2010 International Conference on Intelligent Networking and Collaborative Systems

Crowd-sourcing and Data Mashups challenges

A mini case study for assisting and solving a disaster management scenario

Stelios Sotiriadis[1], Nik Bessis[1], Eleana Asimakopoulou[1,2], Fatos Xhafa[3]
[1,2]Department of Computer Science and Technology
University of Bedfordshire, United Kingdom
[3]Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya, Barcelona, Spain
[1]{stelios.sotiriadis, nik.bessis}@beds.ac.uk, [2]eleana.asimakopoulou@googlemail.com, [3]fatos@lsi.upc.edu

Abstract— **During the past few years much effort has been put into developing community-based methods to capture and analyse a large amount of data in a systematic manner. The concept of crowd-sourcing (also known as citizen science) becomes quite popular. This is mainly due to its distributed service orientation and its group intelligence problem solving production model. Parallel to this emerging technology, Data Mashups join information from different data related services or applications available on the web for presenting a combination of functionalities among two or more services. The large amount of data generating and gathering from the client applications are hosted online and published in APIs so they facilitate an easy to integrate mechanism for future developers. Therefore, a new vision for scientists is the data information processing of Data Mashups alongside with the crowd-sourcing framework in the direction of how to store, process and benefit from collected data. In this exploratory paper we present a case-based scenario of a disaster management by highlighting how the aforementioned standards may play a significant role for solving such problems. Moreover, we discuss the emergence of paradigms and next generation technologies including Grids, Clouds and Ubiquitous and address issues related to their use in crowd-sourcing problem solver.**

Keywords-component; Crowd-sourcing, Google Maps Data Mashups, Grid Computing, Cloud Computing, Ubiquitous Computing

I. INTRODUCTION

Crowd-sourcing aims to identify highlight and solve problems within an unknown community – but bounded to a common interest, activity or practice – of distributed members. The users which are indeed the crowd of problem solvers are typically forming an online neighbourhood with usually common characteristic. We suggest that the crowd community is emerging as a potential solution to such problems. As an inceptive the winning member could be awarded with some recognition price. In parallel to the crowd-sourcing, Data Mashups offer a significant opportunity that is say the possibility for improving and adding new functionality to the community of members by mixing and matching several services aiming to create new representation of the service. The combination of a variety of Data Mashups offers a new service that was not formerly supplied by the provider [18]. However, as different contexts and service oriented purposes exist, several concerns are encountered including the manipulation of the large number of data extracted from the crowd-sourcing community collaboration. Recent studies [5, 14] discuss that Grid and Cloud computing aim to provide seamless and flexible support for processing these data by offering resource discovery, allocation and scheduling mechanisms. More specifically, their spontaneous organisation structure of cooperative heterogeneous and geographically distributed members without a fixed topology, provide computing resources on demand.

In addition, traditional large data sets including Data Mashups [18], until recently have not been satisfied by any pre-existing infrastructure [15], so the design of a computational analysis model with a rigorous performance it was a complex decision. On the other hand, the emerging paradigm of crowd-sourcing highlights the need for building such frameworks. At a glance, [16] suggests that problems can be explored at a reasonable cost as most of the people are willing to share globally their ideas towards problem solving. Furthermore, the most talented members may be awarded (offer an inceptive) so the community can feel a brand-building relationship, while the ownership belongs to the problem solver. The benefits for the whole organisation or either the service providers can be quite remarkable. Through collaboration the contributed members can extend their internal knowledge as well as organisations gain payback by knowing the customers needs and desires. Recent literature studies [17] shown that government, urban and transit planning are the major areas of crowd-sourcing and tend to provide solutions for amateurs, experts and businesses.

It is obvious that the accessibility factor as well as trust issues are always important issues, and members in many cases are mistrustful for the quality of their neighbouring participants. For this reason there are different strategies for awarding best members such as the community voting. So, the potentials of crowd-sourcing and mashups data are huge and this exploratory paper aims to clarify their usage and purposes. More importantly, we draw attention to a set of challenges aiming at collecting intelligence and knowledge from members.

In this research study, we first describe the motivation (Section II), then the related technologies definitions and their applicability to our problem study (Section III). Next,

978-0-7695-4278-2/10 $26.00 © 2010 IEEE
DOI 10.1109/INCOS.2010.44

we discuss the inter-cooperation model of Grids and Clouds and the collective intelligence application (Section IV, V). Finally, we present a case study of Clouds and Grids as a mean of forming the crowd-sourcing and Google Maps Data Mashups collaboration feature (Section VI). At last, we conclude our study with the Data Mashups challenges (Section VII) and the future work (Section VIII)

II. MOTIVATION

Most of the emerging technologies (e.g. Grid, Clouds, Uiquitous computing etc.) produce standards for heterogeneous and distributed applications. However, [14] suggest that almost all of these paradigms should be associated with each other so their combination could offer developers significant benefits. More specifically, Pervasive computing, or either known as Ubiquitous computing, alongside with Web 2.0 [13] offer the technology for capturing information, data and also, trends from the members or the people who form the crowd-sourcing community. The means to achieve it is through the everyday life devices of users such as mobile phones, laptops, and biometric sticks, etc. which contain build-in sensors. More recently, several companies have proven their willingness to extend mobile devices functionality and support the concept of collective intelligence. A remarkable example is Apple iPhone device which contains light sensors, as well as wireless antennas and a GPS chip.

As the amount of data which are collected from the device sensors is rising, then it becomes apparent that the systematic capturing, storing and analyzing of data poses significant challenges for the scientific community. In this direction, next generation technologies such as Grids, Clouds, Virtualization and Utility computing can play the role of a useful mediator tool among the service consumers and the providers [5, 11]. In other words, clients may utilize the Grid to solve a problem and Cloud for personalizing the problem solution. So by using crowd-sourcing business practices may change dramatically the way in which they approach clients. It forces companies to approach clients as potential partners [8], take a specific job that was once performed by employees and outsources it in the form of an open call to a large, undefined group of people, generally using the internet. However, it is apparent that the toolkits which are provided by the companies for solving all these important problems can be only utilized by expertise users and problem solvers. In this direction Data Mashups may put forward new solutions by providing simple to use APIs for amateur users. Applications for solving everyday problems are accessing by different people in the form of these APIs. In this way, both providers and consumers increase their benefits. At the one hand, resource providers gain low or non-cost employees as people are always willing to advice and solve general purpose problems. Also, companies by attracting developers they create a centre of attention for members, so they gain profit. On the other hand, members have a cheap way to create and maintain application using free APIs and data sources as well as they can benefit from this functionality in their everyday life. Finally, all these toolkits will offer repeated visitors to the service and will train the future crowd-sourcing community.

In the next section we highlight the aforementioned technology in the way of their purposed collaboration. Moreover, we clarify the use of Data Mashups in a mini case study scenario.

III. RELATED TECHNOLOGIES DEFINITIONS

Here we present a discussion of the aforementioned technologies as well as their definition and correlation. Moreover, selected technologies are illustrated in the mean of their applicability to the crowd-sourcing and Data Mashups disaster management scenario.

The fundamental technology for solving a well defined crowd problem is Grid. It has been proposed by [15] that "*the real and specific problem underlies the Grid concept is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organisations*". A VO is a group of members whose resources function as a unit. This form of distributed computing tends to be distinguished from the conventional systems, by offering a heterogeneous environment of loosely coupled connections. In the same direction [6] suggests that Grid is best described in terms of what it brings to the individuals and communities and the design of appropriate models is build upon their needs. On the other hand, over recent years the notion of Clouds has proven a significant commercial success model [11]. The commercialized distributed resources are spread across the internet and are available for purchasing from the users by offering capabilities of resource management and provisioning. It may be noted that the concepts behind Grid and Clouds can be described from a technological perspective as a novel way for a) achieving inter-collaboration among several users or distributed resources (from the Grid viewpoint) and b) a high quality and on-demand resource provisioning model involving various stakeholders (from the Cloud viewpoint) [7]. In other words, Clouds can utilise enterprise resources in the way of serving multiple users across an inter-cooperated Grid environment. Fundamentally, the service-oriented infrastructures of both technologies aim to provide a virtualization model of entities as services and the seamless interactions and integration of these services. As virtualization we define everything that can be virtualized in any environment separated by the underlying location and spread across the internet, even they are hardware or software. In general [4, 7, 12] suggest that, Cloud technology is derived from Grid, virtualization, and utility computing. Utility computing is a priced service of server's capacity that is accessed over the Grid [11]. In any case, all the aforementioned technologies by some means collaborate with each other, aiming to offer to user resources over the internet.

However, the means to achieve information processing into the everyday life of people is the Ubiquitous computing. The latter field promotes a model for engaging several computational devices and systems simultaneously.

It should be mentioned that Ubiquitous or even known pervasive computing derived from the area of ambient intelligence which refer to an electronic environment that is sensible and responsible to the presence of people. In other words, previous mentioned technology put forward an important solution framework for crowd-sourcing as it can utilize people's everyday life devices for collecting data through sensors of computers, mobile phones, and any information appliances.

However, an important concern in all of these technologies is how to collect and process the data in a wider manner (which naturally will lead to a better solution) open to more and more people. This vision can be fulfilled by the use of Data Mashups which is defined as a set of online APIs that developers use to build their applications by usually tapping different APIs to build that application. The significance of these is that companies such as Google, Yahoo, Amazon, Ebay etc. are putting APIs on the internet that internet developers can access easily, in contrast with the traditional non web based APIs which are limited to the expertise users. Figure 1 demonstrates novel applications which may be produced by an amateur group of people using the online mashups provided by companies.

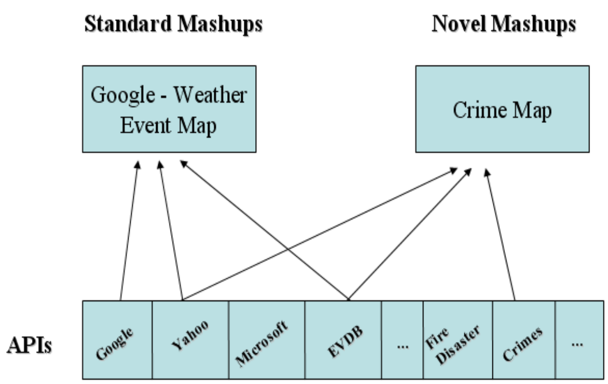

Figure 1. Data Mashups tapping map

More specifically, in Figure 1 we suppose that internet developers access the API for where crimes took place in a specific neighborhoud and use Google, Yahoo and EVDB (a database of events that are about to take place in which people can signup and collect the events) APIs and put them together into a novel API. So a new map is generated that shows where each of the crimes took place in a specific post code. So that's a mashups developer which gets different APIs from mashup websites and merging them or mashing them together in a way that forms a new innovative application that was never before on the web. Future trends show that as more APIs are showing up about new needs, then more of these mashups that tap these different needs show up, too.

Accordingly, all these needs which are raised from the Crowd-sourcing, Data Mashups and Grid, Cloud and Ubiquitous Computing are tapping in the following section which presents the inter-collaborative strategy and the disaster management case study. Collaboration of efforts offers a quick solution and in addition the cooperative environment establishes a new winning team.

IV. THE INTER-COOPERATED STRATEGY

The vision of this exploratory paper is mainly based on the collaboration of different technologies discussed in Section III, in an emerging paradigm, which support the creation of innovative applications resulted from an non-expertise crowd of users. In Figure 2 we demonstrate how technologies fulfil the gaps of each other in a meaningful and collaborative to organisation manner.

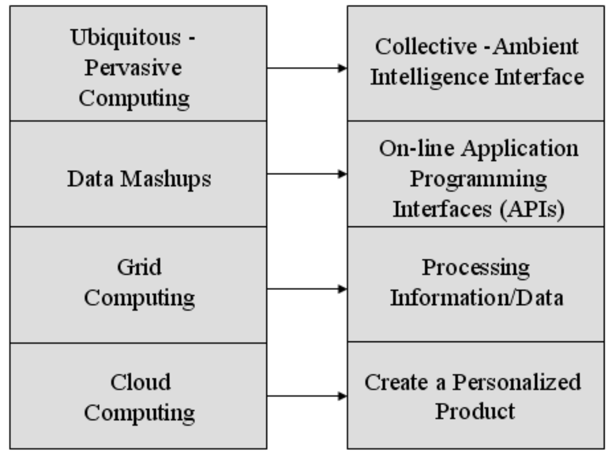

Figure 2. Collaborative technologies

In the next mini case scenario, we utilize the following:

1. A Virtual Organisation (VO), which defines a specific set of individuals or institutions which are bound around a specific set of sharing rules and actions. We extend this definition by assuming that the VO is a crowd-sourcing community of people that are willing to contribute and solve the disaster management problem.
2. A Ubiquitous-pervasive computing as the framework for collecting information from members' everyday life devices (usually mobile devices).
3. Data Mashups as the means of tapping several web-based APIs for creating an innovative application for managing the disaster scenario.
4. Grid computing as the basic framework for submitting the problems or jobs of the crowd-sourcing community VO.
5. Cloud computing as the framework to achieve profit for the companies that are willing to share with low cost their APIs. In other words, by personalizing specific problems and their solution may sell service-level agreements though negotiation between the service providers and the consumers.

In our disaster management prediction scenario mobile phones and mobile sensors can be used by ordinary "citizens" to gather data that could be useful in various settings. Finally, by using crowd-sourcing and Data Mashups we aim to provide newly created services.

V. Mini Case Scenario

Today's commercial ships face up with great number of operational requirements set by both the shipping and global community. The safety issues articulated by shipping, ecological, etc. communities are always in first line of those requirements. As the number of ships becomes denser and their speed becomes higher, the possibility for accidents also increases. In such cases a disaster planning system based on information collected from the web may offer significant benefits. The work reported herein aims to approach the complicated scenario of ships collisions by harvesting and correlating information from several online Data Mashups services. Nevertheless, there are several systems reported in the literature for preventing ship collisions [9], such systems relate different requirements to the modelling capabilities of the routes.

In our scenario we employ the marine traffic ship map [10] for identifying the nearest ships according to some requirements. More especially this API offers the following data:

1. The ship type (passenger or cargo),
2. The image of the ship,
3. The ship speed,
4. The ship length and bread,
5. The destination as well as their received information time.

Figure 3 [10] demonstrates the topological position of the ships in the map as well as Figure 4 [10] the ship characteristics.

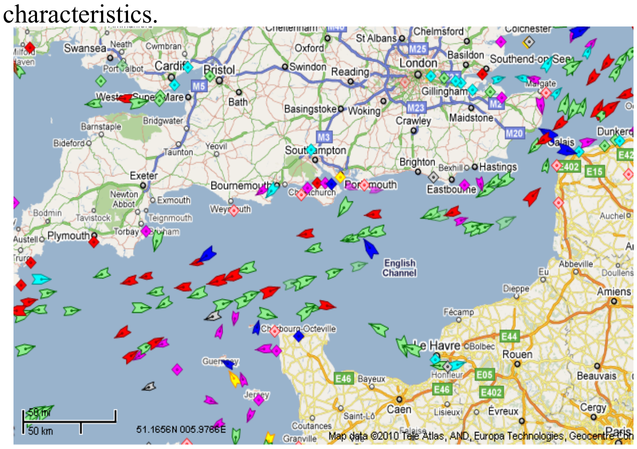

Figure 3. Marine traffic ship map

There are two major concerns that nautical engineers find working on ship collisions. The first one is the simulation of ship collisions and the prediction of their damages [9]. The second one is the identification of collision scenario(s), in order to provide the best candidate solution. The paper addresses the former of the above challenges in a way of approaching solution as a crowd-sourcing problem solving. By using Data Mashups collected from several web-based APIs we aim to collect and distribute information.

In this scenario we utilize several technologies to advice, discuss, or prevent a ship collision disaster from information gathered from members' knowledge. More specifically, we use the following frameworks:

1. Google Maps Data Mashups APIs for collecting information using the Web 2.0 tools.

2. Crowd-sourcing as a framework for achieving collaboration among the ships personnel. Members could come together, communicate and self organise into productive units. Hopefully, the quality of service produced by amateurs essentially reached an equilibrium with those created by professionals.

3. Grid and Cloud computing as the mean of storing and manipulating solutions. In a broader view, a Cloud can be seen as a customized Grid. The members forming the Cloud can access resources, solve problems similar to Grids but in a more structured, scalable and personalised management manner; as well as by be charged with a subscription cost. A Grid VO may offer to Cloud a geographically distributed environment formed under a common policy management scheme either centralized or not. Under these lenses, VO members may utilize resources to solve VO defined problems.

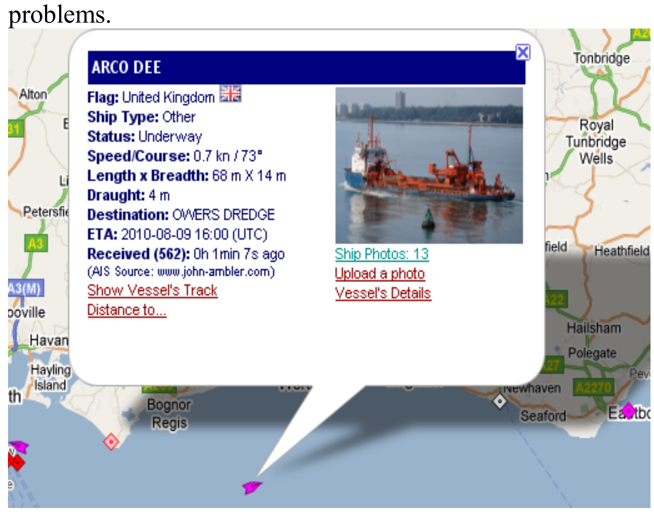

Figure 4. Marine characteristics

In this case we assume that a ship which is travelling in the area of the English Channel between Dover and Calais, which is 21 miles wide. The personnel of the ship carry a wearable body sensor capable to collect information from the environment as well as to monitor their health status. [2, 14] discuss that wearable sensors come in many forms: wrist watches, rings, smart clothes (shirts, shoes etc.), spectacles, plasters, or implanted devices (e.g. subcutaneous); and are known in different contexts or just with different usage under several generic names: wearable or body sensors or ambulatory monitoring, and extended into body sensor networks or body area networks (BANs).

We also assume that there are several sensors within the ship at different positions of the deck which can easily detect smoke, fires, the level of the sea, as well as the temperature of the day. All these data are collected from several devices such as mobile phones and send utilizing a satellite system to a central processing component for manipulation. Eventually, in a case of emergency all this data can be analysed and constructive results can be extracted for preventing accidents. In our scenario we assume that aforementioned information are collected, transmitted, and presented by a data mashup web API. The crowd-sourcing attitude of the ship personnel offers to mashups a great benefit because members of the ships can easily access those mashups in order to get and give useful advices.

The next section (Section VI) aims to discuss the architectural structure for creating a simulation model based on information gathered from the Data Mashups. That simulation model of a ship collision encompasses a number of individual problems, which should be given proper concentration [9]. More specifically the most important issue is the selection of the data, which should be sufficient enough for the results to meet and to reproduce the failure modes and common enough for solving difficult situations in acceptable time. Secondly, the inclusion of the effect of the surrounding water. In any case data should be collected by the sensors. One important issue is also the extent of the ship and its construction materials. So we present the architecture of such systems by utilizing the next Generation Technologies as well as the crowd-sourcing and Data Mashups.

VI. THE MASHUP ARCHITECTURE

Figure 5 illustrates the interactions among each component. In brief, the ship personnel are the people who constitute the crowd-sourcing, who indeed outsource their information and hopefully their problems to an undefined public body of solvers. The personnel also wear several sensors as described in previous section (Section V).

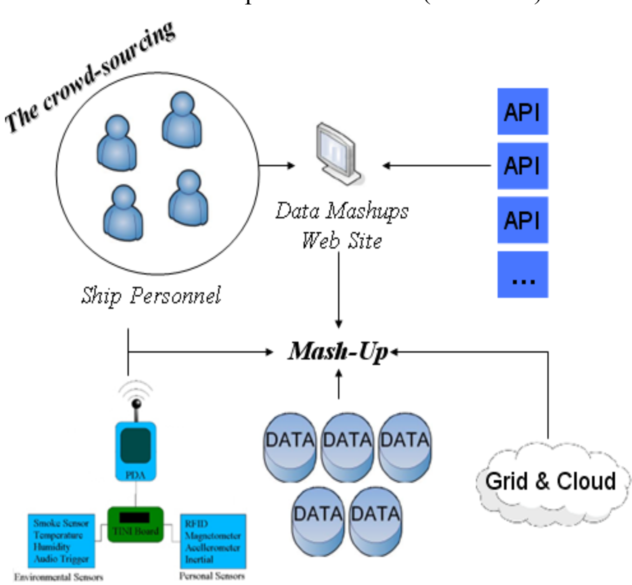

Figure 5. Data Mashups interactions

In our case the sensors send the information which can be environmental, or ship status sensors though a PDA device to the Mashup data centre. More specifically, this includes smoke sensors, temperature, humidity, etc. as well as magnetometer, accelerometer, etc. The Mashup centre server categorize the data collected from each interaction and make them available to the Grid and Cloud users for solving their computational problems. Finally, several well defined APIs are available to either experienced or novice users for creating new applications.

In our scenario we assume that a ship called Ship1 follows a specific route in the English Channel. The area is always a busy place because of the mass number of ships travelling every day. So, by using a simulation model based on information gathered from the data Mashup centre requires knowing the boundaries for making route mistakes or slightly changes to their route. The data gathered from other ships in this area are also available though the Web 2.0 tools so the ship can also predict what can happen if a second ship called Ship2 changes its route. The following low-level flow of events will start to measure data from the personnel and the ship.

1. The ship sensors will measure the ship speed, wind speed, acceleration, temperature, humidity etc. The data then passed to the mashup centre for consideration;
2. The mashup centre will measure the same values for the nearest to ships the Ship1;
3. The data will be passes to the simulation environment of the Grid and Cloud so results could be generated very fast;
4. The human wearable sensors will be able also to count the number of the personnel as well as to measure their health status;
5. The wearable sensors could be also confirm the ship location as also related data such as wind speed from people that are currently to the deck of the Ship1 and Ship2;
6. The mashup centre generates a new API for the specific condition of Ship1, Ship2, etc. which can be accessed by the ship personnel;
7. The Ship1 is awarder with some price.

In such occasions we suggest that information gathered from the aforementioned circumstances could be useful as future data and this poses the challenge of data integration and maintenance at long run. In the case of an emergency yielding to a situation of a predicted accident by the simulator the sensors will be stimulated; and this would lead to the need to create and establish a newly created ad hoc network. Several mobile APIs collect and transfer data to the processing and storage servers of Grids and Clouds. The primary challenging goal in building an ad hoc Grid is supplying each Grid member with specific directions for continuously maintain information related to each community participant, therefore act as an intermediate. Such information will be stored on each VO member public profile and be able for advertising at any time of process.

VII. DATA MASHUPS CHALLENGES

As discussed in previous sections Data Mashups are mainly based on the emerging technologies of the Web 2.0. Through those tools people can create and share everyday life information in a simple way. The big challenge for this data is to enable security and privacy features for sharing sensitive information. This task is made more difficult especially since the targeted audience is or supposed to be a general public and not expert in computing or security [18]. The last authors discuss that the challenging area of the mashup application includes three indications, the content shared from the mashup, the type of sharing and who are the crowd-sourcing users.

In general, the most of the future challenges are in relation to the enterprise Data Mashups. Likewise, as the number of APIs is rapidly increased then it is almost certain that a commonly accepted assembly model is required. This is the reason why the spreadsheets are still so enormous successful by offering a high capable functional environment. In the same direction, the management and support of the end-users is on the first line. Finally, in the case of hundreds new application there is also the need for data storage and maintenance and as suggested in [3] a general support for IT and business units.

VIII. Conclusion

The goal of the previous discussed collective intelligence is to harness the system of self-centred Grids and Clouds to secure a sustainable relationship, so coordinated individuals may solve problems more efficient [1, 14]. In general, collective intelligence of unified and synchronized crowd-sourcing Grid and Cloud communities can offer significant advantages. A clear example in nature is the ants; an individual ant is not very powerful, but a colony of ants can achieve significant results. Collective intelligence can be found in many systems, and it is known as swarm intelligence, ant colony optimization and neural networks. So, we may describe as collective intelligence, an infrastructure or environment in which individuals can do simple operations, however together they can behave in a complicated way with high intelligent level. It is almost certain, that the complexity of the aforementioned environments is high as they are formed from different resource consumers and providers connected in as loosely coupled groups.

In our study we have discussed a sufficient collaboration opportunity among crowd-sourcing, Data Mashups and Grid and Cloud computing, aiming to a collective intelligence model. As Grid is VO members utilizing resources in order to solve VO defined problems, Clouds are about users utilizing these resources for assisting user to solve problems and proposing them solutions for improving their life's. The market players should be ready for this Cloud step forward in order to improve their business by overcoming the obstacles of users' requirements complexity. In this direction the case scenario of preventing a disaster of ship collision illustrates a state of the art application that can easily be created in any case by utilizing the next generation technologies.

IX. References

1. Bessis, N. (2010), 'Using Next Generation Grid Technologies for Advancing Virtual Organisations', Keynote Talk, In: International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2010), 15-18 February 2010, Krakow, Poland
2. Paulos, E. (2009) 'Designing for Doubt: Citizen Science and the Challenge of Change', Proceedings. "Engaging Data", In: 1st International Forum on the application and management of personal information, October 12th-13th, MIT, Cambridge, USA.
3. Asimakopouloiu, E. and Bessis, N. (eds, 2010). Advanced ICTs for Disaster Management and Threat Detection: Collaborative and Distributed Frameworks, IGI Publishing
4. Mc Evoy, G. V. and Schulze, B. 2008. Using Clouds to address Grid limitations. In: *Proceedings of the 6th international Workshop on Middleware For Grid Computing* (Leuven, Belgium, December 01 - 05, 2008). MGC '08. ACM, New York, NY, 1-6.
5. Sotiriadis S., Bessis N., Huang Y., Sant P., Maple C., "From Grids to Clouds: A collective intelligence study for inter-cooperated infrastructures". In: *The Forth International Conference on Advances Engineering Computing and Applications in Science"*, (ADVCOMP), IARIA 2010, October 25th-30th 2010, Florence, Italy (To appear)
6. Winton, L. J., "A simple virtual organisation model and practical implementation", In: Proceedings of the 2005 Australasian workshop on Grid computing and e-research, Vol. 44, pp 57-65, 2005.
7. De Assuncao, M. D., di Costanzo, A., and Buyya, R. 2009. Evaluating the cost-benefit of using Cloud computing to extend the capacity of clusters. In: Proceedings of the 18th ACM international Symposium on High Performance Distributed Computing (Garching, Germany, June 11 - 13, 2009). HPDC '09. ACM, New York, NY, 141-150.
8. Pokharel, M. and Park, J. S. 2009. Cloud computing: future solution for e-governance. In: Proceedings of the 3rd international Conference on theory and Practice of Electronic Governance (Bogota, Colombia, November 10 - 13, 2009). ICEGOV '09, vol. 322. ACM, New York, NY, 409-410.
9. Servis, D., Samuelides, M., "Ship collision analysis using finite elements", Available at: http://research.gl-group.com/Projects/DEXTREMEL/publications/nantes-ntua.pdf, SAFER EURORO Spring Meeting, Nantes 28 April 1999
10. Marine Trafic Data Mashup, Available at: http://www.marinetraffic.com/ais/
11. Schubertt, L., Jeffery, K., Neidecker-Lutz, B. (2010). Expert Group Report, The future of Cloud Computing: Opportunities for European Cloud computing beyond 2010, European Commision, Belgium
12. Wohrer, A., Brezany, P., & Janciak, I. (2004). Virtalisation ofheterogeneous data sources for Grid information systems, available at:http://www.par.univie.ac.at/publications/other/inst_rep_2002-2004.pdf
13. Hintikka, K. A. 2008. Web 2.0 and the collective intelligence. In: *Proceedings of the 12th international Conference on Entertainment and Media in the Ubiquitous Era* (Tampere, Finland, October 07 - 09, 2008). MindTrek '08. ACM, New York, NY, 163-166
14. Bessis, N., Asimakopoulou, E., French, T., Norrington, P., Xhafa, F., "The Big Picture, from Grids and Clouds to Crowds: A Data Collective Computational Intelligence Case Proposal for Managing Disasters", In: Fifth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), November 4-6, 2010, Fukuoka Institute of Technology, Fukuoka, Japan, (To appear)
15. Foster, I., Kesselman, C., & Tuecke, S. (2001). The anatomy of the Grid: enabling scalable virtual organisations. International Journal of Supercomputer Applications, 15(3), 200-222
16. Evans, M., "The Power of Crowdsourcing", Available at: http://www.exinfm.com/board/crowdsourcing.htm
17. Brabham, D., (2009). "Crowdsourcing the Public Participation Process for Planning Projects", *Planning Theory, 8*(3), pp. 242-262.
18. Di Lorenzo, G., Hacid, H., Paik, H., and Benatallah, B. 2009. Data integration in mashups. *SIGMOD Rec.* 38, 1 (Jun. 2009), 59-66.