

The Big Picture, from Grids and Clouds to Crowds: A Data Collective Computational Intelligence Case Proposal for Managing Disasters

Nik Bessis¹, Eleana Asimakopoulou², Tim French¹, Peter Norrington¹ and Fatos Xhafa^{3,4}

^{1,2}Department of Computer Science and Technology
University of Bedfordshire, United Kingdom

³Department of Computer Science and Information Systems
University of London (Birkbeck), London, WC1E 7HX, United Kingdom

⁴Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya, Barcelona, Spain

¹{nik.bessis, tim.french, peter.norrington}@beds.ac.uk, ²eleana.asimakopoulou@googlemail.com, ^{3,4}fatos@lsi.upc.edu

Abstract – Much work is underway within the broad next generation technologies community on issues associated with the development of services to foster collaboration via the integration of distributed and heterogeneous data systems and technologies. Various technology-driven paradigms have emerged, including Web Services, Web 2.0, Pervasive, Grids and Cloud Computing. Recently, some new paradigms have emerged, including Situated Computing and Crowd Sourcing. In this exploratory paper, we aim to be visionary, thus, we offer an overview highlighting relationships between these paradigms; the goal is to present how these fit into the broader picture of IT. More specifically, to discuss how these could help coin and prompt future direction of their usage (integration) in various real-world scenarios. A disaster management scenario is presented to illustrate the big picture's model architecture, as well as briefly discuss the potential impact resulting from the collective computational intelligence approach.

Keywords-component; Grid computing, Cloud computing, Crowd sourcing, Collective intelligence, Ad hoc mobile networks, Disaster management

I. INTRODUCTION

Computer-based developments over the last few decades have facilitated users with numerous collaborative tools to support various levels of enquiry within the environment of their organisation or community. Specifically, the use of collaborative systems has evolved over the years through developments in distributed computational science in a manner which provides applicable intelligence to their problem-solving capabilities.

On the other hand, data integration has long been discussed in distributed computing literature reviews. Many concerns have been encountered, as most of the data systems addressed by individual systems and their applications are both heterogeneous and geographically distributed. The concern is mainly due to the different contexts and purposes for which the data systems were originally built. In other words, the main concern resides on the view that data sources have been originally produced for purposes other than their integration [1]. Another fundamental aspect which also needs to be considered is that nowadays data is available and accessible to and from a wider audience, and thus, data

must support a many-to-many exploitation type of relationship with their owners and/or on-demand users who are also geographically distributed. Thus, data is now utilized in purposes other than what was it originally produced for.

Thus, the ability to make data structures, systems and their stores interoperable remains a crucial factor for the development of these types of systems [21]. One of the challenges for such facilitation is that of data integration, which aims to provide seamless and flexible access to information from multiple autonomous, distributed and heterogeneous data sources through a query interface [16]. Moreover, the combination of large dataset size, geographic distribution of users and resources, and computationally intensive analysis results in complex and stringent performance demands that, until recently, have not been satisfied by any existing computational and data management infrastructure [9]. Rather, various technologies have been developed to address the issue of collaboration, data and resource sharing.

Most of these have emerged with the view of producing frameworks and standards to fully or partially – yet purposefully – support data integration processes within heterogeneous distributed environments. Emerging paradigms and their associated concepts highlighting their benefits include but are not limited to Web Services, Web 2.0, Pervasive, Grids and Cloud computing. Their goal is to enable an approach relevant to collective resource utilization and thus, enhance multi-user participation in functioning as a coherent unit through the use of a Cyber Infrastructure (CI). That is, to purposefully work together, collaborate and solve a well-defined problem of mutual interest from a multi-disciplinary point of view. As such, they typically enable the provision of shared and often real-time access to, centralized or distributed resources, such as applications, data, toolkits and sensors. Recently, new concepts have emerged, including Situated Computing and Crowd Sourcing. Naturally, the emergence of various technologies, which seem to address similar areas, often serves to confuse rather than to clarify their purpose and usage.

In this exploratory paper, we start off with describing our motivation (Section II) and then in Section III, we offer an

overview highlighting how these emerging technologies fit into the broader picture of IT. Most importantly, our aim is to be visionary. To achieve this, we use a disaster management scenario (Section IV) as the means to discuss emerging technologies suitability (Section V) and illustrate the big picture's model architecture (Section VI), as well as briefly discuss the potential impact resulting from the collective computational intelligence approach. Finally, we conclude our exploratory paper with possible future steps and the proposed challenges (Section VII).

II. MOTIVATION

We seek to distinguish data technologies on the basis of their concepts and purpose-wise principles. That is to say there is data over which we have significant capacity for influencing it during its creation cycle (e.g. creating a web or data site, assigning a product price, etc...) and data over which we have little or no influence at all during its creation cycle (e.g. data captured from a sensor, monitoring a patient's data, analyzing users' behaviour, etc).

In the latter approach, data is used for study purposes towards their interpretative purposes with the ultimate aim of understanding the context and thus, take corrective action if feasible and realistic. Usually, the former is used as a demonstratable effort, action and/or claim that we do understand the context. In other words, the latter comes as a prior to the former classification. Finally, we should note the existence of a hybrid approach, which incorporates a combination of the former and latter classifications as a bridge and a means to improvise the context.

Clearly, technologies related to intelligence fell under the latter classification and thus, their aim is to discover and understand the context. In this exploratory paper, we will focus on demonstrating that the emerging paradigms mentioned earlier do fall into the latter classification and thus, their combination could further advance the prospects for intelligence in their specific contexts. The following section aims to present some definitions about these emerging technologies and how they relate to each other.

III. EMERGING TECHNOLOGIES AND DEFINITIONS

Grid computing or Grid technologies have been described as the infrastructure and set of protocols to enable the integrated, collaborative use of distributed heterogeneous resources including high-end computers (nodes), networks, databases, and scientific instruments owned and managed by multiple organizations [9].

The concept of Grid technology has emerged as an important research area differentiated from open systems, clusters and distributed computing [4]. Specifically, open systems such as Unix, Windows or Linux servers, remove dependencies on proprietary nodes (hardware and operating systems), but in most instances are used in isolation. Unlike conventional distributed systems, which are focused on communication between devices and resources, Grid technology leverages of computers connected to a network,

making it possible to compute and to share data resources. Unlike clusters, which have a single administration and are generally geographically localized, Grids have multiple administrators and are usually dispersed over a wide area. But most importantly, clusters have a static architecture, whilst Grids are fluid and dynamic with resources entering and leaving. In brief, Grid can be viewed as a dynamic, enabling paradigm supporting synchronous and asynchronous resource utilization in a c-cube mode (communication, co-operation and collaboration) and it has been purposefully developed for solving well-known scientific problems (mainly by academic researchers) [5].

On the other hand, Web Services aim to provide a service-oriented approach to distributed computing issues, whereas Grid arises from an object-oriented approach. That is to say, Web Services typically provide stateless, persistent services whereas Grids provide state-full, transient instances of objects [4]. In fact, emergence of Web Services with Grid computing has resulted in a service-oriented architecture for the Grid. An important merit of this model is that all components of the environment can be virtualized, an feature which points to what is currently known as Cloud computing.

Various approaches and definitions of Cloud computing exist [5, 8, 17, 19]. All conclude that a Cloud is comprised from Grid, virtualization and Utility computing notions. [6] defines a Cloud as a type of parallel and distributed system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers.

Grids have been developed for solving scientific problems and thus security, reliability and use by non-academics were initially far away from being the primary concerns. On a similar vein, we could also point out that Grids are the first generation of this type of paradigm and thus, it would be completely unfair to expect a fully functional paradigm which would fully meet real-world business requirements. The fact that Clouds are based on Grids demonstrates Grids' sustainable robustness, as well as business value and prospects. Thus, taking a broader view we can define a Cloud as a re-factored business-oriented Grid model. Users forming the Cloud can access resources, solve problems such as in Grids, but in a well-defined robust commercialized context; offering a more structured, scalable and personalized management control; as well as by being charged with a cost [5]. In brief, one can conclude that the goal of Grids and Clouds is to purposefully utilize resources (data, computational power, software, toolkits, expertise, etc) that is available from/to Virtual Organizations (VO) partners so they can more effectively solve mainly scientific (Grid) or commercial (Cloud) problems. However, neither of these technologies should be seen as a panacea and thus, other complementary technologies should be used.

There have been various attempts to define Web 2.0. Mostly they contrast what Web functionality is offered to its users and how it differs with Web 2.0. This distinction is a key to understand where the boundaries are as between 'the Web', as a set of technologies, and 'Web 2.0', as an attempt to conceptualize the significance of a set of outcomes that are enabled by those Web 2.0 technologies [2]. In a similar vein, [14] defines Web 2.0 as a set of economic, social, and technology trends that collectively form the basis for the next generation of the Internet. A more mature, distinctive medium characterized by user participation, openness, and network effects. It is particularly important to realize that Web 2.0 (also seen as an activity regulator) offers a platform where users as individuals or communities are able to communicate online their ideas and feelings on shared topics of interests using available collaborative services (or social software). Web 2.0 and social networking services include but are not limited to wikis, blogs, photosharing, bookmarking and professional networking. In fact, there is little difference in concept with Grid based Virtual Organizations (VO). Mainly, it is the functionality offered by the technology which forms the capabilities, as well as the users' expectations of it.

Moving now to Pervasive (also known as Ubiquitous) computing, one may realize that these technology paradigms have been developed as a means to enable resource computation and utilization in a far more mobile or environmentally-embedded manner. Pervasive computing embeds computing and information technologies into our environments by integrating them seamlessly into our everyday lives [20]. Pervasive computing has many potential real-world applications ranging from health to environmental monitoring systems. It is quite common to involve a number of devices including mobile phones, PDAs, sensors and computers.

Lately, Situated Computing as an emerging paradigm deals with computing devices having the autonomous ability of adapting, detecting, interpreting and responding to the user's environment. Readers are pointed to [11] who gives a solid background of the fundamentals for Situated Computing. Situated Computing makes use of concepts from situated cognition [7].

Thus, where you are and when you are there matters, and that the state you are in affects what you do. The fundamental difference is between encoding all knowledge prior to its use and allowing the knowledge to be developed and grounded in the interaction between the tool and its environment. The effect of this is to provide a computational system such as a tool with experience based on its interaction with its environment. That experience is then used to guide future actions. The effect of this grounded experience is to provide the tool with the capability to respond differently when exposed to the same environment again depending on the experiences it has had between these two exposures. The objective knowledge within the tool remains unchanged, only the knowledge that

is the result of the interaction of the tool with its environment is changed. This provides the basis for computational systems to learn and change their behaviour based on their experiences. The learning is not necessary to improve the performance of the system rather it is designed to customize it to its user [11].

More recently a new technology driven paradigm called Crowd sourcing (also known as Crowd Computing or Citizen Science) has been introduced. Some studies have proven the potential worth of so-called "crowd-sourced" mobile phone data [5, 15]. Some of these pilot studies have shown that mobile phones and mobile sensors can be used by ordinary "citizens" to gather data that could be useful in various settings. [15] has also coined the term "citizen science" for solutions that seek to leverage collective citizen-based collection. However, participatory data collection activities of this kind and their subsequent aggregation and analysis by decision makers pose significant opportunities and challenges.

It is important to briefly define collective intelligence for the purpose of this exploratory paper. The concept of collective intelligence creates a free-flowing system of knowledge with no bureaucratic controller; it also creates an informational free-for-all where no-one decides what knowledge is worthy of contribution and what should be left out [13].

The vision of this exploratory paper is to pose both the challenge and the opportunity in bringing these next generation emerging technologies (Grids, Clouds, Web 2.0 and social networking tools, Pervasive, Situated and Crowd Source Computing) together to compute intelligently (identify data relationships, trends, etc) in a collective manner (not as now) by capturing, integrating, analyzing, mining, annotated and visualized distributed data – made available from various VO and community users – in a meaningful and collaborative for the organization (either as a VO or community) manner, that matters the organizational needs. This will in turn extend the conventional technologies bounded topology to a wider and dynamically co-ordinated community.

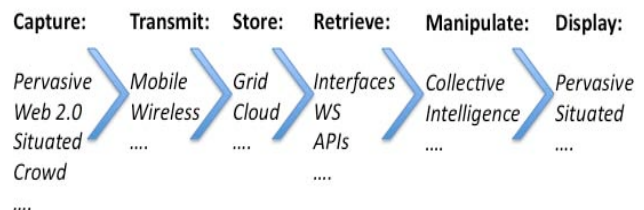


Figure 1. Mapping the Cycle of (Data) Processes with Emerging Technologies Paradigms

The relationship cycle between these emerging technologies paradigms is shown in Figure 1. The view here is that a system should combine technology to enable the processes of capturing, transmitting, storing, retrieving, manipulating and displaying relevant data supporting

peoples' effective and efficient operations, management, and informed decision-making.

IV. MINI-CASE STUDY: MANAGING OCCUPATIONAL HAZARDS

We present here a fictional yet typical disaster management case scenario, which is used throughout the remainder of our exploratory paper as a case.

In an urban area a major earthquake of some significant magnitude on the Richter scale has occurred. The area is highly populated and characterized by multi-storey buildings, such as blocks of offices, malls and other public buildings. The occurrence of the earthquake caused a disastrous situation, as some of the buildings have collapsed and some people have been injured and trapped. Further to this, a number of secondary phenomena follow the occurrence of the main hazard, such as electricity failures, fires and a series of aftershocks.

The area's civil protection department has organised the emergency operation in order to respond to the disaster. According to the area plans and to the emergency calls that reach the emergency services, operational units (OU) have been sent on site to locate and rescue earthquake victims. The members of an OU have to work as a team and to report back to the operation centre about their status and progress. OU members have to find ways to reach trapped victims within the collapsed buildings. This process is dangerous, as the stability of the affected structural elements cannot be easily assessed. Further to this, the fact that aftershocks with different magnitudes and without lead-time occur in the area makes these attempts more difficult and dangerous. For example, imagine that while members of an OU-1 are inside an affected multi-storey block of offices an aftershock occurs, which in turn results in some of the already affected structural elements of the building collapsing. Our assumption leads to a realistic scenario whereby some OU-1 members are injured and trapped inside the building alongside the originally trapped victims. Other OUs (e.g. OU-2 or OU-n) and the operation centre do not know the condition of OU-1 members: if they are alive, seriously injured, as well as their exact condition and location (positioning). The scenario yields even more uncertainties, increased workloads, pressures and problems, as other OUs have to locate and rescue their OU-1 colleagues, help assist in rescuing victims meant to be rescued by the OU-1 team as well as deliver their original rescue plan (issued to them prior to the aftershock) without compromising more lives. Rescuing OU-1 members is considered a top priority as these now-victim members are valuable personnel with significant immediate value and irreplaceable expertise in rescue operations.

V. USING EMERGING TECHNOLOGIES TO MANAGE OCCUPATIONAL HAZARDS

Natural hazard forecasting may inform those responsible for the safety and well-being of a population of the possible

outcomes should a natural event occur; where the event occurring is classed as a disaster, this moves information requirements from speculative planning about the population to actual management for and of the population.

Disaster managers will need to know where people are: there is no benefit in sending OUs to a building, however certain its collapse may be, if there is no one actually there. Equally, there is benefit in sending in an OU to a building not considered at great risk if someone injured is there.

It is easy to construct many plausible scenarios where knowledge (collected from Situated Computing and Crowd sourcing, e.g. other nearby people, sensors, and data made available from various Web 2.0 sources, etc) of the number, whereabouts and health of people in an area struck by a disaster will significantly enhance the ability of disaster managers to respond to the reality of the situation. As mentioned in Section III, various candidate technologies, such Grid, support the integration, storage and processing of data from multiple sources. Furthermore, we suggest the use of a data mashup incorporating GoogleMaps approach in illustrating the actual situation and victims' conditions. A preliminary prototype design is shown in Figure 2. There are also various references to these, the reader is pointed to [3] for further details.

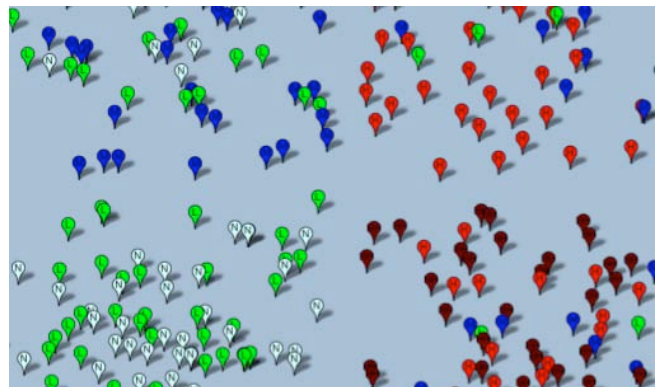


Figure 2. A Data Mashup Illustrating the Actual Situation of a Disaster

The proposal here is for OU members to utilize aforementioned emerging technologies to provide, store, process and assess data on the location and health of victims of an earthquake, namely with wearable or body sensors. 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). [10] report successful modelling for "Intra-Body Communication" (IBC), as 'a short range "wireless" communication technique [that] relies on the conductive property of human tissue to transmit the electric signal [within the] human body'. In contrast, the disposable digital plaster, which won the electronic category

in the 2007 Institution of Engineering Technology (IET) Innovation Engineering Award, “sticks to a patient’s chest and has an ultra low power wireless smart sensor in a silicon chip attached to the plaster, which monitors in real-time a range of vital signs like ECG, body temperature, respiration and physical activity” [18]. The device began clinical trials in November 2009 [12]. Applications of body sensors are now widespread: health care (in hospital and at-home care), military applications (e.g. battlefield monitoring of soldiers’ health for medical attention), athlete training, law enforcement (teams in hostile environments), tracking professional truck driver’s vital signs for fatigue, motion capture in the wider sense of biosensors, environmental monitoring and industrial process control. The signals that body sensors measure include physiological vital signs such as heart rate, temperature, respiration or ECG rate, and other biological signs such as sweat production or glucose levels.

VI. THE MODEL ARCHITECTURE

Figure 3 illustrates a low-level flow of interactions between sensors capable of scanning the environment and establishing an ad hoc mobile network.

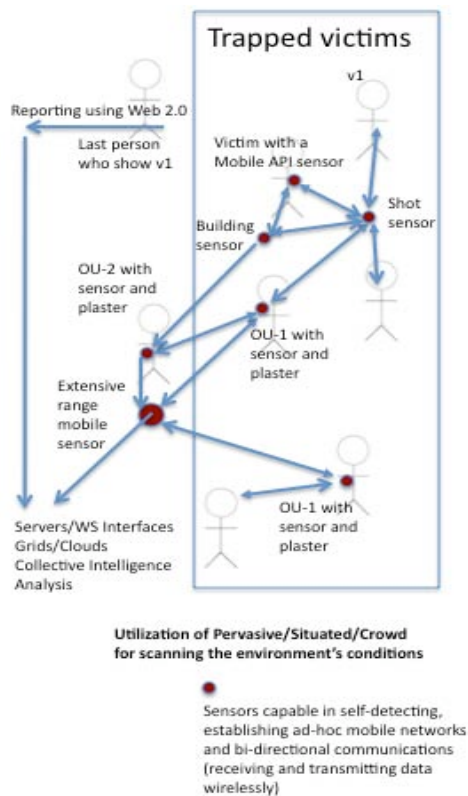


Figure 3. A Low-level Interaction Flow

In brief, the flow takes into account that during a disaster (such as of an occupational hazard) we could usefully leverage various distributed emerging technologies to visualize the status or conditions of victims who have trapped in a structurally damaged building. Specifically, we may expect someone who last saw a victim could inform

about relative positioning and/or condition of a trapped person using Web 2.0 tools. We suggest that every member of an operational rescue unit wears a plaster that records data about individual health condition, as well as a sensor that scans the environment by collecting data about it. We also assume that trapped rescue team members who are in good enough condition to do so could disperse one or more sensors so they can start collecting relevant data about the environment over a range for which their own sensor and plaster could not function and/or detect. We also assume that buildings could have installed sensors and finally, we assume that victims could have installed sensor APIs on their mobile devices. However, we do appreciate that the latter APIs would be limited in data transfer as well as in detecting and capturing a variety of signs.

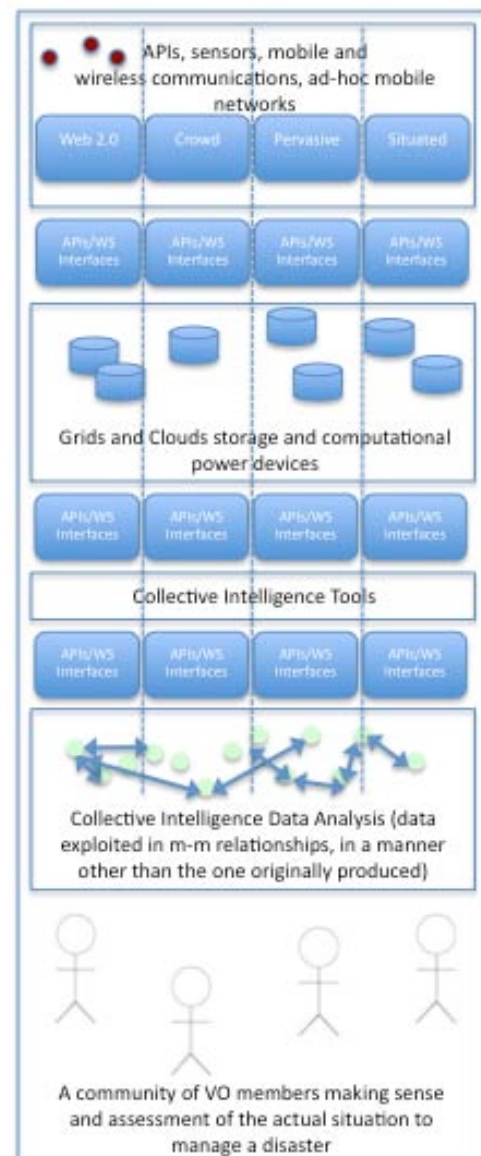


Figure 4. An Emerging Technologies Model Architecture for Managing Disasters

In this way, a trapped OU-1 member's sensor could detect other sensors available in the environment; and this would create and establish a limited ad hoc network, which will enable communication between mobile APIs, sensors and plasters with the view to transferring data across networks residing outside the building. Finally, we suggest the use of Grids and Clouds for data processing and storage, as well as the use of collective intelligence tools for their analysis.

Figure 4 illustrates a more detailed model architecture demonstrating how these emerging technologies relate and impact in realizing, making sense of and ultimately enabling a more informed decision-making based on the actual situation rather than a speculative analysis. The model appreciates that each member from the VO community may have a different domain of specialization, which requires taking into account when managing disasters and occupational hazards. In other words, like the technologies that have been developed with the view of complementing each other, limitations of individual members and their infrastructure may be satisfied from any other member. Since neither everybody nor any technology can perform all tasks, a group encompassing different resources, support technologies and individuals may utilized in a manner which will collectively cover a much larger domain.

VII. CONCLUSION AND FUTURE STEPS

In this exploratory paper, we have discussed a visionary opportunity among various emerging paradigms including Grid, Cloud, Crowd, Pervasive and Situated Computing, to be integrated for a collective intelligence model for disaster management. To achieve this, we have offered a review of emerging technologies and demonstrated the relationship between these and how they could potentially impact decision-making in disaster management scenarios including scenarios of occupational hazards.

Our future steps include the production of a technology-driven roadmap that clearly demonstrates the technical challenges and opportunities in making a realistic and feasible research agenda. Furthermore, we plan to extend this roadmap to include clear steps and be visionary for various critical infrastructures. For example, there is a need to develop complex event processing engines and a detailed specification model for the coordination between these technologies in order to achieve useful and reliable collective computational intelligence. Thus, anonymity, trust management, security and legal issues require high attention.

REFERENCES

- [1] Ahmed, E., Bessis, N., Norrington, P. and Yue, Y. (2010). Managing Inconsistencies in Data Grid Environments: A Practical Approach, *International Journal of Grid and High Performance Computing*, IGI (*in press*)
- [2] Anderson, P. (2007). What is Web 2.0? Ideas, technologies and implications for Education, JISC, available at: www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf
- [3] Asimakopoulou, E. and Bessis, N. (eds, 2010). *Advanced ICTs for Disaster Management and Threat Detection: Collaborative and Distributed Frameworks*, IGI Publishing
- [4] Bessis, N. (ed, 2009). *Grid Technology for Maximizing Collaborative Decision Management and Support: Advancing Effective Virtual Organizations*, IGI Publishing
- [5] Bessis, N. (2010). 'Using Next Generation Grid Technologies for Advancing Virtual Organizations', Keynote Talk in the International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2010), 15-18 February 2010, Krakow, Poland
- [6] Buyya, R. (2008). *Cloudbus Toolkit for Market-Oriented Cloud Computing*, available at: www.buyya.com/papers/Cloudbus-Keynote2009.pdf
- [7] Clancey, W. (1997). *Situated Cognition*, Cambridge University Press, available at: <http://cs.gmu.edu/~jgero/publications/2003/03oGerooCAADRIA03.pdf>
- [8] 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, HPDC '09. ACM, New York, NY, 141-150
- [9] 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
- [10] Gao, Y.M., Pun, S.H., Du, M., Mak, P.U., & Vai, M.I. (2009). Simple electrical model and initial experiments for intra-body communications. *Conf Proc IEEE Eng Med Biol Soc*. 2009: 697-700.
- [11] Gero J. S. (2006). *Situated Computing: A New Paradigm for Design Computing*, available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.91.4545>
- [12] ICL (2009). Digital 'plaster' for monitoring vital signs undergoes first clinical trials. News release. Imperial College London, Faculty of Medicine. 02 November 2009, available at: http://www1.imperial.ac.uk/medicine/news/20091102_digitalplaster
- [13] Lévy, P. (1999). *Collective Intelligence: Mankind's Emerging World in Cyberspace*, Perseus
- [14] O'Reilly R. (2006). *Web 2.0 Principles and Best Practices*, O'Reilly Radar Publishing
- [15] Paulos, E. (2009). Designing for Doubt: Citizen Science and the Challenge of Change, *Proceedings. "Engaging Data"*, 1st International Forum on the application and management of personal information", October 12th-13th, MIT, Cambridge, USA
- [16] Reinoso Castillo, J. A., Silvescu, A., Caragea, D., Pathak, J., & Honavar, V. G. (2004). Information extraction and integration from heterogeneous, distributed, autonomous information sources – a federated ontology – driven query-centric approach. *IEEE International Conference on Information Integration and Reuse*, available at: <http://www.cs.iastate.edu/~honavar/Papers/indusfinal.pdf>
- [17] Schubert, L., Jeffery, K., Neidecker-Lutz, B. (2010). Expert Group Report, *The future of Cloud Computing: Opportunities for European cloud computing beyond 2010*, European Commission, Belgium
- [18] Smith, C. (2007). A little plaster goes a long way, available at: http://www3.imperial.ac.uk/newsandeventspggrp/imperialcollege/newssummary/news_11-12-2007-9-27-28
- [19] Winton, L. J., (2005). A simple virtual organisation model and practical implementation, *Proceedings of the 2005 Australasian workshop on Grid computing and e-research*, Vol. 44, 57-65
- [20] Weiser, M. (2001). The Computer for the Twenty-First Century, *In Scientific American*, 265:3, 94-104
- [21] Wohrer, A., Brezany, P., & Janciak, I. (2004). Virtualisation of heterogeneous data sources for grid information systems, available at: http://www.par.univie.ac.at/publications/other/inst_rep_2002-2004.pdf