# Knowledge society arguments revisited in the semantic technologies era

## Yannis Kalfoglou

Advanced Knowledge Technologies (AKT),
School of Electronics and Computer Science (ECS),
University of Southampton, Highfield Campus,
Southampton SO17 1BJ, UK
Fax: +44-238-059-2865
E-mail: y.kalfoglou@ecs.soton.ac.uk

**Abstract:** In the light of high profile governmental and international efforts to realise the knowledge society, I review the arguments made for and against it from a technology standpoint. I focus on Advanced Knowledge Technologies (AKT) with applications on a large scale and in open-ended environments like the world wide web and its ambitious extension, the semantic web. I argue for a greater role of social networks in a knowledge society and I explore the recent developments in mechanised trust, knowledge certification and speculate on their blending with traditional societal institutions. These form the basis of a sketched roadmap for enabling technologies for a knowledge society.

**Keywords:** semantic technologies; knowledge society; semantic web; semantic interoperability; mechanised trust; emergent semantics; social software.

**Biographical notes:** Yannis Kalfoglou is a Senior Research Fellow. He received his PhD in Artificial Intelligence from the University of Edinburgh. He is working on the Advanced Knowledge Technologies (AKT) project. He was the Principal Investigator of an industrial project funded by Hewlett Packard, CROSI, which explored application of ontology mapping to industry. He is working on semantic web technologies, in particular, semantic interoperability and integration. He has published 52 works in leading journals, conferences and specialised workshops in the areas of the semantic web, artificial intelligence and knowledge engineering. He has served as Member of various programme end editorial committees for international journals and conferences, and refereed national projects for the Greek government. He organised a prestigious Dagstuhl seminar on 'Semantic interoperability and integration' and he is currently a Member of the steering committee of the Ontology Alignment Evaluation Initiative (OAEI), a large international dissemination effort for ontology mapping technology.

# 1 Introduction

The era we are living in is characterised, on one hand, by an unprecedented explosion of information that is digitised and available to large audiences through online open-ended environments, and on the other hand, by an ever more demanding society that is eager to

learn and harness knowledge for lifelong benefits. A consequence of too much information is that people are rendered less capable of making decisions – not knowing what to pay attention to and what information to act on. There is a need to manage this deluge of information and turn it into usable knowledge. This is also driven by major political, economical and societal institutions that have made a plea to realise the *knowledge society*. For example, the prestigious i2010 programme of the European Commission set a strategic goal:

> "to become the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion."[1]

Similarly, the US Government praises the role of knowledge in the 21st century society:

> "[…] knowledge is our strongest insurance for preparedness. Without a combination of old and new knowledge, we cannot develop foresight. As we evolve increasingly into a knowledge-based society, our economic growth, our national security, and our social well-being will depend on the most advanced discoveries in every field. Knowledge is the bedrock."[2]

In the same fashion, governments of other technologically advanced countries have made similar pleas to their people.[3] The idea behind the term *knowledge society* is to stress the fact that the most valuable asset is investment in intangible, human and social capital with knowledge being the main driver for realising that capital.

On the technology front, we also witness a stable growth and establishment of *semantic technologies*. This term is often used to refer to technologies that aim to mechanise the notion of *meaning* which is then embedded into systems in order to automate interoperability and enable intelligent tasks. The most visible example of this sort of work is the ambitious extension of the world wide web, the semantic web (Berners-Lee et al., 2001). Semantic technologies span across a wide range of disciplines and applications and independent surveys advocate that they are already a core part of the Information and Communication Technologies (ICT) industry worldwide:

> "markets for semantic technology products and services will grow 10-fold from 2006 to 2010 to more than \$50 billion worldwide. From 2010 to 2015 the semantic market is expected to grow nearly ten-fold again, fuelling trillion-dollar economic expansions worldwide" (Davis, 2006).

These two much debated aspects of life are, arguably, intertwined in that they support and depend on each other. For example, it is argued that knowledge society is not directly associated with traditional ICT sectors but

> "[...] covers every aspect of the contemporary economy where knowledge is at the heart of value added - from high tech manufacturing and ICTs through knowledge intensive services to the overtly creative industries such as the media and architecture" (EU, 2004a).

And the trends show that this relationship will increase steadily in the near future:

> "[…] up to 30% of the [EU] working population are estimated in the future to be working directly in the production and diffusion of knowledge in the manufacturing, service, financial and creative industries alike."

In this paper, I will elaborate and examine the arguments made for a knowledge society from a technology stance. I will focus on semantic technologies with emphasis on the most prominent products in this category, like ontologies and semantic web applications.

The experiences in using these technologies are drawn from the interdisciplinary research collaboration Advanced Knowledge Technologies (AKT)[4] which is funded by the UK Government. It is a timely choice as AKT represents one of the few nationally funded research projects that aim to investigate the application of knowledge and semantic technologies in the real world, with emphasis on the semantic web environment. The aim of this survey style paper is to identify and sketch a roadmap for semantic and knowledge technologies that could benefit knowledge society and materialise some of its promises.

This paper begins with a review of the arguments made for and against a knowledge society in Section 2. The focal point is to examine the applicability of the desire of policymakers to avoid digital exclusion and enable greater social cohesion (Section 2.1). In Section 3, I look into the application and benefits of semantic technologies by examining their role in the public domain. This leads to an experiences' log where I elaborate on the uses and feedback from applying semantic technologies (Section 3.1) before fleshing out the technology roadmap in Section 4. This study concludes in Section 5. The structure of this paper is depicted diagrammatically in Figure 1.

**Figure 1** The structure of this paper: Section 2 focuses on the arguments made for and against the knowledge society; Section 3 taps on the role of semantic technologies with emphasis on their main deliverable, that of interoperability; Section 4 delves into detail about their contribution to knowledge society
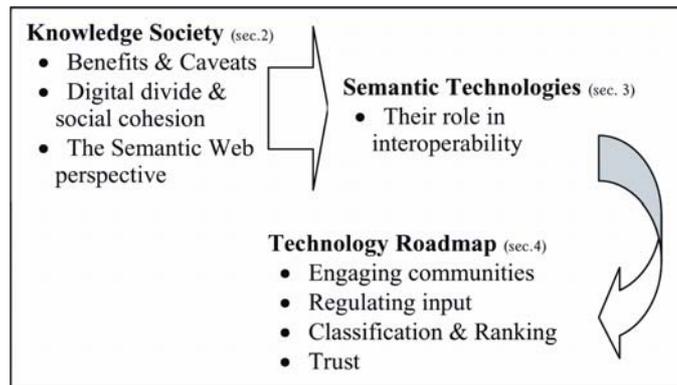


## 2 Knowledge society: benefits and caveats

Knowledge society is an inevitable fortune that has fallen into our lives following the fast development of ICTs in the past decade. As ICTs reached a stage that our way of working and living could be directly influenced and enhanced by using some sort of ICTs (from the mere use of mobile phones to the advanced applications of VPNs), the idea of injecting our society institutions with ICTs brought to bare capabilities that enable the so-called *wired citizenry*:

> "[the public uses Internet more than ever before] 40 million Americans went on-line to look at Federal, State and local government policies, and over 20 million used the Internet to send their views to governments about those policies. This and similar data show that if the U.S. government can harness the power of technology, it will be meeting expectations of an increasingly wired citizenry."[5]

The decreased cost of consumer ICTs made them more affordable for the masses which resulted in an explosion of internet use. That fact alone contributed to the *infosmog* phenomenon (too much information available) and consequently highlights the need for knowledge (turning information to usable knowledge). Policymakers and institutions responded to the increasing engagement and demand of citizenry on internet access with programmes that enable people to access and use new technologies for socially meaningful processes, including microgovernance (e-government). In fact, e-government is a major driver for pushing knowledge onto the internet on behalf of governments. Examples vary from dedicated governmental portals where citizens can pay taxes online, fill in forms and make claims of all sorts[6] to widespread policies enabling governments to pursue and establish an online interface with the electorate,[7] to the personalised features of e-Government services.[8] Governments are not the only organisations that benefited from the increased presence of citizens online. Long time before public organisations began to embrace the idea of knowledge as an asset, the private sector pioneered the use of knowledge as a capital investment (see the ageing but seminal work of Nonaka and Takeuchi (1995)), which gave birth to *knowledge management* technologies. The realisation that knowledge is the biggest capital in an organisation's assets is seen by many as a manifestation driven by demographic changes, not only technological trends. For example, in a survey examining the application of knowledge management technologies in a large US organisation, it was noted that,

> "NASA programme and project managers are retiring in increasing numbers
> [...] US Government figures project that more than half of the civil servant
> corps will be eligible to retire this decade" (Lee et al., 2005).

Common sense dictates that if these demographic statistics hold, then the dearth of experienced personnel equates lack of knowledge and that will have pernicious side effects for an organisation's functioning. The remedy proposed is to pull information out of an organisation's assets, make that information accessible and enrich it with knowledge, thus turning the whole experience of information seeking into a knowledge exploration. This is also the ultimate goal of a knowledge society, to enable knowledge exploration of a domain by its citizens. For that to happen, we need the necessary infrastructure and that is where most of the arguments made against a knowledge society are focused on.

## 2.1 *Digital divide and social cohesion*

When we consider the necessary infrastructure for a knowledge society and the means to achieve this, the most obvious thing that comes to mind is internet access or more specifically, broadband services. According to a recent survey of the European Commission:

> "broadband will be available to more than 95% of urban population in the
> EU25 by 2010, although coverage in rural areas will be much lower (achieving
> 75% by 2013 in the wealthier countries but no more than 35% in most of the
> recent accession countries)" (EU, 2005a).

Although, a credible argument made by policymakers says that the provision of internet access and broadband services should not be the only factor to decide whether citizens

are eligible for participating in a knowledge society; the scarce availability of broadband in rural areas and in low income families is often the reason for the newly emerged phenomenon of *digital divide*. This follows the historic trend on adoption and use of ICTs in general, which were first seen in central cities before dispersing slowly to peripheral cities and rural areas. The digital divide has followed this pattern, with digital services, internet access and broadband facilities being mainly available to urban areas, thereby showing new gaping divisions between digital-literate and digital-illiterate citizens.

The phenomenon of digital divide is taken seriously by policymakers and governments and we have seen high profile programmes like the European Commission's *eInclusion* (EU, 2005b) which aims to ensure that disadvantaged people are not left behind and to avoid new forms of exclusion due to lack of digital literacy or of internet access. On the other hand, the increased access and participation in a knowledge society could lead to greater social cohesion, some studies suggest. For example, one of the benefits of knowledge society and widespread adoption of ICTs is that they could act as a magnet and bring people together, a phenomenon which is evident in areas where access to the internet is restricted; local communities have set up community centres in those areas where access is free or heavily subsidised. Gatherings in those centres contribute to both digital and social inclusion. Another opportunity for social cohesion is the provision of state-of-the-art infrastructure in scarcely populated areas. As these traditionally seem attractive for families with young children living hectic lifestyles in busy urban centres or businesses wishing to expand to virgin markets, accelerating broadband deployment in those areas could bring benefits for both sides. Governments, policymakers and businesses could benefit as service providers whereas people will benefit from a smooth participation in the knowledge society while living in a tranquil countryside.

There are, however, areas which need improvements for those benefits to materialise. Not surprisingly, one of the core areas is that of education. More education and training in the new technologies is needed, especially for those who currently do not have easy access or even the means. Training and equipping people with the right skills to cope with the knowledge society is a necessity. The knowledge society could also bring risks associated with social and human capital. For example, as the opportunities for businesses and people increase exponentially, so does the work load and stress. This intensification of stress and work could bring negative results and damage the image of a knowledge society as people will feel that they need to work harder and longer in an ever more competitive global market.

One of the key issues with digital divide and efforts to achieve social cohesion is concerned with the role of locality. This is in line with some interesting lessons learned from a collateral activity, that of applied knowledge management. In particular, a major survey assessing the impact of knowledge management in the Australian Public Service (APS) (Martin, 2005), concluded that organisational context and local perceptions are key enablers. Especially the issue of locality is important for a knowledge society as the local level is important in fostering participation in a bottom-up fashion. Community building and networking are the keys for boosting participation, but issues like privacy, trust and provision of services should also be addressed.

## 2.2   *The semantic web aspect*

On the technology side of the argument, we will examine in detail in Section 4 those areas that will act as enabling technologies for a truly knowledge society. One of the foundations for most of our experiments has been the emergent semantic web. Although, it is peripheral to this paper, we will briefly list a number of areas where semantic web technologies have not only made a difference in engineering, but also in social perception and participation of the web by lay users. The contributions range from technical, such as: conversion of taxonomies to OWL, use of the underlying Description Logic inference engines, allows for identification of faulty subsumptions and discovery of previously missed ones; use of axiomatic expressions in expressive versions of OWL, allows for stating criteria (rules) which could be used to filter out inappropriate data; use of OWL semantics and build-in constructs to resolve coreferences to the same individual by multiple web references (Kershenbaum et al., 2006); to various uses of semantics and the different levels of formality employed (Sheth et al., 2005); applications of the semantic web in e-Learning (Naeve et al., 2006) and the implications of learning in the knowledge society (Lytras and Sicilia, 2005); digital libraries applications using semantic web technology (Lytras et al., 2005). The semantic web also provides the technology to help us realise a "shift in the social interaction patterns from 'knowledge push' to 'knowledge pull'" (Vossen et al., 2007). This is evident in the newly emerged and trendy social networking websites where slowly but steadily more and more semantic technologies (of which the semantic web is the most prominent) are providing the underlying infrastructure.

From an engineering perspective, it is interesting to see how all the trends identified at the societal level, affect the design, development and deployment of technologies that enable a knowledge society. This is the focus of the next section which will be used in sketching the roadmap where societal and engineering insights will be correlated.

## 3   Semantic technologies

Semantic technologies is a broad term coined recently in the business domain to refer to technologies ranging from ontologies and information extraction on the semantic web to ebXML schemata and SOA-based systems. Its broadness is deliberate to bring under the same roof the work of many communities and disciplines from the academic and industrial realm with a common goal: to inject the descriptions of objects, theories, processes and associations between components in a distributed system with semantics (meaning) that will enable automated semantic interoperability between processes and services.

For the sake of brevity and conciseness, I will focus on a well studied and applicable technology that is the cornerstone for many semantic technologies, that of ontology. Ontology technology promises to enable knowledge sharing through the use of a common, shared conceptualisation of a domain. There is an abundance of theory and uses of ontology references in the literature (see e.g. Davies, 2003; Staab and Studer, 2004) hence, my focus will be on aspects that affect the applicability of this technology on a societal scale.

## 3.1 Interoperability

One of the most targeted areas when deploying ontologies is that of interoperability. As systems become more distributed and disparate within and across organisational boundaries and societal segments, there is a need to preserve the meaning of concepts used in everyday transactions of information sharing. The emergence of the semantic web has made these transactions, arguably, easier to implement and deploy on a large scale in a distributed environment like the internet. However, it also poses interesting challenges: we observe that the demand for knowledge sharing has outstripped the current supply. Moreover, even when knowledge sharing is feasible, this is only within the boundaries of a specific system, when certain assumptions hold, within a specific domain and usually is based on a single underpinning ontologies where the knowledge about the domain is codified. But, in order to realise the benefits of a truly knowledge-based society and enable semantic interoperability between different segments and services in that society, more than one ontology will be needed and this raises the issue of interoperability between ontologies themselves, which is often studied by ontology mapping practitioners. One aspect of ontology mapping is to perform some sort of mapping between ontology constructs. That is, given two ontologies, one should be able to map concepts in one ontology onto those in the other. Further, research suggests that we should also be able to combine ontologies where the product of this combination will be, at the very least, the intersection of the two given ontologies.

In an experiment to assess the feasibility of applying state-of-the-art semantic interoperability solutions, a fictitious but realistic e-government scenario was used (Kalfoglou and Schorlemmer, 2004). The idea was to have a set of agents (human or artificial) tasked to align organisational structures and responsibilities of ministries across different governments. The domain of the experiment was ministerial units in the UK and US governments, in particular, the ministries of the UK Foreign and Commonwealth Office (FCO), the UK Home Office (HO), the US Department of State (DoS) and the US Department of Justice (DoJ). The focal point of this alignment was not only the structural and taxonomic differences of these ministries, but the way in which responsibilities are allocated in different departments and offices within these ministries. The authors gathered information related to these ministries from their websites[9] where they focused on their organisational structures, based on the practical assumption that the meaning of these structures is in accordance to the separation of responsibilities. These structures were extracted from the hierarchical lists of departments, agencies, bureaus, directorates, divisions and offices within these ministries, or organisational charts and organograms publicly available on the web. But as the authors comment: "The extraction of responsibilities and their units though (typically described in a mission statement under a *what we do* hyperlink), requires an intensive manual knowledge acquisition exercise" (Kalfoglou and Schorlemmer, 2004). Upon completion of this labour-intensive knowledge acquisition task, the semantic interoperability solution under investigation (information flow-based ontology mapping system (Kalfoglou and Schorlemmer, 2003) did align a number of similar units between these four ministries. However, the cost of applying such a solution on a large scale could be considerably higher depending on the number of the units to be aligned and their complexity. Furthermore, expert knowledge is required to verify the validity of the proposed alignments as the online resources describing those ministries do not capture the subtle semantics regarding the

functionality of a ministerial unit (a free form text under the  heading 'what we do' is often not enough to understand what a unit does).

The lessons learnt from that experiment highlight the need to tackle interoperability from a holistic point of view. According to David White, Director, European Commission, Enterprize and Industry Directorate General:

> "The realization of i2010 goals will very much depend on platforms, services and applications being able to talk to one another and to build an economic activity on the information retrieved. This is what we understand as interoperability. It is complex, not limited to the infrastructure level but encompasses semantic interoperability, organizational interoperability, and even regulatory interoperability."

The semantic interoperability, White calls for, is where most of the research and development is focused today. But, the call for organisational and even regulatory interoperability is still in its infancy. Despite a relatively large body of knowledge regarding applications of this technology to organisational knowledge (see e.g. the collection on applications of artificial intelligence and semantic web technology to organisational memories – (Dieng-Kuntz and Matta, 2002) we still have little experiences and even fewer success stories of how to bring all those three – semantic, organisational and regulatory interoperability – under the same roof.

This problem is amplified by modern trends in knowledge management: we have to work more and more within distributed and open-ended environments like the web, and its ambitious extension, the semantic web. Although, this is the means to achieve some sort of interoperability, it also has a significant impact on engineering. For example, in ontology engineering, the most prominent changes are:

1    sourcing of ontologies, which is far easier today than it was in the recent past

2    the need to use more than one ontology as a result of aiming to achieve knowledge sharing in an open-ended distributed environment

3    knowledge engineering practice is difficult to enforce when dealing with outsourced ontologies and

4    the socio-political context of an open-ended distributed environment needs to be taken into account when using ontologies.

All these factors could affect the (re) usability of ontologies on such an environment. If we focus on a specific scenario where we outsource an ontology from the public domain and we use it (or more technically speaking, reuse it) in an application; we would like to know specific background information about that ontology. This sort of information could be found (semi-) automatically by using, for example, a catalogue style characterisation of ontologies that was originally proposed by Uschold and Jasper (1999) and further elaborated by Lozano-Tello and Gomez-Perez (2004) and Lewen et al. (2006) or by calling upon advanced techniques, like ontology search and ranking (Alani and Brewster, 2005). Assuming that this step is achievable, then we would like to have some sort of a formal, and mechanised way of verifying the validity of the ontology. That could range from the straightforward check of syntactic soundness of an ontology (which is easily done with popular syntax checkers, like OWL validators) to the much harder conceptual correctness which involves thorny issues of knowledge representation and domain conceptual coverage. While the former is at an advanced

stage with a variety of tools available, the latter is still at an early stage of research (Kalfoglou et al., 2004a). If we further assume that this step will also be achievable, then we would have reached a point where the remaining bits of plugging-in the outsourced ontology are algorithmic and mundane, given the import facilities of most ontology editors and APIs.

The crux of the problem is that the script outlined above is a hypothetical one. It is based on assumptions that are hard to materialise. For instance, we do not have the experiences needed to develop and deploy catalogue style characterisations of ontologies.[10] Even worse, we do not have consensual views on what should go in those catalogues and how ontologies should be classified. Furthermore, the work on search and ranking is at an early stage of research, but it could be a promising one. The most difficult issue, though, seems to be the verification of conceptual correctness of an ontology. It is a problem identified in 1990s referred to as 'metaphysical consistency' in Gomez-Perez (1996) and prevails in today's engineering efforts when we try to achieve interoperable systems. Ensuring that an ontology covers the domain at question adequately and the conceptualisation is a credible reflection of the real world is a task that is difficult to automate, prone to constant changes and should involve practitioners and domain experts, not only ontology engineers. Experience reports highlight the necessity for a collaborative approach, for example, in an interdisciplinary meeting on semantic interoperability (Kalfoglou et al., 2004b), it was argued that: "[…] domain ontologies need to be built and vetted by domain experts and scientists, as those built by computer scientists were usually rejected". Hence, ensuring that the ontology covers adequately the domain at question is an issue that should involve as many stakeholders as possible, ideally the targeted users.

These views confirm what is often regarded as conventional wisdom, that is, you need to engage the end users in order to ensure a high and speedy uptake of the technology you are offering. In the context of knowledge society and the technologies that will enable it, these users are conveniently identified under the term *social networks*. Their increasingly technical role in a knowledge society is best illustrated in the following motto: "Society goes technical, software goes social" (EU, 2005b).

## 4  A technology roadmap

In this section, I will tap on certain areas, which are keys for the realisation of a knowledge society. I will be reviewing them from the technological standpoint, but I will elaborate on their meaning in a societal context, where applicable. An illustrative figure depicting the relationships between these areas is provided in Figure 2. Communities of interest (Section 4.1) are engaged in social networking and we aim for extracting semantically rich information from these interactions. This ideally will provide us with 'emergent semantics' (Section 4.2) that could give us a computational model of these communities. We should regulate the input and classify knowledge that we extract from these communities (Section 4.3) before we apply mechanisms for reasoning about trust of the extracted knowledge (Section 4.4). Ideally, all this should be fed back to the communities in the form of services (possibly powered by semantic web technologies) for the benefit of realising a knowledge society.

## *4.1 Engaging communities*

Terms like *social software* and *emergent semantics* have become popular in a number of disciplines. They are seen an opportunity – if it is properly manifested – in order to regulate a community's vocabulary and help it evolve to a stable version. Examples of communities of that sort are, for instance, web users who had begun using social software such as FOAF,[11] Flickr[12] or del.icio.us[13] few years ago for leisure and as a socialising medium. This kind of software and social networks have seen an unprecedented growth recently with at least a couple of dozens of tools like that. The interesting development that makes this new area appealing for engineers,[14] is that they could be used to alleviate some of the tough problems engineers face today. The premise of the argument is that as communities interact and form up, their members are contributing, knowingly or unknowingly, to communal knowledge assets. Although, this old adage has been acknowledged in the past in neighbouring domains, like communities of practice (Wenger, 2000), today's online social networks on the web has a different form.

**Figure 2**    A technology roadmap for enabling technologies for a knowledge society: *emergent semantics*; *regulating input*; *interoperability (services)* and *trust*
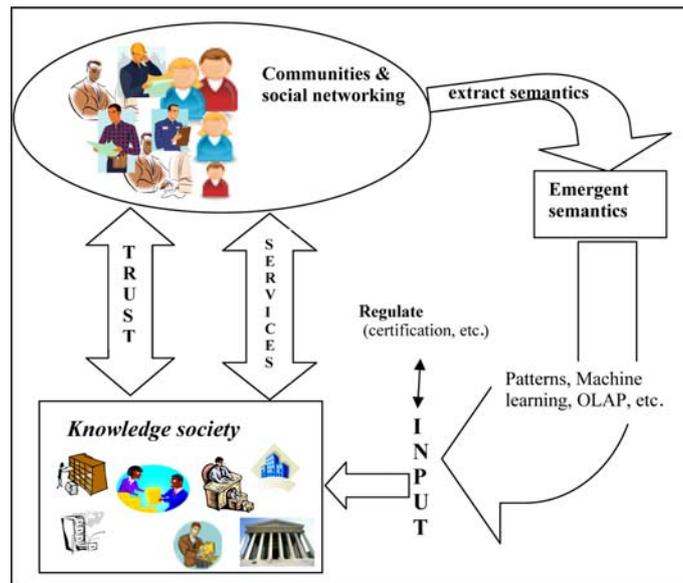


Nevertheless, it is possible to use current technology, like machine learning and text engineering, to extract the meaning (semantics) that underpins information exchange in these networks and then collectively represent it as a community's underlying semantic model. This way of uncovering the semantics used by a community's members is often described as *emergent semantics*. The difference with traditional knowledge engineering is that we no longer have a predefined semantic model (in the form of an ontology) of a domain to which a community adheres to, but rather, we have an emergent model that surfaces while a community functions and its members interact.

This has been put to work in certain scenarios and communities. For example, Zhdanova (2005) presents an approach where input from the community could be used to

validate ontology mapping results, whereas in Dzbor et al. (2005), a range of user-driven applications are discussed. But, to the best of our knowledge, it has not been applied yet to more complex, subtle domains and tasks where a stable and reliable semantics are necessary. However, early steps towards this direction are outlined and examined in Noy et al. (2005).

There are some weak areas of *emergent semantics* that could be potential show-stoppers, if not considered properly. Communities of social software users most often work on the principle of self-organisation and it is more likely to reach local consensus first, before achieving the desired intercommunity interoperability. This is a necessity before any community input can be reused in similar cases. Furthermore, the promise of *emergent semantics* relies on a relatively smooth and uniform representation of a community's interest. But, real world practice in similar domains tells us that communities use a variety of norms and manifestations (Alani et al., 2003). If *emergent semantics* are to be extracted from a community's log, then we would better have a stable log with a standard vocabulary to work with. The more variations we encounter in the semantics used, the more difficult it will be to extract them in an open-ended distributed environment.

Another area where little work has been done is to resolve possible conflicts on representation of common knowledge used by these communities. It is not uncommon for similar communities to use seemingly similar representations of the same domain concepts. Even if the sheer number of members is a strong indicator that the concepts used are prevailing and should be part of a prospective ontology, the fact that there could be slight variations in representing these concepts means that we might have to do some sort of mapping and alignment first before using them. We might even have to identify potentially conflicting concepts and we do not currently have the technology to do conflict resolution in an automatic manner (Alani et al., 2002).

*Emergent semantics* are also heavily engineered with machine learning technology and other statistical techniques, like OLAP. We do not despise the use of these technologies, but we are sceptical about the practicality of deploying machine learning to capture and extract prevailing semantics from a community's log of information exchanges in an environment like the web. Supervised learning methods are probably the most reliable ones, yet, they are the ones that need a lot of input from an engineer to ensure that the right learning data set is fetched into the system, update the learning strategy, maintain the learning rules, etc. All these are time consuming tasks and require specialised knowledge of not only machine learning, but also of the domain at question. We do not see how this could scale up at the level of an e-government portal when we will have to deal with thousands of outsourced domain concept definitions where only a human expert in that domain will be the most qualified resource to interpret the concept correctly and attach the correct semantics to it.

Hence, machine learning practitioners turned to unsupervised learning methods, that require the least possible input. But these have their own limitations as the domains to which they have been applied with considerable success, are well understood and a universal set of initial training data and rules can be easily defined or even outsourced.

Despite the advantages and disadvantages of using *emergent semantics*, we are cautiously optimistic about them. We believe that *emergent semantics* will continue to grow as more and more people will be drawn into these online communities using social software. Even in the relative virgin area of applying *emergent semantics* to knowledge

society, we have seen early work. For example, in Efron (2006) the author describes a system that is based on unsupervised machine learning and uses cocitation analysis to gauge the interest of web bloggers participating in political debates. Although, the proposed algorithm has certain limitations and deficiencies, the key point is that communities' input – in the form of web blogs – can be a valuable tool to use in a knowledge society scenario, that of engagement of citizens in politics.

Communities alone, though, will not be able to provide us with practical input unless we have ways of regulating and vetting their input. Especially, the regulatory aspect is one of the core issues in societal organisations, and we are looking at ways that this could be done on the top of extracted input from communities.

## 4.2   Regulating input

In the knowledge engineering domain, the issue of regulation has been tackled under the term knowledge certification in 1990s (Shadbolt et al., 1999) in the context of certified knowledge bases. Recently, the issue of using ontologies as a commodity, and their commercial interest has been debated (O'Hara and Shadbolt, 2001). We have also witnessed efforts that aim to certify and validate domain specific ontologies, like the work of Eysenbach (2001) with medical ontologies.

All these representative pieces of work emerged from different contexts and application domains, but point to a workable approach: certification could be done by professionals and adhere to standards and practices approved by recognised bodies of prominence. We should also point to efforts that already exist in the commercial world, especially those that apply to the web. For example, the commercial importance of the web and the sheer volume of online trade brought us technologies like SSL certificates for encrypting financially sensitive information and certification mechanisms like VeriSign's 'verified by' trademarked certificates.

Similarly, at the syntactical level, some of the W3C family of languages, and other products related with the consortium's efforts, provide instantly recognisable stickers for web pages that adhere to syntax idioms of these languages (like the 'XHTML checkers') or pointers to syntax validators and checkers where conformance to a standard can be tested instantly on the click of a button.

Despite these activities though, certifying semantic technologies – especially with respect to ontologies – remains an issue largely unresolved and ignored by big standardisation bodies. We might have witnessed high profile efforts in ontology development, like the commercialisation of CyC (Lenat, 1995) or the IEEE sponsored work on SUO,[15] but this does not mean that we have evaluation bodies that provide certificates of ontology quality assurance.

The problem with issuing certificates of ontology quality is twofold: on the technological level, we do not have a clear idea of what quality criteria and tests ontologies should satisfy in order to be accredited. On the political level, there is an issue of authority. Who will certify ontologies and how? How trustworthy will that organisation be and what, if any, will be the costs of certification? Will there be licensing issues and restrictions of use with respect to the ontology? How likely it is to reach a standardisation level when talking about ontology evaluation?

These political issues are also evident in recent work on using governmental standards (where regulation and certification is a core issue) in the UK for the sake of enabling e-government tasks. In Kelly et al. (2006), the authors highlight problems in the

adoption of standards with respect to the e-GIF[16] initiative – a high profile effort for streamlining interoperability across various governmental agencies:

> "the approach to the status of standards [adopted/recommended/under review/for future consideration] is [one dimensional approach] difficult to capture the diversity […] lack of guiding principles: failed to describe underlying principles on which the catalogue is based […] limited discussion."

The authors conclude that there is a need to recognise local, regional and cultural factors which will inform the selection and use of open standards.

Similarly, experience and industry reports on standardisation tells us that standards are hard to debate, difficult to enforce in an open-ended environment, hard to reconcile conflicting commercial interests and take years to materialise. One way of combining the strengths of emergent semantics I reviewed before and ideas from commercial efforts on certification and standards could be to use simple cataloguing techniques like classification.

## 4.3   Classification and ranking

In Alani and Brewster (2005), the authors report on early efforts to come up with ranking mechanisms that allow us to classify ontologies according to their usage. Their domain of application is on searching for appropriate ontologies, but the ranking mechanism is simple and could be adopted to support other semantic technologies. Assuming that a community is willing to participate in a common effort to rank these technologies, such an approach could provide us with a majority's view on what is best and what to avoid. This is the premise of the ranking approach.

We do, however, have certain issues to resolve before making it practical for evaluating semantic technologies, especially with a focus on knowledge society:

1   How to monitor and regulate rankings in an open-ended environment? Reports that examined well-crafted commercial efforts on using communal ranking (e.g. like the ebay feedback mechanism) has shown that it is easy to deceive authoritative systems in order to achieve personal gains (Resnick and Zeckhauser, 2002) (in the case of ebay feedback, a positive one could mean better deals for auctioneers).

2   What sort of features in the technology under consideration users will be called upon to evaluate? This issue is related to knowledge certification discussed above and we see efforts such as in Uschold and Jasper (1999) as an early step towards a consensual set of features that evaluated ontologies should demonstrate. Furthermore, in Lozano-Tello and Gomez-Perez (2004) and Lewen et al. (2006), a more detailed and extensive list of characteristics for ontology classification has been proposed.

3   Will all participating users have equal opinion weighs? For example, in a specialist's case, where a large scale medical ontology used by anatomists was to be evaluated, whose opinion will have greater importance, that of the anatomist (the domain expert) or of the ontology engineer? Common sense might dictate that the anatomist's opinion should take precedence, but there might be subtle knowledge representation issues that only the ontology engineer will be qualified to resolve.

These issues highlight another area that research needs to be focused, that of trust in open-ended networks.

*4.4   Trust*

Trust is a subtle issue. It is hard to conceive it mechanically, especially when we cannot fully explain it in its social form. It is a method of dealing, primarily, with uncertainty. For example, when dealing with independent agents – institutions or providers of resources (including knowledge) – one trusts them if one accepts their characterisation of what they will do. Trust can be a moral notion (X trusts Y to act in X's interests) or not (X trusts Y to perform some task T). Adopting the attitude of trust towards others means that one can plan and cooperate more efficiently, at the cost of greater risk of wasting resources when trust is misplaced.

In the context of semantic technologies, and their application in a knowledge society, the notion of trust takes central stage. We have already seen that semantic web practitioners place trust at the top of the proposed hierarchy,[17] a survey suggested certain policies for trust mechanisms (Matthews and Dimitrakos, 2004) which can be deployed according to certain trust strategies (O'Hara et al., 2004).

For these ideas to fully materialise and deployed on a large scale we still need to work on a variety of issues:

1   *Modelling and representation*: for trust to be used in a mechanised way, it must be first formalised and modelled in an agreed fashion. The first computational model of trust was introduced in Marsh (1994), which inspired much work in this domain. Modelling trust helps to untangle its complexity and understand its main components and evolution. However, the accuracy and completeness of such models is one thing, and their suitability and integration into multiagent systems is a totally different thing. In Giorgini et al. (2004), the authors are calling for security and trust to be added to the systems design process, rather than trying to attach them to already existing systems which might not be compatible. They argue that modelling trust and security should be closely related to where they will be applied and used, which raises questions about the actual applicability of generic trust models.

2   *Sourcing*: there is still no agreement about whether trust information is best gathered and broadcasted in centralised units, or left in a distributed format for agents to collect and reach their own conclusions. In Gil and Ratnakar (2002), the authors developed a centralised system to help users annotate sources to highlight alternative, probably contradictory, sources. User's feedback can then be used to assess information sources. Another example is proposed in Maximilien and Singh (2003) for a centralised agent to measure the reputation of web services by monitoring and collecting outcomes and client feedback, and making this information available to other agents. Such centralised systems raise the question of how trustworthy are the users or agents who provided the feedback in the first place, and why such trust warehouses should be trusted (Ding et al., 2003). In a knowledge society scenario, one would expect that trust should be manifested through centralised systems (a.k.a. e-government scenarios) and the deficiencies of these systems in the computational realm should be taken into consideration.

3   *Measuring*: trust can, and should be measured. There are a number of beneficiary areas when trust is measured, including online trading, and infer trust and reputation electronically (Mui et al., 2002). However, there is no consensus about how this can be done. In Yu and Singh (2002), the authors state the need for agents to combine their own experience with the experience of

other agents when measuring the trustworthiness of a specific agent. This raises the issue of availability and accessibility of such experience and whether trust can still be measured in their absence. Several approaches have been introduced in an attempt to extract trust values from decentralised, interconnected semantic networks. In Gray et al. (2003), the authors argue that trust can be propagated through relatively short paths connecting entities in small world networks. They believe that modern information networks will be decentralised and highly dynamic, where traditional trust control methods will no longer apply. They used an algorithm based on the *small world theory* (Milgram, 1967) to measure trust between entities that are not directly connected. Similarly, in Golbeck et al. (2003), the authors applied techniques to measure trust over a FOAF network, a social software (see Section 2.1) extended with trust relations while in Richardson et al. (2003), the authors applied path algebra and probabilistic algorithms to measure the trustworthiness of users along a network of trust connections explicitly specified by users.

4   *Context*: one major problem with many approaches proposed for measuring trust is the lack of context, which is a basic feature of trust (Ding et al., 2003). You may trust your colleague to write a good project proposal, but you might not trust him to fix your car. Associating trust measurement with specific contexts will inevitably increase its calculation and representation complexity, but nevertheless is crucial for the resulting trust values to be on any use. In a knowledge society scenario, we will see generic trust models which are customised to meet the demands of local societies and in line with their contextual requirements.

Other non-computational insights on the mechanisation of trust and its use in a knowledge society are drawn from socio-political arena. For example, in O'Hara and Stevens (2006), the authors examine the role of the central government in fostering and maintaining a high profile e-Government project with near perfect participation from its citizens. It appears that it is not solely based on a sophisticated and expensive ICT infrastructure, but rather on strong principles about the government's role in the society and the promotion of an ideology about governance. This supports the view that ICTs alone will not enable the envisaged knowledge society, but a careful blending of ICTs with existing societal and governance principles is needed.

## 5   Conclusions

The *blending* I referred to in the previous section offers the closing motto for this paper. In line with this conclusion, a European Commission report on the effectiveness of the knowledge society programme in Europe, commented:

> "the question of digital divide should be reformulated: rather than looking at who has access to ICTs and who does not, the focus should be on the value that technology adds to people's lives […] ICTs should be an outgrowth of people's desires and needs, not an imposition" (EU, 2004b).

Finding that value added is not going to be easy. It is a question that needs to be addressed as various levels; technologists and ICTs providers should make sure that their products are easy to learn and use by the masses; policymakers and regulatory bodies should ensure that phenomena like digital divide are kept to a minimum; businesses

should embrace the idea of knowledge society not only as a fertile ground for corporate profit, but also as an opportunity for innovation, customer feedback and engagement and a greater societal role.

The route to a truly knowledge society, like the one envisaged in the ambitious national and international programmes I referred to in the previous sections, is a long one. However, the rate of development of advanced ICTs provides food for thought and highlights new opportunities. What is missing is to bring closer the technologists, regulators and organisations with an interest in knowledge society. Small scale demonstrations in the context of the ambitious semantic web and an increasing number of e-government projects around the world are the first encouraging steps towards this blending of interests and a common ground for collaborative work among all stakeholders in a knowledge society.

## 6    Acknowledgements

## References

Alani, H. and Brewster, C. (2005) 'Ontology ranking based on analysis of concept structures', *Proceedings of the 3rd International Conference on Knowledge Capture (K-Cap'05),* Banff, Canada, pp 51–58.

Alani, H., Dasmahapatra, S., Gibbins, N., Glasser, H., Harris, S., Kalfoglou, Y., O'Hara, K. and Shadbolt, N. (2002) 'Managing reference: ensuring referential integrity of ontologies for the semantic web', *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW'02)*, Siguenza, Spain, pp.317–334.

Alani, H., O'Hara, K., Dasmahapatra, S. and Shadbolt, N. (2003) 'Identifying communities of practice through ontology network analysis', *IEEE Intelligent Systems*, Vol. 18, No. 2, pp.18–25.

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) 'The semantic web', *Scientific American*, pp.35–43.

Davies, J. (2003) *Towards the Semantic Web: Ontology-Drive Knowledge Management*, John Wiley & Sons, ISBN: 04708486747.

Davis, M. (2006) 'Semantic wave 2006, part-1: executive guide to billion dollar markets', *Semantic Technologies Conference*, Available at: ww.project10x.com.

Dieng-Kuntz, R. and Matta, N. (Eds) (2002) *Knowledge Management and Organizational Memories*, Kluwer Academic Publishers, ISBN: 0792376595.

Ding, L., Zhou, L. and Finin, T. (2003) 'Trust based knowledge outsourcing for semantic web agents', *Proceedings of the 2nd International Conference on Web Intelligence (WI-2003)*, Halifax, Canada. pp.379–387

Dzbor, M., Takeda, H. and Vargas-Vera, M. (Eds) (2005) *End User Aspects of the Semantic Web (UserSWEB'05)*, Vol. 137 of *WS. CEUR.*

Efron, M. (2006) 'Using co-citation information to estimate political orientation in web documents', *Knowledge and Information Systems*, Vol. 9, No. 4, pp.492–511.

EU (2004a) 'Facing the challenge: the Lisbon strategy for growth and employment', in Kok, W. (Ed). *European Commission: Report from the High Level Group Chaired*, ISBN: 9289470542.

EU (2004b) 'ICT and social capital in the knowledge society', *Technical  Report TR EUR 21064 EN, European Commission*, Joint Research Centre.

EU (2005a) 'Digital divide forum report: broadband access and public support in under served areas', *Technical Report, European Commission.*

EU (2005b) 'eInclusion revisited: the local dimension of the information society', *Technical Report, European Commission Staff Working Document (2005)206*, Part I.

Eysenbach, G. (2001) 'An ontology of quality initiatives and a model for decentralized, collaborative quality management on the (semantic) world wide web', *Journal of Medical Internet Research*, Vol. 3, No. 3, p.34.

Farquhar, A., Fikes, R. and Rice, J. (1997) 'The ontolingua server: a tool for collaborative ontology construction', *International Journal of Human–Computer Studies*, Vol. 46, No. 6, pp.707–728.

Gil, Y. and Ratnakar, V. (2002) 'Trusting information sources: one citizen at a time', *Proceedings of the 1st International Conference on the Semantic Web (ISWC'02)*, Sardinia, Italy, pp.162–177.

Giorgini, P., Massaci, F., Mylopoulos, J. and Zannone, N. (2004) 'Requirements engineering meets trust management: model, methodology, and reasoning', *Proceedings of the 2nd International Conference on Trust Management*, Oxford, UK, pp.30–38.

Golbeck, J., Parsia, B. and Hendler, J. (2003) 'Trust networks on the semantic web', *Proceedings of the 7th International Workshop on Cooperative Information Agents (CIA-2003)*, Helsinki, Finland, pp.238–249.

Gomez-Perez, A. (1996) 'Towards a framework to verify knowledge sharing technology', *Expert System with Applications*, Vol. 11, No. 4, pp.519–529.

Gray, E., Seigneur, J., Chen, Y. and Jensen, C. (2003) 'Trust propagation in small world', *Proceedings of 1st International Conference on Trust Management*, Crete, Greece, pp.102–110.

Kalfoglou, Y., Alani, H., Schorlemmer, M. and Walton, C. (2004a) 'On the emergent semantic web and overlooked issues', *Proceedings of the 3rd International Semantic Web Conference (ISWC'04), LNCS 3298*, Hiroshima, Japan, pp.576–591.

Kalfoglou, Y. and Schorlemmer, M. (2003) 'IF-map: an ontology mapping method based on information flow theory', *Journal on Data Semantics*, Vol. 1, pp.98–127, LNCS2800, Springer.

Kalfoglou, Y. and Schorlemmer, M. (2004) 'Formal support for representing and automatic semantic interoperability', *Proceedings of the 1st European Semantic Web Symposium (ESWS'04)*, Crete, Greece, pp.45–60.

Kalfoglou, Y., Schorlemmer, M., Uschold, M., Sheth, A. and Staab, S. (2004b) 'Semantic interoperability and integration. Seminar 04391 – executive summary', *Schloss Dagstuhl – International Conference and Research Centre*.

Kelly, B., Dunning, A., Rahtz, S., Hollins, P. and Phipps, L. (2006) 'A contextual framework for standards', *Proceedings of the WWW'06 Workshop on E-Government: Barriers  and Opportunities (CD-ROM), WWW'06*, Edinburgh, UK.

Kershenbaum, A., Fokoue, A., Patel, C., Welty, C., Schonberg, E., Cimino, J., Ma, L., Srinivas, K., Schloss, R. and J Murdock (2006) 'A view of owl from the field: use-cases and experiences', *Proceedings of the ISWC'06 Workshop on OWL: Experiences and Directions*, Atlanta, USA.

Lee, D., Simmons, J. and Drueen, J. (2005) 'Knowledge sharing in practice: applied storytelling and knowledge communities at NASA', *International Journal on Knowledge and Learning*, Vol. 1, Nos. 1/2, pp.171–180.

Lenat, D. (1995) 'Cyc: a large scale investment in knowledge infrastructure', *Communications of the ACM*, Vol. 38, No. 11.

Lewen, H., Supekar, K., Noy, N. and Musen, M. (2006) 'Topic-specific trust and open rating systems: an approach for ontology evaluation', *Proceedings of WWW'06 4th International EON Workshop Evaluating Ontologies for the Web*, Edinburgh, UK.

Lozano-Tello, A. and Gomez-Perez, A. (2004) 'Ontometric: a method to choose the appropriate ontology', *Journal of Database Management*, Vol. 15, No. 2, pp.1–18.

Lytras, M. and Sicilia, M. (2005) 'Knowledge society: a manifesto for knowledge and learning', *International Journal of Knowledge and Learning*, Vol. 1, No. 1, pp.1–11.

Lytras, M., Sicilia, M., Davies, J. and Kashyap, V. (2005) 'Digital libraries in the knowledge era: knowledge management and semantic web technologies', *Library Management Journal*, Vol. 26, Nos. 4/5, pp.170–176.

Marsh, S. (1994) 'Formalising trust as a computational concept', PhD Thesis, University of Stirling, UK.

Martin, B. (2005) 'Knowledge management at the Australian public service: some lessons learned', *International Journal on Knowledge and Learning*, Vol. 1, Nos. 1/2, pp.146–158.

Matthews, B. and Dimitrakos, T. (2004) 'Deploying trust policies on the semantic web', *Proceedings of 2nd International Conference on Trust Management*, Oxford, UK, pp.67–75.

Maximilien, M. and Singh, P. (2003) 'An ontology for web service ratings and reputations', *Proceedings of the AAMAS 2003 Workshop on Ontologies in Agent Systems (OAS'03)*, Melbourne, Australia.

Milgram, S. (1967) 'The small world problem', *Psychology Today*, pp.60–67.

Mui, L., Mohtashemi, M. and Halberstadt, A. (2002) 'A computational model for trust and reputation', *Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS-35)*, Hawaii, USA, pp.137–145.

Naeve, M., Lytras, M., Nejdl, W., Harding, J. and Balacheff, N. (2006) 'Advances of semantic web for e-learning: expanding learning frontiers', *British Journal of Educational Technology*, Vol. 37, No. 3, pp.321–330.

Nonaka, I. and Takeuchi, H. (1995) *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press, ISBN: 0195092694.

Noy, N., Guha, R. and Musen, M. (2005) 'User ratings of ontologies: who will rate the raters?' *Proceedings of the AAAI 2005 Spring Symposium on Knowledge Collection from Volunteer Contributors*, Stanford, CA.

O'Hara, K., Alani, H., Kalfoglou, Y. and Shadbolt, N. (2004) 'Trust strategies for the semantic web', *Proceedings of the Trust, Security and Reputation Workshop at the ISWC04*, Hiroshima, Japan, pp.78–85.

O'Hara, K. and Shadbolt, N. (2001) 'Issues for an ontology for knowledge valuation', *Proceedings of the IJCAI'01 Workshop on E-Business and the Intelligent Web*, Seattle, WA, pp.35–43.

O'Hara, K. and Stevens, D. (2006) 'Democracy, ideology and process re-engineering: realising the benefits of e-government in Singapore', *Proceedings of the WWW'06 Workshop on E-Government: Barriers and Opportunities (CD-ROM), WWW'06*, Edinburgh, UK.

Resnick, P. and Zeckhauser, R. (2002) 'Trust among strangers in internet transactions: empirical analysis of ebay's reputation system', M.R. Baye (Ed). *Advances in Applied Microelectronics*, Vol. 11, Elsevier.

Richardson, M., Agrawal, R. and Domingos, P. (2003) 'Trust management for the semantic web', *Proceedings of the 2nd International Conference on the Semantic Web (ISWC'03)*, Sanibel Island, FL, pp.351–368.

Shadbolt, N., O'Hara, K. and Crow, L. (1999) 'The experimental evaluation of knowledge acquisition techniques and methods: history, problems, and new directions', *International Journal of Human–Computer Studies*, Vol. 51, pp.729–755.

Sheth, A., Ramakrishnan, C. and Thomas, C. (2005) 'Semantics for the semantic web: the implicit, the formal and the powerful', *International Journal on Semantic Web and Information Systems*, Vol. 1, No. 1, pp.1–18.

Staab, S. and Studer, R. (Eds) (2004) *Handbook on Ontologies*, Springer Verlag, ISBN 3540408347.

Uschold, M. and Jasper, R. (1999) 'A framework for understanding and classifying ontology applications', *Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods(KRR5)*, Stockholm, Sweden.

Vossen, G., Lytras, M. and Koudas, N. (2007) 'Editorial: revisiting the (machine) semantic web: the missing layers for the human semantic web', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19, No. 2, pp.145–148.

Wenger, E. (2000) 'Communities of practice: the key to knowledge strategy', in E. Lesser, M. Fontaine and J. Slusher (Eds). *Knowledge and Communities*, Butterworth-Heinemann, pp.3–20.

Yu, B. and Singh, M. (2002) 'An evidential model of distributed reputation management', *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'02)*, Bologna, Italy, pp.294–301.

Zhdanova, A. (2005) 'Towards community-driven ontology matching', *Proceedings of the 3rd International Conference on Knowledge Capture (K-CAP'05)*, Banff, Canada, pp.221–222.

## Notes

[1]Lisbon European Council – March 2000.

[2]Colwell, R., *Director National Science Foundation*, Available at: http://www.nsf.gov/news/speeches/colwell/rc011119csspacs.htm.

[3]See, for example, the Japanese Government's plea for an 'e-Japan', Available at: http://www.kantei.go.jp/foreign/it/network/0122summary.html.

[4]Available at: www.aktors.org.

[5]US Government quote on the e-Government programme, US Office of Management and Budget, Available at: http://www.whitehouse.gov/omb/egov/index.html.

[6]Like the UK's directgov site, Available at: http://www.direct.gov.uk/.

[7]Like the EU's eEurope2005 Action Plan, Available at: http://europa.eu.int/information_society/eeurope/2005/index_en.htm.

[8]Like the Singapore's My.eCitizen programme, Available at: http://my.ecitizen.gov.sg/.

[9]Accessible from www.homeoffice.gov.uk, www.fco.gov.uk, www.state.gov and www.usdoj.gov.

[10]One could argue that Ontolingua (Farquhar et al., 1997) was such a catalogue, but the ones we are envisaging need to be richer in content.

[11]Available at: www.foaf-project.org.

[12]Available at: www.flickr.com.

[13]Available at: http://del.icio.us/.

[14]See, for example, the first ever dedicated event to social software: Available at: http://sw.deri.ie/~jbreslin/foaf-galway/.

[15]Available at: http://suo.ieee.org/.

[16]Available at: http://www.govtalk.gov.uk/interoperability/egif.asp.

[17]The semantic web layer cake, Available at: http://www.w3.org/2004/Talks/0412-RDF-functions/slide4-0.html.