The Collaborative Semantic Grid

David De Roure, Jeremy Frey, Danius Michaelides and Kevin Page
Schools of Electronics & Computer Science and Chemistry
University of Southampton, Southampton SO17 1BJ, UK
{dder,dtm,krp}@ecs.soton.ac.uk, j.g.frey@soton.ac.uk

ABSTRACT

Grid and Semantic Web technologies both enable heterogeneous resources to be joined up to achieve new functionality and capability, and their combined effectiveness has been demonstrated through a number of 'Semantic Grid' projects. These typically apply Semantic Web technologies in Grid applications, in a datagrid or within the Grid middleware. In this paper we suggest how both approaches also support distributed collaborative endeavours, and explore their combined role. We illustrate this 'Collaborative Semantic Grid' through a number of case studies, and contemplate the changing nature of collaboration as these technologies evolve.

KEYWORDS: Semantic Grid, Semantic Web, e-Science, datagrid, access grid, chemistry.

1. INTRODUCTION

GRID computing has traditionally supported distributed collaborative endeavours. This is reinforced by the contemporary definition of the Grid as being about virtual organisations (VOs), a concept which suggests notions of collaboration, sharing, common understanding and trust:

> "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." [1]

It is interesting to compare this with the ambitions of the Semantic Web [2], described in the W3C Semantic Web Activity Statement as:

> "…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."

The confluence of ideas here is striking. It was the observation that these technologies can be used together to achieve the vision of e-Science that led in 2001 to the Semantic Grid vision – 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 [3-5].

Although collaborations were explicit in this definition, as well as being implied in the Grid and Semantic Web definitions, most of the reported work on Semantic Grid has not focused on this perspective. In the present paper we suggest that the combination of technologies represented by the Semantic Grid can also form a powerful part of a collaboration infrastructure. We conduct this exploration by drawing on three complementary example applications which illustrate different aspects of this discussion.

The paper is structured as follows. In Section 2 we give a brief introduction to the essential ideas of the Semantic Web. Our case studies commence in Section 3 where we look at a 'Semantic Datagrid' created in an e-Science project in combinatorial chemistry, followed in Section 4 by a case study in enhancing meeting support – a kind of

Semantic Access Grid. We then, in Section 5, look at a proposed application area which deploys both technologies in support of a collaborative research community. Section 6 discusses these exemplars and Section 7 summarises our conclusions from this exploration.

2. THE SEMANTIC WEB APPROACH

2.1. The Semantic Web

For the purposes of this discussion, the two key features of the Semantic Web approach are:

1. Machine processable descriptions. Descriptive information (metadata) is represented in an agreed way so that it can be processed by machine, typically as a set of relationships in a simple data model called the Resource Description Framework (RDF), which can itself be represented in an XML format. The shared vocabularies used in the descriptive information are also represented in standard ways (e.g. the OWL Web Ontology Language). This descriptive information is known as semantic annotation and enables the Web to interchange data as well as documents.

2. Sharing and accumulation. As different people create these descriptions they are richly interlinked through sharing the identifiers when the same things are being described, and sharing the identifiers used in the descriptions. Hence knowledge accumulates about the same (real-world) things. This facilitates data interchange and enables us to ask new kinds of questions that draw on the aggregated knowledge. As is the nature of the Web, these descriptions may of course be incomplete, inconsistent and potentially impermanent.

Many tools now exist for working with RDF data, such as editors, browsers and in particular stores (known as RDF triplestores) which can be queried using languages such as the *SPARQL Protocol And RDF Query Language* (SPARQL). Tools also exist for working with the vocabularies (ontologies), for example to check consistency and aid maintenance. However these tools are in an early stage of deployment to a wider community and frequently suffer from less than ideal user interfaces. This has impacted on their uptake outside the Computer Science community.

2.2. Social Networks

To provide an example with a collaborative dimension, consider the CS AKTiveSpace project, which won the award for the 'Semantic Web Challenge' in 2003 [6] and is depicted in Figure 1. CS AKTiveSpace is an integrated Semantic Web application which provides a way to explore the UK Computer Science Research domain across multiple dimensions for multiple stakeholders, from funding agencies to individual researchers' projects. Metadata about publications and projects of academic researchers is harvested from a variety of public data sources into a single RDF store, which is consistent with a reference ontology and can be queried through multiple interfaces.

This is a rich source of information about communities of practice, and it enables researchers to identify other members in order to establish collaborations. Furthermore, ontology-based network analysis techniques can be used to examine the connectivity of instances in the knowledge base with respect to the connections and thereby uncover informal, self-organising communities of practice. Hence CS AKTiveSpace illustrates how Semantic Web can be used to facilitate the Virtual Organisations of the Grid, and it also illustrates the need for a scalable triplestore [7] and an innovative user interface (mSpace [8]) to help make this information available to people as well as machines.

An important complementary example is provided by the FOAF (Friend of a Friend) project which also provides a means for describing social networks [9]. In FOAF, people publish descriptive statements about themselves and others using some standard terms. A unique identifier is used to identify a person, so this descriptive information is linked together by the people it describes. This leads to a model of the social network which is distributed and is maintained directly by the participants in the network.

The Semantic Web clearly involves distributed RDF stores, not single standalone solutions with single ontologies, and this is one of the challenges in the use of Semantic Grid [10]. Most significant to our discussion here, we observe that CS AKTiveSpace could be described as publishing-based, while FOAF is participation-based. In CS AKTiveSpace, knowledge is used to map the community, while FOAF is an example of the using the community to map knowledge.

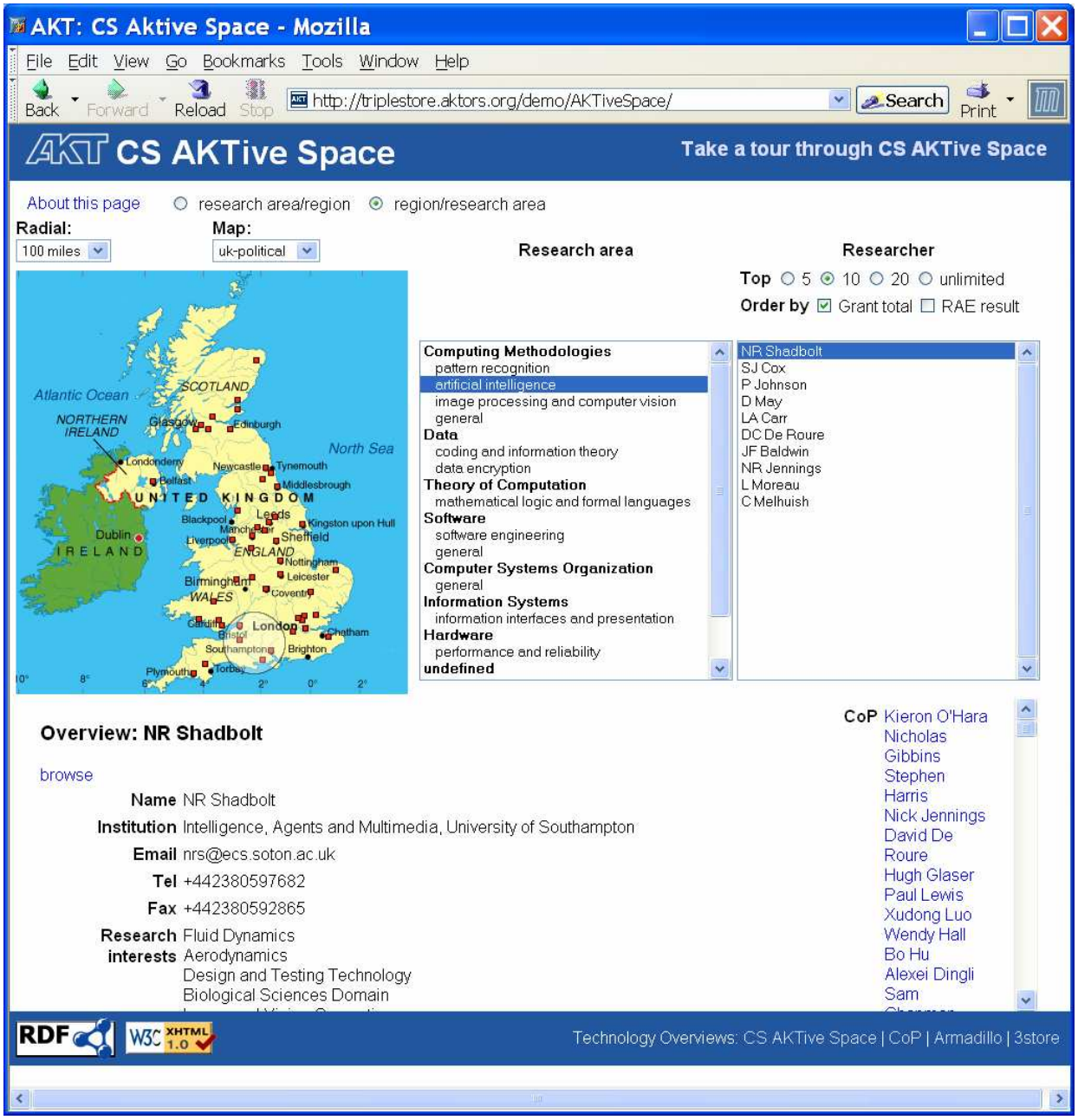

Figure 1. CS AKTive Space

3. COMBECHEM

In the chemistry community a major source of data is the individual scientist working in a traditional chemistry laboratory, supplemented today by the advent of automated high-throughput laboratory technology. Data is generated in laboratories spread throughout the global community, each contributing significant discoveries, which are collated via the worldwide literature. As well as the chemists reviewing their literature, several different organisations attempt to correlate the various material produced across the wide and diverse community. Parallel to the worldwide academic community there are the large islands of semi-isolated but heavily-loaded commercial chemical data sources, which represent the industrial component of the chemical market. Transfers between these two effective VOs is an aspect of the Technology Transfer Process that the UK government is promoting in general between Universities and Industry.

These individual laboratory data sources are an excellent example of a Data Grid, or rather the need for one. Each site acts almost independently but adheres (more-or-less) to standard chemical practice. This means that considerable human effort is required to assemble, assign, correlate, document and standardise to enable the whole community to access the data using conventional relational database mechanisms. While this has been achieved in several sub-domains of chemistry (for example, crystallography), as a whole the required human effort is simply too great and exceeds the resources available. The modern generation of high-throughput technology has in the recent years led to an exponential rise in the amount of chemical information being produced.

A methodology to collect this information in a manner that will enable automated generation of chemical knowledge without imposing unrealistic uniformity on the researchers is a major part of the end-to-end objective of the Combechem project [11]. We call this strategy 'publication at source' – the dissemination not only of experimental data, but also contextual information about the conditions, setup, running and ongoing analysis of experiments. Related projects include the Collaboratory for Multi-scale Chemical Science (CMCS) [12] and the World Wide Molecular Matrix [13].

In all the stages of the chemical data pathway that CombeChem aims to support in the vision described above, data is being exchanged between people and computers, and increasingly between computers and computers. In the current practice the subsequent ability to use the data in interpretation of a chemical reaction mechanism or structure depends heavily on the scientist keeping explicit track of the metadata (for example, which sample compound is associated with a spectra and under what conditions).

Depending on the circumstances, the amount of metadata needed in the interpretation phase can be very considerable. All the relevant metadata, which is usually hidden inside a written laboratory notebook, needs to be made accessible digitally if the data item is to be properly understood. All of the constituent players in the relevant VO in principle need access to this data and metadata, something provided for in the Combechem vision.

The project set out to provide comprehensive semantic support for the whole spectrum of chemistry research in as transparent fashion as possible. A fundamental part of this strategy was to study established practice in the field and to introduce as few changes to normal everyday working practice as possible. This was part of the objective of capturing as much metadata at source (i.e. as it is generated) completely automatically. CombeChem adopted the additional premise that it would be impossible to predict in advance the way that data would be accessed and used, hence flexibility of use was a fundamental objective. This led directly to the requirement that the information infrastructure to hold this data and metadata should be as general as possible.

The design approach adopted four principles:

1. Grounding in established operational practice – the starting point is to study chemists at work;
2. Capturing a rich set of associations between all types of things, expressed pervasively in RDF and hence explicitly addressing the sharing of identifiers;
3. Metadata capture should be automated – the goal is augmentation not disruption;
4. Information will be reused in both anticipated and unanticipated ways.

The result of this exercise is a Semantic Datagrid [14] comprising large numbers of RDF triples across multiple triplestores (45 million triples are cached in the 3store scalable triplestore [7] at the time of writing, which is still a small fraction of the metadata available). Shared identifiers have been provided by the IUPAC INCHIs [15]. The chemists have particularly benefited from the flexibility of working with RDF as opposed to the rigid schema of relational databases – the RDF has been used both to glue together existing datastores and to encode harvested information.

Although, like CS AKTiveSpace, CombeChem includes harvested information, it contrasts CS AKTiveSpace in that it uses multiple stores and does not impose a single ontology. Furthermore a significant part of the content creation is a result of user participation: by working within the CombeChem framework, users gain the added value (network effect) of the accumulating knowledge.

Again, it is essential to build interfaces that meet user requirements. CombeChem starts in the laboratory [16,17] and involves all the users in the lifecycle of the scholarly output [18].

4. COAKTING

The objective of CoAKTinG ('Collaborative Advanced Knowledge Technologies in the Grid') has been to advance the state of the art in collaborative mediated spaces for distributed e-Science through the novel application of advanced knowledge technologies. It comprises four tools: meeting capture and replay, instant messaging and presence notification (BuddySpace), graphical meeting and group memory capture

(Compendium) and intelligent 'to-do' lists (I-X Process Panels).

Each of these individual tools has proved successful performing specific tasks supporting collaborative environments (such as the Access Grid); collectively they also provide metadata sources which, through use of a shared OWL ontology to exchange the structure they gather, promote enhanced process tracking and navigation of resources before, during, and after a meeting occurs. 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.

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: each of the CoAKTinG tools can be thought of as extracting structure from the collaboration process, as illustrated in Figure 2. This structure, mediated through the ontology, is used to construct the Meeting Replay tool – a hypertext navigation of the video, centred around an annotated interactive timeline.

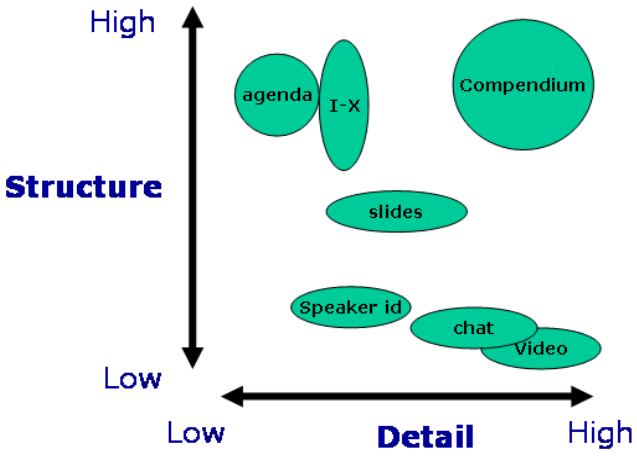

Figure 2. Structure vs. Detail in CoAKTinG

In its latter phase CoAKTinG has conducted trials with NASA and with CombeChem [20], and this has emphasised the requirements of interworking annotations across projects.

The tools provide support for the e-Science process in Comb-e-Chem and they also enable the digitisation of 'missing links' in the processing chain which form part of the typical collaborative scientific processes that we are attempting to enhance: support of the experimental process, tracking and awareness of people and machine states, capturing of the discussions about data as well as the traditional metadata, and enriched metadata regarding these components to support interlinking.

CoAKTinG has provided proof of concept, with successful trials, and it has also highlighted research challenges. Specifically, capture and retrieval of live metadata stresses the existing semantic web infrastructure (e.g. RDF triplestores are not engineered for real-time update and for querying over space and time), the CoAKTinG tools work with a small number of ontologies and domain-specific ontologies cannot be plugged in, and the problems of distributed collaborative annotation (such as privacy and conflicts) have not been fully addressed.

The Memetic project continues the development of some of the tools and ideas in the CoAKTinG project with a particular emphasis on deploying in Access Grid nodes with end-users [21]. A participatory design process was used in a series of workshops to provide end-users with a familiarity of technology and allowed them to develop an emergent wish-list, which lead to a series of core requirements. The project focusses on the meeting replay and Compendium tools and combines them with a meeting management portal and components to record and replay Access Grid sessions. This allows users to make structural annotations about meetings they attend and upload them to a meeting repository and thereby continue the collaborative process.

5. MUSIC INFORMATION RETRIEVAL

Music Information Retrieval is the science of extracting features from musical content, such as melody and tempo, to facilitate tasks such as analysis and music retrieval. This research community is highly interdisciplinary, ranging from signal processing experts to music scholars and digital libraries experts. The community has established an array of software tools to support this work. [22]

In 2005 the 1st Music Information Retrieval Evaluation eXchange (MIREX) took place during the 6th International Conference on Music Information Retrieval (ISMIR). MIREX aims to compare state-of-the-art algorithms and systems relevant for Music Information Retrieval. By way of example, the evaluations included: artist identification, drum detection, genre classification, melody extraction, tempo extraction, key finding and melodic similarity.

To achieve this, people submit their code to a central site where it is executed against a standard database of musical content. The centralisation is necessary because

the content itself cannot be distributed, for licensing reasons.

Running MIREX has raised a number of technical challenges which this community wishes to address, including:

- Ease of execution of remote code by the community (a Web Services solution is being considered);
- The computational challenge (the codes can take a long time to run);
- Sharing of intermediate data used by the algorithms (since some algorithms perform the same pre-processing stages – for example, Fast Fourier Transforms);
- Creation and sharing of annotations: establishing the ground truth against which the algorithms are compared (currently achieved using experts but this could extend to the broader community); and sharing of automatically extracted annotations; sharing of performance results from running software components and algorithms.

The last point is significant, since the annotations are formed by a diverse community of users. This is an example of 'social tagging', where the community create the metadata. The collaborations within the community, and the sharing of resources and services, can be seen as formation of social networks or virtual organisations.

For people who want to make use of the algorithms in their research, a future research environment is envisaged in which there may be multiple services available which perform a variety of feature extraction functions. Additionally, there will be many applications of these techniques in professional and consumer systems, which raise a related set of requirements outside the research context. Application scenarios include Content based retrieval (e.g. taking a clip of music and finding pieces that match), analysing, organising and generating personal playlists for portable audio players, music recommender systems based on musical selections (e.g. collections, playlists) of multiple users and detection of plagiarism of musical content.

These challenges exhibit some classic Grid requirements – accelerated computation, working with large volumes of data, and remote execution. However the data sharing and annotation are classic Semantic Grid problems, as is the description and discovery of resources and services within the research or commercial environment. The MIREX exercise involves evaluation of algorithms and generation of intermediate results which provide additional forms of information within the system.

There are interesting distributed and collaborative dimensions. If multiple people create annotations about the same pieces of music then it should be possible for this knowledge to accumulate, providing a rich network of metacontent which is itself the basis for new functionality (e.g. in search). To achieve this requires establishing ways of uniquely identifying content and sharing those identifiers, and describing the relationships between the "same piece of music" appearing in various forms. At the level of describing pieces of music there are already classifications in use; for example, in online music catalogues. These vary – there is not a unique underlying conceptualisation of this domain.

The steps needed to apply Semantic Grid in this context would involve establishing a shared identifier scheme for musical content and descriptions of the relationships between musical content, and establishing metadata schema to describe derived data, features (ground truth annotations and extracted features), feature extraction techniques and their evaluation, software components and deployed services.

The latter points are perhaps an area of the Semantic Grid that has not been explored before – the use of Semantic Web technologies to describe software components, algorithms, tests and results, in support of the developer community.

6. DISCUSSION

Each of these case studies illustrates different aspects of the Collaborative Semantic Grid. Significantly, we note that they all illustrate a trend provision of content, metadata and services by multiple participants, leading to an evolving, self-organising ecosystem rather than a traditional publishing model of providers and consumers. This trend is intrinsically collaborative

CombeChem illustrates this trend from the traditional publishing models towards a participation approach by adopting philosophies such as 'publication at source' (and, from a Semantic Web viewpoint, 'annotation at source') which is making all the users first class citizens within the information and knowledge environment. CombeChem illustrates exploitation of provenance, opening up the community to more rapid dissemination of full descriptions of investigations, enabling different communities to combine understanding and exchange data. Collaborations are built while supporting and retaining the dynamic nature of research discovery as exemplified by the ease with which new information and descriptions can be applied and spread within the community.

CoAKTinG and Memetic extend this to synchronous collaboration, illustrating the value of shared annotation in realtime but also in subsequent reuse – it illustrates a 'record and reuse' paradigm, which is also a way of looking at CombeChem. In fact Semantic Web technologies have been developed in the context of fairly static content and the real-time nature of the annotations in CoAKTinG do pose an engineering challenge for these technologies.

MIREX is a compelling example of how Grid and Semantic Web separately support the collaborative requirements of a given community – to share code, to show results and potentially to establish 'ground truth'. It also hints at a future environment where there are many services available and the community will collaborate over the creation, sharing and use of these services.

This last point is important, as it suggests something about future collaboration. The Semantic Grid offers a vision of automatic service creation by composing services and resources in a 'service soup', as for example advocated in the Service Oriented Knowledge Utility vision [23]. This semantically-described service-oriented architecture is itself a collaborative infrastructure in the sense that it facilitates sharing of services, i.e. it permits multiple participants to produce, discover and consume services.

This is not a traditional form of collaboration, but we would suggest that sharing services will be part of future collaborative behaviour as well as sharing content and other resources. It is interesting to reflect on how in the future there will be exchange of services in addition to the exchange of content that we see today. How do collaborative behaviours relating to content extend to services? Semantic Service Oriented architecture research has focused on automation rather than collaboration.

By tracking provenance of research results, CombeChem facilitates the interpretation of scientific data but it also provides a basis for establishing intellectual property ownership and probity. All of our examples have privacy and digital rights management issues, which are largely unaddressed at this stage. The scientific discourse captured in CoAKTinG is not usually placed in the public domain in the same way as the scholarly publications that may result. Meanwhile it is the digital rights management aspect of MIREX which imposes the need for taking computation to data. Some of the added-value of the Semantic Web results from its interlinked content, and restrictions to sharing play against this. Furthermore, we have the distinction between metadata privacy and data privacy. All these issues demand further exploration.

7. CONCLUSION

We have shown that Semantic Web technologies can be used to enhance both asynchronous and synchronous collaboration in the Grid, supporting the creation and operation of virtual communities.

We have observed two important trends in the kind of collaboration supported by the Semantic Grid: from publishing-based to participation-based, and from content-based to service-based. All three of our scenarios suggest that the participation-based approach is a particularly effective strategy.

The set of research issues that arise from this include collaboration over services as we enter a world with increasing numbers of services available, dealing with incomplete, inconsistent and dynamic knowledge, as well as privacy and digital rights management.

We also encourage the engagement of the social network analysis community with the Grid community to explore the evolving notion of Virtual Organisations and to develop tools and concepts which take advantage of the symbiosis between social networks and VOs.

ACKNOWLEDGMENT

This paper builds on a talk given at the NSF Social Networks and Cyberinfrastructure (SNAC) workshop at NCSA in November 2005, and thanks are especially due to Noshir Contractor for our participation in this event. The Music IR scenario is due to J. Stephen Downie. This research was partially supported by Combechem and CoAKTinG (e-Science programme EPSRC grants GR/R67729/01 and GR/R85143/01), and the authors wish to extend thanks to the entire project teams and to colleagues in the Advanced Knowledge Technologies IRC (GR/N15764/01), especially Nigel Shadbolt, Stephen Harris, Nicholas Gibbins and Hugh Glaser. eBank and Memetic are funded by UK JISC under the Semantic Grid and Autonomic Computing Programme and the Virtual Research Environments programme.

REFERENCES

[1] Foster, I., Kesselman, C. and Tuecke, S. "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," *International Journal of Supercomputer Applications*, vol. 15, 2001.

[2] Berners-Lee, T., Hendler, J. and Lassila, O. "The Semantic Web," *Scientific American*, 284(5), 2001, pp. 34-43.

[3] De Roure, D., Jennings, N.R., and Shadbolt, N.R. “Research Agenda for the Semantic Grid: A future e-science infrastructure,” Technical Report UKeS-2002-02, National e-Science Centre, December 2001.

[4] Goble, C.A., De Roure, D., Shadbolt, N.R. and Fernandes, A.A.A. “Enhancing services and applications with knowledge and semantics,” in I. Foster and C Kesselman, editors, The Grid 2: Blueprint for a New Computing Infrastructure, Morgan-Kaufmann, 2004, pp. 431-458.

[5] De Roure, D., Jennings, N.R., and Shadbolt, N.R. “The Semantic Grid: Past, Present, and Future,” *Proceedings of the IEEE*, Volume 93, Issue 3, March 2005, pp. 669-681.

[6] Shadbolt, N.R., Gibbins, N.M., Glaser, H., Harris, S. and schraefel, m.c. “CS AKTive space or how we stopped worrying and learned to love the semantic web,” *IEEE Intell. Syst*., vol. 19, no. 3, May–June 2004, pp. 41-47.

[7] Harris, S and Gibbins, N. “3store: Efficient Bulk RDF Storage,” Proceedings of the First International Workshop on Practical and Scalable Semantic Web Systems (PSSS2003), Sanibel Island, Florida, USA.

[8] The FOAF Project, see http://www.foaf-project.org/

[9] schraefel, m. c., Smith, D. A., Owens, A., Russell, A., Harris, C., and Wilson, M. 2005. “The evolving mSpace platform: leveraging the semantic web on the trail of the memex,” Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia (Salzburg, Austria, September 06 - 09, 2005). HYPERTEXT '05. ACM Press, New York, NY, pp. 174-183.

[10] De Roure, D. and Hendler, J.A. “E-Science: The Grid and the Semantic Web,” *IEEE Intelligent Systems*, Special Issue on E-Science, Volume 19, Issue 1 (January/February), 2004, pp. 65-71.

[11] Frey, J.G. et al. “Combinatorial Chemistry and the Grid,” in ‘Grid Computing: Making the Global Infrastructure a Reality’, edited by F.Berman, G.Fox and T.Hey, Wiley 2003.

[12] Myers, J.D. et al, “A Collaborative Informatics Infrastructure for Multi-scale Science,” Challenges of Large Applications in Distributed Environments (CLADE) Workshop, Honolulu, 2004.

[13] Murray-Rust, P. “The World Wide Molecular Matrix – a peer-to-peer XML repository for molecules and properties,” presented at EuroWeb2002, Oxford, UK, 2002.

[14] Taylor, K., Gledhill, R., Essex, J.W., Frey, J.G., Harris, S.W. and De Roure, D. “A Semantic Datagrid for Combinatorial Chemistry”, Proceedings of IEEE Grid Computing Workshop at SC05, IEEE, Seattle, WA. November 2005.

[15] InChI International Chemical Identifier, see http://www.iupac.org/inchi/

[16] Hughes, G., Mills, H., De Roure, D., Frey, J.G.. Moreau, L., schraefel, m.c., Smith, G. and Zaluska, E. “The semantic smart laboratory: A system for supporting the chemical e-Scientist”, *Org. Biomol. Chem*. Vol. 2, No. 22, 2004, pp. 3284-3293.

[17] schraefel, m.c., Hughes, G., Mills, H., Smith, G., Payne , T. and Frey, J.G. “Breaking the Book: Translating the Chemistry Lab Book to a Pervasive Computing Environment,” Proceedings of the Conference on Human Factors (CHI), 2004.

[18] Duke, M., Day, M., Heery, R., Carr, L.A. and Coles, S.J. “Enhancing access to research data: the challenge of crystallography,” Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital libraries. 2005, pp. 46-55.

[19] Page, K. R., Michaelides, D. T., Buckingham Shum, S. J., Chen-Burger, Y., Dalton, J., De Roure, D. C., Eisenstadt, M., Potter, S., Shadbolt, N. R., Tate, A., Bachler, M., and Kozmak, J. “Collaboration in the Semantic Grid: a Basis for e-Learning.” In *Journal of Applied Artificial Intelligence* 19(9-10), 2005, pp. 881-904.

[20] Bachler, M., Buckingham Shum, S., Chen-Burger, Y., Dalton, J., De Roure, D., Eisenstadt, M., Frey, J., Komzak, J., Michaelides, D., Page, K., Potter, S., Shadbolt, N. and Tate, A. “Chain ReAKTing: Collaborative Advanced Knowledge Technologies in the Comb-e-Chem Grid,” UK E-Science All Hands Meeting, Nottingham, UK. 2004.

[21] Buckingham Shum, S., Slack, R., Daw, M., Juby, B., Rowley, A., Bachler, M., Mancini, C., Michaelides, D., Procter, R., De Roure, D., Chown, T., and Hewitt, T. “Memetic: An infrastructure for Meeting Memory,” Proceedings 7th International Conference on the Design of Cooperative Systems, Carrey-le-Rouet, France, 2006.

[22] Futrelle, J. and Downie, J.S. “Interdisciplinary Research Issues in Music Information Retrieval: ISMIR 2000-2002”, *Journal of New Music Research*, 2003, Vol. 32, No. 2, pp. 121-131.

[23] NGG Expert Group. “Future for European Grids: GRIDs and Service Oriented Knowledge Utilities,” Next Generation Grids Expert Group Report 3. Available from http://www.cordis.lu/ist/grids/