. et al.: Ontology evaluation through text classification. *Advances in Web and Network Technologies, and Information Management*. 210-221 (2009).
  88. Neuman, W.L.: *Social Research Methods: Quantitative and Qualitative Methods*. Allyn & Bacon (2010).
  89. Norheim-Hagtun, I., Meier, P.: Crowdsourcing for Crisis Mapping in Haiti. *Innovations: Technology, Governance, Globalization*. 5, 4, 81-89 (2010).
  90. Noy, N.F. et al.: *Ontology development 101: A guide to creating your first ontology*. (2001).
  91. Van Nuffelen, B. et al.: *Creating Knowledge out of Interlinked Data: The Integrated LOD2 Tool Stack*. (2011).
  92. Okolloh, O.: Ushahidi, or “testimony”: Web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action*. 59, 1, 65-70 (2009).
  93. OReilly, T.: *What is Web 2.0: Design patterns and business models for the next generation of software*. (2007).
  94. Ortmann, J. et al.: Crowdsourcing Linked Open Data for Disaster Management. *Terra Cognita 2011 Workshop on Foundations, Technologies and Applications of the Geospatial Web*. 798, 11-22 (2011).
  95. Palen, L. et al.: Online forums supporting grassroots participation in emergency preparedness and response. *Communications of the ACM*. 50, 3, 54-58 (2007).
  96. Palen, L., Liu, S.B.: Citizen communications in crisis: anticipating a future of ICT-supported public participation. *Proceedings of the SIGCHI conference on Human factors in computing systems*. 727-736 (2007).
  97. Panofsky, E. et al.: *Studies in iconology: humanistic themes in the art of the Renaissance*. Harper & Row New York (1972).
  98. Powers, R.: *Introduction to Disasters and disaster nursing*.
  99. Pultar, E. et al.: Dynamic GIS case studies: Wildfire evacuation and volunteered geographic information. *Transactions in GIS*. 13, 85-104 (2009).
  100. Radev, D.R. et al.: Centroid-based summarization of multiple documents. *Information Processing & Management*. 40, 6, 919-938 (2004).
  101. Sahu, S.: *Guidebook on technologies for disaster preparedness and mitigation*. Asian and Pacific Centre for Transfer of Technology.
  102. Shadbolt, N. et al.: The semantic web revisited. *Intelligent Systems, IEEE*. 21, 3, 96-101 (2006).
  103. Shamoug, A., Juric, R.: Addressing Interoperability through the Semantic of Information Highway in Managing Responses in Humanitarian Crises. *Proceedings of the 8th International ISCRAM Conference-Lisbon*. 1, (2011).

104. Shatford, S.: Analyzing the subject of a picture: a theoretical approach. *Cataloging & classification quarterly*. 6, 3, 39-62 (1986).
105. Stankovic, M. et al.: Looking for experts? what can linked data do for you. *Linked Data on the Web (LDOW 2010)*. CEUR Workshop Proceedings (April 2010). (2010).
106. Steets, J. et al.: Cluster Approach Evaluation 2 Synthesis Report. IASC Cluster Approach Evaluation 2nd Phase, Groupe URD and the Global Public Policy Institute, April. (2010).
107. Supekar, K.: A peer-review approach for ontology evaluation. *8th International Protege Conference*. 77-79 (2005).
108. Tapia, A.H. et al.: Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations. *Proceedings of the 8th International ISCRAM Conference*. 1-10 (2011).
109. UN: Pakistan Floods : Rapid Response Plan. (2011).
110. UNICEF: Emergency Field Handbook: A guide for UNICEF staff. UNICEF Division of Communication (2005).
111. UNOCHA: 3W:Who What Where/Contact Management Directory , Quick Start Administrators Guide. (2007).
112. UNOCHA: Assessment and Classification of emergencies (ACE) project: Mapping of key emergency needs assessment and analysis initiatives. Online at [www.humanitarianinfo.org/iasc/downloaddoc.aspx?docID=4927](http://www.humanitarianinfo.org/iasc/downloaddoc.aspx?docID=4927). (2009).
113. UNOCHA: Guidelines for OCHA Field Information Management. Office for the Coordination of Humanitarian Affairs (OCHA), United Nations, New York, United Nations (2006).
114. UNOCHA: Operational Guidance on Responsibilities of Sector Cluster Leads and OCHA in Information Management. (2009).
115. UNOCHA: United Nations Office for the Coordination of Humanitarian Affairs Symposium on Best Practices in Humanitarian Information Exchange Palais des Nations Geneva, Switzerland 5 – 8 February 2002. (2002).
116. UNOCHA, Bennett, C.: Office for the Coordination of Humanitarian (OCHA) Affairs Orientation Handbook on Complex Emergencies. New York: United Nations (2002).
117. Verity, A.: Collaboration with Volunteer & Technical Community (V&TC) for Libya and Japan OCHA's Lessons Learned. Online at [https://docs.google.com/document/d/1wut8oDRo9BYS1c0hQ34Ng8qQ-pLVGIRO95WOvR3MN78/edit?hl=en\\_US&pli=1#](https://docs.google.com/document/d/1wut8oDRo9BYS1c0hQ34Ng8qQ-pLVGIRO95WOvR3MN78/edit?hl=en_US&pli=1#) (accessed on 21 Feb 2012). (2011).
118. Van de Walle, B. et al.: Humanitarian information management and systems. *Mobile Response*. 12-21 (2009).
119. Williams, C.: Data Preparedness and Emergency response: No Good Decisions without Good Information. Paper presented at Seventeenth United Nations Regional Cartographic Conference for Asia and the Pacific Bangkok, 18-22 September 2006 Item 7 of the provisional agenda. (2006).

## Bibliography

---

120. Wisner, B., Adams, J.: Environmental health in emergencies and disasters: a practical guide. World Health Organization (2002).
121. Witzel, A.: The problem-centered interview. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*. 1, 1, (2000).
122. Ziesche, S.: Social-networking web systems: Opportunities for humanitarian information management. *Journal of Humanitarian Assistance*. (2007).
123. Zook, M. et al.: Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy*. 2, 2, 7 (2010).

## 10. Appendices:

### 10.1 Appendix 1: Triplification Sample

```

<http://http://ifgibox.de/m_limb03/moac/L425>
<http://purl.org/dc/elements/1.1/identifier> "L425" .

<http://http://ifgibox.de/m_limb03/moac/L425>
<http://www.loc.gov/marc/countries/Country>
"Haiti"^^<http://www.w3.org/2001/XMLSchema#string> .

<http://http://ifgibox.de/m_limb03/moac/L425> <http://purl.org/dc/elements/1.1/title>
"Shelter response"^^<http://www.w3.org/2001/XMLSchema#string> .

<http://http://ifgibox.de/m_limb03/moac/L425>
<http://observedchange.com/moac/ns#Agency>
"CARE"^^<http://www.w3.org/2001/XMLSchema#string> .

<http://http://ifgibox.de/m_limb03/moac/L425> <http://purl.org/dc/elements/1.1/subject>
<http://observedchange.com/moac/ns#ShelterOffered> .

<http://http://ifgibox.de/m_limb03/moac/L425>
<http://purl.org/dc/elements/1.1/coverage>
"Leogane"^^<http://www.w3.org/2001/XMLSchema#string> .

<http://http://ifgibox.de/m_limb03/moac/L425>
<http://observedchange.com/moac/ns#verified>
"true"^^<http://www.w3.org/2001/XMLSchema#boolean> .

<http://http://ifgibox.de/m_limb03/moac/L425> <http://purl.org/dc/elements/1.1/date>
"2010-02-17T00:00:00-05:00"^^<http://www.w3.org/2001/XMLSchema#dateTime> .

<http://http://ifgibox.de/m_limb03/moac/L425>
<http://observedchange.com/moac/ns#Cluster>
"ShelterCluster"^^<http://www.w3.org/2001/XMLSchema#string> .

<http://http://ifgibox.de/m_limb03/moac/L425> <http://rdfs.org/sioc/ns#content> "Care
distributed Shelter relief materials in
Leogane"^^<http://www.w3.org/2001/XMLSchema#string> .

<http://http://ifgibox.de/m_limb03/moac/L425>
<http://observedchange.com/moac/ns#latlon> "18.5235,-
72.57463333"^^<http://www.w3.org/2001/XMLSchema#string> .

<http://http://ifgibox.de/m_limb03/moac/L425>
<http://www.w3.org/2003/01/geo/wgs84_pos#latitude>
"18.5235"^^<http://www.w3.org/2001/XMLSchema#float> .

<http://http://ifgibox.de/m_limb03/moac/L425>
<http://www.w3.org/2003/01/geo/wgs84_pos#longitude> "-72.57463333"
^^<http://www.w3.org/2001/XMLSchema#float> .

```

## ***10.2 Appendix 2: Vocabulary Evaluation Survey Questionnaires***

### ***10.2.1 Survey 1: Sample of Linked Data Enthusiasts Evaluation Questionnaires***

**[Thesis Abstract- DELETED]**

Linked Data Crowd Experts Questionnaires:

1. When (Date) was the first Emergency shelter Needs reported and where?
  - a. option 1: Port Au Prince, 14 Jan 2010
  - b. Option 2. 67 Thor Ave, Port-au-Prince, Haiti and Jacmel, Haiti, 13 Jan 2010
  - c. Option 3. Jacmel, Haiti, 14 Jan 2010
  
2. Based on the given data, please indicate which was the first agency to provide shelter, where (Location) and when (Time)?
  - a. option1: Agency name: Food for the Hungry / Shelter provided at Location: Delmas /Date 13 Jan 2010
  - b. option2: Agency name: Oxfam / Shelter provided at Location: Petit-Goave /Date 14 Jan 2010
  - c. option3: Agency name: Food for the Hungry / Shelter provided at Location: Delmas /Date 14 Jan 2010
  - d. option4: Agency name: Oxfam / Shelter provided at Location: Petit-Goave /Date 15 Jan 2010
  
3. In the given reporting period, please indicate the number of locations where Oxfam (Agency) provided shelter assistance.
  - a. option1: only in 1 location
  - b. option2: more than 5 locations
  - c. option3: not sure
  - d. option4: Stupid visualization, it doesn't make sense
  
4. on 14 Jan 2010 (date), how many shelter needs were reported and where?
  - a. option1: 1 shelter need report / location: 1 location at Jacmel
  - b. option2: 4 shelter need reports / location: 1 location at Port Au Prince
  - c. option3: 2 shelter need reports / location: 2 locations at Port Au Prince and jacmel
  - d. option4: Not sure
  
5. From timeline graph at the bottom of the map, can you indicate how many days there were NO reported shelter offered on the first week of Haiti Disaster i.e from 12 Jan 2010 - 19 Jan 2010?(Please uncheck all the options and click on ShelterOffered)
  - a. option1: In the first week No reported shelter distribution for: 3 days - Jan 15, Jan 16 and Jan 17
  - b. option2: In the first week No reported shelter distribution for: 1 day - Jan 13 2010
  - c. option3: In the first week No reported shelter distribution for: Shelter Offered everyday
  - d. option4: In the first week No reported shelter distribution for: Visualization doesnt provide clear indication

6. From the timeline graph, please indicate the first reported date of shelter distribution (Cluster Response:ShelterOffered) in Loegane (Location)?
  - a. option1: First reported shelter distribution in Loegane: 13 Jan 2010
  - b. option2: First reported shelter distribution in Loegane: 14 Jan 2010
  - c. option3: First reported shelter distribution in Loegane: 18 Jan 2010
  - d. option4: First reported shelter distribution in Loegane: 22 Jan 2010
  
7. From the timeline graph, please indicate the first reported date of shelter needs (Cluster Response:ShelterNeeded) in Loegane (Location)?
  - a. option1: First reported shelter distribution in Loegane: 13 Jan 2010
  - b. option2: First reported shelter distribution in Loegane: 15 Jan 2010
  - c. option3: First reported shelter distribution in Loegane: 21 Jan 2010
  - d. option4: First reported shelter distribution in Loegane: 22 Jan 2010
  
8. From the timeline graph, please indicate the first reported date of shelter distribution (Cluster Response:ShelterOffered) in Jacmel (Location)?
  
9. From the timeline graph, please indicate the first reported date of shelter needs (Cluster Response:ShelterNeeded) in Jacmel (Location)?

### ***10.2.2 Survey 2: Domain Expert Interview Survey and Questionnaires***

Primary Humanitarian Information Management Experts Questionnaires:

**[Thesis Abstract - DELETED]**

#### **Expert Survey: Information Management General Practices:**

1. Information in early stages of a disaster:
  - a. According to you, to support informed decision making process, what are the most important information necessary in the first few weeks on the immediate onset of a disaster?
  
  - b. How do you rate the important of (5 best 1 worst)
    - i. Needs and Gaps information
    - ii. Who What Where and When - Agency Specific
    - iii. Who What Where and When - Affected Population Specific
  
  - c. Source of information: While Compiling information what are the different information sources you use:
    - i. Government
    - ii. Traditional humanitarian agency
    - iii. Directly from affected Population
    - iv. Social media or crowdsourced eg. Uhahidi, Shahana
    - v. Others (Comments):

- d. Trust of information source: How much do you trust the given source of information in the first week of a disaster? (rate 1 worst 5 best)
    - i. Government -
    - ii. Traditional humanitarian agency -
    - iii. Directly from affected Population -
    - iv. Social media or crowdsourced eg. Uhahidi, Shahana
    - v. Others (Comments):
  
  - e. Formats of information collection and dissemination:
    - i. In what format do you collect data?
    - ii. In what format do you disseminate data?
  
  - f. If the government provide you a right set of data or right format of data, how much of an impact will it have in your work i.e. by what percentage?
2. Do you also provide raw data for partners, if yes in what format?
3. Information Collection interoperability and heterogeneity:
- a. When you compile information, have you faced problems where different agencies had different definitions of indicators, for example - "CollapseStructure" or "AffectedPopulation" etc?
    - i. Yes: Can you please elaborate
    - ii. NO. It is part of the game.
  - b. How much of your personal workload will be save if all the agencies were using indicators that would bear similar meaning throughout humanitarian response?

**Expert Survey: Two new Sources of information**

4. Information from Affected Population:
- a. If the affected population are trying to directly communicate with you, how do you integrate affected population request for help in terms of what their needs are?
5. Crowd information:
- a. Do you also use the social media reports or crowd sourced technologies to support your information compilation
    - i. No. Why?
    - ii. Yes. How?
  
  - b. In the absence of a reliable government data or any other humanitarian agency reports, would you consider using crowdsourced or crowd generated or social media compiled information?
    - i. Yes. Why:
    - ii. No. Why:

## Expert Survey: Integration of crowdsourced information with tradition crisis and Disaster management Information using Linked Data

6. Affected Population, Crowd communities and traditional humanitarian information integration: A joint report published by UNOCHA, Harvard Humanitarian Initiative, UN Foundation titled " Disaster Relief 2.0: The future of information sharing in humanitarian emergencies " pointed out that " The international humanitarian system (in Haiti Earthquake response)was not tooled to handle these two new information fire hoses—one from the **disaster-affected community** and one from **a mobilized swarm of global volunteers**. [47]page 10) " . Integration of such new found information is the core research question of this master thesis.
  - a. As a humanitarian information manager how important is it for you to know what affected population are in need of? (1less important -5very important)
  - b. If a crowd of volunteers (somewhere in the world),including but not limited to, expert disaster information managers and humanitarian officers with academicas provide you with a solution on how to integrate crowd generated data with that of your traditional data, would you be interested to use it as a part of your information analysis of needs and gaps in the first week of a disaster (when credible information is very low)?
  - c. It a known fact that in a disaster - information manager works for 98% of the time and sleeps for the rest 2%. To tap into new sources of information (affected population direct reports and global network of volunteers ), if there is a way to visualize in almost real time who/what/where the affected population are , would you be willing to explore such platform?
  
7. To facilitate such integration this research proposes the concept of linking data using principle of Linked Data [15]. Based on such needs, I am proposing following "Humanitarian Needs and Response visualization". The research proposes a prototype to integrate information from the Crowd i.e. Ushahidi platform with that of traditional humanitarian agency like Shelter Cluster Partners. Please check the given demonstration at [http://ifgibox.de/m\\_limb03//minu-haiti-response/HTMLPage1.htm](http://ifgibox.de/m_limb03//minu-haiti-response/HTMLPage1.htm)
  - i. "shelterNeeded" data are captured from Social Network (Ushahidi)
  - ii. "ShelterOffered" data represented by Shelter Cluster Agency reports
  - a. Does the information of population needs in connection to where the shelter have been provided make sense to you?
  - b. The whole visualization is based on a set of disaster management vocabulary called Management of a Crisis (MOAC) Vocabulary.
   
[http://ifgibox.de/m\\_limb03/moac/moac2012.html](http://ifgibox.de/m_limb03/moac/moac2012.html)
    - Does the vocabulary cover basic "humanitarian" taxonomies based on "Humanitarian Cluster Approach"?
    - Does the vocabulary cover basic humanitarian need and response taxonomies on "who what where" and when?
    - Does the vocabulary cover basic humanitarian need and response taxonomies related to Shelter Cluster Needs and Response?

### **10.3 Appendix 3: Profiles of Research Results Evaluators**

#### **10.3.1 Linked Data Enthusiasts**

**Ahmed Mahmoud AbdelMouty (Egypt)** has worked as teaching assistant at the department of information systems, university of Zagazig, Egypt. Currently he is enrolled as a Masters student at the Institute of Geoinformatics, University of Munster, Germany. Linked Open Data is one of his major subject.

**Alber Sánchez (Colombia)** is a Colombian Cadastral and Geodetic engineer, with a Msc on Information Sciences and certified Project Manager Professional who nowadays is a master student at the Intitute for Geoinformatics at the University of Münster, Germany. His background covers the geographic information system conceptualization, design, implementation and management for the public sector. His most recent experience was a GIS officer at the Coordination Center for Territorial Recovery (CCAI for its abbreviation in Spanish) of the Republic of Colombia which is in charge of coordinating other government agencies in the process of assuring security and peace for democracy strengthening in favour of free exercise of civil rights and human development at specific troubled zones of the country.

**Dong Wang (China)** is 25 years old and studies Geoinformatics as a PhD student at the University of Münster, Germany since October 2010. He finished his Bachelor and Master separately in July 2007 and July 2010 from the China University of Mining and Technology. His research is mainly about semantic interoperability in terms of transformations among semantic reference systems. He has worked with the Author for a workshop paper with respect to Linking Crowdsourcing information by the technology of Linked Open Data.

**Jim Jones (Brazil)** is polyglot (speaks Portuguese, Spanish, English and German) born in Recife, Brazil is a Software Engineer graduated by the UNIBRATEC - Universidade Brasileira de Tecnologia (Brazilian Technological University) and is currently doing a MSc. in Geoinformatics at the University of Münster in Germany. During his professional life he focused his software development expertises for the embedded systems and machine communication applications, mostly acquired as a Development Engineer in the American glass industry Owens-Illinois, where he also worked as an Oracle DBA (Database Administrator), due to the experience gained at official trainings at Oracle University. Currently he is working as a Test Engineer at Olympus Microscopy with softwares in the microscopy field, but most of attributions are solidly related to Oracle databases testing

**Sven Harpering (Germany)** is 24 years old and studies Geoinformatics at the University of Münster, Germany since October 2007. He finished his Bachelor degree in September 2010 and started his postgraduate studies right after in October 2010. He will finish his Master degree in September 2012. During his studies Sven Harpering gained a lot of experience in using and developing GIS functionalities for urban planning. Another field of experience is software engineering with a special focus on planning and implementing applications for Android smart phones.

### ***10.3.2 Domain Experts***

#### **Humanitarian Relief Workers:**

***Catherine Lefebvre (France):*** Mrs. Lefebvre is a senior Humanitarian Information Manager who has more than 10 years of field experience. She was deployed in Haïti from the immediate onset of the earthquake in January 2010, working initially with UNOCHA and later on with the emergency shelter cluster leads, ie International Federation of Red Cross (IFRC) and UN Habitat. Initially she was based in the capital Port-au-Prince, but worked mainly in Jacmel and Leogane. till mid 2011. For further information, she can be reached at cathlefebvre@yahoo.fr.

***Cecilia Utas (Sweden):*** Ms. Utas is one of the humanitarian information manager who has been deployed to many of the emergencies in the last few year that includes, but limited to, Philippines 2009, Haiti 2010, Kenya 2011. She also supported the Haiti earthquake response as UNOCHA Senior Information Management Officer in 2010. She is currently based in Pakistan as Inter Agency Liaisons Officer with UNOCHA where her main role is to facilitate coordinated assessments and information sharing in support of informed humanitarian decision making processes. For any further information, she could be reached at utas@un.org.

***Raul Ecay Torres (Spain), raecto@cruzroja.es:*** "I'm Spanish Red Cross volunteer since 1992 into de departamental Crisi center ans tecnico and coordinator. Since 2001 involved into the International department as IT/ICT technician as volunteer. First international mission as IFRC Telecom for an emergency deployment during social unrest in Haiti on March 2004 and Tsunami Operation on 2005 in Ache, Indonesia. From 2006 to 2011 as part of the Disaster management Unit of the International department of the Spanish Red Cross as IT/Telecom focal point and IT & Telecom ERU Coordinator. Deployed on January 2010 as IT & Telecom ERU Team Leader for two month, I came back to Haiti on June 2010 to November 2010 as Information Manager of the Shelter Cluster. During 2011 as Independent consultant developing the Rooster management database at global level of the Shelter Cluster. From April to July 2011 as ICT delegate, working on the development of the Disaster Management-Information Centre of the Indonesia Red Cross in Jakarta. October 2011 to January 2012 in El Salvador as Information Manager and Acting Coordinator of the Shelter Cluster. Tropical Depression 12E."

#### **Social Media Volunteer:**

***Christoph Dennenmoser (Germany) christoph@humanityroad.org*** is a German Red Cross Disaster Management expert. He is also the member of the Humanity Road Inc, a social media volunteer network where he holds post of "Team Leader Urgent Needs" focal pont.

*10.4 Appendix 4: Expert Survey and Linked Data Enthusiasts Samples*

*10.4.1 Summary of Expert Survey for Section 4*

*10.4.2 Sample of Domain Expert Evaluation for Section 7*

*10.4.3 Sample of Linked Data Enthusiasts Evaluation for Section 7*

NOTE: Please note that, with the permission of the experts and volunteers, the entire session of the survey and peer interviews were recorded. With proper permissions from the individual interviewee or reviewer, the audio files can be made available. Please contact the author for more details. Thank you.

# 17 [responses](#)

## 10. .4.1 Summary of Expert Survey for Section 4

### Summary [See complete responses](#)

**Question ID:**

W0R1D W0R1D W0R1D WOR1D W0R1D WOR1D WOR1D W0R1D W0R1D W0R1D W0R1D WOR1D WOR1D  
 W0R1D W0R1D W0R1D andrejverity@gmail.com

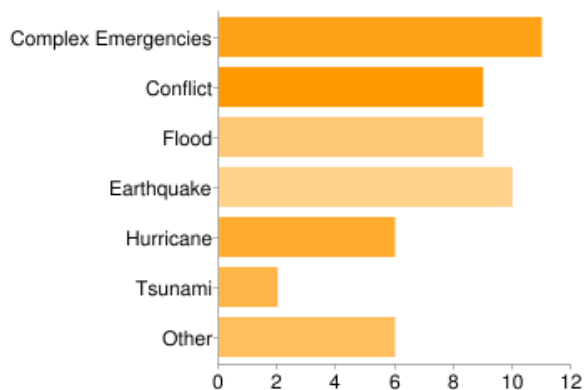
**1. Please indicate your field(s) of technical expertise:**



Monitoring and Evaluation	2	13%
Information Management (GIS/IMIT)	11	73%
Disaster Management	5	33%
Coordination	6	40%
Logistics	2	13%
Public Information/Reporting	3	20%
Other	1	7%

People may select more than one checkbox, so percentages may add up to more than 100%.

**2. Please indicate your field(s) of experience:**

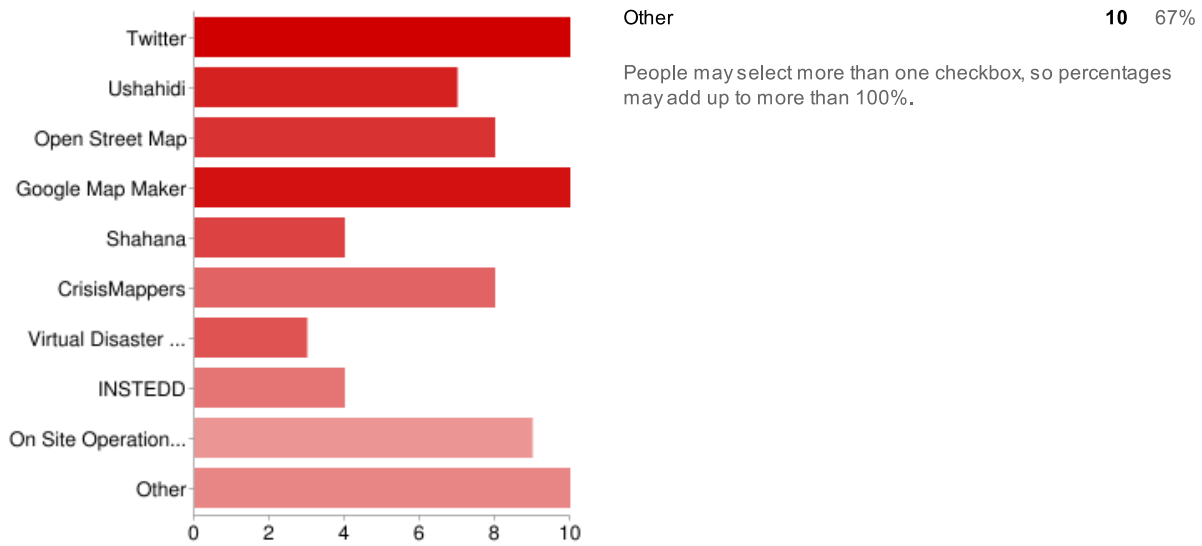


Complex Emergencies	11	73%
Conflict	9	60%
Flood	9	60%
Earthquake	10	67%
Hurricane	6	40%
Tsunami	2	13%
Other	6	40%

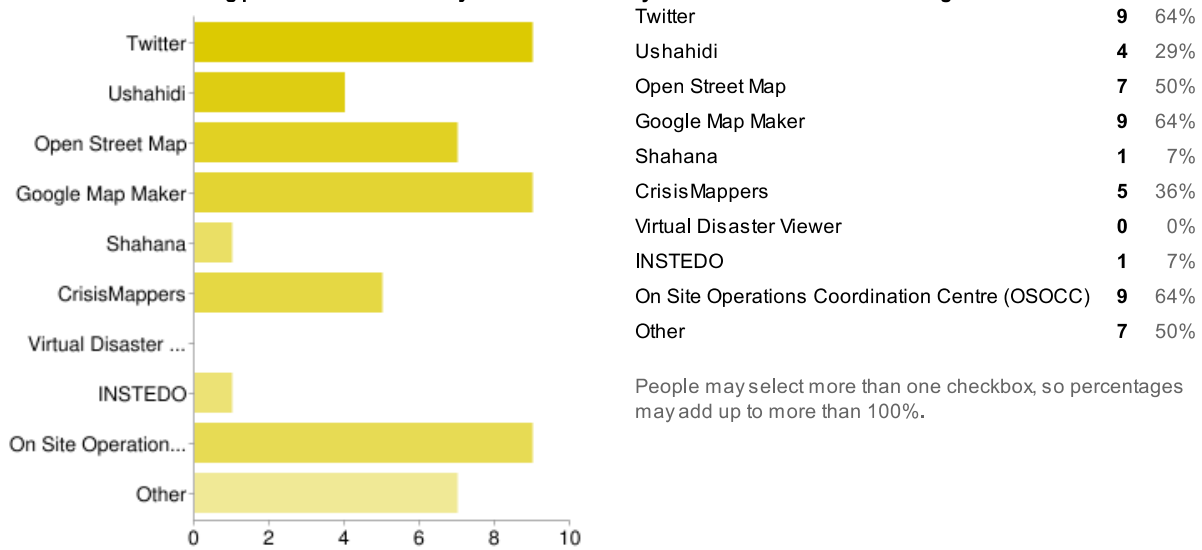
People may select more than one checkbox, so percentages may add up to more than 100%.

**3. According to you, which of the following platforms/services allow non-experts to contribute information relevant to disaster management?**

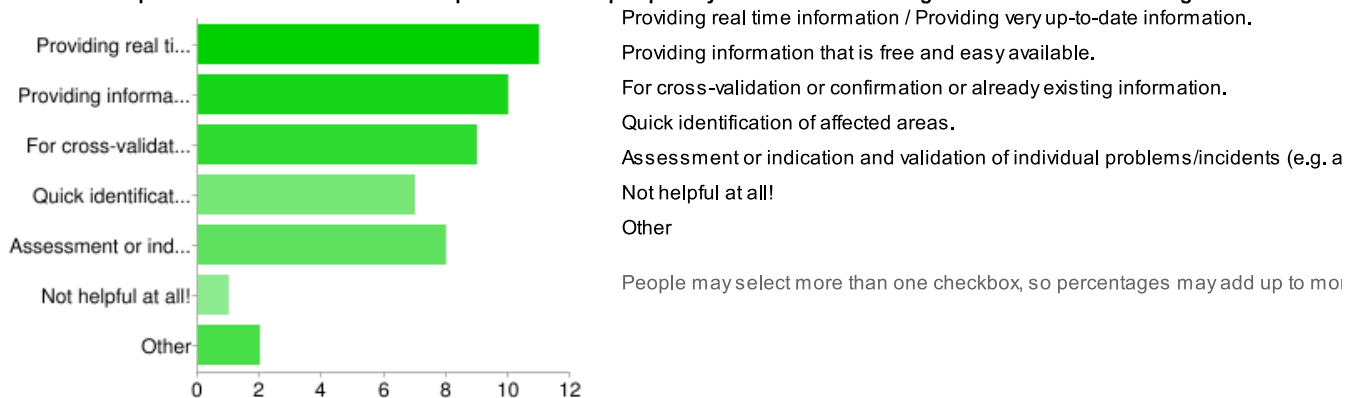
Twitter	10	67%
Ushahidi	7	47%
Open Street Map	8	53%
Google Map Maker	10	67%
Shahana	4	27%
CrisisMappers	8	53%
Virtual Disaster Viewer	3	20%
INSTEDD	4	27%
On Site Operations Coordination Centre (OSOCC)	9	60%



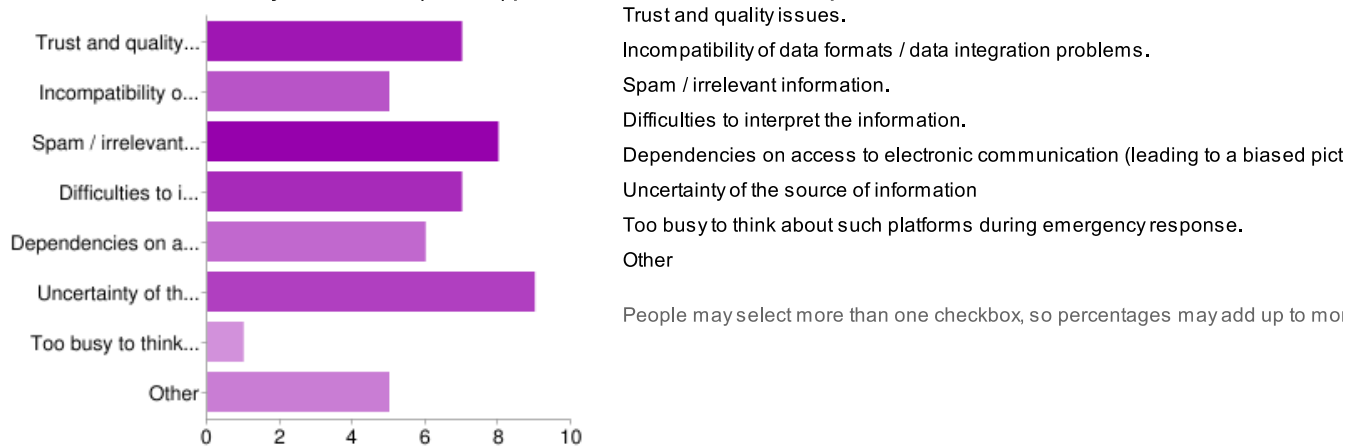
**4. Which of the following platforms/services did you or members of your team use in disaster management?**



**5. How do the platforms/services mentioned in question no.4 help improve your information management and decision making?**

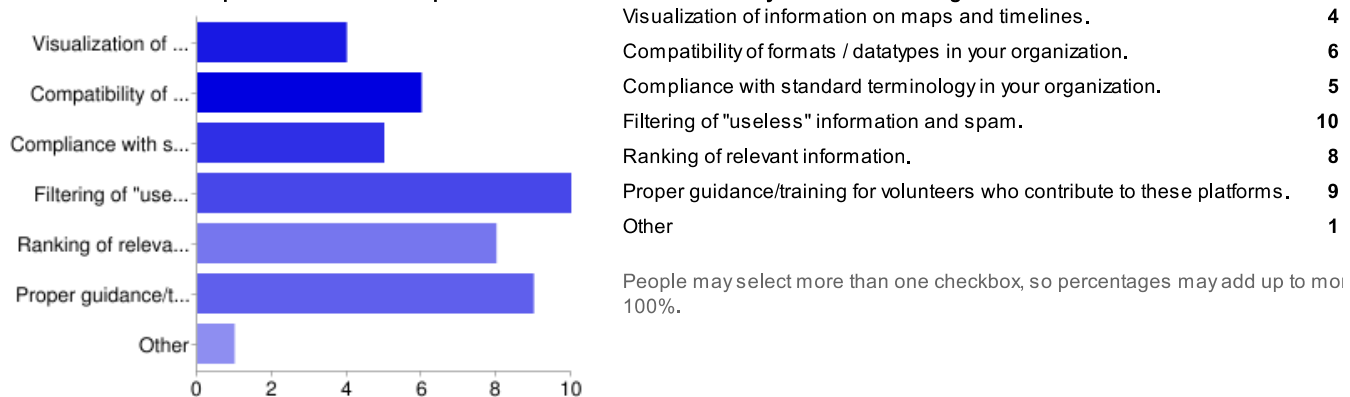


**6. What are the reasons for you NOT to use (some of) platforms/services mentioned in question no.4?**



People may select more than one checkbox, so percentages may add up to more than 100%.

**7. What needs to be improved to make these platforms/services more useful for your disaster management tasks?**

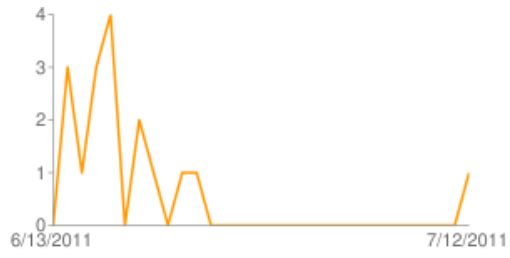


People may select more than one checkbox, so percentages may add up to more than 100%.

**8. Please reflect on your own experiences, problems and suggestions with respect to the mentioned platforms/services and the importance of linking data across different platforms/agencies:**

While the emergence of these Volunteer technical committees is exciting and represents an opportunity, I have yet to see an example of any operationally interesting or useful information being produced. It all seems to be a lot of "noise" It looks to me like the designers of this survey already have an idea of the problems and potential solutions pertaining to social data gathering. To me the short of the problem is that socially collected data is unverifiable and unstandardized - the platforms themselves are quite valuable if the community can be controlled, trained, or closed. In emergency ...

**Number of daily responses**





d. Trust of information source: How much do you trust the given source of information in the first week of a disaster? (rate 1 worst 5 best)

Say Hashi  
13:10

- i. Government -
- ii. Traditional humanitarian agency -
- iii. Directly from affected Population -
- iv. Social media or crowdsourced eg. Ushahidi, Shaha
- v. Others (Comments):

core info. (4/5)  
no choice  
family / place  
represent what  
pass info to  
concerned

11:00

① Standard data → common data sets before survey → definition

e. Formats of information collection and dissemination:

- i. In what format do you collect data?
- ii. In what format do you disseminate data?

depends on user.

17 mins.

\* Qualitative → anything email, anything  
Quantitative → words / excel → "nightmare"

depends on medium  
pictures / colors

f. If the government provide you a right set of data or right format of data, how much of an impact will it have in your work i.e. by what percentage?

20 mins

↳ methodology is clear  
→ if trust "80%" → depends on who you work with

link "toemile"

using excel  
→ no data in 2 weeks?

3 star data

2. Do you also provide raw data for partners, if yes in what format?

COCHA → "Yes" → raw data Pop data  
→ minimum data, geo boundaries → also shv.

24 mins

3. Information Collection interoperability and heterogeneity:

a. When you compile information, have you faced problems where different agencies had different definitions of indicators, for example - "CollapseStructure" or "AffectedPopulation" etc?

- i. Yes: Can you please elaborate
- ii. NO. It is part of the game.

documentation in more

\* common definition before you do survey

b. How much of your personal workload will be save if all the agencies were using indicators that would bear similar meaning throughout humanitarian response?

29 mins.

→ 40% → 39, HH data vs. persons

→ 60% → cleaning data → 25 mins

more direct impact in work. reduction → validity

\* 3000 HH, 3000 people

Raw data problem with "property rights" / protection

Expert Survey: Two new Sources of information

more exploration  
how to handle massive info!

4. Information from Affected Population:

a. If the affected population are trying to directly communicate with you, how do you integrate affected population request for help in terms of what their needs are?

27:-  
Mark: best first and foremost baseline info "fantastic". Quote. 28mins.

manage with vt time consumption

5. Crowd information:

a. Do you also use the social media reports or crowd sourced technologies to support your information compilation

- i. No. Why?  
ii. Yes. How?

Major Shift 28:00

crowd help

b. In the absence of a reliable government data or any other humanitarian agency reports, would you consider using crowdsourced or crowd generated or social media compiled information?

- i. Yes. Why?  
ii. No. Why?

No  
Even reliable -> still  
"Quick and Dirty"  
30mins assessments -> priority

Eq. ranging total of population

Expert Survey: Integration of crowdsourced information with tradition crisis and Disaster management Information using Linked Data

6. Affected Population, Crowd communities and traditional humanitarian information integration: A joint report published by UNOCHA, Harvard Humanitarian Initiative, UN Foundation titled "Disaster Relief 2.0: The future of information sharing in humanitarian emergencies" pointed out that "The international humanitarian system (in Haiti Earthquake response) was not tooled to handle these two new information fire hoses—one from the disaster-affected community and one from a mobilized swarm of global volunteers. (Harvard Humanitarian Initiative et al. 2011)page 10". Integration of such new found information is the core research question of this master thesis.

a. As a humanitarian information manager how important is it for you to know what affected population are in need of? (1less important -5very important)

→ "6" → quote

→ with in OCHA

Care →

depends! — "Context specific"

Reliability

Eq. UNHCR. working on similar approach. Human right violation

31mins

cannot rely but needs checking  
"6"

- b. If a crowd of volunteers (somewhere in the world), including but not limited to, expert disaster information managers and humanitarian officers with academics provide you with a solution on how to integrate crowd generated data with that of your traditional data, would you be interested to use it as a part of your information analysis of needs and gaps in the first week of a disaster (when credible information is very low)?

"Yes;"

→ OSINT and others a lot on geo info  
 37 mins. → a lot of students also helped with databases construction

- c. It is a known fact that in a disaster - information manager works for 98% of the time and sleeps for the rest 2%. To tap into new sources of information (affected population direct reports and global network of volunteers), if there is a way to visualize in almost real time who/what/where the affected population are, would you be willing to explore such platform?

49 mins  
 "draw"

→ monitored internet is working ⇒ "Tool" if available → great help.

7. To facilitate such integration this research proposes the concept of linking data using principle of Linked Data (Berners-Lee 1998). Based on such needs, I am proposing following "Humanitarian Needs and Response visualization". The research proposes a prototype to integrate information from the Crowd i.e. Ushahidi platform with that of traditional humanitarian agency like Shelter Cluster Partners. Please check the given demonstration at [http://ifgibox.de/m\\_limb03//minu-haiti-response/HTMLPage1.htm](http://ifgibox.de/m_limb03//minu-haiti-response/HTMLPage1.htm)

tested with students as well.

- i. "shelterNeeded" data are captured from Social Network (Ushahidi)
- ii. "ShelterOffered" data represented by Shelter Cluster Agency reports

- a. Does the information of population needs in connection to where the shelter have been provided make sense to you?

① problem was road/landslide → 3-4 days city was cut off  
 ② quite early the need ← Port Au Prince sucking in

logistic.

- b. The whole visualization is based on a set of disaster management vocabulary called Management of a Crisis (MOAC) Vocabulary.

[http://ifgibox.de/m\\_limb03/moac/moac2012.html](http://ifgibox.de/m_limb03/moac/moac2012.html)

- Does the vocabulary cover basic taxonomies based on "Humanitarian Cluster Approach"?
- Does the vocabulary cover basic humanitarian need and response taxonomies on "who what where" and when?
- Does the vocabulary cover basic humanitarian need and response taxonomies related to Shelter Cluster Needs and Response?

\* expert question on how to use? I'm interested!  
 → test with students

eg. → WFP → after 1 hrs / people started to start moving out  
 → "hot meal" day 1 → tents → help  
 → "health clinic" on day 1

Mustor / PhD

3W

4W

NEED/RESPONSE  
GAPS

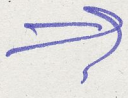
LINKED DATA

ENTHUSIAST

Linked Data Enthusiasts  
Survey SAMPLE

NO: 03.

Questions:



## Integration of Crowdsourced Information with Traditional Crisis and Disaster Management Information using Linked Data

This paper reviews to what extent Linked Data can support information management in a humanitarian decision making processes during early stages of a disaster response. On the immediate onset of a disaster, there is little or no reliable information especially about the affected population and the presence of the relief agencies including its services. At the cost of time and people in need of help, decision makers are faced with uncomfortable choices between acting on less reliable information or waiting longer for additional verifications. This study proposes to link data using specific semantic vocabularies for the crowd generated content as well as information extracted from agency specific assessment reports. The research explain the basic technology and knowledge about how disaster information managers can identify 3W that is which organizations (Who) are carrying out what activities (What) in which locations (Where) using linked data. The study also discuss the underlying concepts of corwdsourcing 3W linked data, including quality issues and give some concrete examples using emergency shelter cluster coordination during Haiti earthquake 2010 and Libya Crises 2011. Finally the research explores the future of disaster information management and crowdsourcing linked data, the usefulness of management of a crisis vocabulary and gives some concluding remarks.

Linked Data Crowd Experts Questionnaires:

Needs/Response

1) When was the first Emergency shelter Needs reported and where?

✓ option 1: Port Au Prince

Option 2. 67 Thor Ave, Port-au-Prince, Haiti and Jacmel, Haiti

Option 3. Jacmel, Haiti

(NEED)

2) Based on the given data, please indicate which was the first agency to provide shelter, where (Location) and when (Time)?

✓ option 1: Agency name: Food for the Hungry / Shelter provided at Location: Delmas /Date 13 Jan 2010

option 2: Agency name: Oxfam / Shelter provided at Location: Petit-Goave /Date 14 Jan 2010

option 3: Agency name: Food for the Hungry / Shelter provided at Location: Delmas /Date 14 Jan 2010

option 4: Agency name: Oxfam / Shelter provided at Location: Petit-Goave /Date 15 Jan 2010

(RESPONSE)

3) In the given reporting period, please indicate the number of locations where Oxfam (Agency) provided shelter assistance.

option 1: only in 1 location

✓ option 2: more than 5 locations

option 3: not sure

option 4: Stupid visualization, it doesnt make sense

WHO WHAT WHERE

4) on 14 Jan 2010 (date), how many shelter needs were reported and where?

TARGETED NEEDS

→ BY DATE

- option1: 1 shelter need report / location: 1 location at Jacmel
- option2: 4 shelter need reports / location: 1 location at Port Au Prince
- option3: 2 shelter need reports / location: 2 locations at Port Au Prince and Jacmel
- option4: Not sure

*Need/Response gaps*

5 From timeline graph at the bottom of the map, can you indicate how many days there were NO reported shelter offered on the first week of Haiti Disaster i.e from 12 Jan 2010 - 19 Jan 2010?  
(Please uncheck all the options and click on ShelterOffered)

- option1: In the first week No reported shelter distribution for: 3 days - Jan 15, Jan 16 and Jan 17
- option2: In the first week No reported shelter distribution for: 1 day - Jan 13 2010
- option3: In the first week No reported shelter distribution for: Shelter Offered everyday
- option4: In the first week No reported shelter distribution for: Visualization doesnt provide clear indication

6 From the timeline graph, please indicate the first reported date of shelter distribution (Cluster Response:ShelterOffered) in Leogane (Location)?

- option1: First reported shelter distribution in Leogane: 13 Jan 2010
- option2: First reported shelter distribution in Leogane: 14 Jan 2010
- option3: First reported shelter distribution in Leogane: 18 Jan 2010
- option4: First reported shelter distribution in Leogane: 22 Jan 2010

7 From the timeline graph, please indicate the first reported date of shelter needs (Cluster Response:ShelterNeeded) in Leogane (Location)?

- option1: First reported shelter distribution in Leogane: 13 Jan 2010
- option2: First reported shelter distribution in Leogane: 15 Jan 2010
- option3: First reported shelter distribution in Leogane: 21 Jan 2010
- option4: First reported shelter distribution in Leogane: 22 Jan 2010

8 From the timeline graph, please indicate the first reported date of shelter distribution (Cluster Response:ShelterOffered) in Jacmel (Location)?

*2017 - Feb 15 year?*

9 From the timeline graph, please indicate the first reported date of shelter needs (Cluster Response:ShelterNeeded) in Jacmel (Location)?

*2017  
Feb 08  
year?*

- No year in the timeline
- It'd be nice to have filters for locations
- Extended text in tooltips for abbreviations

## 11. Declaration of Originality

This is to confirm and certify that the Master of Science Geospatial Technologies thesis titled "*Integration of Crowdsourced Information with Traditional Crisis and Disaster Management Information using Linked Data*" is completed under the supervision of indicated supervisors from the three Erasmus Mundus consortium Universities in Portugal, Spain and Germany (in acknowledgement). To the best of my knowledge and belief, I confirm that this is the author's original contribution. Efforts have been made to clearly acknowledge works of other researchers or/and individuals by including citations and if/when possible by providing URL links of both published and unpublished sources. Should you find similar conceptual similarities, it is pure coincidence and unintentional.

With this declaration, I also confirm that the given thesis topic has not been previously submitted to other universities or to any other institution. However, attempts will be made in future to publish the finding of the research in a journal or a conference proceedings/workshops.

Thank you so much in advance for your good wishes, for your understanding and for your kind cooperation.

Sincerely,

Minu Kumar Limbu /minulimbu@gmail.com

Signed \_\_\_\_\_

Date:





Masters  
Program  
in **Geospatial  
Technologies**

**Minu Kumar LIMBU**

Supported by:



Education and Culture

**Erasmus Mundus**