The Semantic Grid: Past, Present, and Future

DAVID DE ROURE, NICHOLAS R. JENNINGS, SENIOR MEMBER, IEEE, AND NIGEL R. SHADBOLT

Invited Paper

Grid computing offers significant enhancements to our capabilities for computation, information processing, and collaboration, and has exciting ambitions in many fields of endeavor. In this paper, we argue that the full richness of the Grid vision, with its application in e-Science, e-Research, or e-Business, requires the "Semantic Grid." The Semantic Grid is an extension of the current Grid in which information and services are given well-defined meaning, better enabling computers and people to work in cooperation. To this end, we outline the requirements of the Semantic Grid, discuss the state of the art in achieving them, and identify the key research challenges in realizing this vision.

***Keywords**—Cooperative systems, distributed computing, grid computing, knowledge representation, pervasive computing, Semantic Grid, Semantic Web, software agents.*

I. INTRODUCTION

Fundamentally, Grid computing is about bringing resources together in order to achieve something that was not possible before. In the mid-1990s there was an emphasis on combining resources in pursuit of computational power and very large scale data processing, such as high-speed wide area networking of supercomputers and clusters. This new power enabled researchers to address exciting problems that would previously have taken lifetimes, and it encouraged collaborative scientific endeavors. To characterize this movement, the term "Grid" was chosen to draw an analogy with the way in which the electricity power grid brought about a revolutionary change from the use of local electricity generators [1]. In this view, computational resources, data, and expensive scientific instruments can be regarded as utilities to be delivered over the network.

As Grid computing has evolved, it continues to be about bringing resources together, but the emphasis has shifted from the earlier view—caricatured now as "big iron and fat pipes"—to the notion of Virtual Organizations, defined by Foster *et al.* in [2]:

> The real and specific problem that underlies the Grid concept is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations. The sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources, as is required by a range of collaborative problem-solving and resource brokering strategies emerging in industry, science, and engineering.

Against this background, we introduced the notion of the "*Semantic Grid*" in 2001 [3]. Through undertaking research at the intersection of the Semantic Web, Grid and software agent communities, we observed the gap between aspiration and practice in Grid computing. Our report, entitled "The Semantic Grid: A Future e-Science Infrastructure," stated:

> e-Science offers a promising vision of how computer and communication technology can support and enhance the scientific process. It does this by enabling scientists to generate, analyze, share and discuss their insights, experiments and results in a more effective manner. The underlying computer infrastructure that provides these facilities is commonly referred to as the Grid. At this time, there are a number of Grid applications being developed and there is a whole raft of computer technologies that provide fragments of the necessary functionality. However there is currently a major gap between these endeavors and the vision of e-Science in which there is a high degree of easy-to-use and seamless automation and in which there are flexible collaborations and computations on a global scale.

We recognized that this emerging vision of the Grid was closely related to that of the Semantic Web—which is also,

fundamentally, about joining things up. The Semantic Web is an initiative of the Worldwide Web Consortium (W3C).

> ... to create a universal medium for the exchange of data. It is envisaged to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data. Facilities to put machine-understandable data on the Web are quickly becoming a high priority for many organizations, individuals and communities. The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people. For the Web to scale, tomorrow's programs must be able to share and process data even when these programs have been designed totally independently [4].

To researchers aware of both worlds, the value of applying Semantic Web technologies to the information and knowledge in Grid applications was immediately apparent.

At that time the service-oriented architecture of the Grid was also foreseen, and we advocated the application of the ideas of agent-based computing (principally autonomous problem solvers that can act and interact in flexible ways) to achieve the necessary degree of flexibility and automation within the machinery of the Grid. Thus the vision of the Semantic Grid became established as the application of Semantic Web technologies both *on* and *in* the Grid [5].

In the three years that have followed since our original report, many of these ideas have been put into practice and the Semantic Grid research and development community continues to grow. In particular, the Semantic Web Resource Description Framework (RDF) and the Web Ontology Language (OWL) became W3C recommendations, and tools which support these are increasingly available off-the-shelf. Consequently, a variety of Semantic Web applications and services are now starting to appear that embody and exploit these standards [6]. Moreover, as Grid developers have found a need for interoperable metadata they too are turning to RDF, and Grid application developers in domains such as life sciences are already working with ontologies [7]—shared vocabularies which can be expressed in OWL.

The experience of working with real applications and real users has also highlighted the issues of the interface between the Grid and the physical world. Hence, a number of researchers are now beginning to explore this interface for the Semantic Grid, for example through focusing on sensor networks, hand-held devices, and human interaction with these.

The notion of agents as the entities that procure and deliver services (under some form of service level agreement) in dynamic and uncertain environments has been articulated both for the Semantic Web [8] and for complex, distributed systems in general [9]. At this time, the requirements which motivate an agent-based approach in the Grid are partly achieved through the application of Semantic Web technologies within the service-oriented architecture. The recognition that resources require automated negotiation is now also gaining attention, for example in the emerging WS-Agreement specification [10].

Meanwhile, many aspects of the Semantic Grid still contain significant research challenges, in some cases requiring the bridging of research communities to achieve them [11].

Given this background, this paper provides an update on the original Semantic Grid report, capturing the new activity and the evolution in thinking. In Section II we revisit our vision of the Semantic Grid and in Section III we discuss the requirements. Section IV provides a review of the state of the art. Semantic Grid case studies are increasingly available in the literature, and in Section V we highlight four projects which illustrate some of the new thinking in the field. After discussion in Section VI, we conclude in Section VII with a revised research agenda for the Semantic Grid.

II. THE SEMANTIC GRID VISION

The Semantic Grid vision is driven by practical requirements. The Grid is not an end in itself but a means to an end—its ultimate purpose is to realize new capabilities for the benefit of its users. Hence, the vision for the Grid is best described in terms of what it brings to the individuals and communities that use it, and the middleware design is informed by their needs. Success is indicated ultimately by advances in science, engineering, business or arts and humanities research, as well as successful middleware developments.

This user-driven, application-led approach was adopted in the $500 million, five-year, UK e-Science Programme which has funded over 100 separate e-Science projects, all of which involve one or more forms of distributed data, computation, and collaboration [12]. The requirements for the Semantic Grid were largely apparent at the outset of the programme just considering individual projects, but they have been further reinforced during the programme itself. Specifically, the coexistence of these diverse projects over a common resource and network infrastructure has created a clear understanding of the role of the middleware. Significantly, this intense execution of a large number of projects has shifted the focus away from one project at a time and emphasizes the need to maximize reuse of software, services, information, and knowledge. Thus, while Grid middleware was originally conceived to hide the heterogeneity of computational resources so that they may work together, a new Grid problem is now apparent—interoperability across time as well as space, both for anticipated and unanticipated reuse of services, information, and knowledge.

In fact, the e-Science programme has not been confined to the scientific disciplines of natural sciences, engineering, and medicine: it has also encompassed e-Social Science, both quantitative and qualitative, and extends now into arts and humanities. Internationally, the Humanities, Arts, and Social Sciences activity in the Global Grid Forum (GGF) has recognized the need for the Semantic Grid. Hence, e-Science might better be termed "e-Research" and it is characterized by the increasing scale and complexity of research endeavors as collaborations grow larger, become more geographically distributed, and involve a wider range of disciplines. This increasing multidisciplinary diversity

also emphasizes Semantic Grid requirements. Of course, collaboration is not mandatory: "lone researchers" also stand to benefit from improved resources.

Furthermore, e-Science is not confined to academic research (for example, more than 80 companies are engaged with the UK e-Science programme). This interest stems from the fact that e-Science can support competitive industrial research and development, and the middleware requirements of e-Science are closely related to those of e-business (the sharing of computational and data resources for and as part of routine business is a common problem). Moreover, as e-Science has extended to education, the need to integrate with learning management systems, or collaboration and learning environments, has also become recognized [13].

Another lesson from e-Science is the need to facilitate the deployment, configuration, test, and maintenance of tools and services. Considerable effort and expertise has been expended on this through e-Science and Grid computing research projects. Thus, as well as the focus on the e-Scientist as user, it is important to address the needs of the "user" installing and administrating the e-Science software. There is an underlying principle of liberating humans from the mundane interactions of computer systems, so that they can get on with what they are good at. In turn, the systems are becoming increasingly automated, which is how the increasing scale and complexity is addressed. However, to fully achieve this automation requires much more to be "machine-processable" and much less to need human intervention. The process of software deployment, configuration, and maintenance is itself a candidate for tools and for automation which require the application of Semantic Web technologies.

As can be seen, humans are very much part of virtual organizations and the Semantic Grid has to facilitate their collaboration, both in establishing the appropriate coalitions and in supporting interaction within them. Moreover, machine-processable knowledge about individuals enables the formation of communities of practice that may stretch over vast distances and be large scale. Both of these characteristics, supported by the Semantic Grid, enable people to achieve results that were not possible before this facilitating infrastructure.

In short, then, the Semantic Grid vision is to achieve a high degree of easy-to-use and seamless automation to facilitate flexible collaborations and computations on a global scale, by means of machine-processable knowledge both on and in the Grid.

III. Requirements

Our understanding of the requirements for the Semantic Grid has deepened as we have gained experience across a range of e-Science applications. Thus, we are now in a position to move beyond our initial analysis in [2], where we identified a set of requirements motivated by a generic scenario in which a sample is analyzed and the results feed into a sequence of analysis and processing steps. Broadly speaking, these requirements can be positioned on a spectrum: one end can be characterized by automation, virtual organizations of services, and the digital world and the other can be characterized by interaction, virtual organizations of people, and the physical world.

Underlying these requirements (at both ends of the spectrum) is the issue of scale. As the scale of the virtual organizations increases, so does the scale of computation, bandwidth, storage, and complexity of relationships between services and information. Scale demands automation, and automation demands explicit knowledge.

Given these drivers, we identify the following key requirements of the Semantic Grid.

1) Resource Description, Discovery, and Use: The system must store and process potentially huge volumes of distributed content in a timely and efficient fashion, perhaps through the federation of resources, and hence be able to identify content, services, computational resources, Grid-enabled instruments, and so on. It must be able to discover and locate these resources efficiently, and to negotiate access. It also requires generation and processing of job descriptions, and on-demand and dynamically planned use of resources in order to meet quality of service requirements and achieve efficient resource utilization.

2) Process Description and Enactment: To support the creation of virtual organizations of services, the system needs descriptions (such as workflows) to facilitate composition of multiple resources, and mechanisms for creating and enacting these in a distributed manner.

3) Autonomic Behavior: Systems should auto-configure to meet the needs of their multiple users in dynamically changing circumstances, and "self-heal" in the presence of faults (thus, the systems will appear to be reliable but, in practice, they may conceal failures and exception handling at various levels). The system should also support evolutionary growth as new content and services become available.

4) Security and Trust: There are authentication, encryption, and privacy requirements, with multiple organizations involved, and a requirement for these to be handled with minimal manual intervention. Indeed from one perspective this is what really defines virtual organizations. Related to this are issues of charging: different stakeholders need to be able to retain ownership of their own content and processing capabilities, allowing others access under the appropriate terms and conditions. By their nature, policies need to be represented such that they can be applied to multiple resources and with consistent interpretation.

5) Annotation: From logging a sample through to publishing the analysis, it is necessary to have annotations that enrich the description of any digital content. This metacontent may apply to data, information, or knowledge and depends on agreed interpretations. Ideally, in many cases, annotations will be acquired automatically. This also supports the important notion of provenance, whereby sufficient information is stored so that it is possible to repeat the experiment, reuse the results, or provide evidence that this information was indeed produced at this time (the latter may involve a third party). Obtaining annotations is only half the story however; we also need to make use of them. Examples of such usage include: finding papers, finding people,

finding a previous experimental design (these queries may involve inference), annotating the uploaded analysis, and configuring a smart space to suit its occupants. Annotation may thus be distributed and collaborative.

6) Information Integration: The ability to make meaningful queries over disparate information stores, and to make use of content in ways which may or may not have been anticipated, requires interoperability of information. For example, this may involve mapping between terminologies used in different domains. This is the classical role of Semantic Web technologies.

7) Synchronous Information Streams and Fusion: In addition to persistent stores, there is a requirement for working with streams of information in a timely fashion. These streams could be data from instruments, video, or any data stream such as that resulting from interaction. Both live and replayed streams have a role, especially where the stream is enriched by associated temporal metacontent. Notifications are also streams—the arrival of new information prompts notifications to users and initiates automatic processing. Given this, support is needed both for recording and replaying streams, and making it easy to synthesize new artefacts from them. Moreover, content needs to be able to be combined from multiple sources in unpredictable ways according to the users' needs; descriptions of the sources and content, and annotations, will be used to combine content meaningfully.

8) Context-Aware Decision Support: Information needs to be presented to users at the right time, in the right format, on the right device, and with the right level of intrusiveness. In short, it needs to be sensitive to the context and, in particular, the tasks at hand. This is an adaptive information systems task and is emphasized by the scale of Grid endeavors. It is therefore a generalization of portals such as [14]

9) Communities: Users should be able to form, maintain, and disband communities of practice with restricted membership criteria and rules of operation. This involves identifying the set of individuals in a virtual organization through collaborative tools, and exploiting knowledge about communities of practice across disciplines.

10) Smart Environments: The environment should display a degree of ambient intelligence. For example, the equipment detects the sample (e.g., by barcode or RFID tag), the scientist may use portable devices for note-taking, the Access Grid node may achieve speaker recognition, and visualizations may be available on a variety of displays.

To these original requirements we add two new ones which reflect configuration and deployment issues:

11) Ease of Configuration and Deployment: Grid applications should be deployable by nonspecialist users (rather than requiring teams of postdoctoral researchers).

12) Integration with Legacy IT Systems: The interoperability challenge of interworking with established business processes, and with learning management systems, needs to be addressed in a more satisfactory manner since Semantic Grid environments are necessarily brown field developments.

Although many of the above requirements would be regarded as traditional Grid ones, we believe that all of them stand to benefit from some aspect of the Semantic Web. To us, therefore, this is an effective illustration of the way in which the Semantic Web thoroughly permeates the machinery of the Grid.

IV. STATE OF THE ART

Now that the requirements have been identified, we turn to five of the key technologies that are being used to address them.

A. Web Services

The key to bringing structured content to life is to run services over it. What these might look like can begin to be seen in a variety of Semantic Web applications (and we will also encounter them in the case studies in the next section). In more detail, recent efforts around Simple Object Access Protocol (SOAP), Web Services Description Language (WSDL), and Universal Description, Discovery, and Integration (UDDI) enable software applications to be accessed and executed via the Web based on the idea of Web Services. Such Web Services significantly increase the Web architecture's potential, by providing a way of automated program communication, discovery of services, etc. In this view, Web Services connect computers and devices with each other using the Internet to exchange and combine data in new ways. Effectively, Web Services provide on-the-fly software composition through the use of loosely coupled, reusable software components (in contrast to the tightly coupled solutions of the past).

At the same time as the Web community began to embrace Semantic Web technologies, Web Services were achieving increasing prominence as an industry-led solution to a service-oriented architecture for e-business. Ten companies submitted v1.1 of the specification for SOAP to the W3C in 2000. Ariba, IBM, and Microsoft then published a specification for the WSDL, defining an XML grammar for describing network services, and also for the UDDI service registry.

Building on this, the Open Grid Services Architecture (OGSA) was published in 2002 [15], describing a service-oriented architecture for the Grid, and the Open Grid Services Infrastructure (OGSI) working group in GGF specified the conventions to which a Web Service must cohere to be a Grid Service. The enhancements for Grid services include the creation of new services on demand, service lifetime management, service groups, state handling, and notification.

Conceived as an enhancement of Web Services to meet the special requirements of Grid computing, OGSI was seen by some researchers to diverge from important Web Service practices, particularly in the approach to stateful service interactions [16], and a more natural mapping to Web Services was sought. Subsequently, IBM and Globus, together with a number of other companies, presented a proposal for an evolution of OGSI based on a set of new Web Service specifications called WS-ResourceFramework (WSRF) and WS-Notification (WSN) (see www.globus.org/wsrf).

B. Software Agents

Multiagent systems research addresses a problem space which is closely aligned to that of the Semantic Grid. In particular, software agents bring the dynamic decision-making, decentralization, coordination, and autonomous behavior needed to realize virtual organizations. Fundamentally, agent-based computing is a service-oriented model [17]—agents are producers, consumers, and brokers of services—and, hence, there is a close relationship between agent-based computing and Web Services, which maps directly into service-oriented Grids [18]. Given this, there are several ways in which work in the agent research community informs solutions to Semantic Grid computing challenges, and, moreover, a Grid does not need to be based on an agent framework to profit from notions of agency and the fruits of the research in this community (see [19] for a detailed discussion of the key agent concepts that can be used to develop Grid applications with the desired degree of flexibility and richness).

First, the notion of autonomy needs to be brought into the conception of Grid services. Thus, services are not always available to be invoked by any entity within the system. This autonomy means that some service invocations may fail—because, for example, the entity delivering the service is unable or unwilling to provide the service at the current moment or under the proposed terms and conditions. This view of services as autonomous is a fundamental mind shift from the present conception, but is essential if Grids are to operate effectively in resource-constrained or open systems.

Second, and following on from the autonomy of the services, is the fact that the de facto means of provisioning a service will be some form of negotiation (a process where the relevant parties attempt to come to a mutually acceptable agreement about the terms and conditions of the service's execution). This view is now starting to be recognized in the Grid and Web communities (through developments such as WS-Agreement), but there is still much to be done. To this end, there has been considerable research in the multiagent systems research community on various forms of negotiation and auctions that can be used in such contexts [20]. Such research deals with two main facets: 1) how to structure the encounter such that the ensuing negotiation process and outcome have particular properties (e.g., maximal efficiency, maximum social welfare, fairness) and 2) given a particular mechanism, what strategy should the agent employ in order to achieve its negotiation objectives. In the former case, considerable attention has been placed on various forms of auctions since these are known to be an effective means of allocating resources in decentralized and open systems.

Third, the notion of virtual organizations as dynamically formed teams that have a particular collective aim has long been studied in agent-based computing. This has resulted in a large number of models, methods, and techniques for establishing cooperation between autonomous problem solvers, for ensuring the actions of collectives are appropriately coordinated, for selecting an appropriate set of partners to participate in the team, and for modeling trust and reputation in open systems.

C. Metadata

The growing body of literature on the Semantic Web has a substantial component that deals with issues of ontologies and reasoning, reflecting the burgeoning research activity in this area (indeed the Semantic Web is sometimes perceived as being synonymous with ontologies). However, before we get to ontological reasoning, there is a significant, but apparently mundane, step of moving into a metadata-enabled world.

Fundamentally, much of the Semantic Web's added value comes from accumulating descriptive information about the various artefacts and resources in the application domain. As different stages of the scientific process work with the same referents—perhaps a sample for analysis, a piece of equipment, a chemical compound, a person, or a publication—metadata can be recorded in various stores, in databases or on Web sites. Thus, this distributed metadata is effectively interlinked by the objects it describes. This, in turn, enables us to ask new kinds of questions, which draw on the aggregated knowledge (e.g., [21]). The naming problem is facilitated in some areas by existing standards, such as the Life Sciences Identifier, which is the standardized naming schema for biological entities in the Life Sciences domains, and the IUPAC Chemical Identifier (IChI) for chemistry.

Building on this, the scaling of the Semantic Web depends on a network effect being reached in "information space"—allowing the sharing and linking of machine readable content, and gaining power by linking to, extending, or even disagreeing with, that specified in another Semantic Web document. However, to achieve this effect we need shared, unique URIs for the objects (real and virtual), and appropriate assertions of the relationships (including equivalence) between them. Thus, for Semantic Web technologies to take hold, more communities must recognize the importance of linking their resources and make more of them nameable on the Web.

D. Ontologies and Reasoning

An ontology determines the extension of terms and the relationships between them. For most practical purposes an ontology is simply a published, more or less agreed, conceptualization of an area of content. The ontology may describe objects, processes, resources, capabilities, or whatever. Given this, it can be seen that ontologies provide the basis of metadata. Thus, any kind of content can be "enriched" by the addition of ontological annotations (e.g., it may indicate the origin of content, its provenance, value, or longevity). The Semantic Grid requires ontologies as a fundamental building block.

We can see increasing adoption both within specifically Grid-based projects [22], [5], and, more widely, throughout a range of science and technology efforts: UMLS (see www.nlm.nih.gov/research/umls), Gene Ontology (see www.geneontology.org), CS Research (see

www.aktors.org/publications/ontology), and the Military Coalition Ontology (see ontology.coginst.uwf.edu).

Moreover, a number of ontologies are emerging as a consequence of commercial imperatives where vertical marketplaces need to share common descriptions. Here, relevant examples include the Common Business Library (CBL, see www.xcbl.org), Commerce XML (cXML, see www.cxml.org), Standardized Material and Service Classification (see ecl@ss), the Open Applications Group Integration Specification (OAGIS, see www.openapplications.org), Open Catalog Format (OCF, see www.martsoft.com/ocp), the Open Financial Exchange (OFX, see www.ofx.net), Real Estate Transaction Markup Language (RETML, see www.rets.org), RosettaNet (see www.rosettanet.org), United Nations Standard Products and Services Code (UN/SPSC, see www.unspsc.org), and the Universal Content Extended Classification System (UCEC).

Of course, a significant set of challenges are encountered in developing, deploying, and maintaining ontologies. These include the fact that ontologies are often highly implicit in scientific and business practice and that they vary as the task or role varies. Furthermore, integrating across multiple ontologies is difficult, as is their maintenance in the face of changing characterizations of a domain. Nevertheless, the upside is that they clearly facilitate interoperability, both for machines and people, they do enhance reuse, and they are evidently becoming part of the distributed scientific infrastructure.

However, providing content enrichment and metadata is only the first phase in exploiting the common conceptualization that is an ontology. Since ontologies encode relationships between classes of object, inferences can be drawn between instances of these classes. To this end, reasoning has been effected in the OWL standard using a variety of description logic inference engines (e.g., [23]). For ontologies distributed across locations and containing many thousands of instances it becomes likely that, in addition to rule based reasoning over this content, it will be necessary to exploit probabilistic and stochastic methods. Thus, in general, reasoning can be regarded as a special case of a Semantic Web service and it is to developments in this area that we now turn.

E. Semantic Web Services

The level of abstraction currently involved in invoking a Web Service is relatively low. Thus, technology around UDDI, WSDL, and SOAP only provides limited support in mechanizing service recognition, service configuration and combination, service comparison, and automated negotiation. The ambition for Semantic Web services, therefore, is to raise the level of description such that services are detailed in a way that indicates their capabilities and task achieving character.

To this end, the Web Ontology Language for Services (OWL-S) [24] encodes rich semantic service descriptions [25] in a way that builds naturally upon OWL. The Semantic Web Services Initiative (SWSI, see www.swsi.org) extends this work by relaxing the constraint of using a description logic formalism for defining service workflow, instead

using a first-order logic based language. The Web Services Modeling Framework (WSMF) [26] is an alternative approach for semantically annotating Web Services, aimed at resolving semantic and protocol interoperability problems faced by Web Service composition. Extending earlier work on the Unified Problem Solving Method Development Language (UPML) framework [27], logical expressions defined in goals, mediators, ontologies, and Web Services are expressed using frame logic.

UPML distinguishes between *domain models, task models, problem solving methods,* and *bridges*, and is also the basis of the Internet Reasoning Service (IRS) [28]. A knowledge-based approach to Semantic Web Services, IRS provides a means for ontology-based Web Service selection using reasoning, by describing them semantically. Here, domain models are effectively the domain ontology, while the task models provide a generic description of tasks to be solved. Problem-solving methods provide implementation-independent descriptions of tasks, while the bridges map between the various components. It takes a task-centric view, where the client asks for a task to be achieved, and the IRS broker calls the appropriate problem-solving method.

Now with such descriptions in place, automatic brokering and composition of services become possible. This, in turn, draws upon agent-based technology (as above) in order to bring together and coordinate the discovery, composition, and enacting of such services.

V. CASE STUDIES

A number of Semantic Grid projects are now in progress and these activities have been reported through the Semantic Grid Workshops held by the Global Grid Forum (see www.semanticgrid.org). For now, however, we consider a number of vignettes taken from projects in which we are directly involved. These projects were chosen in order to try and illustrate different aspects of the ongoing Semantic Grid vision.

A. CombeChem/eBank

Some Grid applications are motivated by the sheer volume of data that can be produced with modern experimental techniques which massively accelerate, or even parallelize, the experimental process. For example, a single DNA microarray can provide information on thousands of genes simultaneously, a significant leap from one gene per experiment. Similarly, in the field of combinatorial chemistry, the chemist produces mixtures of large numbers of different compounds simultaneously. The synthesis of new chemical compounds by combinatorial methods provides major opportunities for the generation of large volumes of new chemical knowledge, and this is the principal drive behind the CombeChem e-Science pilot project [29]. The project aims to enhance the correlation and prediction of chemical structures and properties, by increasing the amount of knowledge about materials via synthesis and analysis of large compound libraries.

Automation of measurement and analysis is required in order to do this efficiently and reliably, and is a clear case for

making knowledge explicit and machine processable through the application of Semantic Web technologies. However, the project takes this further with its objective to achieve a complete end-to-end connection between the laboratory bench and the intellectual chemical knowledge that is published as a result of the investigation—this is described as "publication at source" [30]. The creation of original data is accompanied by information about the experimental conditions in which it is created. There then follows a chain of processing such as aggregation of experimental data, selection of a particular data subset, statistical analysis, or modeling and simulation. The handling of this information may include annotation of a diagram or editing of a digital image. All of this generates secondary data, accompanied by the information that describes the process that produced it. Through publication at source, all this data is made available for subsequent reuse in support of the scientific process, subject to appropriate access control.

Hence, one role of Semantic Web technologies in this project is to establish this complete chain of interlinked digital information all the way from the experiment through to publication. This starts in the smart laboratory and Grid-enabled instrumentation [31]. By studying chemists within the laboratory, technology has been introduced to facilitate the information capture at this earliest stage [32]. Additionally, pervasive computing devices are used to capture live metadata as it is created at the laboratory bench, relieving the chemist of the burden of metadata creation. This data then feeds into the scientific data processing. All usage of the data through the chain of processing is effectively an annotation upon it. By making sure everything is linked up through shared URIs, or assertion of equivalence and other relationships between URIs, scientists wishing to use these experimental results in the future can chase back to the source (i.e., the provenance is explicit). This is achieved by using RDF triplestores to interlink the diverse legacy relational databases and datastores.

The output is typically a scholarly publication, which may be self-archived in an institutional repository or published in a digital library. The move toward self-archiving and linking of research data has recently been encouraged in the U.K. [33]. This is not the end of the process, since research and learning "close the loop" and feed back into further experiments. The scholarly process supports the further interpretation of outputs, and the suggestion and investigation of alternative theories. Publication at source also supports the interlinking of published knowledge to facilitate the process, and enables automated processing. Given the throughput of knowledge created through combinatorial chemistry, it is not plausible for every new compound to be the subject of a traditional scholarly publication by a scientist since this would introduce a massive bottleneck—perhaps 80% of data would be left unprocessed. Thus, this is an example of the Semantic Grid enabling a significant culture shift in the scientific process within this discipline.

Some of these ideas are also demonstrated in the World Wide Molecular Matrix [34] and the Collaboratory for Multiscale Chemical Science (CMCS) [35].
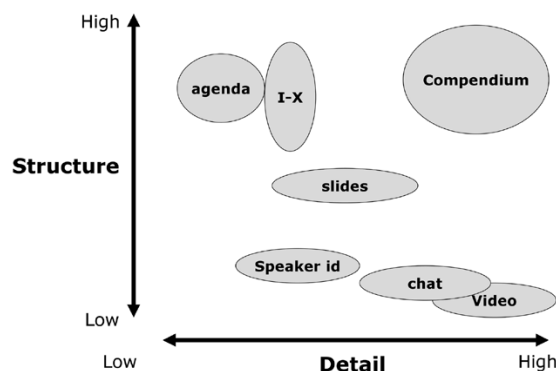

Fig. 1. Structure versus detail in CoAKTinG.

B. CoAKTinG

The CoAKTinG project [36] has applied Semantic Web technologies in novel ways to advance the state of the art in collaborative mediated spaces for distributed e-Science. It comprises four tools: instant messaging and presence notification (BuddySpace), graphical meeting and group memory capture (Compendium), intelligent "to-do" lists (Process Panels), and meeting capture and replay. These are integrated into existing collaborative environments (such as the Access Grid), and through use of a shared ontology to exchange structure, promote enhanced process tracking and navigation of resources before, after, and while a meeting occurs.

Each of the CoAKTinG tools can be thought of as extracting structure from the collaboration process. The full record of any collaboration (e.g., a video recording of a meeting) is rich in detail, but to be useful we must extract resources which are rich in structure. This is represented in Fig. 1. In this context, collaboration as an activity can be seen as a resource in itself, which with the right tools can be used to enhance and aid future collaboration and work. CoAKTinG is also an example of a system which supports recording and reuse facilitated by distributed collaborative semantic annotation—this is a paradigm which can be generically applied across a spectrum of e-Research scenarios.

C. MIAKT

Breast cancer screening also involves an annotation task, with a large volume of content being generated by mandatory screening programmes. This process consists of the capturing of an X-ray mammogram and any areas considered abnormal on the mammogram are assessed by means of pathology tests (biopsies). Data from the radiologist, responsible for the mammogram, the histopathologist, responsible for the interpretation of biopsy results, and the clinician, with knowledge of the history of the patient, are brought together to make a consultative appraisal of each particular case. This is known as the Triple Assessment Procedure and we have undertaken the Medical Imaging with Advanced Knowledge Technologies (MIAKT) project [37] to support this collaborative meeting and the knowledge that goes with it, using the Semantic Grid technologies.
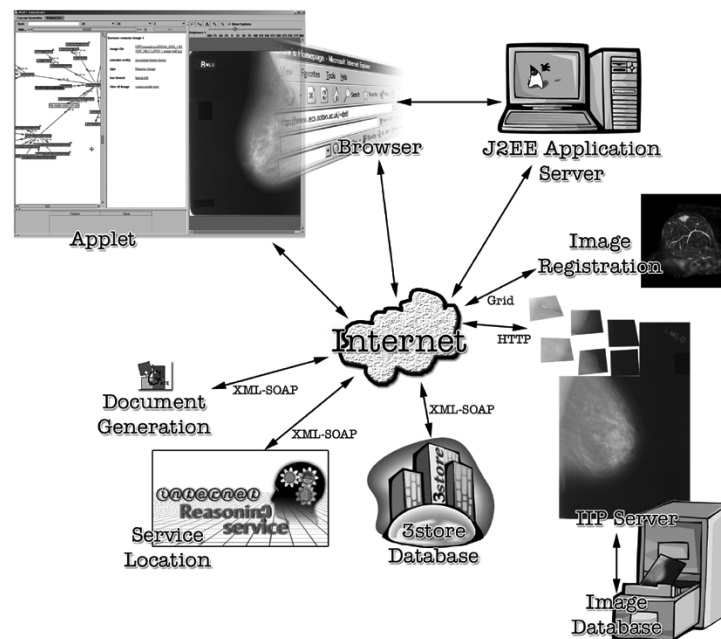
Fig. 2. MIAKT Mixed Web and Grid Services Architecture.

The MIAKT application is built around a distributed architecture (represented in Fig. 2) which uses Web and Grid based services to provide discrete and disparate functionality to a generic client application. The architecture is deliberately abstracted from any particular application domain and its description, providing a powerful structure for rapidly prototyping new knowledge management applications in new domains. In this respect, MIAKT becomes a particular application of this architecture.

In terms of the Semantic Grid vision, services are presented to the client application according to the application-specific *application ontology*. The server instantiates the framework using this ontology and provides a simple and homogeneous API to the application's web-services which may be running over various protocols.

The MIAKT system provides knowledge management for the data that the screening process generates, as well as providing a means for medical staff to investigate, annotate, and analyze the data using Web and Grid services. Use of ontologies to store knowledge facilitates a mechanism for providing ancillary diagnosis to the consultation. The application software allows viewing and annotation of various types of images, from X-ray mammograms to three-dimensional MRI scans, provides for searching of patient data, and supports invocation of services on the Web for image analysis, data analysis, and natural language report generation. Some of these services are computationally intensive and use explicit Grid services.

D. Medical Devices

Grid and Pervasive Computing come together in the Grid Based Medical Devices for Everyday Health project in which patients who have left hospital are monitored using wearable computing technology. Since the patient is mobile, position and motion information is gathered (using devices such as accelerometers) to provide the necessary contextual information in which to interpret the physiological signals. The signal processing occurs on the Grid and medics are alerted—by pervasive computing—when the patients experience episodes that need attention.

The interesting infrastructure research question is to what extent the Grid services paradigm can be deployed in the direction of the devices. The project has been conducted using Globus Toolkit 3. The devices and sensors that we are dealing with typically have limited computational power and storage, and they only have intermittent network connectivity. Although in some cases they may be capable of hosting Grid Services, generally the devices interface via a Grid service proxy. In turn, users can access the information through a portal, and this can itself be accessed by pervasive devices.

The additional contextual information provided by the wearables—GPS and accelerometer data—is essential to interpretation of the physiological signals. Modeling of context, reasoning about it, and managing it, calls once again upon Semantic Web technologies.

This is one example of a sensor network. Other sensor networks demand grid processing due to the volume of data collected at greater spatial or temporal density, reflecting the evolution of sensor technology, and again for visualization. As in this medical scenario, the primary purpose is to obtain information, which can be represented using Semantic Web technologies to facilitate reuse and interoperability.

VI. DISCUSSION

The notion of the Semantic Grid came about in 2001. Within our research groups we were jointly conducting work in the areas of the Grid, the Semantic Web, Web Services, and agents. Meanwhile, Grid practitioners were also thinking in terms of services, a view reinforced by the Grid "Anatomy" paper [2] published in 2001 and then made
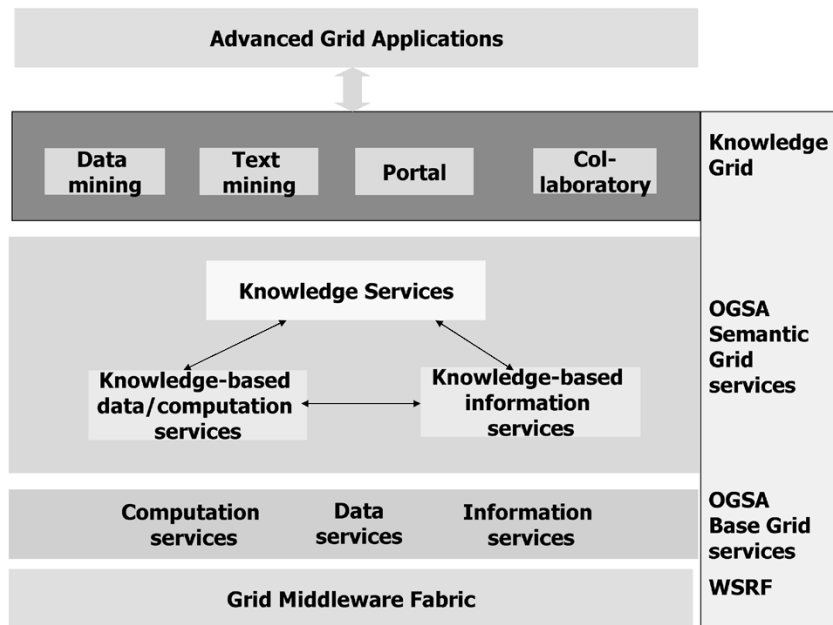
Fig. 3. Semantic Grid Architecture (due to Goble *et al.*).

explicit by the "Physiology" paper [38] in 2002. The Web community had been developing Web Services and formulating requirements for the Semantic Web. The samizdat publication of our report to the e-Science community in July 2001 had an impact on a number of the projects emerging in the e-Science programme, and the first Semantic Grid papers were written the following year [39], [40].

Today we see an increasing volume of activity involving Grid and Semantic Web. The Semantic Grid Research Group in the Global Grid Forum has held workshops, other communities are holding Semantic Grid events, and the literature in this area is growing. Now, three years later, we can reflect on the evolution of the Semantic Grid. As discussed in Section II, the e-Science programme has reinforced the practical need for the Semantic Grid—Grid middleware is needed to join up computers, while Semantic Grid middleware is needed to join up projects.

There are also results flowing from Semantic Web research that will assist the research agenda advocated here. Work has led to new standards, such as OWL, for expressing ontologies in a way that supports interoperability between systems. Tools are now appearing that facilitate the construction and verification of ontologies. Ontologies are a vital element in enabling a common conceptualization to hold between the various elements (human and machine) of a Grid-based collaboration. The Semantic Web effort is also producing tools to support annotation, linking, search, and browsing of content. We are also beginning to see integrated applications that exploit semantic annotations and metadata [41]. A pressing need is to develop standards and methods to describe the knowledge services themselves, and to facilitate the composition of services into larger aggregates and negotiated workflows. An important element of this is likely to be protocols and frameworks that emerge out of the agent community.

That we advocated an agent-oriented approach, rather than adoption of software agent frameworks per se, was a reflection of our conviction that agent-based computing had very important techniques to offer, but also an awareness that Grid-scale agent deployments were not evident at that time. Hence, it is natural to see these ideas coming through in Web Services, and, for example, in WS-Agreement which aims to define a language and a protocol for advertising the capabilities of providers, creating agreements and runtime monitoring of agreement compliance. The volume of work now evident in workflow confirms the focus on automation. However, we are only a step along the path with respect to agency.

1) Although agents are producers, consumers and brokers of services, functionality which we can see reflected in Web Services and Semantic Web Services, they also have a very important notion of autonomy. This is, we believe, still something that the Semantic Grid needs, especially to realize the more autonomic infrastructure.
2) There is a wealth of expertise in agent negotiation which needs to be applied within the Semantic Grid context, to the creation of agile virtual organizations. Again, this is an area where progress has been somewhat slower than we expected.

In these three years, however, probably the most significant evolution in thinking—or at least in presentation—has been the appreciation of the role of the Semantic Web *within* the Grid infrastructure. This has resulted in a move away from the initial three-layer "Grids" architecture [42]. Valuable though this model has been, it fails to convey the role of knowledge services within the Grid infrastructure. To this end, we believe a more recent architecture, due to Goble [11], better captures the interactions in the various levels (see Fig. 3).

As our vignettes highlight, at some point the digital, automated world of the Grid must meet the physical, interactive world of the users. Now in many cases this may be through the devices with which the users are interacting: PDAs, wearable computers, displays, Access Grid, or virtual reality. Nevertheless, the general question of the relationship between the Grid, the physical world and its inhabitants remains.

To start answering this question, we believe it is important to consider the area of pervasive or ubiquitous computing (which is about devices everywhere; e.g., in everyday artefacts, in our clothes and surroundings, and in the external environment). The term ubiquitous was adopted by Xerox through the work of Mark Weiser [43], who emphasized the "calm" aspects, where the computing is everywhere but "stays out of the way." In Europe, pervasive computing is part of the Ambient Intelligence vision.

Moore's Law tells us that if you keep the box the same size (the desktop PC, for example) then a series of computers will get increasingly powerful over time. However, if you only want the same power then you can work with smaller and smaller devices, and more of them. Broadly then, this gives us the world of the Grid and the world of pervasive computing, respectively. Both are important and inevitable technological trends that therefore need to be considered together, but we suggest that Grid and pervasive computing have another very important relationship: pervasive computing provides the manifestation of the Grid in the physical world.

Sometimes the Grid application demands the pervasive computing and sometimes the pervasive computing demands the Grid. In the former category would be the "Grid-enabled" devices in a laboratory, the pieces of scientific equipment and "grid appliances" connected directly to the Grid—and also the handheld devices with which users gather information and access results. Devices, such as novel interfaces, may also be deployed to support the collaborative environment, perhaps facilitating visualization and annotation of data. In the latter category we have the sensor networks—as sensors and sensor arrays evolve, we can acquire data with higher temporal or spatial resolution, and this increasing bulk of (often realtime) data demands the computational power of the Grid. Meanwhile, many pervasive deployments are currently small scale, due to small numbers of devices or small numbers of users, but will demand more Grid processing as numbers inevitably scale up.

Given this, we can see that Grid and Pervasive computing are each about large numbers of distributed processing elements. At an appropriate layer of abstraction, they both involve similar computer science challenges in distributed systems. Specifically, these include service description, discovery and composition, issues of availability and mobility of resources, autonomic behavior, and of course security, authentication, and trust. Both need ease of dynamic assembly of components, and both rely on interoperability to achieve their goals. The peer-to-peer paradigm is also relevant across the picture.
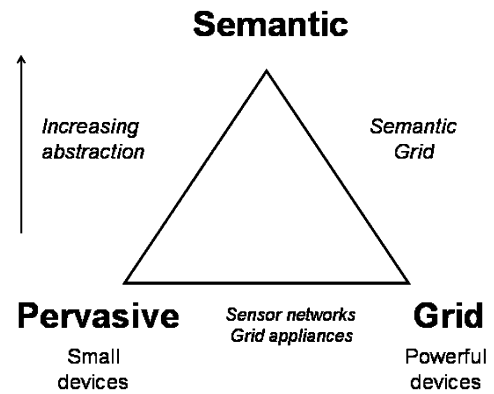

Fig. 4. Semantic-Pervasive-Grid triangle.

In common with the Grid, we can also argue that the full richness of the pervasive vision needs the Semantic Web technologies. Again, this is about semantic interoperability: we need service description, discovery and composition, and indeed research areas such as Semantic Web Services are being applied both to Grid and to Pervasive computing. Hence, the Semantic approach sits above the large-scale distributed systems of Pervasive and Grid computing, as illustrated in Fig. 4.

A key motivation for the semantic interoperability is the need to assemble new applications with ease. Essentially, we have lots of distributed bits and pieces that need to work together to provide the requisite global behavior, and we wish this to happen as far as appropriate without manual intervention. This move toward a more autonomous approach is best achieved through the techniques of agent-based computing, and in the future we look toward self-organization. Taking this view to its logical conclusion leads us to envisage a self-organizing Semantic Grid which behaves like a constantly evolving organism, with ongoing, autonomous processing rather than on-demand processing—an organic Grid which itself can generate new processes and new knowledge [44], manifest in the physical world through ambient intelligence.

VII. Research Agenda

Achieving the full richness of the Semantic Grid vision brings with it many significant research challenges. The following ten, updated from our original report, identify areas (in no particular order) where we believe research needs to be targeted.

1) Automated Virtual Organization Formation and Management: In many cases it will not be possible to tell *a priori* exactly what services users will need in a given system. Thus, new services (virtual organizations) need to be composed as and when they are required. Moreover, given the issues of scale and timeliness, it is important that automated tools are available to support or even enact this endeavor. Thus, research is required to develop general models that enable the need for new virtual organizations to be determined, the ability to discover appropriate services, the ability to select the best set of participants to deliver the desired service, the

ability to divide the work in the most effective way between the participants, the ability to monitor the ongoing operation of the virtual organization and to alter it for the better when appropriate, and the ability to dissolve such partnerships when they are no longer sustainable.

2) Service Negotiation and Contracts: Further research is needed to determine the types of contract that are appropriate for Grid systems, how contracts can be policed and enforced, the way in which negotiations can be organized to be efficient and effective in Grid systems, and the types of strategies that agents can be used to attain their negotiation objectives. Another facet of this is that contracts need to be represented in an interoperable manner.

3) Security, Trust, and Provenance: Further research is needed to understand the processes, methods, and techniques for establishing computational trust and determining the provenance and quality of content and services in Grid systems. This also extends to the issue of digital rights management in making content available. Policies also need to be represented and demand interoperability.

4) Metadata and Annotation: While the basic metadata infrastructure already exists in the shape of RDF, metadata issues have not been fully addressed in current Grid deployments. It is relatively straightforward to deploy some of the technology in this area, and this should be promoted. RDF, for example, is already encoding metadata and annotations as shared vocabularies or ontologies. However, there is still a need for extensive work in the area of tools and methods to support the design and deployment of ontologies. Annotation tools and methods need to be developed so that emerging metadata and ontologies can be applied to the large amount of content that will be present in Grid applications.

5) Content Processing and Curation: Research is required into incorporating a wide range of media into the e-Research infrastructure. This will include video, audio, domain-specific data, and a wide range of imaging methods. Research is also needed into the association of metadata and annotation with these various media forms, including distributed collaborative annotation. As curation becomes increasingly automated, we look toward what might be termed "autonomic curation."

6) Knowledge Technologies: In addition to the requirement for the research in metadata and annotation, there is a need for a range of other knowledge technologies to be developed and customized for use in e-Science contexts. These include knowledge capture tools and methods, dynamic content linking, annotation based search, annotated reuse repositories, natural language processing methods (for content tagging, mark-up, generation, and summarization), data mining, machine learning, and Internet reasoning services. These technologies will need shared ontologies and service description languages if they are to be integrated into the e-Science workflow. The problem of composing task-achieving services from collections of other services is an important research area. These technologies will also need to be incorporated into the pervasive devices and smart laboratory contexts that will emerge in e-Research.

7) Design and Deploy: Research is needed to make it easier to design, configure, and deploy applications. Specifically, Grid infrastructure and applications need to be easy to use by all users in the lifecycle—from design, through development, deployment, operation, maintenance, and reuse.

8) Interaction: Research is required into methods and techniques that allow information to be visualized in ways consistent with the e-Research collaborative effort. This will also involve customizing content in ways that reflect localized context and should allow for personalization and adaptation. The intersection of the Semantic Grid with the physical world exposes semantics at the interaction interface.

9) Collaboration: Much more needs to be done to understand the workflow of current and future e-Science collaborations. Users should be able to form, maintain, and disband communities of practice with restricted membership criteria and rules of operation. Currently, most studies focus on the e-Science infrastructure behind the socket on the wall. However, this infrastructure will not be used unless it fits in with the working environment of the e-Scientists. This process has not been studied explicitly and there is a pressing need to gather and understand these requirements. There is also a need to collect real requirements from users, to collect use cases, and to engage in some evaluative and comparative work. Finally, we need to understand more fully the process of collaboration in e-Science.

10) Pervasive Computing: Currently, most references and discussions about Grids imply that their primary task is to enable global access to huge amounts of computational power. Generically, however, we believe Grids should be thought of as the means of providing seamless and transparent access from and to a diverse set of networked resources. These resources can range from PDAs to supercomputers and from sensors and smart laboratories to satellite feeds. We believe that for e-Science to be successful and for the Grid to be effectively exploited, much more attention needs to focus on how laboratories need to be instrumented and augmented. For example, we believe it is essential to develop infrastructure that allows a range of equipment to advertise its presence, be linked together, and annotate/markup content it is receiving or producing.

Moreover, we also believe that many of the issues, technologies and solutions developed in the context of e-Research can be exploited in other domains where groups of diverse stakeholders need to come together electronically and interact in flexible ways. Thus we believe that it is important that relationships are established and exploitation routes are explored with domains such as e-Business, e-Commerce, e-Education, and e-Entertainment.

To sum up, three years of progress have confirmed the value of the Semantic Grid vision and this emerging community is achieving significant momentum. There are still many challenges. Some of the technical ones have been highlighted. However, others arise from the need to bring together the research communities to achieve the Semantic Grid ambitions. This can be viewed as building bridges in order to build bridges. We need to bring communities together to create the Semantic Grid, which can then be used for flexible

collaborations and computations on a global scale—for the creation of new scientific results, new business, and even new research disciplines.

REFERENCES

[1] I. Foster and C. Kesselman, *The Grid: Blueprint for a New Computing Infrastructure*. San Mateo, CA: Morgan Kaufmann, 1999.

[2] I. Foster, C. Kesselman, and S. Tuecke, "The anatomy of the grid: Enabling scalable virtual organizations," *Int. J. High Perform. Comput. Appl.*, vol. 15, no. 3, pp. 200–222, Fall 2001.

[3] D. De Roure, N. R. Jennings, and N. R. Shadbolt, "Research Agenda for the Semantic Grid: A Future e-Science Infrastructure," National e-Science Centre, Edinburgh, U.K., UKeS-2002-02, Dec. 2001.

[4] *Semantic Web Activity Statement*, 2004. W3C, Worldwide Web Consortium.

[5] C. A. Goble, D. De Roure, N. R. Shadbolt, and A. A. A. Fernandes, "Enhancing services and applications with knowledge and semantics," in *The Grid 2: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, Eds. San Mateo, CA: Morgan Kaufmann, 2004, pp. 431–458.

[6] D. Fensel and M. A. Musen, "Special Issue on the Semantic Web," *IEEE Intell. Syst. Mag.*, vol. 16, no. 2, pp. 24–25, Mar.–Apr. 2001.

[7] C. Wroe, R. Stevens, C. Goble, A. Roberts, and M. Greenwood, "A suite of DAML + OIL ontologies to describe bioinformatics web services and data," *Int. J. Coop. Inf. Syst.*, vol. 12, pp. 197–224, 2002.

[8] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Sci. Amer.*, vol. 284, no. 5, pp. 34–43, May 2001.

[9] N. R. Jennings, "On agent-based software engineering," *Artif. Intell.*, vol. 117, pp. 277–296, 2000.

[10] A. Andrieux, K. Czajkowski, A. Dan, K. Keahey, H. Ludwig, J. Pruyne, J. Rofrano, S. Tuecke, and M. Xu, "Web Services Agreement Specification (WS-Agreement)," Global Grid Forum, May 2004.

[11] C. Goble and D. De Roure, "The semantic grid: Myth busting and bridge building," presented at the 16th Eur. Conf. Artificial Intelligence (ECAI-2004), Valencia, Spain, 2004.

[12] T. Hey and A. E. Trefethen, "The UK e-science core programme and the grid," *Future Gen. Comput. Syst.*, vol. 18, pp. 1017–1031, 2002.

[13] *Roadmap for a UK Virtual Research Environment*, JISC, Joint Information Systems Committee, 2004.

[14] D. Gannon, G. Fox, M. Pierce, B. Plale, G. V. Laszewski, C. Severance, J. Hardin, J. Alameda, M. Thomas, and J. Boisseau, "Grid Portals: A Scientist's Access Point for Grid Services," Global Grid Forum, Sep. 2003.

[15] I. Foster, C. Kesselman, J. Nick, and S. Tuecke, "Grid services for distributed system integration," *Computer*, vol. 35, no. 6, pp. 37–46, Jun. 2002.

[16] S. Parastatidis, J. Webber, P. Watson, and T. Rischbeck, "A Grid Application Framework Based on Web Services Specifications and Practices," North-East Regional e-Science Centre, Aug. 2003.

[17] N. R. Jennings, "An agent-based approach for building complex software systems," *Commun. ACM*, vol. 44, pp. 35–41, 2001.

[18] L. Moreau, "Agents for the grid: A comparison with web services (Part 1: The transport layer)," presented at the 2nd IEEE/ACM Int. Symp. Cluster Computing and the Grid (CCGRID 2002), Berlin, Germany, 2002.

[19] I. Foster, N. R. Jennings, and C. Kesselman, "Brain meets brawn: Why Grid and agents need each other," presented at the Proc. 3rd Int. Conf. Autonomous Agents and Multi-Agent Systems, New York, 2004.

[20] N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, C. Sierra, and M. Wooldridge, "Automated negotiation: Prospects, methods, and challenges," *Int. J. Group Dec. Negot.*, vol. 10, pp. 199–215, 2001.

[21] N. R. Shadbolt, N. Gibbins, H. Glaser, S. Harris, and M. C. Schraefel, "CS AKTive space or how we stopped worrying and learned to love the semantic web," *IEEE Intell. Syst.*, vol. 19, no. 3, pp. 41–47, May–Jun. 2004.

[22] D. De Roure, Y. Gil, and J. Hendler, "Special Issue on E-Science," *IEEE Intell. Syst.*, vol. 19, no. 1, pp. 1–89, Jan.–Feb. 2004.

[23] V. Haarslev and R. Moller, *RACER User's Guide and Reference Manual*, Sept. 2003.

[24] D. Martin, M. Paolucci, S. McIlraith, M. Burstein, D. McDermott, D. McGuinness, B. Parsia, T. Payne, M. Sabou, M. Solanki, N. Srinivasan, and K. Sycara, "Bringing semantics to web services: The OWL-S approach," presented at the First Int. Workshop on Semantic Web Services and Web Process Composition (SWSWPC 2004), San Diego, CA, 2004.

[25] T. Payne and O. Lassila, "Semantic web services," *IEEE Intell. Syst.*, vol. 19, no. 4, pp. 14–15, Jul.–Aug. 2004.

[26] D. Fensel and C. Bussler, "The web service modeling framework WSMF," *Electron. Commerce: Res. Appl.*, vol. 1, pp. 113–137, 2002.

[27] D. Fensel, V. R. Benjamins, E. Motta, and B. Wielinga, "UPML: A framework for knowledge system reuse," presented at the Int. Joint Conf. Artificial Intelligence (IJCAI-99), Stockholm, Sweden, 1999.

[28] E. Motta, J. Domingue, L. Cabral, and M. Gaspari, "IRS-II: A framework and infrastructure for semantic web services," presented at the 2nd Int. Semantic Web Conf. (ISWC2003), Sanibel Island, FL, 2003.

[29] J. G. Frey, M. Bradley, J. W. Essex, M. B. Hursthouse, S. M. Lewis, M. M. Luck, L. Moreau, D. C. De Roure, M. Surridge, and A. Welsh, "Combinatorial chemistry and the grid," in *Grid Computing—Making the Global Infrastructure a Reality*, F. Berman, G. Fox, and T. Hey, Eds. New York: Wiley, 2003, Wiley Series in Communications Networking and Distributed Systems, pp. 945–962.

[30] J. G. Frey, D. De Roure, and L. A. Carr, "Publication at source: Scientific communication from a publication web to a data grid," *Euroweb 2002 Conf., The Web and the GRID: From e-Science to e-Business*, 2002.

[31] G. Hughes, H. Mills, D. D. Roure, J. G. Frey, L. Moreau, M. Schraefel, G. Smith, and E. Zaluska, "The semantic smart laboratory: A system for supporting the chemical escientist," *Org. Biomol. Chem.*, vol. 2, no. 22, pp. 3284–3293, 2004.

[32] M. C. Schraefel, G. Hughes, H. Mills, G. Smith, T. Payne, and J. Frey, "Breaking the book: Translating the chemistry lab book into a pervasive computing lab environment," presented at the 2004 Conf. Human Factors in Computing Systems (CHI 2004), Vienna, Austria.

[33] "Scientific Publications: Free for All? Tenth Report of Session 2003–04," Science and Technology Committee, UK Parliament HC 399-1, 2004.

[34] P. Murray-Rust, "The world wide molecular matrix—A peer-to-peer XML repository for molecules and properties," presented at the EuroWeb2002, Oxford, U.K..

[35] J. D. Myers, T. C. Allison, S. Bittner, B. Didier, M. Frenklach, J. William, H. Green, Y.-L. Ho, J. Hewson, W. Koegler, C. Lansing, D. Leahy, M. Lee, R. McCoy, M. Minkoff, S. Nijsure, G. V. Laszewski, D. Montoya, C. Pancerell, R. Pinzon, W. Pitz, L. A. Rahn, B. Ruscic, K. Schuchardt, E. Stephan, A. Wagner, T. Windus, and C. Yang, "A collaborative informatics infrastructure for multi-scale science," presented at the Challenges of Large Applications in Distributed Environments (CLADE) Workshop, Honolulu, HI, 2004.

[36] S. Buckingham Shum, D. De Roure, M. Eisenstadt, N. Shadbolt, and A. Tate, "CoAKTinG: Collaborative advanced knowledge technologies in the grid," presented at the Second Workshop on Advanced Collaborative Environments, Edinburgh, U.K., 2002.

[37] N. Shadbolt, P. Lewis, S. Dasmahapatra, D. Dupplaw, B. Hu, and H. Lewis, "MIAKT: Combining grid and web services for collaborative medical decision making," presented at the UK e-Science All Hands Meeting, Nottingham, U.K., 2004.

[38] I. Foster, C. Kesselman, J. Nick, and S. Tuecke, "Grid services for distributed system integration," *Computer*, vol. 35, pp. 37–46, 2002.

[39] C. A. Goble and D. De Roure, "The grid: An application of the semantic web," *ACM SIGMOD Rec.*, vol. 31, pp. 65–70, 2002.

[40] C. Goble and D. De Roure, "The semantic web and grid computing," in *Real World Semantic Web Applications, Frontiers in Artificial Intelligence and Applications*, V. Kashyap and L. Shklar, Eds. Amsterdam, The Netherlands: IOS, 2002, vol. 92.

[41] M. Klein and U. Visser, "Semantic web challenge 2003," *IEEE Intell. Syst.*, vol. 19, no. 3, pp. 31–33, May–Jun. 2004.

[42] K. G. Jeffery, "Knowledge, Information and Data," CLRC Information Technology Dept., Sep. 1999.

[43] M. Weiser, "The computer for the twenty-first century," *Sci. Amer.*, vol. 265, pp. 94–104, Sep. 1991.

[44] D. De Roure, "On self-organization and the semantic grid," *IEEE Intell. Syst.*, vol. 18, no. 4, pp. 77–79, Jul.–Aug. 2003.

David De Roure graduated in mathematics with physics from the University of Southampton, Southampton, U.K., in 1984 and received the Ph.D. degree in distributed systems in 1990.

He is a Full Professor in the School of Electronics and Computer Science at the University of Southampton, where he was a founding member of the Intelligence, Agents, Multimedia Group and is currently head of the Grid and Pervasive Computing Group. He is a co-director of the Open Middleware Infrastructure Institute and the Southampton Regional e-Science Centre, and Director of the Centre for Pervasive Computing in the Environment funded by the UK Department of Trade and Industry. His research interest is in the application of knowledge technologies to Grid and pervasive computing. He chairs the Semantic Grid Research Group in the Global Grid Forum and sits on the Grid Forum Steering Group as an Area Director. He is a member of the W3C Advisory Committee and national committees including the e-Science Architecture Task Force and the JISC Committee for Support of Research.

Prof. De Roure is a Fellow of the British Computer Society.

Nicholas R. Jennings (Senior Member, IEEE) received the first class degree in computer science from Exeter University, Exeter, U.K., in 1988 and the Ph.D. degree in artificial intelligence from the University of London, London, U.K., in 1992.

He is a full Professor in the School of Electronics and Computer Science at the University of Southampton, Southampton, U.K., where he carries out basic and applied research in agent-based computing. He is Deputy Head of School (Research), Head of the Intelligence, Agents, Multimedia Group, and is also the Chief Scientific Officer for Lost Wax. He has published more than 200 articles and eight books on various facets of agent-based computing and holds two patents (with three more pending).

Prof. Jennings received the Computers and Thought Award (the premier award for a young AI scientist) in 1999, an IEE Achievement Medal in 2000, and the ACM Autonomous Agents Research Award in 2003. He is a Fellow of the British Computer Society, the Institution of Electrical Engineers, and the European Artificial Intelligence Association (ECCAI). He is a member of the UK Computing Research Committee (UKCRC).

Nigel R. Shadbolt received the first class degree in philosophy and psychology from the University of Newcastle, Newcastle upon Tyne, U.K., in 1978 and the Ph.D. degree in artificial intelligence from the University of Edinburgh, Edinburgh, U.K., in 1981.

He is Professor of Artificial Intelligence at the School of Electronics and Computer Science (ECS), University of Southampton, Southampton, U.K. He is a member of the Intelligence, Agents, Multimedia Group, Head of the BIO@ECS Group, and Director of Interdisciplinary Research within ECS. His research concentrates on two ends of the spectrum of AI—namely, Knowledge Technologies and Biorobotics. In 2000 he led a consortium of five Universities that secured an EPSRC Interdisciplinary Research Collaboration in Advanced Knowledge Technologies (AKT). He is the Director of this multimillion pound, six-year research programme that is pursuing basic and applied research in the provision of technologies to support Knowledge Management and realize the promise of the Semantic Web.

He is Editor in Chief of *IEEE Intelligent Systems*, an Associate Editor of the *International Journal for Human Computer Systems*, and on the editorial board of the *Knowledge Engineering Review* and the *Computer Journal*. He is a member of various national committees including the UK e-Science Technical Advisory Committee (TAG) and the UK EPSRC Strategic Advisory Team (SAT) for ICT, and is currently a Vice President of the British Computer Society, where he chairs its Knowledge Services Board. In 2003, he was elected Fellow of the British Computer Society. In 2004, he was elected Fellow of the European Artificial Intelligence Association (ECCAI).