UNIVERSITY OF LEEDS

This is a repository copy of *Enabling e-Research in combustion research community* .

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/1898/

Book Section:
Pham, T.V., Dew, P.M., Lau, L.M.S. et al. (1 more author) (2006) Enabling e-Research in combustion research community. In: Second IEEE International Conference on e-Science and Grid Computing, 2006, 4-6 December 2006. IEEE , p. 130. ISBN 0-7695-2734-5

https://doi.org/10.1109/E-SCIENCE.2006.261063

Reuse
See Attached

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Enabling e-Research in Combustion Research Community

Tran Vu Pham
School of Computing
University of Leeds
Leeds, LS2 9JT, UK
tranp@comp.leeds.ac.uk

Peter M Dew
Informatics Institute
School of Computing
University of Leeds
Leeds, LS2 9JT, UK
dew@comp.leeds.ac.uk

Lydia MS Lau
School of Computing
University of Leeds
Leeds, LS2 9JT, UK
llau@comp.leeds.ac.uk

Michael J Pilling
School of Chemistry
University of Leeds
Leeds, LS2 9JT, UK
m.j.pilling@leeds.ac.uk

Abstract

This paper proposes an application of the Collaborative e-Science Architecture (CeSA) to enable e-Research in combustion research community. A major problem of the community is that data required for constructing modelling might already exist but scattered and improperly evaluated. That makes the collection of data for constructing models difficult and time-consuming. The decentralised P2P collaborative environment of the CeSA is well suited to solve this distributed problem. It opens up access to scattered data and turns them to valuable resources. Other issues of the community addressed here are the needs for computational resources, storages and interoperability amongst different data formats can also be addressed by the use of Grid environment in the CeSA.

1. Introduction

e-Science is about global collaborations amongst key areas of science and the infrastructures that enable it [21]. The introduction of e-Science is the result of the increasing demand for distributed global online collaborations to support international collaborations amongst scientific communities. The concept of e-Science is now broadening to e-Research to encompass the social sciences, the arts and humanities [1]. Generally, e-Research *"encapsulates research activities that use a spectrum of advanced ICT*[1] *capabilities and embrace new research methodologies"*, which emerge for increasing access to network, hardware and software infrastructures together with application and collaboration tools [6]. The kind of collaborations addressed by e-Science, and hence e-Research, includes the exchange of scientific results, potentially in huge volume of data, amongst scientists from across disciplines and the sharing of back-end computational resources, such as CPU cycles, memory and storages.

Research in combustion is a typical example. The main subject of studies in combustion is the modelling of chemical reaction mechanisms and their applications in combustion processes such as burning of fuel in combustion engines. The modelling of chemical reaction is a complicated process. It requires knowledge, expertise and data from a range of research areas in chemistry such as quantum chemistry, thermodynamics, measurement and calculation of reaction rate coefficients. The modelling of chemical reactions is a central area in reaction kinetics, with applications not only in combustion but also in atmospheric chemistry. Research activities in these three areas are interrelated through this modelling process. Therefore, there is the need for collaborative research activities to transfer research data, knowledge and expertise within research disciplines and communities in applied reaction kinetics. There is also considerable advantage in developing communities across disciplines, e.g. amongst combustion, reaction kinetics and atmospheric chemistry (as illustrated in Fig. 1).

Grids and Grid technologies are currently seen as the most promising infrastructures to support distributed collaborations. It deals with the coordinated sharing of huge amount of scientific data and computational power across

[1] Information and Communication Technology

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

IEEE COMPUTER SOCIETY

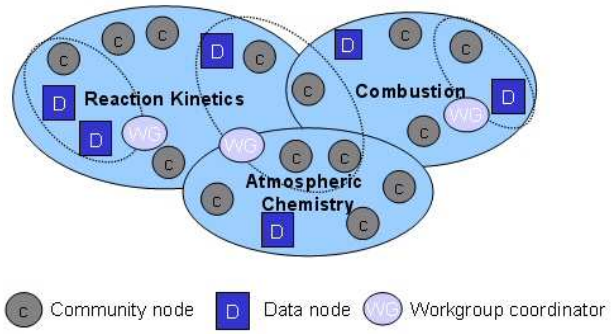

Figure 1. Collaborative research activities amongst inter-related research communities of combustion, reaction kinetics and atmospheric chemistry

organisational boundaries and disciplines [7, 9, 8]. An example in the domain of bioinformatics is myGrid, which deals with similar requirements for collaborations [19][12]. myGrid focuses on the middleware components that enable collaborations using distributed Web Services and composition of these services into workflows. This approach is workgroup oriented. The collaboration amongst end users at a community scale has not been addressed. In general, the Grid model alone is often project oriented and does not scale well with respect to the granularity and heterogeneity of community [17].

Web-based collaborative portals are commonly used as environments for end users of distributed communities to get involved in the collaborations, such as the Virtual Knowledge Park [31] and the British Atmospheric Data Centre[2]. However, the centralised Web-based approach has shown to be inflexible for distributed and loosely coupled communities as it does not well support spontaneous collaborations[29].

Peer-to-Peer (P2P) is a computing model that supports the sharing of smaller computing resources. This model has been proved successful in a number of commercial desktop file-sharing applications such as Napster [20] and Kazaa [16]. This success has motivated the use of P2P model in scientific projects such as Triana [30] and Chinook [13]. P2P is a decentralised network-computing model, where computation is taken place at the edges. That makes it scale well with respect to the community structure. In addition, P2P applications often provide means for real-time communication. It is highly suitable for direct and spontaneous collaborations amongst users. In an essence, the decentralised model of P2P can efficiently support user collaborations at a higher level of granularity.

Given the pros and cons of each of the architectural models discussed above, there is considerable advantage in the integration of P2P and Grid. This paper discusses a proposed application of a Collaborative e-Science Architecture (CeSA) [22], which is an integration of Grid infrastructure and a P2P collaborative environment, for the combustion research community. It firstly reports on the requirements collected from the research community. The paper then describes the CeSA and earlier work on its prototype implementation and evaluation. Further work on the CeSA together with an assessment of related work will be presented.

2 Research in Combustion

Research in combustion chemistry is centred around reaction models [24]. The basic component of a reaction model is a chemical reaction mechanism, which consists of a series of steps called elementary reactions in which chemical species are inter-converted. Each elementary reaction is associated with involved species (reactants and products) and a rate coefficient, which determines the rate at which the reaction occurs. The elementary reactions and their associated rate coefficients are investigated in laboratory. It is also feasible to calculate some rate coefficients using quantum theory. The computing resource needed for this approach is substantial. This mechanism can then be used to construct a model that consists of a set of ordinary differential equations that represent the rates at which the concentration of individual species in the mechanism changes with time.

Application of reaction models in combustion, for example, involves the interaction between chemistry and fluid dynamics. This adds a further stage of complexity and requires an additional set of scientists with specific expertise. This stage is essential in applications to real systems, such as the design of engines. It is often divorced from the chemical developments, for practical reasons, but it should ideally be incorporated in the overall set of interactions, so that feedback is feasible between all elements.

2.1 The Three Stage Modelling Process

The reaction modelling process, the central activity in combustion research, consists of three stages: gathering and evaluating data, generating mechanisms and models, and publishing and archiving models, as summarised in Fig. 2.

i. *Gathering and evaluating data*: Data gathering and evaluation is an essential stage of model construction. In this stage, elementary reactions are identified together with their reaction rate coefficients and thermodynamic data required for the reaction mechanism of the model. These data are produced by different research groups in the community and scattered over various sources. The data need to be collected and evalu-

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE
IEEE COMPUTER SOCIETY

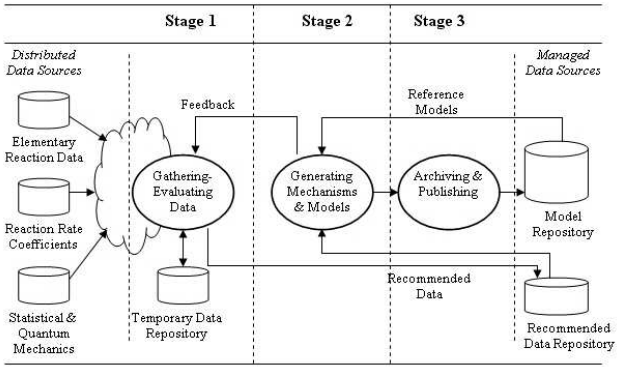

Figure 2. The three stage modelling process

ated. Good datasets are then archived as recommended data for use in later stages.

ii. *Generating mechanisms and models*: In this stage, relevant elementary reactions and their associated parameters gathered and evaluated in the previous stage are put together to build a mechanism. The resulting mechanism is then put into a model. This comprises ordinary differential equations describing the chemistry and partial differential equations describing the fluid dynamics. The model can then be tested in a variety of ways, for example, through experiments on a flame, in which the concentrations of some of the species are directly measured and checked against those simulated using the mechanism. A sensitivity analysis can also be conducted. This is to determine the sensitivity of an important observable to the mechanism components, e.g. the rate coefficients, to allow the experiments to be targeted at key features of the mechanism. The result of sensitivity analysis is essential feedback to the overall model development process.

iii. *Publishing and archiving new models*: The resulting model from this process needs to be published so that other researchers and potential model consumers are aware of its availability. The new model is also archived for later use by application engineers or by other modellers as a referenced model. For archiving purposes, standards for data formats are essential, so that the archived models can be easily retrieved and used without the need for any conversion.

2.2 Current Limitations and Issues

One major problem that the combustion research community is facing is that the data required for generation of reaction mechanisms and models is scattered in the community and often inadequately evaluated. There are different groups working on different reactions and aspects of data required for the modelling process. There might also be two or more different groups working on the same reactions and set of data. Currently, there is no coordination across the group. The organisation of research topics amongst these groups is unstructured. Therefore, it is often the case that a dataset produced from by one group is different from another group, even though they are both working on the same reaction. There is the need for evaluation to select the best datasets as recommendations for use in later stage of the modelling process.

International panels consisting of experts in the field are often set up to evaluate these kinds of data. Typically, they meet together once a year in a peer-review meeting. During the meeting, participants discuss and recommend rate data generated by experiments. The outcomes of such a meeting are recommended datasets which are then deposited to a database and used as reference data. This process is time-consuming and costly.

There are examples of evaluated databases in combustion [3] and atmospheric chemistry [14]. The former is currently only available in hard copy, while the latter is accessible from a website. However, such efforts are still patchy. Recently, a Process Informatics Model (PrIMe), set up by the initiative of experts in combustion, has started to coordinate the development of predictive chemical reaction models [25]. In PrIMe, the evaluation is broken down to smaller sets of reactions. Each set is evaluated by a workgroup, operating remotely. In turn, other workgroups will eventually cover all aspects of the three stage process. This is described in section 2.1. The PrIMe project, however, is currently focused on the development of data repositories.

Another issue that concerns the community is the existence of many different formats for the same set of data. This is also a result of a lack of collaboration and coordination amongst research groups in combustion. Many different tools (e.g. for simulations, analyses or editing model data) are used in the community, and are often self-built by individual groups to meet their own needs. The formats of data are therefore customised to the habits and conventions of the groups that build the tools. Furthermore, different versions of the same tool might also be used at the same time. As a result, different researchers, or research groups may use and produce data using different formats and standards. That makes the data transfer from one group to another group more difficult. Format conversion tools are often necessary. There is the need for a lightweight ontology.

Finally, the construction of a reaction model often involves hundreds to thousands species and requires large sets of ordinary differential equations. Solution of these equations at a large number of "computational grid points" in a

IEEE COMPUTER SOCIETY

computational fluid dynamic simulation requires a substantial amount of computational resources. Currently, the platforms on which these tools are running are often personal computers or workstations in clusters. As executing large amount of ordinary differential equations, sometimes, takes hours to days or even weeks, these computers may not be adequate.

2.3 Requirements for Combustion Research

The combustion research community needs a collaborative infrastructure for their distributed collaborations. In particular, the infrastructure needs to be able to:

- Allow scientists who are working on the same or similar research activities to dynamically form working groups (small focused groups) to make the data transfer process from one research group to another easier and smoother.
- Provide efficient support for timely collaborations within and across working groups in the community for sharing expert knowledge, day-to-day working data, such as experimental data, chemical reaction mechanisms and related input data for reaction modelling to speed up the data collection and evaluation process.
- Provide easy access to computational intensive resources for time and resource consuming simulations and analyses and for storage of large amount of experimental data deal with large amount of calculation and storage required by the community.

3 The Collaborative e-Science Architecture

Previous work on the Collaborative e-Science Architecture (CeSA) which has detailed the prototyping, an early use experiment and evaluation [22]. This section briefly outlines the CeSA and the earlier work.

3.1 Goals

The CeSA is aimed at a better support for collaborations within distributed scientific communities by providing an integrated collaborative environment. In order to sufficiently support the community, the CeSA:

- is scalable with respect to decentralised nature of scientific communities.
- supports scientific collaborations at different levels of granularity, so that different collaboration activities such as those required in section 2.3 could be performed.
- provides access and enables back-end computationally intensive resources for complex computation and storage requirements.

3.2 A High Level View of the CeSA

A high level view of the CeSA is shown in Fig. 3. The CeSA consists of two layers: a P2P collaborative environment on top of a Grid environment.

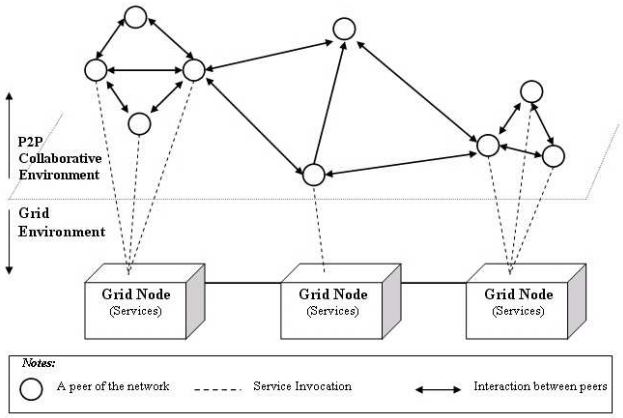

Figure 3. High level view of the Collaborative e-Science Architecture

The Grid environment may consist of one or more Grids, represented by Grid Nodes in Fig. 3. They may or may not be connected together. Grids handle computational intensive resources (e.g. CPU cycles, memory, network bandwidth and storage) and scientific data (e.g. data generated by experiments). The kind of collaborations supported by the Grid environment is often heavyweight. The Grid architecture adopted for the CeSA is specified by the Open Grid Services Architecture [8, 10], as shown in Fig. 4.

The P2P collaborative environment consists of P2P applications. Collaborations supported by the P2P environment are smaller granularity and lightweight processes. For example, a direct exchange of information about a Grid resource or a working dataset between two researchers. An important property of the P2P in the CeSA is its ability to scale well with the size of the community. In addition, cross working group, discipline and community boundary collaboration is easier and transparent in a P2P environment.

This combination will leverage the capability of providing computationally intensive resources of the Grids and the advantage of P2P computing environments in enabling

IEEE COMPUTER SOCIETY

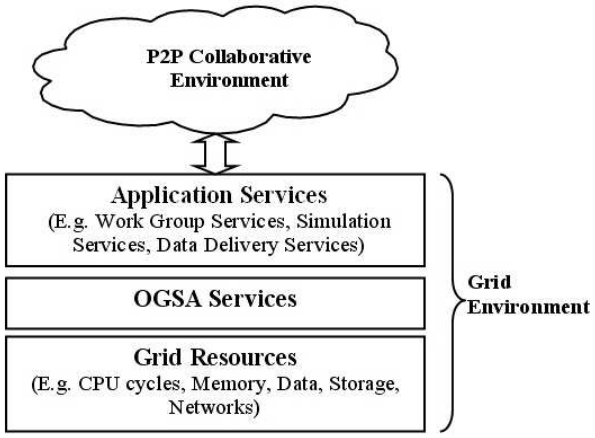

Figure 4. Grid architecture for the CeSA

lightweight and direct scientific collaborations within scientific communities.

This integration is done is a way that the Grid and P2P environments are loosely coupled together. The P2P collaborative environment can operate independently from the Grid to maintain its openness and autonomy to attract a wider user communities. Secured management within the Grid environment is separated from day-to-day collaborations in the autonomous P2P environment.

The loose coupling of the Grid and P2P environments is facilitated by a service oriented architecture. In a service oriented view, Grid resource providers enclose the resources in services. The information about the services are published in the P2P collaborative environment. Resource consumers in the P2P environment can perform resource (in form of services) discovery and then gain access to Grid resources via published service interfaces (as shown in Fig. 4). Services from the Grid environment are application specific. Examples of the application services are simulation services and data delivery services.

3.3 Prototype Implementation

The P2P application was developed using JXTA P2P technology [15]. Services implemented in the prototype system were Grid services specified by Open Grid Service Architecture (OGSA) [8].

The P2P application prototype consisted of generic tools for collaborations, such as file sharing, chatting and group formation. It also provided a service client for executing simulation and analysis services provided from the Grid environment and a service publication and discovery agent for publishing and discovery of information about services that exist in the P2P environment.

On the Grid side, a few applications used in the community for simulations and analyses of chemical reaction mechanisms were wrapped into Grid services. These applications usually use files as input and output. When wrapping these programs into Grid services using Java, input and output were mapped to the input and output parameters of Grid services. These Grid services conformed to the unified service interface. After being wrapped into Grid services using Java, these new services were deployed into a Grid service container provided in Globus Toolkit version 3.0.2.

3.4 Early Experiment and Evaluation

The main aim of this experiment was to collect feedback from a sample of three potential users for School of Chemistry, University of Leeds on the functionalities of the prototype based on the CeSA. These potential users were guided to walk through the prototype. There comments, collected using questionnaires, were invited to compare this new way of working with their current practice.

The result of the experiment on CeSA prototype system showed that P2P environment was preferable for collaborations within user community than the centralised web-based approach. The reason was that the users had more sense of control over shared resources. The users were also more willing to share as the resources can be shared directly from their computers without moving around to any third party servers. Resource consumers could also instantly get access to most updated shared by their partners. The ability to provide direct instant communication between users in a community was also an advantage of a P2P environment. The result also showed that running scientific simulation on computational intensive remote computers via Grid services would release computing resources on small user desktop computers for other day-to-day work. Easy access to Grid services via a unified service interface was also an advantage of the architecture. The following were general feedback recorded:

> *"A fully working system would benefit the atmospheric chemistry group provided it was widely accepted by the whole community"*

> *"I think that our group would certainly use such a system if it proved to be the way forward in e-Science (which I feel it is) and the community embraced the use of such a system"*

Despite of these advantages, issues about security, change documentation and connectivity of P2P applications were raised as concerns. These issues should be carefully considered in the next version of the CeSA.

IEEE COMPUTER SOCIETY

3.5 An Ontology-based Adaptive Resource Discovery Approach for the CeSA

An ontology-based adaptive approach to resource discovery for P2P collaborative environment of the CeSA has been developed to address requirements of scientific research communities. These requirements are currently not efficiently addressed by common P2P discovery methods based on flooding technique, used in Gnutella-like systems [11], and indexing using distributed hash table [26, 27, 28].

This approach separates the routing of queries from query matching mechanism so that it can efficiently route search queries in decentralised P2P environment while can still support any type of query matching techniques. Three properties of scientific research communities provide the grounding for the method: the existence of common interest groups, the willingness to share resources of common interests and the transitive relationship in the sharing behaviour. By exploiting these properties, search queries can be efficiently forwarded to those who are more likely to have the answers to improve the quality of search results and to reduce the network traffic. This discovery method uses a classification ontology to describe interests of peers. The collection of ontology descriptions of peer interests forms the basis for query routing.

Simulation results have shown that query hit rate of adaptive approach quickly got over 90% while, with the same configuration, the hit rate returned flooding approach was constantly at about 30% as reported in [23].

4 Application of the CeSA for Combustion Research Community

The result of the experiment above has motivated a further exploration on the CeSA, especially the P2P environment for collaborations within the user community and the access to remote simulations and analyses in form of services. An application of the CeSA for the combustion research community and a further user evaluation using this community are being proposed.

4.1 Mapping the CeSA

The Fig. 5 shows an overview of a realisation of the CeSA on the combustion and its overlapping research communities. In those communities, there can be members that are more data oriented (denoted as Data Nodes in the figured), while others may be ordinary researchers (Community Nodes). As shown in the figure, different research communities, i.e. combustion, reaction kinetics and atmospheric chemistry, can jointly operate in one P2P collaborative environment. Working groups can be formed in the environment, even across community boundaries. Usually, each working group has an group coordinator (shown as Workgroup Coordinators in the figure). In the P2P environment, a member can seamlessly communicate with another member across working groups and communities (illustrated by straight thin lines).

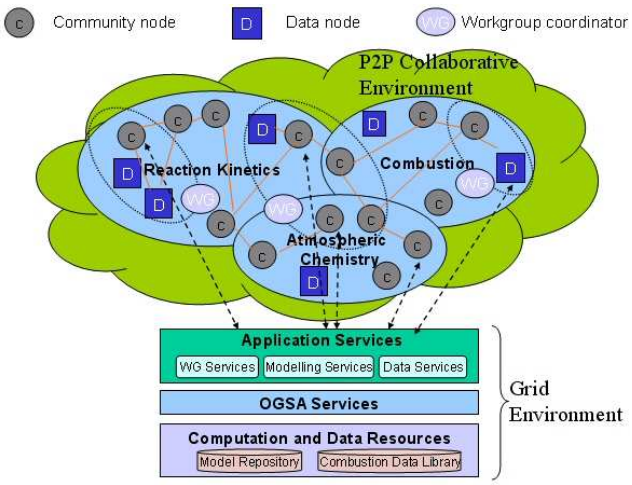

Figure 5. Application of the CeSA for combustion research community

The underlying Grid supports the community with backend computational and data resources. Access to these Grid secured resources is made possible via services designed specific for the research communities (illustrated by dashed bidirectional arrows).

4.2 Addressing the Limitations and Issues

Making scattered data easily accessible. Data scattering problem, discussed earlier, can be addressed by the P2P discovery function of the CeSA. Datasets produced by working groups or individuals usually reside on the groups' or individuals' personal storages. Only the final, well-prepared version of selected data are published. Majority of the earlier datasets are hidden from outsiders, although these datasets may be very important to some other groups. Through P2P discovery process, these datasets can become visible to others. As a result, the data gathered for a modelling process is much richer. In a reversed direction, data publication is also easier in a P2P environment. Newly produced data can be made ware of and brought into use by other members of the community.

Identifying expertise for potential collaborations. In addition, data held on a researcher's storage reflects his/her

IEEE COMPUTER SOCIETY

interests and expertise. Similarly, it reflects working areas of the working group(s) where the researcher is a member. Knowing the interests of researchers and/or working groups before they actually make publications of their research data can bring about potential collaborations in the early stages of research processes. This speeds up the collaboration process, particularly in terms of data transferring and resource sharing across community boundaries.

For example, if a group working on modelling of a chemical reaction model (i.e. process of burning methane), they will need to know the research data about elementary reactions involved in the burning process, such as thermodynamic data, structural properties, the reaction rates etc. If they know that other working groups are working on the related data, a collaboration of the related groups can be set up. The benefit for the modelling group is that they can produce the most up-to-date models. The groups working on the input data also benefit from the modelling group as their data is validated and used early in the research process. Time required for review and validation is likely to be reduced.

Supporting the modelling process with computational and data resources. Computational, storage and data resources for the modelling process are provided to the communities in form of services (available in the P2P environment). As shown in Fig. 5, potential services are Workgroup (WG) Services, Modelling Services and Data services:

WG Services are services that support collaborations amongst members of community in the P2P environment. The collaboration mainly happens in the P2P environment. However, there is still the need for support from Grid. Examples of WG Services are Shared Storage Services, WG Information Services and WG Authentication Services

Modelling Services provide computational capability for constructing reaction models. Examples of Modelling Services are Model Simulation Services, Model Optimisation Services, Model Reduction Services and Model Verification Services.

Data Services handle Grid data resources involved in the modelling process, such as experimental data, reaction rates, mechanics statistics and combustion models. Capability of Data Services is enhanced by metadata standards and ontology. Examples of Data Services are Data Publishing Services, Data Archiving Services and Data Validation Services.

The use of common sets of shared services together with standardised metadata and shared ontologies will also help to reduce the number of data formats. It will make the data transferred across platforms, working groups and communities easier and smoother. The time and efforts required for unnecessary conversion will be reduced.

In summary, the functionalities provided by the CeSA has the potential to deal with various requirements of the combustion research community. Especially, with the P2P collaborative environment, the distributed collaborations within combustion and its overlapping research communities can easily be extended across the community boundaries. In this environment, researchers can easily contribute their own resources to the community, while still maintain control over the resources.

5 Related Work

In addition to the myGrid project, which was discussed in the Introduction, the following projects are also closely related to this work.

The Process Informatics Model [25], which was proposed as a new approach to supporting the model building process by combustion community, consisting of combustion experts from the UK, other parts of Europe and the US. At the current stage, the current focus has been on a data warehouse for archiving of combustion data.

Collaboratory for Multi-Scale Chemical Science (CMCS) project [4][18] used a combination of Grid infrastructure, for management of large datasets, and web-based portal, for user's collaborations. CMCS also used PrIMe as its pilot.

CombeChem, part of the UK e-Science programme, addressed a similar chemical science research process as PrIMe, but in the domain of combinatorial chemistry [5]. It is based on the concept publication at source and pervasive computing, covering the whole research process in combinatorial chemistry from collecting of raw experimental data from laboratories, to processing, archiving data and publication of experimental results. CombeChem was built on the Grid and the Semantic Grid infrastructures.

The focus of PrIMe, CMCS and CombeChem has been extensively on semantic technologies (i.e. metadata, ontology) for describing chemical data and services to support experiments in combustion or combinatorial chemistry. The focus of the CeSA is slightly different. It is aimed at providing a scalable infrastructure to support collaborations within distributed scientific communities at different levels of granularity.

6 Conclusion and Future Work

The paper has discussed the current limitations and requirements for combustion research community. These requirements included the support for distributed scientific collaborations amongst combustion scientists and the need for access to large-scale computing resources. The combustion modelling community is a representative example of scientific communities. Hence, the CeSA has wider applicability.

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

IEEE COMPUTER SOCIETY

In order to move forwards, it is necessary to address a number of technical qualities and usability issues. Technical qualities include security, connectivity and scalability of resource discovery of P2P application. There is a need for further work on the proposed resource discovery method and management of ontologies within decentralised P2P environments. The usability issues have not been addressed but would need to be addressed before further evaluation study could be undertaken.

Acknowledgement

Many thanks to Professor Peter Jimmack and Mr Chris Martin at School of Computing, University of Leeds for their help with collecting combustion modelling requirements.

References

[1] R. Allan, A. Allden, D. Boyd, R. Crouchley, N. Harris, L. Lyon, A. Robiette, D. De Roure, and S. Wilson. Roadmap for a uk virtual research environment. Technical report, JCSR VRE Working Group, 2005.

[2] BADC. British atmospheric data centre, http://badc.nerc.ac.uk/, 2006.

[3] D. L. Baulch, C. T. Bowman, C. J. Cobos, R. A. Cox, T. Just, J. A. Kerr, M. J. Pilling, D. Stocker, J. Troe, W. Tsang, R. W. Walker, and J. Warnatz. Evaluated kinetic data for comubstion modeling: Supplement ii. *Journal of Physical and Chemical Reference Data*, 34(3):757–1397, 2005.

[4] CMCS. Collaboratory for multi-scale chemical science, http://cmcs.org/home.php, 2005.

[5] CombeChem. http://www.combechem.org/, 2006.

[6] e-Research. http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/e_research_consult/, 2006.

[7] I. Foster and C. Kesselman. Computational grids. In I. Foster and C. Kesselman, editors, *The Grid: Blueprint for a New Computing Infrastructure*, chapter 2, pages 15–51. Morgan-Kaufman, 1999.

[8] I. Foster, C. Kesselman, J. Nick, and S. Tuecke. *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*. Open Grid Service Infrastructure WG, Global Grid Forum, June 2002.

[9] I. Foster, C. Kesselman, and S. Tuecke. The anatomy of the grid: Enabling scalable virtual organisations. *International J. Supercomputer Applications*, 15(3), 2001.

[10] I. Foster, H. Kishimoto, A. Savva, D. Berry, A. Djaoui, A. Grimshaw, B. Horn, F. Maciel, F. Siebenlist, R. Subramaniam, J. Treadwell, and J. V. Reich. The open grid services architecture, version 1.0. Technical report, Global Grid Forum (GGF), 2005.

[11] Gnutella. http://www.gnutella.com, 2006.

[12] C. A. Goble, S. Pettifer, R. Stevens, and C. Greenhalgh. Knowledge integration: In silico experiments in bioinformatics. In I. Foster and C. Kesselman, editors, *The Grid: Blueprint for a New Computing Infrastructure*. Morgan-Kaufman, 2 edition, November 2003.

[13] GSC-Chinook. http://www.bcgsc.bc.ca/chinook/, 2005.

[14] IUPAC. Iupac subcommittee for gas kinetic data evaluation: http://iupac-kinetic.ch.cam.ac.uk/, 2006.

[15] JXTA. http://www.jxta.org, 2003.

[16] Kazaa. http://www.kazaa.com/, 2006.

[17] J. Myers, T. C. Allison, S. Bittner, B. Didier, M. Frenklach, W. H. Green Jr., Y.-L. Ho, J. Hewson, W. Koegler, C. Lansing, D. Leahy, M. Lee, R. Mccoy, M. Minkoff, S. Nijsure, G. V. Laszewski, D. Montoya, L. Oluwole, C. Pancerella, 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. *Cluster Computing*, 8:243–253, 2005.

[18] J. D. Myers and et al. A collaborative informatics infrastructure for multi-scale science. In *Proceedings of the Challenges of Large Applications in Distributed Environments (CLADE) Workshop*, Honolulu, USA, June 2004.

[19] myGrid. http://www.mygrid.org.uk, 2006.

[20] Napster. http://www.napster.com/, 2005.

[21] National e-Science Centre. http://www.nesc.ac.uk/, 2005.

[22] T. V. Pham, L. M. Lau, P. M. Dew, and M. J. Pilling. Collaborative e-Science architecture for reaction kinetics research community. In *Proceedings of the Challenges of Large Applications in Distributed Environments (CLADE) workshop*, pages 13–22, North Carolina, USA, July 2005. IEEE Computer Society Press.

[23] T. V. Pham, L. M. S. Lau, and P. M. Dew. An adaptive approach to p2p resource discovery in distributed scientific research communities. In *Sixth International Workshop on Global and Peer-to-Peer Computing (GP2P) in conjunction with IEEE/ACM International Symposium on Cluster Computing and the Grid 2006*, Singapore, 2006.

[24] M. J. Pilling, editor. *Low-temperature combustion and autoignition*, volume 35 of *Comprehensive Chemical Kinetics*. Elsevier, Amsterdam, 1997.

[25] PrIMe. http://prime.citris-uc.org/, 2006.

[26] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A scalable content-addressable network. In *Proceedings of the ACM SIGCOMM 2001*, pages 161–172, 2001.

[27] A. Rowstron and P. Drusche. Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems. In *Proc. of the 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware 2001)*, Heidelberg, Germany, November 2001.

[28] I. Stoica, R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In *Proceedings of the ACM SIGCOMM 2001*, pages 149–160, 2001.

[29] Y. Tian, L. M. S. Lau, and P. M. Dew. A peer-to-peer knowledge sharing approach for a networked research community. In *Proceedings of the Fifth International Conference on Enterprise Information Systems*, pages 642–645, Angers, France, Apr. 2003.

[30] Triana. http://www.trianacode.org/index.html, 2005.

[31] VKP. Virtual knowledge park, http://www.leeds.ac.uk/vkp/intro.html, 2005.

IEEE COMPUTER SOCIETY