DRAFT

E-Research Middleware:

The Missing Link in Australia's e-Research Agenda

Discussion Whitepaper on E-Research Middleware

for submission to

The Commonwealth of Australia
Department of Education, Science and Training
National Research Infrastructure Taskforce

March 2004

Editorial Team

Distributed Systems Technology Centre (DSTC)
Jane Hunter, Rob Cook, Simon Pope

Acknowledgements

The following people have generously provided feedback and comments on drafts of this paper. However, the overall responsibility for the contents of this document remains with the editorial team.

UK e-Science Programme
Anne Trefethen

APAC
John O'Callahan

GrangeNet
Paul Davis

Monash University
David Abramson

QPSF
Kevin Burrage
Bernard Pailthorpe

Macquarie University
James Dalziel

Table of Contents

1. Introduction

The purpose of this document is twofold: firstly to introduce the concepts of e-Research middleware and to articulate the reasons for its critical importance to the Australian research community; and secondly to propose the establishment of an Open Middleware program within Australia and possible approaches for the implementation such a program.

The remainder of this document is structured as follows:

- Section 2 examines the role of middleware in research infrastructure along with the benefits it brings by facilitating seamless access to distributed research information, tools and facilities, and in fostering collaboration between Australian researchers and their world-wide colleagues.

- Section 3 explores the current state of e-Research middleware in Australia with an emphasis on defining what is needed to close the gap between what we have today, and what Australian research will need in order to compete on equal terms in the international research and development community in future.

- Section 4 describes related e-Research middleware initiatives which have been established overseas and how Australia might leverage such existing programs;

- Section 5 proposes the establishment of an Open Middleware Program within Australia as a key component of the National Research Infrastructure and suggests mechanisms for developing and implementing such a program.

- Section 6 presents a number of problem issues and questions which need to be resolved in order to successfully implement such a program;

- Section 7 concludes with suggestions for how to proceed from here.

This document has been written for the Department of Education, Science and Training at the request of the National Research Infrastructure Taskforce and is the result of a collective effort. It has been compiled and edited by the Distributed Systems Technology CRC, with input from key e-Research organizations across Australia.

2. Introduction to E-Research Infrastructure

2.1 Overview and Terminology

The growing use of information and computer technologies, online instruments and high speed networks, is having a profound impact on the way research is being conducted and is leading to the emerging concept of e-Science or e-Research [2,3].

E-Science is the term that is given to large scale science that is increasingly being carried out through distributed, global collaborations, enabled by the Internet and related technologies. Typically, a feature of these collaborative scientific enterprises is their need to access very large data collections, unique scientific facilities, very large scale computing resources, and high performance analysis, modelling and visualisation by scientists. Another important aspect of this large-scale research is the high value placed on cross-fertilisation of ideas through collaboration between researchers and across disciplines.

E-Research is the more general version of e-Science and includes non-scientific research, such as humanities and social sciences, and is also characterised by the need to use distributed computing resources for collaboration and sharing of knowledge.

Grid Technologies play an important role in the development of e-Science and e-Research. In much the same way that consumers and businesses are able to access their electricity supply, *Grids* enable researchers and research institutions networked access to distributed data repositories, specialised scientific equipment, knowledge services, and computing power in an 'on demand' way. They enable flexible, secure resource sharing and coordinated problem solving amongst dynamic collections of individuals, institutions and resources – often referred to as *Virtual Organizations*.

C*yberinfrastructure* refers to the distributed computer, information and communication technologies that provide the platform on which to build the new types of scientific and engineering knowledge environments which will enable research to be conducted in new ways and with increased efficiency.

2.2 E-Research Middleware

As researchers solve complex scientific and engineering problems through new knowledge mining techniques, interactive modelling and simulation tools and collaboration, their infrastructure requirements are changing. Many of these emerging infrastructure requirements are common to scientific and engineering problems across domains. Astronomy, biology, the geosciences, public health and nano-materials, all have common needs with regard to information integration, data analysis and the secure sharing of resources. They all require secure, interoperable, seamless access to physical facilities (e.g. computer power, data storage, instruments and equipment), data and information, (large data sets, commercial and scientific databases, information and software repositories, digital libraries) and individual specialists.

E-Research Middleware refers to the software that provides standard community tools and services for knowledge management, knowledge sharing, collaboration and interoperability between applications, computing resources, institutions, and individuals across the cyberinfrastructure. It is an essential component of the e-Research infrastructure required to provide a long-term platform to empower the modern scientific endeavour.

Conceptually, middleware sits between the users or applications and the resources which they use to solve complex scientific and engineering problems (see Figure 1, below). The Middleware provides the common set of services and tools that allows researchers and applications to treat the computing, data repositories and other disparate resources as if they were one large virtual facility. Middleware moves the set of core services required by the applications into a standard, ubiquitous pool. This common service provisioning simplifies the development of applications, increases their robustness and interoperability, reduces duplication of effort and delivers widespread efficiencies in operation.

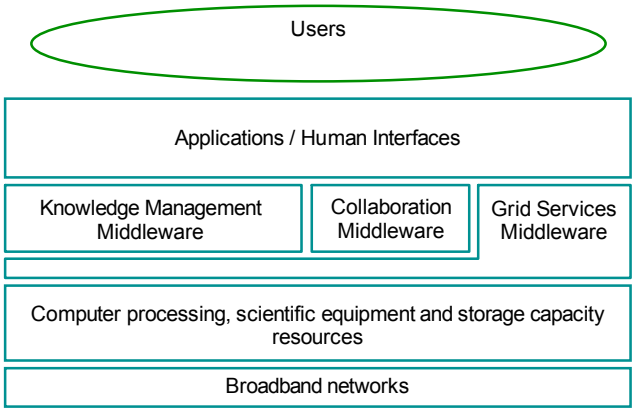


Figure 1: Key components of e-Research Cyberinfrastructure

Although we have chosen to classify E-Research Middleware into three major categories of services and tools, there are other legitimate ways to partition the middleware space. In addition, certain components (e.g., security, semantics, and provenance) actually span all three categories.

- **Grid Services/Resource Management middleware**, which includes the Open Grid Services Infrastructure (OGSI), provides essential access, communication, accounting, security, trust, and co-ordination services between the (computational and data) resources of the grid and the higher-level services that use them. Computational and data grids rely on Grid Services middleware – we also refer to this layer as Resource Management Middleware.

- **Knowledge Management (KM) middleware** provides tools and services that enable the indexing, archival, discovery, analysis, integration, management and preservation of large heterogeneous distributed data repositories and digital archives. It covers tools that enable automated indexing and assimilation of multidisciplinary datasets, interactive analysis, modelling and visualisation, and shared annotations and the mining, capture and publishing of new layers of knowledge.

- **Collaborative middleware** provides tools and services to support formal and informal, real-time and offline collaborative activities between remotely located researchers, research communities, and resources (dynamic, scalable virtual organizations).

Table 1, below, provides more detailed examples of the middleware services and tools typically required and currently being developed for eResearch. The majority of such middleware is not commercially available through off-the-shelf software and needs to be developed by researchers for specific projects. Together, these three types of middleware enable *virtual organizations* to solve highly complex problems collaboratively using remote resources, such as scientific instruments or computers, as if they were all physically located in the one institution.

Middleware type	Examples of middleware services and tools
Grid Services/ Resource Management	Resource provisioning Access control Federated identity, trust and authentication Resource scheduling, monitoring and management Workload management Virtual Organization Management Resource discovery Authentication, Authorisation Accounting and billing Data transfer and synchronisation Data Indexing and query Workflow services Event notification Recovery and fail-over
Knowledge Management	Workflows, Audit Trails, Provenance Metadata Schemas, Metadata Capture Tools Ontologies, Ontology management Inference, logic and reasoning tools Semantic indexing and querying Semantic Web/Grid Services Intelligent Brokering and Service Discovery Recommender services Annotation Tools Data fusion, Information Integration Query mediation Search engines/portals Visualisation, Presentation, Virtual Environments Simulation, Modelling Digital rights management Preservation and migration tools Knowledge Extraction and Capture Decision Support, Hypothesis Testing Virtual Organization Management tools Knowledge Publishing ELearning services and tools
Collaboration	Face-to-face collaboration e-Meeting and e-Seminar services Access Grid Nodes Peer-to-peer data sharing Document sharing Application sharing Collaborative Annotation tools Collaborative Simulation Steering Collaborative Ontology Development Group/collective knowledge capture Offline discussion tools

Table 1: Examples of three categories of cyberinfrastructure middleware

3. Australian Research Infrastructure

3.1 Current eResearch Activities in Australia

Through a variety of funding mechanisms, the Australian Government currently supports physical assets (a synchrotron[1], telescopes, microscopes), broadband networks (GrangeNet, CENTIE, AARNet, AREN), advanced computing and supercomputer facilities (APAC), access grid nodes, data repositories, digital libraries (SII) and people networks (ARC Research Networks). In addition, there are a number of research groups in Australia actively engaging in e-Science in specific domains or application areas.

Appendix A attempts to list the key Australian organizations and individuals currently participating in e-Research activities or contributing to e-Research infrastructure within Australia. This list is by no means complete but it provides an initial snapshot of activities that is expected to expand as a result of further input and feedback.

3.2 Problems and Gaps in Australian eResearch Middleware

Despite the significant Government investments in cyberinfrastructure in Australia, members of the Australian research community report that effectively utilising the resources across the existing cyberinfrastructure is often time-consuming and laborious. They are forced to use unreliable, manual methods to find suitable resources; individually negotiate access with their owners; utilise them through inefficient, expensive or time-consuming methods, often involving interstate travel. Lack of awareness of resources, facilities, services and data accessible through high speed networks, leads to lost opportunities. Users are often uncertain of the security of the systems or the protection of their research results against unauthorized access. In the absence of standardized, supported and robust interfaces, researchers are forced to invest additional time and resources in supporting and maintaining software.

A new level of software services – middleware – is required to provide the appropriate trusted, collaborative and interoperable environment required by Australian research communities. Although many of the middleware services and technologies (listed in Section 2.2) are being investigated by ICT researchers, discipline-specific research groups and industry within Australia – they are being developed without central coordination and motivated by a particular application or near-term demonstration. More co-ordination of the middleware effort is needed both within Australia, and with other international middleware programs. The current system of funding for Australian middleware infrastructure is limited and fragmented, lending itself to duplication and inefficiencies in the conduct of research and gaps in the provision of such infrastructure.

An Australian Open Middleware Program would provide the overall coordination and integration of such activities. It would extend and modify existing legacy systems to comply with the architecture of the Open Middleware Program and to provide discipline-specific services. The Program would also identify and fill gaps in existing technologies and take "proof-of-concept" research project software and upgrade it to "production strength" for use by e-Research communities.

Existing Grid services middleware, such as single-sign on, identity management, trust and access control, provisioning, scheduling and notification services, have been found to be brittle or

1 The Synchotron is primarily funded by the Victorian Government and is due to open in 2007.

unreliable when implemented across existing infrastructure. The Grid Services components require engineering effort to make them more robust and reliable so that users have seamless transparent access to a shared pool of instruments and computing and data resources. Australia will need to expand its Grid Services middleware efforts – to make it more standardised, robust and usable.

One of the key objectives of the Open Middleware Program is to resolve and streamline the interface between OGSA grid services; application level middleware developed for the Web; digital library/information management services and the emerging knowledge management services. In the past twelve months the Grid infrastructure specifications being developed by the Global Grid Forum (GGF) [11] (e.g., the Globus Toolkit [12] and the Open Grid Services Architecture (OGSA) [13]) have converged with Web services in the form of WS - Resource Framework (WSRF), recently announced by the Globus Alliance, HP and IBM. This will allow the Grid community to leverage technologies and tools developed by W3C and OASIS, that have already seen a great deal of industrial investment. WSRF and related specifications are by no means industry standards at this time and one role of the OMP will be to track these developments and ensure that they are reflected and understood in the Australian eResearch/Grid community.

To summarize:

- Existing middleware tools and services need to be identified and made more robust;
- Existing middleware tools and services need to be made interoperable, shareable, customisable and possibly integrated within a larger framework or within Grid environments;
- New middleware tools and services need to be developed. Specific areas which are lacking include: grid security, grid management and deployment, quality of service adaptation, workflow engines to streamline data capture and provenance, tools for collaborative ontology development, semantic indexing of multimedia, intelligent service discovery, decision support and hypothesis testing software, data/knowledge validation and calibration tools, automated preservation systems, collaborative visualization and simulation, and high-level grid service interfaces for application scientists [21];
- Large heterogenous research datasets (spatial, temporal, image, video, audio, 3D, spectral, graphical, multimedia) within domain-specific scientific databases need to be made accessible, shared and integrated with information resources from other domains, digital libraries (e.g., publications, journals, digital theses) and the Internet (web sites)
- A Knowledge Grid layer needs to be added to existing computational or data grids. This will involve defining the interface between knowledge management services and grid services and integrating knowledge management services into the Grid environment;
- Greater coordination of effort and increased funding is required to prevent duplication and to fill the gaps.

4. Middleware world-wide

The development of significant e-Research infrastructure in the USA, UK, Europe, Japan and elsewhere is expected to produce increased long-term economic prosperity to these countries, as a result of the increased knowledge distribution power that the infrastructure will provide [4-8]. Many of these programs have developed significant middleware and are collaborating with other

national programs to exchange and jointly develop middleware for use across global programs. The United States, through the National Science Foundation, recently recommended that they invest an additional one billion US dollars each year, for the establishment and development of an *Advanced Cyberinfrastructure* program. Of this, roughly one-third (AUD 395 million) will be invested in middleware-related technologies and activities.

Table 3, below, lists some of the more significant e-Research infrastructure programs along with their approximate annual middleware expenditure (in approx.$AUD).

Table 3: International e-Research expenditure (in $AUD)

Country / Region	Name of program(s)	Expenditure [1]
European Union	EGEE, GÉANT (FP6) + *others*[2]	$486M over 4 years [25], [26]
Japan	Information Technology Based Laboratory (ITBL) National Research Grid (NAREGI) Grid Technology Research Centre (GTRC)	$1B over 5 years[3] [22], [23], [24]
United Kingdom	UK e-Science Programme	$300M over 3 years [27][4]
United States of America	Advanced Cyberinfrastructure Internet2 Middleware NSF Middleware Initiative	$1B over 3 years [9], [28]

An Australian *Cyberinfrastructure/eResearch* Program is required to enable Australia to leverage international programs, and to ensure long-term interoperability between Australian research and those of its international partners. Given the more advanced and much better funded Grid and eScience middleware programs established in the USA, the UK and Europe, it is essential that Australian developers of middleware technologies are able to actively communicate and collaborate with the international community in this area by:

- becoming members of and participating in relevant international standards initiatives (e.g., GGF, W3C, IETF, OASIS).
- attending meetings/workshops of eResearch middleware initiatives in the US and UK (e.g., NSF Cyberinfrastructure activities, Internet2 NSF Middleware Initiative, NSF Digital Libraries Initiative, UK eScience workshops, EU EGEE Enabling Grids for E-Science in Europe).
- establishing international links through collaborative joint projects, fellowships and the exchange of researchers with major international research centres/projects in e-Science e.g., the UK e-Science Centres and the US's Argonne National Laboratory, San Diego Super-computing Centre (SDSC) and the National Center for Supercomputing Applications (NCSA) at the University of Illinois.

In addition, assuming an Australian eResearch Middleware program is established, then the proposed component organizations (i.e., the Open Middleware Institute (OMI) and Grid Operations Centre (GOC) proposed in Section 5.3) would also want to establish links with their international equivalents (e.g., NSF Middleware Initiative, GRIDS Center and DOE Science Grid in the US and the OMII and Grid Operations Centre in the UK).

---

[1] Expenditures in Australian dollars (AUD) are approximate values only.

[2] Excludes separate national programs (e.g., *UK E-Science Programme*) - includes networks, middleware and applications.

[3] Funding for Japanese projects includes some network infrastructure funding

[4] Middleware components only.

5. Proposed Australian e-Research Middleware program

5.1 Principles for an Australian e-Research Middleware program

Australia needs to build on existing research infrastructure and begin a co-ordinated and collaborative effort to develop and deploy e-Research middleware. An Australian middleware program should be inclusive of all relevant research, industry, government and academic communities in Australia. An Open Middleware Program will need to have a significant outreach program to ensure communication of the proposed e-Research middleware infrastructure to all relevant communities. It will need to be developed in parallel with the development of applications aimed at solving particular complex scientific and engineering problems.

The program would also hold great potential for industry and engagement with industry should be a key objective of an Open Middleware Program. For example, the proposed infrastructure would allow the secure federation of databases across enterprises, enable companies to share the resources of one or more partners or outsource elements of their IT environment to complete important tasks more cost-effectively. In addition, software vendor companies, such as IBM, HP, SUN and SGI, have already invested a great deal in "Web services" and are currently extending their investment to "Grid services" - an integral part of a cyberinfrastructure. An Open Middleware Program will help Australian industries to compete internationally by being closely involved in the development of these leading edge technologies. Industry involvement will also help ensure sustainability of the developed middleware technologies.

To achieve its objectives, an Australian e-Research Middleware program would need to:

- Work closely with relevant research, industry, government and academic communities in Australia to:
    - determine the commonalities in tools, services and ICT infrastructure required by researchers across all disciplines;
    - identify, co-ordinate and integrate existing relevant activities, upgrade and extend "proof-of-concept" prototypes to production-strength software, identify and fill technology gaps
- Take advantage of the more advanced e-Research programs overseas (particularly in the UK and USA) by funding involvement in international standards activities and collaborative projects which will generate key middleware components and leverage the existing middleware that is available through these international programs.
- Include an education and outreach program which will conduct activities such as seminars, workshops and conferences to educate and train researchers, inform communities about standards and best practices, promote the program results and encourage their dissemination, deployment and adoption by the Australian research communities and inform users about expected future changes in direction and technology. A significant amount of effort will be required to motivate involvement and support from all of the necessary participants.
- Collaborate with a number of specific research communities, (e.g., particle physics, astronomy, engineering, biomedical, environmental, social sciences/humanities), to develop pilots or specific testbed projects. The pilots will serve both to evaluate the middleware technologies and to assist with their customisation, adoption and deployment.

5.2 Proposed Process

It is important that the development of any Australian middleware program should follow an open, consultative, and inclusive process. This document serves as a starting point for the development of such a program, and constructive, critical input is actively encouraged.

An Australian e-Research Middleware program would need to harness the capabilities of key ICT research organisations. In addition, many different research and academic institutions, commercial organisations and government bodies currently involved in the acquisition, development or integration of middleware in Australia, need to be brought onboard. Industry partners from the mining, metallurgical, engineering, chemical and biosciences arenas should be actively involved in defining and developing the applications or pilot projects which will be built on top of the middleware. A common work program that is distributed and co-ordinated across all of these organisations is seen to be critical in ensuring any program's success.

The generation of a more detailed analysis of such a Program will require:

- A review of relevant international activities;
- The engagement of key players and organizations through a series of meetings/workshops.

Consultative input should be actively sought from all stakeholders including, but not limited to:

- Government bodies (e.g., DEST, NOIE, DoCITA, State Governments);
- Research organisations (e.g. DSTC, CSIRO, NICTA, DSTO)
- Advanced computing facilities (e.g. APAC, QPSF);
- Advanced network facilities (e.g. GrangeNet, AARNET, CENTIE);
- Industry organisations (e.g., IBM, Sun, HP, Cisco, Telstra, SGI, BHP Billiton, Comalco, Boeing, Energex, GMHolden etc.);
- Representatives from major libraries, archives, scientific, cultural and educational institutions; and,
- End-user representatives from major research organizations and domain-specific research communities with whom collaborative pilots or test-bed projects will be developed, building on the provided middleware infrastructure.

In addition, a single individual or organization with the right vision and drive will need to be appointed as the "champion" of the eResearch Middleware Program in Australia. Particularly during the consultative phase, there will be the need for a leader to chair the process and make the final overriding decisions based on input from the key stakeholders in order to move forward.

5.3 Proposed Structure

An analysis of the structure of the most significant international eScience initiatives is provided in Appendix B. Although each of these program structures differs in size and detail, there is a common composition - they each have a support service component, a development component and an outreach component. The UK model (Appendix B.1) is the most efficient, focussed and successful of the international initiatives. EScience in the US is more fragmented and less coordinated – with more players and greater duplication. This is primarily due to the larger pools of funding available in the US. US initiatives include: the NSF Middleware Initiative (NMI), the

Internet2 Middleware Initiative, the DOE (Dept. of Energy) Science Grid, NASA's Information Power Grid, the GRIDS Center, the GrADS (Grid Application Development Software) project and the NIH BIRN (Biomedical Informatics Research Network) (Appendix B.2). Europe is similarly fragmented but mainly due to the number and range of independent national programs – although the EGEE project is attempting to rectify this by coordinating eScience middleware development within the European Union (Appendix B.3).

Given the limited funding available in Australia and existing organizational structures, the most efficient approach within Australia would be to adopt a similar approach to the UK – establishing an Open Middleware Institute (OMI) to provide the coordination, middleware development and outreach components and a Grid Operations Centre (GOC) to provide the support service component. This model has proven highly successful in the UK and would allow Australia to maximize the amount of expertise and technologies which can be leveraged from the UK experience. We envisage that both the OMI and GOC will be virtual organizations whose activities are distributed between a number of key, existing organizations.

An OMI would work with government, research, industry and academic organisations to define, develop and coordinate the middleware infrastructure across existing Australian research facilities and programs, such as the Australian Foundation Facilities, Australian Landmark Facilities, Australian Major Research Facilities, Australian Research Sector Facilities, and ARC networks. The OMI would also work closely with specific research communities, (e.g., particle physics, astronomy, engineering, biomedical, environmental, social sciences/humanities), on the development of pilots or specific testbed projects which serve to evaluate the middleware technologies and to assist with their customisation, adoption and deployment. The OMI would also be responsible for delivering the outreach, education and training program - organizing events such as workshops, seminars, conferences and training, and developing and disseminating supporting documentation, best practise guides and training materials. We envisage that it would also coordinate and facilitate international collaboration activities including involvement in international standards bodies, attendance at international meetings and workshops, collaborative international projects, fellowships and exchanges and establishing links with the equivalent international organizations (e.g., NSF Middleware Initiative, GRIDS Center, UK Open Middleware Infrastructure Institute).

Advanced research activities (in the areas described in Table 1), will feed directly into the OMI's middleware development and engineering activities. Hence it is essential that related research activities are tightly integrated with the engineering being carried out within the OMI. Proof-of-concept prototypes developed within research projects should provide the basis for more robust middleware products to be developed for use within scientific and engineering applications.

The GOC would manage the operation, resources and infrastructure of the national research Grid. The GOC's responsibilities would include implementing security policies, monitoring networks, providing a suitable transition environment for testing new middleware tools, operating a helpdesk and providing support, maintenance, service and tools to user communities. This role is an extension to that already managed by the Australian Partnership for Advanced Computing (APAC) which plans to provide a Grid infrastructure for the High Performance Compute facilities in Australia.

We anticipate that further elements of an Australian Program will become defined as a thorough analysis of the e-Research middleware Infrastructure is carried out. For example, a number of independent Regional Centres which would provide regional support and outreach and assist with pilot projects for specific research communities and industries (using the middleware developed by the OMI) may need to be established. Ideally each of these centres would be co-located with existing key e-Science facilities and communities, e.g., University of Queensland, ANU, Monash, University of Sydney.

5.4 Proposed Funding

At this stage it is envisaged that all funding for an Open Middleware Program would be the responsibility of DEST and would come from existing programs. Specific funding mechanisms would need to be determined for funding the middleware effort and the organizations selected to do the work. For example, mechanisms for funding the pilot projects should follow an open, competitive grant application process such as ARC Discovery Grants or ARC Research Networks. All recognised Australian research organisations should be eligible to apply and receive funding for developing the pilot projects, which would ideally be aligned with specific research communities (e.g. those funded through the ARC Research Networks or Special Research Initiatives Programs).

Funding mechanisms would need to address the following major categories:

- Provisioning and operations of the Open Middleware Institute (OMI)
    - Coordination of activities;
    - Fundamental engineering and applied research to develop Open Middleware Services;
    - Advanced 'blue-sky' research to develop more sophisticated domain-specific knowledge management services;
    - Outreach and education program;
    - International collaborations.
- Grid Operations Centre (GOC);
- Regional Grid Centres (RGCs) and the development of the pilot projects.

6. Questions & Open Issues to be addressed

- What existing funding mechanisms can be drawn upon to fund an Australian eResearch Middleware program?
- Who will be the "champion" for such an initiative in Australia? Someone to provide the leadership and final decision-making role during the consultative process is essential.
- Who will coordinate the proposed Middleware program? Who will run the GOC?
- Should the proposed OMI and GOC be centralized or virtual distributed organizations, consisting of a number of regional centres, or combination of both?
- A significant proportion of the middleware development is not ICT research but basic software engineering – who will develop this and where will the funding come from?
- How will the long term research issues be determined, addressed and feed into the middleware and application development?
- What is the best way to implement and facilitate the "research, development, deployment, user feedback" cycle?
- How do you get the domain-specific research communities to adopt the middleware which is developed or to adopt recommended standards?

- How will the pilot projects be selected and funded? What is their relationship to ARC Research Networks?
- Who will carry out the customisation of the middleware for particular domains?
- What would be the involvement of State Governments?

7. Conclusions

Current and future Australian research infrastructure requirements have expanded beyond the hardware and communications networks that connect facilities. A new level of software services – middleware – is required to provide the appropriate trusted, collaborative and interoperable environment that Australia needs to remain competitive. Significant and highly-funded e-Research middleware programs are already established in the USA, UK, Europe, China, Japan, Singapore, and other countries. Timely development of Australian e-Research infrastructure is essential for ensuring that Australian research remains world-class.

If Australia does not adopt a coordinated approach to the development of common middleware infrastructure for e-Research environments, quickly and with sufficient levels of investment, the negative impact will be felt economically, socially and educationally. Consequences include: wasteful redundant duplication of systems development among research communities and industry; incompatibility and lack of interoperability among data formats, databases, tools, services and instruments; permanent loss of expensive observational data, archives and knowledge; increased fragmentation and isolation of research communities; reduced ability to collaborate nationally and internationally; lack of synergy amongst IT research, the IT industry and researchers; poor quality research results and a lack of commercialisation opportunities. Without such middleware infrastructure, Australia will fall behind our competitors in both research and industry.

Australia needs to establish a national e-Research Middleware program that would define and deliver the cyberinfrastructure on which to build the new types of knowledge environments that will enable research to be conducted in new ways, and with increased efficiency. It will add value to existing government research expenditure, extend the life of legacy systems and facilities and enable Australia to actively participate in the 'knowledge economy' that will help ensure our long term economic well-being. It will revolutionise the way in which research is conducted – changing *what* researchers can do, *how* they do it, and *who* participates – and it will help to ensure Australia's place in the world as a research leader. An Open Middleware Program would also fulfil many of the recommendations outlined in the Houghton report [2] for developing a sustainable research information infrastructure, collaborative research environment and scholarly communication system within Australia, as research practises change with the use of ICT and digital technologies.

To conclude, this discussion paper provides an outline for the proposed Open Middleware Program and a starting point for further discussions. The generation of a more detailed analysis of such a Program will require:

- A review of relevant international activities;
- The engagement of key players and organizations through a series of meetings/workshops.

The objective of these meetings would be to:

- Further define the scope and objectives of the Open Middleware Program by:
    - determining the commonalities in tools, services and ICT infrastructure required by researchers across all disciplines and
    - classifying and prioritising them into short, medium and long-term milestones;
- Determine the best approach to establishing, running and maintaining the Program;
- Determine how best to liaise with and satisfy the needs of research communities, education and industry;
- Estimate the funding required to develop and implement the Program.

Glossary

Cyberinfrastructure – refers to infrastructure based upon distributed computer, information and communication technology, required for a knowledge economy

Grid – infrastructure which enables computing resources, compute cycles, data storage, scientific facilities, instruments and software to be accessed on demand like electric power utilities. It enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources. In this context, a resource includes computational systems and data storage and specialized experimental facilities.

Grid Service – The Grid service concept extends Web services with additional capabilities and interfaces such as an information port and notification port.

GGF – Global Grid Forum

IETF – The Internet Engineering Task Force

OGSA – Open Grid Services Architecture

OMI – Open Middleware Institute

e-Science - refers to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists.

e-Research – more general version of e-Science, including non-scientific research e.g., humanities and social sciences, which is enabled by the routine use of distributed computing resources

Metadata – Metadata is information about resources (data, services, people) to enable the discovery, machine understanding, processing, filtering and matching of those resources

Middleware – software that connects two otherwise separate applications, services or devices and passes data between them

Semantic Grid – An infrastructure which draws on research and development in both the Grid and Semantic Web areas, to support e-Science through a service-oriented approach

Semantic Web – An extension of the current web in which information is given well-defined meaning, better enabling people and machines to work cooperatively

W3C – World Wide Web Consortium

Web Service – a software application identified by a URI (Uniform Resource Identifier) whose interfaces and bindings are defined using XML, enabling direct interactions with other software applications via XML messages and internet protocols

Workflow – automation of a process, during which information or tasks are passed from one participant to another for action, according to a set of procedural rules

References

[1] National Research Infrastructure Taskforce http://www.dest.gov.au/highered/ri_taskforce/

[2] Houghton, J., Steele, C. and Henty, M. 2003, 'Changing Research Practices in the Digital Information & Communication Environment' Draft, Canberra. http://www.dest.gov.au/highered/respubs/changing_res_prac/c_res_pract.pdf

[3] Sargent M., 'A Framework for an Australian Research and Education Network: The final report of the systemic infrastructure initiative', Higher Education Bandwidth Advisory Committee. http://www.dest.gov.au/highered/research/pdf/aren.pdf

[4] Hey T., Trefethen A.E., "The UK e-Science Core Programme and the Grid", Future Generation Computer Systems, Vol 18 (8) (2002) pp. 1017-1031

[5] EGEE: Enabling Grids for E-Science in Europe http://egee-intranet.web.cern.ch/egee-intranet/gateway.html

[6] Internet2 Middleware Initiative http://middleware.internet2.edu/

[7] National Science Foundation Middleware Initiative http://www.nsf-middleware.org/

[8] NSF Announces Continuing Steps To Enhance Cyberinfrastructure, Oct 28, 2003 http://www.cise.nsf.gov/news/cybr/cybr2.cfm

[9] Atkins D.E. et al, NSF Blue Ribbon Advisory Panel on Cyberinfrastructure, "Revolutionizing Science and Engineering Through Cyberinfrastructure", January 2003 http://www.communitytechnology.org/nsf_ci_report/

[10] Shipp, 2002, 'Research Information Infrastructure Framework for Australian Higher Education', Report of the DEST Systemic Infrastructure Initiative Information Infrastructure Advisory Committee, Canberra.

[11] Global Grid Forum http://www.ggf.org/

[12] The Globus Alliance http://www.globus.org/

[13] The Open Grid Services Architecture http://www.globus.org/research/papers/ogsa.pdf

[14] W3C The World Wide Web Consortium http://www.w3.org/

[15] OASIS http://www.oasis-open.org/

[16] ARC Research Networks http://www.arc.gov.au/grant_programs/centres_networks/research_networks.htm

[17] The Internet Engineering Task Force (IETF) http://www.ietf.org/

[18] Innovation Access Program http://backingaus.innovation.gov.au/commercial/access.htm

[19] National Research Infrastructure Taskforce Discussion Paper October 2003 http://www.dest.gov.au/highered/ri_taskforce/dis_paper.pdf

[20] Changing the way research is done – The user perspective on e-Infrastructure http://www.cordis.lu/ist/rn/rn-brochure2.htm

[21] Fox G., Walker D., 'e-Science Gap Analysis', Community Grids Lab, June 2003

[22] Miura, K. ‘Presentation data in SC2003’, NAREGI, 2003 http://www.naregi.org/data/SC2003Miura.pdf

[23] Sekiguchi, S. ‘Grid Projects Status Overview in Japan’, 2003 http://www.hlrs.de/news-events/events/2003/metacomputing/slides/sekiguchi.pdf

[24] Watase, Y. ‘HEP Data GRID in Japan’, HEP Data Grid Workshop 2002, http://hep2.knu.ac.kr/HDGWS/ihdgw/talks/watase.ppt

[25] Resch , M, 'European GRID Perspectives', 5th Pacific Rim Applications & Grid Middleware Assembly Workshop (PRAGMA 5), Hsinchu/Taiwan, October 23, 2003.

[26] Gagliardi, F. ‘European Grid Programme Update’, 2002, http://www.hicb.org/hicb/HICB-021013/Fab-October13-EU-presentation.ppt

[27] UK E-Science Programme, http://www.escience-grid.org.uk/

[28] ‘NSF Middleware Initiative (NMI) Program Announcement’, 2003 http://www.nsf.gov/pubs/2003/nsf03513/nsf03513.htm

Appendix A: Overview of Australian eResearch Activities

eResearch Activity	Facility / Project	Nominated Contact	Public Funding Sources
Research Networks	AARNET/AREN	Chris Hancock (AARNET) chris.hancock@aarnet.edu.au	Aust. Unis, CSIRO
	GrangeNet	Paul Davis (GrangeNet) paul.davis@grangenet.net	BITS-ANP
	CENTIE	Terence Percival (CSIRO) Terry.Percival@csiro.au	BITS-ANP, CSIRO, UNSW, UTS, IVEC
Advanced Computing Facilities	APAC	John O'Callaghan (APAC) john.ocallaghan@apac.edu.au	SII
	Qld Parallel Supercomputing Foundation (QPSF)	Bernard Pailthorpe (UQ) bap@uq.edu.au	QLD Govt
	ANUSF	Bob Gingold (ANU)	
	Griffith University QPSF Node	Andrew Lewis (Griffith) a.lewis@griffith.edu.au	ARC-LIEF, QPSF, Griffith
	James Cook University High Performance Computing Unit	Ian Atkinson (JCU) ian.atkinson@jcu.edu.au	ARC-Linkage, QPSF, JCU
	Monash University High Performance Computing Facility	Bill Appelbe (RMIT) bill@cs.rmit.edu.au David Abramson (Monash) davida@csse.monash.edu.au	VPAC, ARC
	CSIRO Higher Performance Scientific Computing (HPSC)	Rhys Francis (CSIRO) rhys.francis@csiro.au	CSIRO
	National Computing Facility for Lattice Gauge Theory	Tony Williams (Adelaide) anthony.williams@adelaide.edu.au	ARC-RIEF
	Ormond Facility and Access University of Melbourne/RMIT	Dirk van der Knijff (Melbourne) dirk@its.unimelb.edu.au	
	High Performance Computing and Research Support Group	Dr Joe Young (QUT) j.young@qut.edu.au	QUT, ARC-*various*
	Swinburne Centre for Astrophysics and Supercomputing	Prof Matthew Bailes (Swinburne) mbailes@swin.edu.au	MNRF, Swinburne
	ACESS – Australian Computational Earth Systems Simulator	Peter Mora (UQ) p.mora@quakes.uq.edu.au	MNRF, UQ
Advanced Networked Instrumentation	Nanostructural Analysis Network Organisation (NANO)	Simon Ringer (USYD) simon@emu.usyd.edu.au	MNRF
	Australian Telescope National Facility	Brian Boyle (CSIRO) brian.boyle@csiro.au	MNRF
	Synchrotron	Garry Seaborne http://www.synchrotron.vic.gov.au/	Victorian Govt $157mill
People Networks/ Research Communities	http://www.arc.gov.au/pdf/Successful_SRI_web121203.pdf		ARC Research Networks

Grid Computing/ Grid Services	Integrating Australia into Global eScience	Kevin Burrage (Uni of Qld) kb@maths.uq.edu.au	Innovation Access Programme – International Science and Technology
	Gridbus Project	Rajkumar Buyya (Macquarie) raj@cs.mu.oz.au	Melbourne, VPAC, ARC
	Australian Nimrod Testbed	David Abramson (DSTC, Monash) davida@csse.monash.edu.au	DSTC
	PAGIS Grid Application Environment	Andrew Wendelborn (Adelaide) andrew.wendelborn@adelaide.edu.au	
Digital Libraries/ Research Repositories	Towards an Australian Partnership for Sustainable Repositories (APSR)	Robin Stanton (ANU) robin.stanton@cs.anu.edu.au	SII, ARIIC
	Australian Research Repositories Online to the World (ARROW)	Cathrine Harboe-Ree (Monash) cathrine.harboe-ree@lib.monash.edu.au	SII, ARIIC
	Australian Digital Theses Program Expansion and Redevelopment	Marian Bate, (UNSW) m.bate@unsw.edu.au	SII, ARIIC
	Meta Access Management System (MAMS)	James Dalziel (Macquarie – MELCOE) james@melcoe.mq.edu.au	SII, ARIIC
	Harmony International Digital Library Project	Jane Hunter (DSTC) jane@dstc.edu.au	NSF Int'l Digital Library, DSTC
	PictureAustralia, MusicAustralia	National Library of Australia eedmonds@nla.gov.au	NLA
Data Mining	Monash Data Mining Centre	Trevor Dix trevor@csse.monash.edu.au	Monash
	ANU Data Mining Group	Markus Hegland markus.hegland@anu.edu.au	
Knowledge Management/ Semantic Grid	DSTC	Jane Hunter (UQ) jane@dstc.edu.au	DSTC
	NICTA	Arun Sharma (UNSW) arun.sharma@nicta.com.au	NICTA
	CSIRO ICT Centre	Dave Abel (CSIRO) dave.abel@csiro.au	CSIRO, QLD Govt
	DSTO	-	DSTO
Visualization, Simulation, Interactive Virtual Environments	Sydney ViSLAB	Masahiro Takatsuka (ViLAB) masa@vislab.usyd.edu.au	USYD, NICTA, APAC
	Sydney University Information Visualisation Research Group	Peter Eades (Sydney University) peter@it.usyd.edu.au	
	AC3 – Australian Centre for Advanced Computing and Communications	Phil McCrea (AC3) philip.mccrea@ac3.com.au	*Seed funding at establishment only*
	Interactive Virtual Environments Centre (IVEC)	Karen Haines (UWA) karen@csse.uwa.edu.au	APAC, WA Govt
	CSIRO Virtual Environments IMVS Group	Duncan Stevenson (CSIRO) duncan.stevenson@cmis.csiro.au	CSIRO
	Advanced Computational Modelling Centre (ACMC)	Kevin Burrage (UQ) kb@maths.uq.edu.au	DEST, ARC, UQ
	Advanced Displays, Scientific Visualization	Bernard Pailthorpe (UQ) bap@uq.edu.au	
Applications	***Computational Chemistry***		

Centre for Computational Molecular Science (CCMS)	Sean Smith (UQ) s.smith@chemistry.uq.edu.au	UQ
Fujitsu Computational Chemistry	Roger Amos (ANU) roger.amos@anusf.anu.edu.au	
Bioinformatics		
Institute for Molecular Bioscience (IMB)	Mark Ragan (UQ) m.ragan@imb.uq.edu.au	
Centre for Biocomplexity Science	Simon Easteal (ANU) simon.easteal@anu.edu.au	ARC
Victorian Bioinformatics Consortium	Ross Coppell (Monash) ross.coppel@med.monash.edu.au	STII, NHMRC, ARC
Nano-materials and Mesoscale physics		
Nano-materials and Mesoscale physics	Simon Ringer (USYD) simon@emu.usyd.edu.au	ARC
	John Drennan (UQ) j.drennan@mailbox.uq.edu.au	DSTC, UQ
Environmental Modelling / Climate Prediction		
Qld Parallel Supercomputing Foundation (QPSF)	Bernard Pailthorpe (UQ) bap@uq.edu.au	Qld Govt
Earth Sciences / Solid Earth and Environment Grid (SEEGrid)		
Australian Government Geoscience Australia	Lesley Wyborn lesley.wyborn@agso.gov.au	DITR
Queensland University Advanced Centre for Earthquake Studies (QUAKES)	Peter Mora (UQ) morap@quakes.uq.edu.au	UQ
CSIRO Solid Mechanics Research Group	Louis Moresi (CSIRO) louis.moresi@csiro.au	CSIRO
Marine and Oceanographic Modelling		
Tasmanian Partnership for Advanced Computing (TPAC)	Nathan Bindoff n.bindoff@utas.edu.au	APAC, UTAS
Computational Engineering		
James Cook University Mechanical Engineering	Jeff Loughran (JCU) Jeffrey.Loughran@jcu.edu.au	JCU, ARC
Victorian Partnership for Advanced Computing (VPAC)		APAC, VIC Govt
Astronomy and Computational Astrophysics		
Australian Virtual Observatory (MACHO)	Peter Lamb (CSIRO) peter.lamb@csiro.au	CSIRO, ARC-LIEF
Australian Consortium for Interferometric Gravitational Astronomy (ACIGA)	David McClelland (ANU) david.mcclelland@anu.edu.au	ARC-LIEF, UWA, ADELU, ANU, CSIRO, Monash
Victorian Partnership for Advanced Computing (VPAC)		APAC, VIC Govt
Particle Physics		
Belle Analysis Data Grid (BADG)	Lyle Winton (UMEL) l.winton@physics.unimelb.edu.au	
Mesh-based computation		
South Australian Partnership for Advanced Computing (SAPAC)	Anthony Williams, (UADEL) anthony.williams@adelaide.edu.au	SA Govt, UADEL
Financial Modelling		

Australian Centre for Advanced Computing and Communications (AC3)	Tony Hall (UTS) tony.hall@uts.edu.au	*Seed funding at establishment only*
E-Learning		
Education Network Australia (EdNA)	Garry Putland garry.putland@edna.edu.au	FOLP
Learning Federation	Stuart Tait stuart.tait@thelearningfederation.edu.au	
education.au	Jon Mason jmason@educationau.edu.au	
Macquarie E-learning Centre of Excellence	James Dalziel james@melcoe.mq.edu.au	SII, ARIIC
E-Health		
HealthConnect	Paul Fitzgerald nhimac.secretariat@health.gov.au	DHA, State Govts
MediConnect	mediconnect@health.gov.au	DHA, HIC
HealthInsite	healthinsite.feedback@health.gov.au	DHA
National Health Information Advisory Council		
Computational Linguistics		
Language Technology Research Group	Steven Bird (UMEL) sb@csse.unimelb.edu.au	
Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)	Linda Barwick (USYD) linda.barwick@paradisec.org.au	ANU, ARC, USYD, UMEL, ANU
Social Sciences – Data Analysis		
Centre for Social Research in Communication (CSRComm)	Cindy Gallois (UQ) c.gallois@uq.edu.au	
Australian Consortium for Social and Political Research Incorporated (ACSPRI)	Deborah Mitchell (ANU) deborah.mitchell@anu.edu.au	
Australian Social Science Data Archive (ASSDA)	Sophie Holloway sophie.holloway@anu.edu.au	ANU, ACSPRI
Centre for Research into Sustainable Urban and Regional Futures (CR-SURF)	Robert Stimson r.stimson@uq.edu.au	UQ, ARC

Key to Funding Sources:

APAC Australian Partnership for Advanced Computing
ARC Australian Research Council (Australian Government)
ARIIC Australian Research Information Infrastructure Committee (Australian Government)
BITS Building on IT Strengths Initiative (DoCITA)
CSIRO Commonwealth Scientific and Industrial Research Organisation (Australian Government)
DSTC Distributed Systems Technology Centre
DEST Department of Education, Science and Technology (Australian Government)
DHA Department of Health and Ageing (Australian Government)
DITR Department of Industry, Tourism and Resources (Australian Government)
DoCITA Department of Communication, Information Technology and the Arts (Australian Government)
FOLP Framework for Open Learning Program (DEST)
HIC Health Insurance Commission (HIC)
IVEC Interactive Virtual Environments Centre
LIEF Linkage Infrastructure Equipment and Facilities (ARC)
MNRF Major National Research Facilities (DEST)
NHMRC National Health and Medical Research Council (DHA)
NICTA National ICT Australia (DoCITA)
NLA National Library of Australia (Australian Government)
NSF National Science Foundation (United States of America)
RIEF Research Infrastructure Equipment and Facilities Scheme (ARC)
SII Systemic Infrastructure Initiative (DEST)
STII Victorian Science, Technology and Innovation Initiative (Victorian Government)
VPAC Victorian Partnership for Advanced Computing (Victorian Government)

Appendix B: Structure of International eScience Initiatives

B.1 Structure of the UK eScience Middleware Program

The UK eScience initiative has three main components:

1. An Open Middleware Infrastructure Institute (OMII); and,
2. A Grid Support Centre (GSC).
3. A number of Regional Centres

The OMII and GSC work in close partnership to facilitate the acquisition, coordination, development and integration of the middleware required by e-Research communities and to support pilot projects and communities in the Regional Centres which use the middleware and Grid environments.

The responsibilities of the OMII include:

- To work with government, research, industry and academic organisations to define, develop and coordinate the middleware infrastructure across existing research facilities and programs.
- To fill gaps in existing middleware by either 'productising' existing research codes or developing new components. The development of new components could be done either in-house or subcontracted to an appropriate specialist group. The OMII would be responsible for the quality of the services developed.
- To deliver the associated outreach and education program. As such, the OMII organizes outreach events (workshops, seminars, conferences and training), develop and disseminate supporting documentation, best practise guides and training materials.
- To work closely with specific research communities, (e.g., particle physics, astronomy, engineering, biomedical, environmental, social sciences/humanities), on the development of pilots or specific testbed projects which serve to evaluate the middleware technologies and to assist with their customisation, adoption and deployment. Incentives, in the form of software engineers seconded from the Open Middleware Infrastructure Institute (see Section 5), encourage communities to adopt the technologies recommended or developed by the Open Middleware Program.
- Be responsible for coordinating and facilitating international collaboration activities such as involvement in international standards bodies, attendance at international meetings and workshops, collaborative international projects, fellowships and exchanges and establishing links with the equivalent international organizations (e.g., NSF Middleware Initiative, GRIDS Center, UK Open Middleware Infrastructure Institute).

The GSC has the complementary role of managing the operation, resources and infrastructure of the national research Grid. The GSC's responsibilities include:

- Issuing of Certificate Authorities and defining and implementing security policies;
- Network Monitoring;

- Providing user services and tools;
- Running, maintaining, supporting production level infrastructure (e.g., Access Grid Support);
- Testing, implementing and integrating new middleware tools (provide a suitable transition environment);
- Operating a helpdesk and providing support and maintenance to user communities.

A number of independent Regional Centres provide regional support and outreach and assist with the development of pilot projects for specific research communities and industries, using the middleware developed by the OMII. Each of these centres is co-located with existing key e-Science facilities and research communities and they provide a network for diffusion of knowledge and technologies from both the OMII and GSC.

B.2 USA eResearch Structure

A number of significant eScience middleware initiatives are being funded in the US. Below we describe the three main initiatives:

1. The NSF Middleware Initiative
2. Internet2 Middleware Initiative
3. The DOE Science Grid

In addition there are a number of other significant relevant US projects which we do not have the space to cover here including:

- NASA's Information Power Grid,
- the GrADS (Grid Application Development Software) project and
- the NIH (National Institute's of Health) BIRN (Biomedical Informatics Research Network)

B.2.1 The NSF Middleware Initiative (NMI)

NSF established the NMI in 2001 to define, develop and support an integrated national middleware infrastructure. NMI activities facilitate sharing of scientific resources such as telescopes, supercomputing systems and linear accelerators, as well as common resources such as databases, directories and calendars. One important emphasis is to explore ways in which grid computing can be integrated with enterprise computing on university campuses. In particular, the NMI funds:

- The Grid Research Integration Deployment and Support (GRIDS) Center; and,
- The Enterprise and Desktop Integration Technologies (EDIT) consortium.

The GRIDS Center and EDIT Consortium, working closely with the NSF Partnerships for Advanced Computational Infrastructure and private industry, are defining, developing, deploying and supporting an integrated national middleware infrastructure for science and engineering applications. To date they have issued four releases of production-quality open-source and open-standards middleware tools (NMI-R4). Freely available, the components developed at universities and national laboratories, are designed to support or fill functions needed by the research and education community in such areas as user authentication and authorization, resource identification and allocation, job management, and scheduling. NMI tools and procedures also ensure that the middleware is deployable and supportable on a wide variety of systems, including production environments at campus and laboratories. The NMI teams also provide support and training to ensure the success of early adopters and new user communities.

NSF has also funded the NMI Integration Testbed, consisting of eight universities that coordinate closely to deploy and evaluate NMI middleware in production environments. The testbed sites use and evaluate software, services and architectures that facilitate access to distributed electronic resources for faculty and campus projects. Their efforts gauge the middleware's practicality, emphasizing factors such as performance, ease of use, robustness and technical support.

- NMI Integration Testbed: http://www.nsf-middleware.org/testbed/

B2.2 Internet2 Middleware Initiative (I2-MI)

I2-MI is one activity of the Internet2 consortium. - 206 universities working in partnership with industry and government to develop and deploy advanced network applications and technologies. Through Internet2 Working Groups, Internet2 members are collaborating on the following technical areas: Applications, Backbone Network Infrastructure, End-to-End, Engineering, and Middleware. Each area has an assigned Director who is responsible for the activities in their area. Members who do the work of the groups come from Internet2 member institutions and partners. Support for the Working Groups is provided by Internet2 and Internet2 staff.

I2-MI is collaborating with the NMI but is also working on many middleware developments and integration activities of their own, including:

- MACE Middleware Architecture Committee for Education
- FOO Federating Organizations Organization (FOO)
- Shibboleth - web access control software
- OpenSAML
- MW-E2ED Middleware End-To-End Diagnostics Advisory Group
- Integrated Infrastructure for Instant Messaging (I2IM)
- Video Middleware (VidMid)
    - Videoconferencing (VidMid-VC)
    - Video on Demand (VidMid-VoD)
- Multicampus Middleware
- Internet2/MACE Uniform Resource Name (URN) Registry
- Medical Middleware (MedMid)

B2.3 DOE (Department of Energy) Science Grid

The principle goal of the DOE Science Grid is to provide an integrated, advanced Grid cyberinfrastructure which uses persistent, scalable, standards-based Grid services to support DOE's large scale collaborative science projects. Major accomplishments to date include:

- Construction of a Grid across five major DOE facilities (LBNL, ANL, ORNL, NERSC and PNNL) with an initial complement of computing and data resources
- Integration of DOE's Office of Science supercomputing center's (NERSC) production large-scale storage systems into the Grid;
- Design and deployment of a Grid security infrastructure that is facilitating collaboration between US and European High Energy Physics Projects and within the US Magnetic Fusion community. This work is being used by the DOE's Particle Physics Data Grid (PPDG), Earth Systems Grid and Fusion Grid projects.
- A resource monitoring and debugging infrastructure that facilitates managing this widely distributed system and building high performance distributed science applications
- Development and deployment partnerships with several key vendors;
- Use of Grid infrastructure by applications from several disciplines – computational chemistry, ground water transport, climate modelling, bioinformatics, etc.

Details of middleware being developed by the DOE Science Grid project can be found at the project web site: http://doesciencegrid.org/

B.3 European Approaches to Structure

The EGEE project (Enabling Grids for E-Science in Europe) aims to integrate current national, regional and thematic Grid efforts, in order to create a seamless European Grid infrastructure for the support of European eScience research. This infrastructure will be built on the European Research Network (GÉANT) and exploit Grid expertise that has been generated by projects such as the European DataGrid project and the national Grid initiatives such as UK e-Science, INFN Grid, Nordugrid and others. Until very recently there was no single project or organization that was responsible for coordinating the delivery of E-Science infrastructure across the whole of Europe. The EGEE is intended to coordinate existing and new national, regional and thematic programs, rather than replace them.

The EGEE is a two-year project that was conceived as part of a four-year program. Two pilot application areas, High-Energy Physics and Biomedical Grids, have been selected to guide the implementation and certify the performance and functionality of the evolving European Grid. The EGEE project has a three-fold mission:

1. To deliver production level Grid services, the essential elements of which are manageability, robustness, resilience to failure, and a consistent security model. In addition, scalability is essential to rapidly absorb new resources as these become available and to ensure the long-term viability of the infrastructure.
2. To carry out a professional Grid middleware re-engineering activity in support of the production services. This will support and continuously upgrade a suite of software tools capable of providing production level Grid services to a base of users which is anticipated to rapidly grow and diversify.

3. To ensure an outreach and training effort that can proactively market Grid services to new research communities in academia and industry, and determine new e-Science applications**.**

For efficient management within EGEE, the European Grid community has been structured into ten partner regions or 'federations': CERN, Central Europe, France, Germany/Switzerland, Ireland/UK, Italy, Northern Europe, Russia, South-East Europe, and South-West Europe. This structuring has led to several of these partners integrating their regional Grid efforts in order to provide coordinated resources to the overall EGEE project. In addition, the United States are participating as an unfunded partner in the project and are considering establishing a US-Europe federation. Similar participation from Asia-Pacific (including Japan and Australia) is seen by the European Union as being highly desirable.

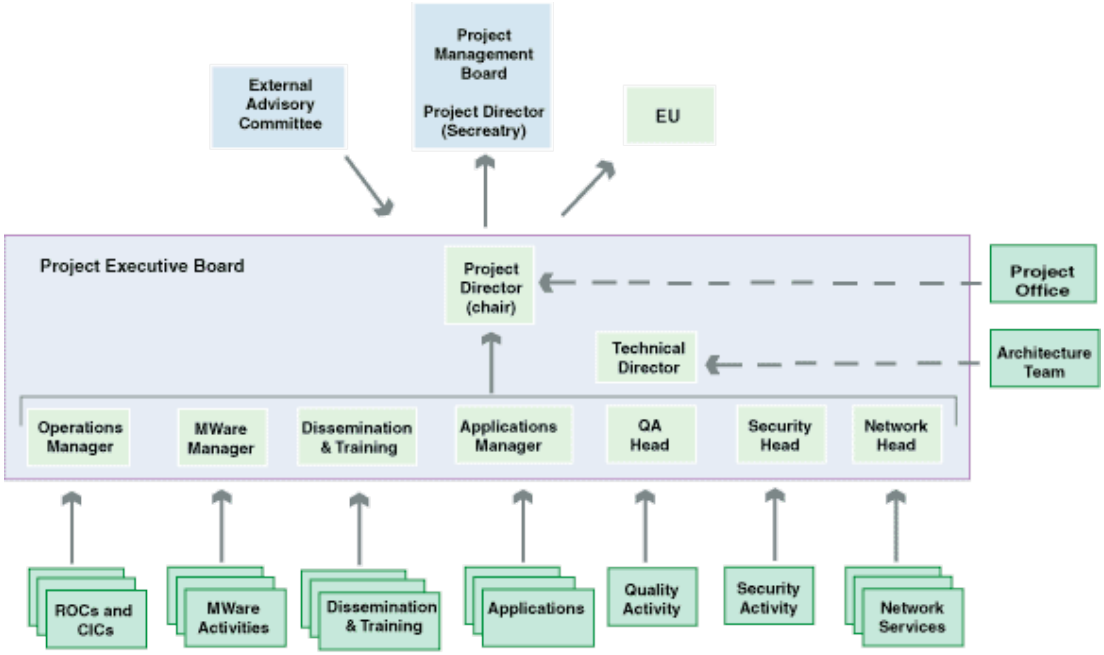


Figure 1: Management structure of the EGEE project[1]

Management of the EGEE comprises a core management unit, staffed by representatives from CERN, and a number of activity-specific subunits, staffed by representatives from their partner institutions in the EGEE project. Each of these subunits reports through to the project executive board through a designated manager within the core management unit (see Figure 1).

Each subunit is responsible for distinct activities, including Dissemination and Outreach; User Training and Induction; Application Identification and Support; International Cooperation; European Grid Support, Operation and Management; Network Resource Provision; Middleware Re-engineering and Integration; Quality Assurance; Security; and Network Services Development. Typically each federation is represented in each sub-unit activity through one or more participants from their geographical area.

[1] http://egee-intranet.web.cern.ch/egee-intranet/Project-Structure/project-structure.html

B.4 Japanese eScience Structure

The Table[1] below provides an overview of the major Grid-related and eScience projects underway within Japan. In the next two sub-sections we describe Japan's two major eScience projects (NAREGI and ITBL) in more detail.

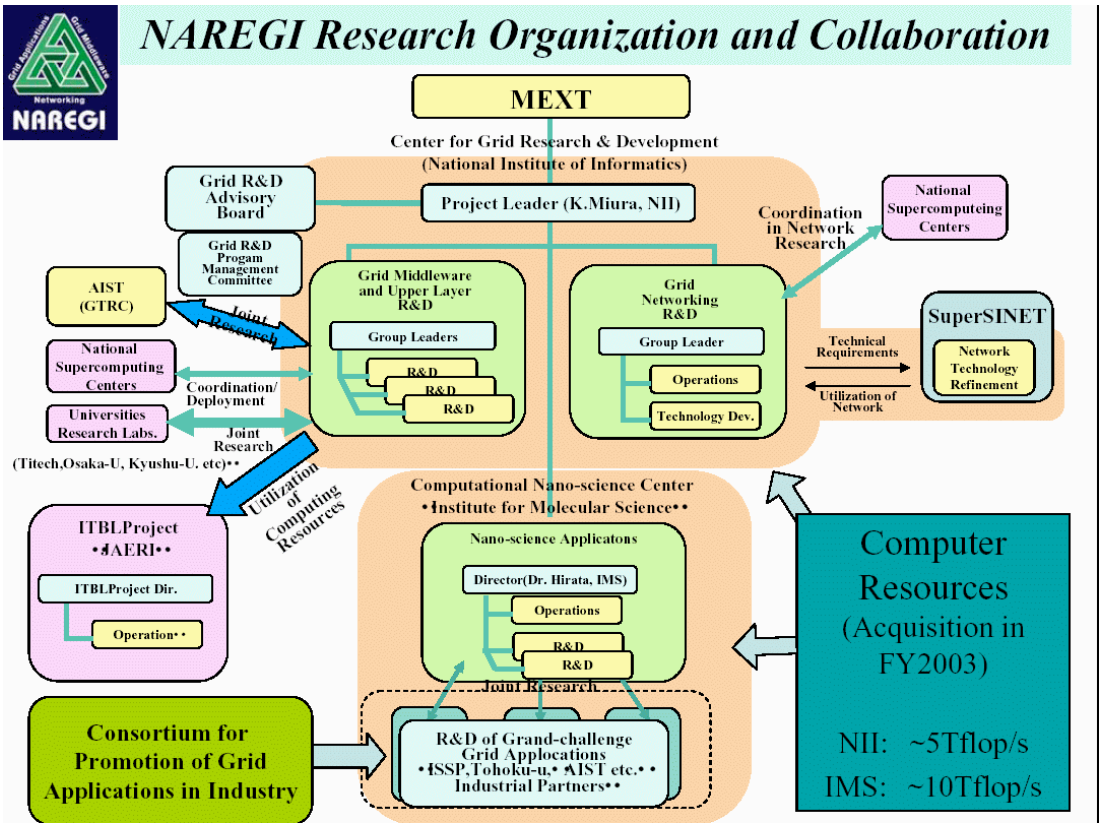
Project	Participants	Funding Agency, Period, Budget	Focus
ITBL http://www.itbl.jp/	JAERI, RIKEN, NAL, JST, NIED, NIMS	MEXT 2001-2005 AU$245M	Application for supercomputer ensemble
NAREGI http://www.naregi.org/	University National Labs	MEXT 2002-2007 US$360M	National Research Grid Infrastructure
Grid Technology Research Center (GTRC) http://www.gtrc.aist.go.jp/en	AIST	METI 2003 – 2007 US$140M	Grid middleware, applications, coordination
Super SINET http://www.nii.ac.jp/	Universities	MEXT 2001-2004	10 G backbone
Ninf http://ninf.apgrid.org/	AIST TITECH Uni. Tsukuba Kyoto Uni	JST, METI, IPA 1994- US$6M	Develop Grid middleware and applications (Ninf-G/Grid RPC)
ApGrid http://www.apgrid.org/	Grid Partnership in Asia Pacific Region	MEXT 2002-2004 US$0.9M	Resource sharing, grid development, mutual support for applications development
Grid Data Farm http://datafarm.apgrid.org/	AIST, TITECH, KEK, Uni Tokyo, NEC	METI 2002-2002 US$2.3M	Construction of a peta-scale parallel filesystem exploiting local storage of PCs over the world wide Grid
Grid Drug Design	AIST, Uni Tsukuba, Tokushima Uni	JST 2001-2003 US$1.2M	Drug design system on grids
AIST Grid	AIST Osaka Uni	AIST 2002-2004 US$4M	Access Grids, medical applications, telescience
BioGrid http://www.biogrid.jp/	Osaka Uni Pharmacy Industry NEC, AIST	MEXT 2002-2006	Bio-application deployment
Quantum Chemistry Grid	AIST	AIST $US1M	Provide user friendly portals
Grid Consortium Japan	AIST Business Sector 36 companies, 150 members	US$0.1M	Venues for info exchange, GGF franchise
Business Grid	IT vendors AIST	METI 2003-2005 2.8 B JPY	Business Applications

B.4.1 The National Research Grid Initiative (NAREGI)

NAREGI is a major Grid R&D project launched by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) in 2003 that brings together a number of smaller grid initiatives as well as major cluster and grid-related software development projects in Japan. Funded for five years from 2003-2007, its budget in 2003 was approx. $US17mill. Different from the Information Technology based Laboratories (ITBL) project which was conceived in 2001 around research organizations attached to the former Science and Technology Agency (STA), NAREGI is mainly focused on university computing centres and former inter-university research centres, such as the National Institute of Informatics (NII) and the Institute for Molecular Science

[1] http://www.hlrs.de/news-events/events/2003/metacomputing/slides/sekiguchi.pdf

(IMS). NAREGI is also linked to the SuperSINET research network sponsored by MEXT and managed by NII. NAREGI was perceived almost entirely as a software research effort. While a large grid test bed will be built as part of the project, NAREGI is not a production grid project. In terms of applications, nanoscience and nanoscale simulations in materials science, chemistry and life sciences have been selected as the major application focuses of the project. The overall focus of NAREGI is on computational grid, rather than data or knowledge grids, as well as developing grid infrastructures for large-scale simulations. The figure below illustrates the collaborative structure of the NAREGI project[1].



B.4.2 Information Technology Based Laboratories (ITBL)

Started in April 2001, the objective of the ITBL project is to realize a virtual joint research environment which connects 100 or more supercomputers located in Japan over SuperSINET to build a system enabling joint use over the Internet of large-scale experimental facilities. ITBL will also develop and provide software utilizing databases and supercomputers, promoting a variety of types of R&D and improving usage efficiency of various facilities, including supercomputers. A total of six institutions have been members since the first year: National Institute for Materials Science (NIMS); National Research Institute for Earth Science and Disaster Prevention (NIED); National Aerospace Laboratory of Japan (NAL); Institute of Physical and Chemical Research (RIKEN); Japan Atomic Energy Research Institute (JAERI); and Japan Science and Technology Corporation (JST). Among these institutions, JAERI and RIKEN are handling the development of the common basic technology required for implementing the ITBL plan. In addition, each of the six research institutions is handling different aspects of the development of application software that can be used through ITBL.

[1] http://www.naregi.org/data/SC2003Miura.pdf