



By Alan Blatecky, Ann West, and Mary Spada

Middleware

The New Frontier

What is *middleware* anyway? A standing joke is that if you ask five people, you will get six definitions. Unfortunately, this is quite true. One of the few things that people can agree on is that middleware is the “stuff between” the application and the network. That is, the application people say that it is something they need but that it is outside their area of expertise and effort; they thus look to someone else to take responsibility. Likewise, the network people say that middleware is something they need but that it is outside their area of expertise and effort; they also look to someone else to take responsibility. But the “stuff between” definition doesn’t really help other than to show that middleware covers a very wide expanse of software, infrastructure, and even policy.

Alan Blatecky is Program Director at NSF with responsibilities for the NSF Middleware Initiative and advanced networking. Ann West manages the NMI-EDIT outreach and education activities for the NSF Middleware Initiative and works for EDUCAUSE and Internet2. Mary Spada is Program Manager, Strategic Initiatives, at Argonne National Laboratory MCS Division and the Computation Institute of the University of Chicago; she helps lead the outreach and participation efforts for NMI through the GRIDS Center.

Another way to approach the definition of middleware is to look at what is popularly being called *cyberinfrastructure*. Traditional infrastructure includes roads, sewer and water systems, buildings, bridges, and power plants. Infrastructure for the electronic world, or cyberinfrastructure, includes Grid technologies, computing resources, various types of networks, data repositories, storage, and so forth. In this environment, middleware is often being called the “glue” that makes these elements of the cyberinfrastructure work together. As computing and networking capabilities continue to grow, the need for this glue becomes ever more important. And, since cyberinfrastructure is a largely unexplored and undeveloped territory, middleware quite literally becomes the “new frontier” of the electronic age.

The Function of Middleware

Perhaps the best way to define middleware is to look at how it functions in two different examples: (1) leading e-science activities, and (2) production-level campus support and services.

E-science is large-scale science that studies complex micro- to macro-scale problems across time and space and conducts research using an array of cyberinfrastructure capabilities including Grid technologies, computational power, large data repositories, high-speed networks, remote instruments, and extensive ad hoc and distributed sensor networks and arrays. These capabilities use entirely new tools that allow collaboration and the sharing of resources to advance science and research. In e-science, the cyberinfrastructure itself becomes a third scientific method. The first scientific method is to form theories. The second is to perform laboratory experiments. And now the third is to utilize the cyberinfrastructure to conduct simulations and modeling through the deployment of Grids, managed sets of

complex distributed instruments and resources. In this scenario, middleware is concerned with managing these resources so that they work in an integrated fashion and the scientist does not have to make separate arrangements to locate or use remote resources, share data, or ensure levels of security. Middleware takes care of the underlying details and provides the glue, allowing the scientist to focus on the research and science rather than on the supporting technologies.

In production-level campus support and services, middleware plays a similar role: managing and organizing enterprise resources and networks. In this environment, middleware is concerned with such things as common directory services (or interoperating directories), ways to provide authentication and authorization of users, methods to manage resources in terms of financial accountability and control, conventions for naming, and so forth. Middleware also manages the external connections to the Internet and the world so that a campus can become part of dynamic virtual organizations and multi-site activities and collaborations. As in e-

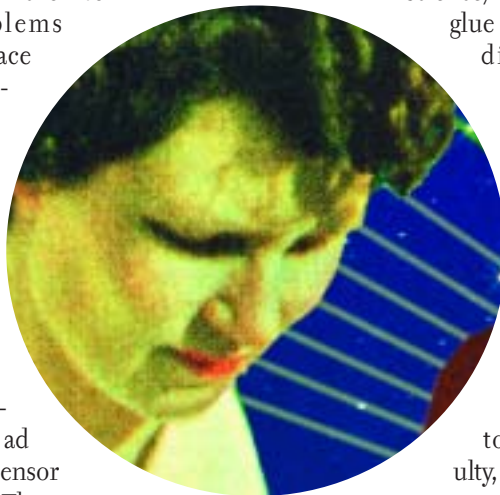
science, middleware is the glue that manages the diverse components of computers, networks, and other electronic resources for a campus so that these resources and capabilities are easily available to students, faculty, and staff.

Middleware as a Discipline

The primary focus of the technology research community for the last two decades has been on developing faster networks and faster and larger computing capabilities. Network speeds have thus increased by orders of magnitude and have become more extensive and ubiquitous; likewise, computing capabilities and performance have continued to follow Moore's law and look to stay on that curve for several more years. Costs per unit in networking, computing, and storage have continued to decrease rapidly. As a result of this explosion in technological magnitude and capabilities, education, research, and science are all facing new opportunities today.

In response, the scientific community has pushed the edge of scientific research further in many disciplines and across geographic boundaries as never before. The deployment of Grid technologies and projects continues to grow; currently, there are more than seventy-five major Grid projects being developed across the world. These Grid projects involve collaboration among many institutions in different countries and across many disciplines, from physics and astronomy to engineering and bioinformatics.

At the same time, however, there has been very little focus on middleware development or research; the focus has been on networking and on computing hardware. As a result, unless middleware issues and problems are seriously addressed, not only will technology advances be limited but the emerging e-science applications will be jeopardized. Members of the science and research community will not be able to effectively utilize the high-performance networks and computing resources to do their work. For example, modern genome



Middleware is often called the “glue” that makes the elements of the cyberinfrastructure work together.

research is impossible without access to remote databases, and astronomical research depends on a series of interconnected telescope sites around the world. Routine work with colleagues at remote or geographically dispersed locations can be accomplished only by a robust, reliable suite of middleware tools and cyber-infrastructure technologies.

Currently, none of these issues are being addressed in any systematic way. In some cases, competing solutions that do not interoperate are being pursued, which is both costly in terms of time and antithetical to building systems that bridge multiple domains and capabilities. In other cases, piecemeal solutions are being developed to meet immediate needs; again, the result is a significant duplication of effort at best and limited success in terms of performance or scalability.

The NSF Middleware Initiative

These challenges and problems were the primary reasons behind the formation of the National Science Foundation (NSF) Middleware Initiative. The NSF Middleware Initiative (NMI) is focused on establishing a critical mass of expertise, software, technology, and approaches to adequately address and solve the large, multi-domain, multi-discipline systems and architectures and the ubiquitous middleware issues facing education and research over the next decade.

NMI grew out of a series of workshops and white papers identifying middleware as a gap that needed to be addressed and supported by NSF. In response, NSF issued the first middleware program announcement, with proposals due on May 10, 2001. The program was established at \$10 million per year for at least three years. The stated goal is to assemble and identify existing pieces of middleware and to identify gaps in middleware knowledge and software. The desired outcome is to create a growing set of software releases, using open-source and open-standards approaches, which will lead to and provide production-level middleware.

As a result of the program announcement, the NSF executed three cooperative agreements to form NMI. The cooperative agreements established a team of Grid and campus enterprise experts who serve as

NMI is focused on establishing a critical mass of expertise to address the ubiquitous middleware issues facing education and research.



the middleware System Integrator and Service Provider for NMI. The two NMI team members are the GRIDS (Grids Research Integration Deployment and Support) Center (the Information Sciences Institute at the University of Southern California, the University of Chicago, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the University of California at San Diego, and the University of Wisconsin at Madison) and EDIT (Enterprise and Desktop Integration Technologies) Consortium (Internet2, EDUCAUSE, and the Southeastern Universities Research Association). The members of the NMI team started working together in September 2001. Proposals for the second

year of the program were due on March 1, 2002; the emphasis for the second year is to encourage new participants and approaches in order to leverage and broaden NMI efforts and activities.

The team outlined the purpose of the NMI program: to design, develop, deploy, and support a set of reusable, expandable middleware functions and services that will benefit many applications in a networked environment and that will

- facilitate scientific productivity,
- increase research collaboration through shared data, computing, code, facilities, and applications,
- support the education enterprise,
- encourage the participation of industry partners, government labs, and agencies for more extensive development and wider adoption and deployment,
- establish a level of persistence and availability so that other applications developers and disciplines can take advantage of the middleware,
- encourage and support the development of standards and open-source approaches, and
- enable scaling and sustainability to support the larger research and education communities.

Since middleware covers a very wide variety of software, capabilities, and architectures, the NMI team also identified three middleware capabilities that could be used to guide the activities and development of the NMI program:

1. To allow scientists and engineers to transparently use and share distributed resources such as computers, data, networks, and remote instruments
2. To develop effective collaboration and communication tools such as

Grid technologies, desktop video, and other advanced services to expedite research and education

3. To develop a working architecture and approach that can be extended to the larger set of Internet users and network users

In brief, then, middleware makes the process transparent to the end users and must provide consistency, security, and privacy capabilities in order to be useful and successful.

NMI Releases

The first results of NMI's work were released on May 7, 2002: six software packages, three object classes, numerous "best practices" and policy documents, and several white papers offering current views and a perspective on what will be addressed next by the NMI team.

To ensure a robust baseline of usability, all software packages must have been in production use before being incorporated and integrated into the NMI release. Six software packages have been included in the first release:

- *Globus Toolkit*: an open-source software toolkit enabling Grid computing, coordinated resource sharing, and problem solving in multi-institutional virtual organizations
- *Condor-G*: software that works in conjunction with the Globus Toolkit to provide high-throughput computing on large collections of distributed workstations
- *Network Weather Service*: a distributed system that periodically monitors and dynamically forecasts network performance
- *KX.509 and KCA*: software that provides a bridge between security infrastructures using Kerberos and PKI
- *CPM*: Certificate Profile Maker, which makes certificate profiles in XML formats
- *Pubcookie*: software that authenticates Web-based services across multiple Web servers

The software packages are thoroughly tested and debugged by NMI team members so that they can be easily deployed by various campuses and institutions.

The NMI software will also be used and further developed by several major Grid projects, such as NEES (Network for Earthquake Engineering Simulation), GriPhyN (Grid Physics Network), and iVDGL (international Virtual Data Grid Laboratory). In addition, the NMI testbeds will deploy the packages both to address campus and enterprise-wide issues and to help with integration and next-generation features. NMI provides a limited amount of help-desk and other support for the software.

NMI also released three object classes that help facilitate directory-enabled sharing and exchanging of information to support authentication and authorization between campuses and institutions:



- *eduPerson 1.5*: an enterprise directory object class for interrealm authentication and authorization
- *eduOrg 1.0*: a class of institutional attributes including account management policies, security, etc.
- *commObject*: a generic superclass to support videoconferencing or IP telephony

These object classes are fundamental building blocks for middleware deployment and services. It is difficult, if not impossible, to share data and resources across campus or national boundaries without basic authentication and authorization procedures and capabilities in place.

Also released in May 2002 were several "best practices" and policy documents. Since the whole middleware field is very young, there are few standards, and even if standards are established, they will change quickly because of the rapid evolution in the middleware software and approaches. Thus these documents can provide insight into what will likely happen in the future. The best practices are useful for those campuses and enterprises that want to get involved as early adopters or to provide leadership. The following documents are included in the first NMI release:

- *Practices in Directory Groups 1.0*: experiments and early experiences with authorization applications
- *LDAP Recipe 2.0*: common directory development within higher education
- *Metadirectory Practices for the Enterprise Directory in Higher Education 1.0*: alternative approaches to metadirectory functions such as identity management and information flow among connected systems

Since the whole middleware field is very young, there are few standards, and even if standards are established, they will change quickly.

- *Shibboleth Architecture v.4*: an architecture for the secure exchange of authorization information that can be used to decide who can access a protected Web resource
- *Campus Certificate Policy for Use at the Higher Education Bridge Certification Authority (HEBCA)*: policy requirements for campuses to connect to HEBCA
- *Lightweight Campus Certificate Policy and Practice Statement*: Entry-level, lightweight certificate policy to use when deploying “PKI-lite” for existing applications
- *Sample Campus Account Management Policy*: example of a campus account policy

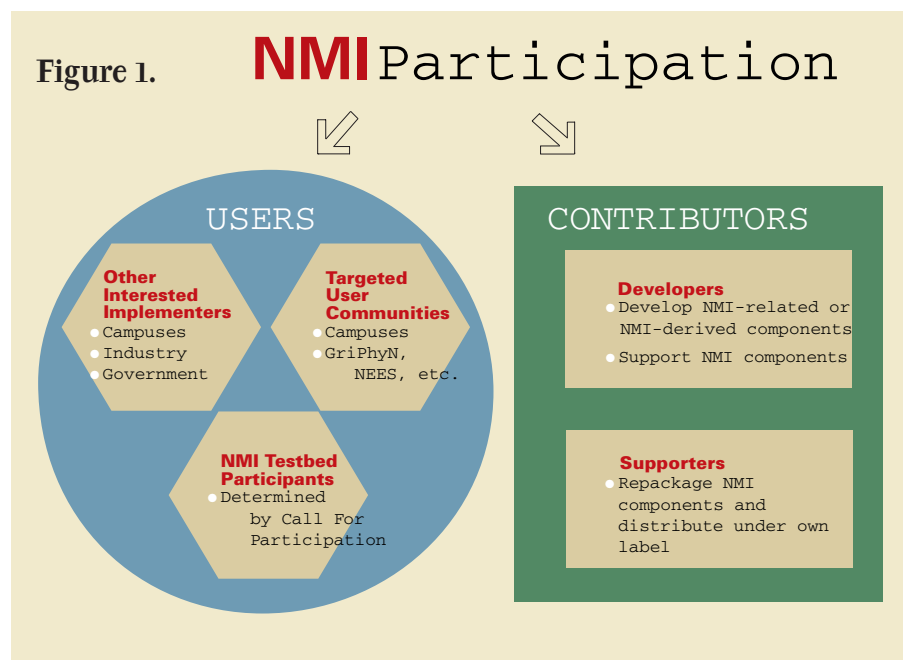
These practices and policies have been used or are being deployed by several campuses and institutions. Each has also been the subject of considerable multi-campus reviews and working groups.

Finally, the NMI release also includes several works in progress, representing a summary of the current thinking and a preview of future NMI work. These white papers are intended to spur discussion on the infrastructure components underpinning collaborative applications. Topics covered include the architectural issues in offering secure and authenticated videoconferencing and the role of directories in video-on-demand.

Additional releases are scheduled for at least twice a year. The next set will contain the latest versions of software and upgrades. New types of software will also be released as they are tested and deployed, along with new objects classes, best practices and policy documents, and white papers.

NMI Participation

The success of NMI and middleware depends on a variety of factors. The effort required for the project is so extensive that substantial participation, input, and support is needed from the research and education community, the private sector (including industry), and other government agencies, as well as collaboration and involvement with international participants. Since middleware is, by definition, the “space between” the network and the application, it is very broad. As soon as two people want to collaborate at



different locations, middleware becomes essential. Thus, NMI will be successful only if it is able to encourage, facilitate, coordinate, and enable a variety of participants to participate in a meaningful way in NMI. The outreach program, a major component of NMI, has devised an initial model for participants (see figure 1).

The NMI participation model consists of two primary classes of participants: users and contributors. Users consist of formal testbed participants, targeted user communities (campuses and scientific communities using Grid technologies), and other interested implementers from user groups that are interested in deploying middleware technologies. The contributors consist of developers and supporters and are self-explanatory. In each instance—user or contributor—the participant can

come from the public or private sector.

It is important to note that NMI is not a standards body or organization for middleware. Although standards will certainly become essential as middleware begins to enter the mainstream of the market, at this point in the lifecycle of middleware, the focus is on consensus and open-source development. In some cases, “consensus” may simply be the tacit agreement of a major Grid project to adopt various pieces of software and conventions to enable collaboration and interoperability. In other cases, “consensus” may be the willingness of several campuses to deploy similar approaches so that data and resources can be readily shared. In all cases, an emphasis on open-source development allows and encourages everyone to participate and collaborate as they like.

NMI will be successful only if it is able to encourage, facilitate, coordinate, and enable a variety of participants to participate in a meaningful way.

Campus Impacts

A subtle change is occurring because of middleware, a change that will significantly affect campus operations and policies in the near future. Although the rapid and widespread deployment of the Internet and use of the Web quickly became standard on college and university campuses over the last several years, this next transition will be less noticeable. It will actually require more radical changes in campus-wide policies and structures than occurred because of the Web. Part of this is due to the rapidly growing capabilities of the cyberinfrastructure, part is due to the growing use of advanced information technology for research and science, and part is due to the need for increased security and management of resources. Several scenarios illustrate the change.

- *Scenario one:* A research scientist on a campus uses remote facilities located at another site (e.g., a telescope in Peru, the Large Hadron Collider at CERN in

Switzerland, the supercomputer at San Diego) to conduct research and/or to explore new areas of science. Access to and use of the remote facilities form a critical component of the research itself.

- *Scenario two:* A student is doing a project that involves getting data from a set of distributed weather probes to address the impact of regional weather changes on crop production. Access to these networked resources is essential.
- *Scenario three:* Two or more universities decide to pool their resources and collaborate on a major groundwater pollution problem emerging in their region. This collaboration will require significant sharing of local, regional, and national resources.

Many physicists, biologists, bioinformatic scientists, earthquake engineers, material scientists, and geologists, among others, are developing Grid projects and plan to use these technologies as a way to conduct their scientific work as part of a

production environment. As these capabilities become more and more “routine” and easy to use, the number and type of users will quickly increase. The impact on campus operations will be profound. What is being done on an ad-hoc basis today will need to be done on a policy-based operational basis tomorrow.

The use of remote shared resources presents several new issues and requirements for each college and university. First, the simple act of sharing expensive resources introduces the need for accountability. Accountability will extend beyond the geographic or administrative boundaries of a department or campus and will be reciprocal: campus A will require certain types of accountability from campus B to meet its fiscal and/or legal obligations, and vice versa.

Accountability in turn requires authentication, authorization, and security. Authentication (some sort of verification about the person—e.g., enrolled student, faculty member) and authorization (permission or authority to use the resource requested) depend on established cyberinfrastructure. This includes capabilities such as accessible common directories that interoperate with remote applications and institutions. The directories must contain the information required by the remote resources and must be maintained with current, up-to-date information. In addition, campus procedures and policies must be secure and must maintain the level of privacy required by the institution. These administrative guidelines must be informed by the conditions-of-use of remote facilities and resources. Since sharing is a hallmark of e-science activities, campus policies should consider these external requirements and should iterate their local policies as appropriate if they want to participate in Grids or collaboration across administrative domains. The use of resources located in foreign countries raises even more security issues.

Because of the driving requirement to use remote resources, leading scientists and researchers are developing “work around” solutions to the problems of authentication, authorization, and security. Although this is understandable and permissible when the number of participants is small, it is not a solution that can

What is being done on an ad-hoc basis today will need to be done on a policy-based operational basis tomorrow.

scale. Worse, if these basic middleware issues are not addressed at the outset by campuses, it is quite likely that multiple solutions will be deployed and that these solutions will not interoperate. This will be costly both in terms of the financial impact at the campus level and in terms of the lost opportunities for science, research, and education.

Summary

The tremendous advances of information technology have revealed a “new frontier”: middleware. Using middleware as the glue to hold together the various ele-

ments of the cyberinfrastructure, including Grid technologies and collaborative tools, promises to significantly change how we do research and education. Scientists and educators will be able to share resources such as distributed computers, large data repositories, and remote instruments without regard to geographic location. They will be able to collaborate through virtual organizations for joint scientific research with colleagues around the world. And they will be able to use these technologies and capabilities to solve problems that cannot be addressed in any other way. *e*

