

## Cover Page

---

U.S. Department of Energy Office of Science  
Scientific Discovery through Advanced Computation Solicitation LAB 01-06  
National Collaboratories and High Performance Networks

# DOE Science Grid: Enabling and Deploying the SciDAC Collaboratory Software Environment

A Collaboratory Pilot  
For the period June 1, 2001 – September 31, 2004

### Principal Investigator

**William E. Johnston**  
Senior Scientist  
Lawrence Berkeley National Laboratory  
One Cyclotron Road, MS: 50B-2239  
Berkeley, CA 94720  
(510) 486-5014 (Voice)  
(603) 719-1356 (Fax)  
WEJohnston@lbl.gov

### Official Signing for LBNL

**Horst D. Simon, Director**  
NERSC Division Director  
Lawrence Berkeley National Laboratory  
(510) 486-7377 (Voice)  
(510) 486-4300 (Fax)  
HDSimon@lbl.gov

---

PI Signature and Date

---

Official Signature and Date

### Co-Investigators

---

**Ray Bair**  
**Ian Foster**  
**Al Geist**  
**William Kramer**

Pacific Northwest National Laboratory  
Argonne National Laboratory  
Oak Ridge National Laboratory  
LBNL / NERSC

RayBair@pnl.gov  
foster@mcs.anl.gov  
gst@ornl.gov  
WTKramer@lbl.gov

# Table of Contents

---

Cover Page .....	i
Table of Contents .....	ii
Abstract .....	iii
<b>1 Narrative .....</b>	<b>1</b>
1.1 Background and Significance .....	1
1.1.1 The Opportunity: Creating a Collaboratory Software Environment.....	2
1.1.2 The State of the Art: Grid Technologies and Grid Infrastructures .....	4
1.1.3 Our Proposal: Creating a DOE Science Grid.....	7
1.2 Preliminary Studies .....	9
1.2.1 Application Studies .....	9
1.2.2 Prototyping Grid Infrastructures .....	11
1.3 Research Design and Methods .....	12
1.3.1 Central Directory and Security Services.....	12
1.3.2 Creation of a Multi-Lab Grid Prototype for a Global-Scale Science Grid .....	13
1.3.3 R&D Tasks: Extending the Grid Technology Base.....	17
1.3.4 User Support Services .....	19
1.3.5 Standards Advocacy and Enablement.....	20
1.3.6 Use of Collaboratory Technologies .....	20
1.3.7 Tasks and Milestones .....	20
1.3.8 Technology Transfer and Application .....	22
1.3.9 Connections.....	22
1.3.10Evaluation Criteria .....	23
1.4 Subcontract or Consortium Arrangements.....	<b>Error! Bookmark not defined.</b>
<b>2 Literature Cited .....</b>	<b>25</b>
<b>3 Biographical Sketches.....</b>	<b>29</b>
3.1 William Johnston .....	29
3.2 Ray Bair.....	29
3.3 Ian Foster.....	30
3.4 Al Geist .....	32
3.5 William Kramer .....	33
<b>4 Description of Facilities and Resources .....</b>	<b>35</b>

## Abstract

---

We propose a multi-laboratory Collaboratory Pilot aimed at defining, integrating, deploying, supporting, evaluating, refining, and developing (as necessary), the persistent Grid services needed for a scalable, robust, high-performance DOE Science Grid, thus creating the underpinnings for a DOE Science Grid Collaboratory Software Environment. This Grid will provide to applications and workflow systems persistent services for security, resource discovery, resource access, system monitoring, and so on. By thus reducing barriers to the use of remote resources and to the use of advanced Collaboratory services, we will make a significant contribution to SciDAC wide software standards and resources. Integrated activities in deployment, research and development, and application outreach will allow us to refine the tools and their deployment and support processes, providing the capabilities that will enable the DOE Science Grid to be cost-effectively scaled to arbitrary size. Close collaborations with a variety of application projects will ensure relevance to SciDAC goals and enable innovative approaches to scientific computing, via secure remote access to online facilities, distance collaboration, shared petabyte datasets, and large-scale distributed computation. The eventual outcome should be revolutionary changes in a wide range of scientific disciplines across DOE. The project team includes internationally recognized leaders in Collaboratory and Grid technologies, two of whom (WJ and IF) are on the executive committee of the Global Grid Forum, an international effort to define and standardize Grid services.

# 1 Narrative

---

## 1.1 Background and Significance

---

Grids [1] are a unified and integrated approach to building distributed, scientific computing environments that incorporate computation, data management, scientific instruments, and human collaboration. The vision for such Grids is that this unity and integration will revolutionize the use of computing in DOE's science.

The result of this of this revolution will be the ability to routinely and easily build and use large-scale, multi-institutional, and dynamic, distributed application environments for doing science that is not possible, or is difficult and expensive, today. Examples of such environments include:

- o Computational modeling, multi-disciplinary simulation, and scientific data analysis with a world-wide scope of participants and the use of computing and data resources at many sites.
  - the High Energy Physics data analysis that involves hundreds of collaborators, and tens of institutions providing data and computing resources
  - observational cosmology that involves data collection from a world-wide collection of instruments, analysis of that data to re-target the instruments, and subsequent comparison of the observational data with simulation results
  - climate modeling that involves coupling simulations running on different supercomputers
- o Real-time data analysis and collaboration involving on-line instruments, especially those that are unique national resources – e.g. LBNL's and ANL's synchrotron light sources, PNNL's high field NMR machines, etc.
- o Generation, management, and use of very large, complex data archives that are shared across global science communities – e.g. high energy physics data, earth environment data, human genome data
- o Collaborative, interactive analysis and visualization of massive datasets – e.g. DOE's Combustion Corridor project
- o Multi-disciplinary R&D that integrates the computing and data aspects of the different scientific disciplines.

We believe that this revolution will come about as a result of fundamental changes and improvements in access to powerful computing systems, large-scale data archives, scientific instruments, and collaboration tools. These changes will be in the form of services that are integrated with the user's work environment, and that enable uniform and highly capable access to these computers, data, and instruments, regardless of the location or exact nature of these resources. This applies both to transient-use resources like computing systems and data caches, e.g. as they are needed to perform a simulation or analyze data from a single experiment, and to persistent-use resources, such as databases and archives, scientific instruments, and collaborators, whose use will exist for the lifetime of a project, or longer.

### *What is the Role of "Grids?"*

The role of a Grid is to provide standardized middleware services within organizationally and geographically dispersed environments that greatly simplify the construction and operation of application systems that interconnect the computing, data, instrumentation, and human components of complex science collaborations. The basic Grid services will locate and schedule resources, provide uniform and secure access to heterogeneous resources, provide the global naming and scoping mechanisms to support formation and management of virtual organizations / collaborations with a potentially world wide extent, and provide the monitoring, auditing, and fault recovery mechanisms needed for resilience and accountability / allocation management.

On top of these basic services will be built both generic and discipline specific workflow management systems (e.g. as contained in the "framework" in the figure) that will carry out the human defined protocols for, e.g., multi-disciplinary simulations and data analysis; global data cataloguing and replica management systems needed to

manage the data for these scenarios, etc. That is, the services needed directly by scientific and engineering problem solvers.

Finally, all of these services will be available through Web / desktop interfaces (as illustrated by the “problem solving framework” in the figure) in order to produce a highly usable environment in which problem solving protocols may be formulated, controlled, modified, and integrated with other aspects of the work environment, and shared securely with collaborators.

### 1.1.1 The Opportunity: Creating a Collaboratory<sup>a</sup> Software Environment

DOE Office of Science laboratories operate a wide range of unique resources, from light sources to supercomputers and petabyte storage systems, that are intended to be used by a large distributed user community. The laboratories’ geographically distributed staff are frequently faced with scientific and engineering problems of great complexity, the solution of which requires the creation and effective operation of large multidisciplinary teams. The problems to be addressed are large and highly challenging, often greatly exceeding the limits of traditional computing and information systems approaches.

Recognizing these challenges, the SciDAC call for proposals in National Collaboratories and High Performance Networks, echoing the document “Scientific Discovery through Advanced Computing,” calls for the creation of “a *Collaboratory Software Environment* to enable geographically separated scientists to effectively work together as a team and to facilitate remote access to both facilities and data.”

A Collaboratory Software Environment must, by definition, span multiple institutions and link numerous, geographically distributed resources of different types. These characteristics introduce two unique challenges that have, until recently, made the development of collaboratory applications very difficult. First, the diverse resources involved typically feature vendor and site-specific software and management mechanisms; writing code that can deal with the variety of different mechanisms that arise in even a small collaboratory can become prohibitively difficult. Second, the sheer scale of the resources involved, and the fact that they span multiple administrative domains, make such fundamental issues as resource discovery, authentication, and fault detection very difficult—often beyond the abilities of a typical collaboratory application developer.

A new class of distributed infrastructure called Grids [1, 3] address these two challenges directly by (a) defining and deploying standard protocols and services that provide a uniform look and feel for a wide variety of computing and data resources, and (b) providing global services that provide essential resource discovery, authentication, and fault detection capabilities. Together, these capabilities reduce the barriers to the large-scale coordinated sharing and use of distributed resources that is at the heart of collaboratory applications. Grid capabilities are thus fundamental to the efficient construction, management, and use of widely distributed application systems; human collaboration and remote access to, and operation of, scientific and engineering instrumentation systems, and the management and securing of the computing and data infrastructure. For these reasons, emerging Grid technologies have been adopted by major collaboratory and supercomputer center access projects worldwide, such as the NSF National Technology Grid [4], NASA’s Information Power Grid [5], the European Union’s Data Grid [6], the NSF’s Network for Earthquake Engineering Simulation Grid [7], and the DOE ASCI DISCOM project [8]. The Global Grid Forum ([www.gridforum.org](http://www.gridforum.org)) [9] provides an international coordinating and standards definition body, analogous to the Internet’s IETF.

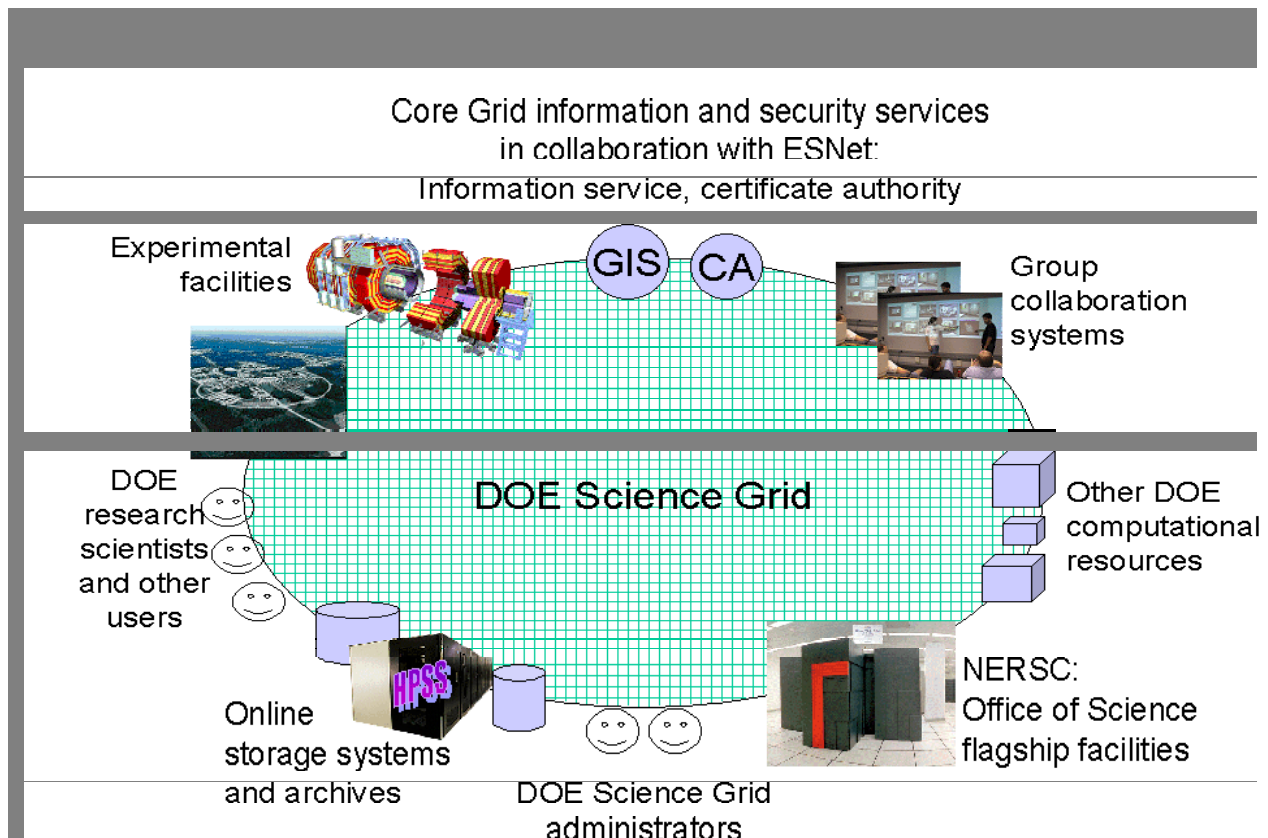
DOE’s Office of Science has consistently invested in the information technology infrastructure required to support great science and, where necessary, the research required to create that infrastructure. This spirit of innovation has, over the years, produced for example the first national supercomputer facility (CTRCC, subsequently NERSC), the first network devoted to science (MFEnet, subsequently ESnet), and High Performance Computing Research

---

<sup>a</sup> “... combining the interests of the scientific community at large with those of the computer science and engineering community to create integrated, tool-oriented computing and communication systems to support scientific collaboration. Such systems can be called *collaboratories*.” . "National Collaboratories - Applying Information Technology for Scientific Research," 1993, Committee on a National Collaboratory - National Research Council.

Centers. It has also motivated research and development efforts such as the Cray Timesharing System in the 1970s, the grand challenge program, and today's DOE2000 program of collaboratory and numerical toolkits research.

The creation of a DOE-wide distributed computing infrastructure, or *DOE Science Grid* (Figure 1) represents an initiative of similar ambition and, we believe, opportunity for impact. The DOE Science Grid that we describe here is designed to reduce or eliminate barriers to the coordinated use of DOE resources, regardless of the physical location of those resources and the users that access them. For the first time, we will provide a persistent and supported set of Grid infrastructure and deployable services in the DOE community. This infrastructure, as we explain below, will provide a range of new Grid Services addressing issues of resource discovery, security, resource management, instrumentation, and the like. These services will in turn be used to create a range of innovative Grid tools targeting specific application classes. Our goal in creating this infrastructure and tools will be to enable innovative approaches to scientific computing based on such concepts as secure remote access to online facilities, distance collaboration, shared petabyte datasets, and large-scale distributed computation; the eventual outcome should be revolutionary changes in a wide range of scientific disciplines across DOE.



**Figure 1. The DOE Science Grid provides uniform access to a wide range of DOE resources, as well as standard services for authentication, resource discovery, and the like**

The following simple example illustrates the opportunities of Grids. Imagine a “Fusion Collaboratory” designed to allow fusion science community to both pool computational resources and access advanced simulation codes from a desktop, with these codes being executed on any idle pooled resource. Once data is created, they wish to allow distributed, collaborative visualization and analysis.

In the absence of Grid infrastructure, the creation of such a system would be very difficult. In order to pool resources, we require mechanisms for creating distributed directories of computational, storage, and other systems. In order to select resources, we require mechanisms for determining the characteristics (e.g., availability) of any

device in the Collaboratory. In order to access resources, we need to be able to negotiate the diverse access mechanisms and local idiosyncratic features of different systems. And all this needs to be done with high security, which means we need secure mechanisms for establishing identity, authorization, and so forth. Building these various mechanisms from scratch is a lot of work, and typically (because of the inexperience and limited resources of a single development group) results in a fragile, insecure, inadequate product. Yet if a DOE Science Grid is in place, Collaboratory developers and users will be able to take them for granted. Furthermore, the resulting Fusion Collaboratory will be interoperable with our DOE Office of Science Collaboratories, due to the use of standard mechanisms.

DOE researchers have developed much of the basic technology for Grids and we feel that building a Grid for DOE Collaboratory science will give DOE the benefits of this work, and give DOE computer science researchers to opportunity to identify and address the many outstanding issues for building workable Grids, especially Grids designed to support collaboratories.

## **1.1.2 The State of the Art: Grid Technologies and Grid Infrastructures**

### **1.1.2.1 *The DOE Collaboratory and HPC Environment***

As noted, the DOE Science Labs and their partners represent a diverse and dynamic Collaboratory and high performance computing (HPC) environment. They involve numerous machine types, architectures, and operating systems, from ultra-computers (in particular at NERSC, but also e.g., at ORNL and PNNL) to commodity clusters, storage facilities, specialized visualization hardware, and a wide range of scientific instrumentation. Such resources are available at each of the labs in this proposal, spanning both development and production capabilities.

One of the biggest barriers impeding the broad development of collaboratories is the lack of distributed computing infrastructure including uniform standards for security and resources access. Collaboratories and HPC communities need to remotely access HPC resources and large volumes of data, to connect and manage scientific instruments, to coordinate the activities of the people involved in scientific experiments, and to integrate a diverse set of resources and tools into problem-solving environments for specific problems. In the process of addressing large and complex simulation problems, a wide range of calculations are performed using compute and data resources, both local and remote. Hence the rising demand for computational power, which DOE laboratories address primarily through terascale supercomputers, is an important driver of uniform distributed computing technologies in the HPC community.

Elements of the modeling and simulation community also have significant interest in coupling multiple computers across the country to solve a single problem. Another emerging requirement is to facilitate computing resource “load leveling” within communities that have common allocation policies, thereby effectively improving time-to-solution, enabling “on-demand” access to computing, and increasing the computing resources available to an individual by providing uniform access to all systems within the community. All of these capabilities are difficult, expensive and time-consuming to develop without a persistent infrastructure with standard interfaces.

The complex and evolutionary nature of the scientific environment requires general services that can be combined in many different ways to support different types of collaboratory applications and support the changes in those applications so that the collaboratory can evolve along with the scientific understanding of the problem. Resource management for such a dynamic and widely distributed environments requires global naming and authorization services, scalability and fault-tolerance well beyond the scope of existing systems and standards. Collaboratories generate and operate on data sets multiple terabytes in size, and services are needed to manage this data. For example, the management of auxiliary resources, such as network bandwidth, archival storage bandwidth, and network caches, must be addressed to simultaneously utilize multiple, remote, high performance resources. Uniform and interoperable security is a critical issue for collaboratories for the management of access rights, the protection of proprietary data, and general protection against hacker shenanigans. Security services and tools must provide for secure and standardized authentication and authorization mechanisms for all resource interactions.

### **1.1.2.2 The Promise of Grids**

The considerable collective experience of the PIs with collaboratories and Grid technology (e.g., see [10], [11], [12], [13], [14], and also Section 1.2) persuades us that Grid technologies are an essential path forward to meet these requirements for DOE collaboratory science and HPC access. Grids provide the common infrastructure capabilities needed for:

- o creating and managing collaboratories and other “virtual organizations” (communities) within which resource sharing is required;
- o building and managing complex, multi-component distributed application systems; and
- o making resources accessible through common access methods,

all of which are needed to build collaboratories.

The uniformity provided by Grids for diverse computing platform environments also provides for readily incorporating emerging technologies and new platform capabilities, at least for those aspects of the problem solution that are not architecture dependent.

Many Grid capabilities are currently R&D topics and the focus of a substantial and growing R&D community. However, as attested to by the Global Grid Forum [9] standards work, there is both available technology and a scientific community consensus for building prototype production Grids that will support a variety of collaboratory and HPC service delivery models. Further, the current state-of-the-art is capable of providing persistent and usable infrastructure that many other projects can build on, though the architecture and the details of large, multi-site Grids are still evolving.

We note that our views on Grids are shared by many of our colleagues across the DOE laboratory system, as attested by the attached letters of support.

### **1.1.2.3 Technology Assessment**

As we approach the creation of a DOE Science Grid, we need not start from scratch. On the contrary, there is a substantial body of technology and experience on which we can build: much, it turns out, developed by the PIs on this proposal and their collaborators worldwide. Hence, our goal of constructing a DOE Science Grid can proceed efficiently, building on (and extending) this existing technology base and profiting from prior experience.

Figure 2 presents a logical view of Grid technologies, identifying services concerned with managing underlying resources, generic services concerned with security and so forth, and specialized middleware that supports different styles of usage, such as different programming paradigms and access methods. Services marked “Globus” or “Condor” are currently available, though not necessarily complete. Most of the other services are in various stages of R&D, and some of these are the subjects of this proposal.

Reviewing each of the layers in Figure 2 in turn, we point out what we know how to do and where major R&D work remains to be done.

At the resource level, we have our various computers, storage systems, and the like. Condor [15] provides mechanisms for managing resources within workstation pools. Further work is required to integrate reservation [16, 17] and monitoring, but much of what we need is in place.

At the service level, the Globus Toolkit [18] provides us with mechanisms for authentication, limited authorization (e.g., Akenti [19] and the IETF Generic Authorization and Access control API (GAA) [20], the topic of a SciDAC proposal [21]), resource information access [22], network characterization (e.g., Network Weather Service [23]), and for accessing compute and storage resources [24, 25]. At this level, there are also significant gaps, for example with respect to monitoring, accounting, and fault management. Significant future developments in security and data management technologies are also required (these are the subject of ANL-led SciDAC Collaboratory Middleware proposals). However, again, there is sufficient technology here to make progress.

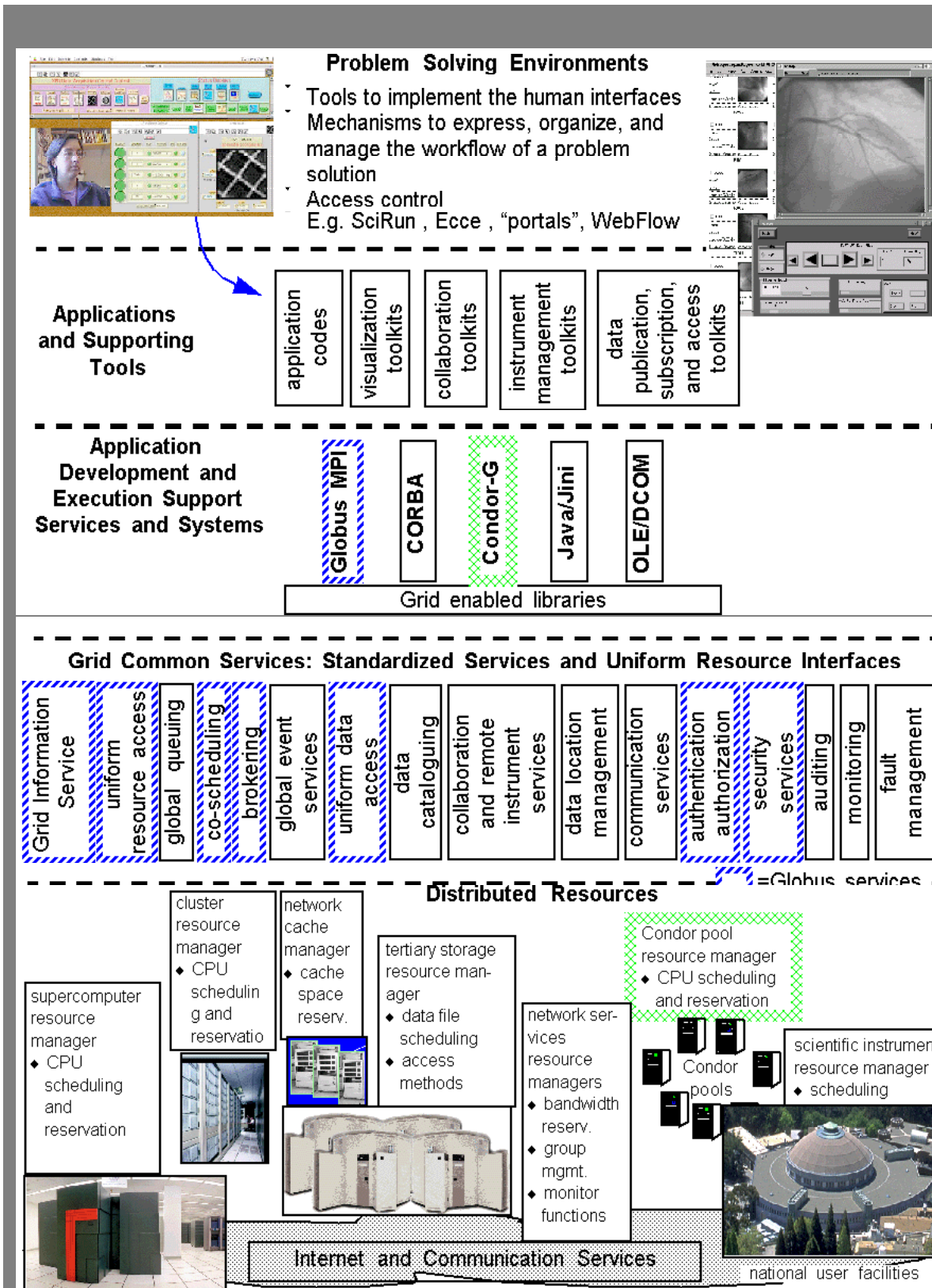


Figure 2. A detailed view of Grid architecture, showing the primary required services

At the applications level, we see a wide variety of tools, focused on different user communities and at different levels of development. For example, we see a Grid-enabled Message Passing Interface implementation (MPICH-G [26]), Condor-G [27] for job management, NetSolve [28], application-level schedulers [29, 30], replica management tools [31], to name just a few.

At the top of the Grid architecture are the frameworks that allow humans to organize the solution of multi-step computing and data problems in terms of the relationships between data generation and storage, the processing that generates and transforms data, access control, import and export of visual exploration and interaction interfaces, etc. Such integrations include various Web portals for managing user “sessions” (usually for computing resources), e.g., the PACI Portal [32], Cactus Portal, and some experimental systems for software component-based systems such as Indiana’s CCAT [33]. There are several projects underway to build the CoG toolkits for integrating Grid services into the Web environment [34, 35] as well as into the Python environment (a SciDAC proposal) [36]. Work on applicable, general workflow engines and the global event services needed to support them are addressed in another SciDAC proposal [37].

Collectively, these different services and tools represent a solid technology base that, while far from complete, represents a solid basis on which to establish a DOE Science Grid. As we explain in more detail in Section 1.2.2 below, many of these services have been used on a large scale in other Grid deployment efforts, and applied in applications.

#### **1.1.2.4 Creating a DOE Science Grid**

We now return to the question of why it is important to create a DOE Science Grid. A Grid, as we indicated earlier, is a set of persistent infrastructure services designed to support Collaboratory and HPC access applications. As such, it supports community access to a set of resources that are of interest to that community. It is important to emphasize that a DOE Science Grid is thus a critical development project in support of DOE science. The fact that NASA, NSF, and others are creating their own Grids to support their own science should be encouraging to us, but in no way obviates the need for a DOE Science Grid. Only a DOE Science Grid can provide DOE scientists with Grid-enabled access to DOE resources.

Many of the basic Grid services exist and Grids are being built using these services, but there is no concerted effort that enables a broad array of projects and resources for DOE Office of Science missions. There also remains considerable R&D work to be done, and building an operational collaboratory Grid such as is proposed here will expose elements of importance to collaboratories. To be effective, Grid capabilities must be standard across institutions, agencies and nations. This is well recognized in the technical community, and there is considerable momentum behind the Global Grid Forum, and Grid services are now being refined and standardized in an international venue.

The clear need, and the goal of this proposal, is to provide services that enable an integrated collaboratory and HPC Grid environment for the DOE community based on the previous experience in building and deploying collaboratories and working with HPC users. The central issue of this goal is to define a core set of such services, deploy the services, and then develop the operational aspects needed to make these services persistent. Some of the technology and experience exists, and what does not will have to be developed. Operational issues for a large, multi-site Grid must also be identified and addressed in this project. This effort complements ongoing Grid deployment and application efforts funded by other agencies and governments. Not only can it make the benefits of Grid technologies (many developed within DOE laboratories) available to DOE researchers; it can also provide a basis for interoperability and resource sharing with these other Grid infrastructures, at both national and international levels.

#### **1.1.3 Our Proposal: Creating a DOE Science Grid**

We propose here a multi-laboratory Collaboratory Pilot aimed at defining integrating, deploying, supporting, evaluating, refining, and developing (as necessary), the persistent Grid services needed for a scalable, robust, high-performance DOE Science Grid, thus creating the underpinnings for a DOE Science Grid Collaboratory Software Environment (Science Grid-CSE). By thus reducing barriers to the use of remote resources and to the use of advanced security, resource discovery, and other services, we will make a significant contribution to the realization

of the overall goals of the SciDAC program. Close collaborations with a variety of application projects will ensure relevance to SciDAC application goals. The experience gained will also allow us to refine the tools and their deployment and support processes, providing the capabilities that will enable the DOE Science Grid to be cost-effectively scaled to many institutions, resources, and users.

The DOE Science Grid Collaboratory Pilot will undertake activities that we will, for simplicity, label “deployment,” “research and development,” and “application outreach,” although in practice we expect to find these different activities intermingling as, for example, application outreach activities provide experiences that guide further refinements to technologies and deployment strategies.

Deployment activities will focus on the creation of the DOE Science Grid, a persistent Grid infrastructure that embraces unique DOE resources and designed to support DOE science. As illustrated in Figure 1 and discussed in detail in Section 1.3, the DOE Science Grid will:

- o Provide advanced services for authentication, resource discovery, and the like, based on the Globus Toolkit [38].
- o Provide secure, uniform access to advanced resources at multiple DOE resource sites: initially, computers and storage systems at ANL, LBNL, NERSC, ORNL, and PNNL; later, we hope to also incorporate other resource types (e.g., networks) and resources at other laboratories and universities.
- o Provide management infrastructure that allows for monitoring of various aspects of DOE Science Grid operation.

In addition to deploying and operating these services at DOE laboratories, the DOE Science Grid team will produce packaged versions of Science Grid-CSE software for deployment at other sites, e.g., other DOE laboratories and university collaborators.

As we explain below, the technologies to be deployed in this effort will come from two primary sources: existing commercial and open source technologies (in particular, elements of the Globus Toolkit and Condor system), and new technologies developed concurrently with this project in other Grid R&D efforts, in particular efforts funded under the SciDAC Network and Collaboratory Middleware call. The deployment activity will hence be designed to inject a steady stream of new technologies into the SciDAC Collaboratory Software Environment. In so doing, it will contribute to the important goal of integrating diverse DOE-sponsored Collaboratory technologies into a common framework (see Section 1.3.5).

Research and development activities will address a small set of R&D issues that we believe are critical to the creation of a broadly usable DOE Science Grid and that are not being addressed in other efforts. Specifically, we will:

- o Develop, deploy, and operate a DOE Science Grid-wide monitoring and auditing infrastructure designed to identify “faults” (whether failures or performance bottlenecks) in various resources (networks, computing, storage). Our intention is that by logging selective information from these monitors we will also be able to account for usage of these resources by SciDAC users. The third part of this task is to develop easy to use tools so users can access their allocation status as well as monitor their DOE Science Grid applications.
- o Investigate techniques for detecting faults and reacting to faults that are the result of the Science Grid or that affect jobs submitted through the Science Grid. Our initial approach will be to detect faults or variations from resource requests, and report these problems back to the user and the Science Grid log. This monitoring information will also be fed back into the Grid scheduling components. Subsequently, we will explore automatic adaptation techniques.
- o Investigate, in collaboration with ESNNet, the issues of providing mechanism and management for a rich, global namespace for virtual organizations, for data catalogues, and for resources – a global Grid Information Service.
- o Develop a Grid security model for open scientific computing environments, and investigate the issues of implementing such a model.

Application outreach and support activities will simultaneously engage DOE application and collaboratory groups in the use of DOE Science Grid services and ensure early and regular feedback from those groups to guide ongoing R&D and deployment activities. We envision the outreach activities comprising the following specific tasks:

- o Development of detailed documentation for DOE Science Grid services, as well as case studies illustrating how these services can be used in various types of application.
- o Regular tutorials (conducted both in person and over Access Grid facilities) aimed at communicating various aspects of DOE Science Grid services, and focused “bring your application” workshops designed to jumpstart use of the DOE Science Grid by application groups.
- o Focused support for key DOE application groups, to accelerate their migration to the use of DOE Science Grid services.

We note that the ambitious vision of this project and relatively modest budget means that there are many desirable tasks that we must omit from this proposal. If additional funds are available in future years, it will be straightforward to expand the scope in terms of the range of technologies deployed, the range of resources supported, the set of new technologies developed, and the amount of support provided to application groups. Space does not permit us to describe these additional tasks here, but we would be delighted to provide more details if requested.

## 1.2 Preliminary Studies

---

The personnel involved in this project have a great deal of experience with the creation of advanced networked applications, some of which we outline briefly below. In addition, we provide details on early work conducted at ANL and LBNL on the creation of Grid systems, which has guided our planning for the DOE Science Grid project.

### 1.2.1 Application Studies

**High Energy and Nuclear Physics.** The management and analysis of the extremely large quantities of data produced by leading high energy and nuclear physics experiments (e.g., BaBar, D0, RHIC, CMS, ATLAS) represents an unprecedented information technology challenge. There is a broad realization within these communities that the computational and storage resources needed for data management and analysis cannot realistically be gathered at a single location, and that future computational environments must hence be “Data Grids”: distributed collections of storage systems and compute farms that are operated in a coordinated fashion [39] [40]. Over the past two years, the DOE-funded Particle Physics Data Grid (PPDG) [41] project has explored the use of Grid technologies for distributed management and analysis of data from experiments such as BaBar, D0, CMS, and ATLAS. PPDG participants have demonstrated high-speed transfers of physics data using GridFTP, the use of [42], technologies such as Condor for distributed data analysis, and the use of replica catalogs for data management. These experiments have been highly successful in that a solid understanding and considerable consensus concerning Data Grid requirements and architecture has emerged. The PPDG-2 project (a proposed SciDAC Collaboratory Pilot) now aims to build on this consensus and understanding to establish production Data Grids for a range of DOE-funded physics experiments.

The adoption of Grid concepts by this community represents a tremendous endorsement of the technology, but also introduces significant challenges. Specifically, they now face the need to deploy production services for authorization, resource discovery, resource access, etc. If left to their own devices, they will inevitably be forced to create their own Certification Authorities and other core services, which will (unless care is taken) not be interoperable with other DOE Grid components. The DOE Science Grid project hence becomes doubly important: it can help them with the task of creating production Data Grids, by providing key services, and it can help ensure that we end up with a single consistent set of Grid services across all DOE resources, not multiple inconsistent and non-interoperable Grids.

**Chemistry:** Computational chemistry is a nearly ideal application for Grid technologies. A large fraction of all scientific computing cycles is devoted to computational chemistry, and computational chemists often use high performance computers at distant institutions to perform their work. Although there are many computational chemistry codes, the vast majority of computing cycles are used by a small number of closely related codes for

electronic structure and molecular dynamics computations (e.g., Gaussian, Gamess-US/UK, NWChem, Amber, Charm, Gromos). Therefore, a small number of Grid implementations of chemistry applications can leverage a large number of compute cycles, and serve as models for many other similar codes. These codes carry out computational experiments on many different kinds of molecular systems by simply changing the input parameters (and not the code itself). For any given study, computational chemists may run the same code on a many systems, ranging from their desktop computer, to departmental computers (local and remote), to distant terascale systems.

Managing this increasingly complex array of computations and resources has always been a problem, traditionally mediated by hand-crafted scripts and paper records. Problem-Solving Environments (PSEs) provide interactive tools for integrating and managing these computations, supporting the discovery process, and managing databases of computational results. For example, the Extensible Computational Chemistry Environment (Ecce) [43] manages calculations for both Gaussian and NWChem. (Gaussian is arguably the world's most popular electronic structure code, and NWChem is used at hundreds of sites for large scalable computations.) However, the impact of these PSEs is limited without standard tools and interfaces for locating resources, authorizing and authenticating access, transferring data, launching jobs, etc. Because there is no standard way to access remote computer resources, Ecce offers multiple methods to launch jobs, including Globus, Remote Shell, Secure Shell, Telnet and FTP. The Grid (Globus) approach is the only one that promises to have the flexibility needed for present and future generations of computational Chemistry codes and their associated Collaborative Problem-Solving Environments. The availability of a persistent Grid is crucial to provide the stable environment for the community developing chemistry codes and their PSEs.

**Materials.** The Materials Microcharacterization Collaboratory (MMC), a joint effort between ORNL, ANL, and LBL, was one of two collaboratory pilots funded by the DOE2000 Program. Over the past four years this project has put in place a collaboratory framework across DOE laboratories that caters to the needs of the microscopy community [44]. The primary focus of the MMC was remote instrument control (microscopes and beamline experiments). Their framework is now used routinely for remote access to lab resources by industry and other DOE researchers. The MMC also pushed the frontiers of electronic lab notebooks as a means to collaborate and share research in a production environment. Nevertheless, much of the underlying infrastructure was created specifically for this project. The success of the MMC collaboratory pilot emphasizes the need for a DOE Science Grid to avoid replication of collaboratory framework efforts across the second round of collaboratory pilots.

A number of research projects in wide area distributed computing have been done that help define framework requirements. In one demonstration, scientists at ORNL combined an Alpha cluster in California, a supercomputer in New Mexico, and a supercomputer in Tennessee in order to simulate the combustion of a methane flame. Cumulvs was used to visualize and steer this distributed combustion computation in real time from Washington DC [45]. Because of the thousands of processors involved and the length of the run, monitoring of resources was critical to the project. The combustion application was modified to allow it to adapt to resource faults and network degradation. This research was a preliminary study into the fault detection, auditing, and adaptation that will be required in a DOE Science Grid. Another wide area distributed computing study highlighted the conflict between high-speed communication and security. Low-level ATM protocols (AAL5) were used to transfer data inside a materials science electronic structures calculation, which was distributed across supercomputers at ORNL and Sandia [46]. While AAL5 allowed much higher communication bandwidth through ESnet, it initially had no provisions for authentication or security. Routinely allowing insecure low-level (high-speed) communication protocols to be used between labs could present a security hole that would allow hackers to compromise the labs computing resources. After six months of research a multiprotocol, multichannel solution was found that satisfied the security groups at all the sites and the single application was run on an aggregate of nearly 3000 processors. This preliminary study illustrates the kinds of security services that a DOE Science Grid will need to provide for research scientists to run their SciDAC applications.

**Climate: Earth System Grid.** The need to evaluate climate change scenarios under the Kyoto accord makes climate modeling a mission critical application area for DOE. The climate modeling component of DOE's SciDAC program seeks to address this need through the creation of an advanced climate simulation program that will accelerate the execution of climate models one hundred-fold by 2005 relative to the execution rate of today. High-resolution, long-duration simulations performed with these models will produce tens of petabytes of output. However, to be useful, this output in turn must be made available to global change impacts researchers nationwide, both at national laboratories and at universities, other research laboratories, and other institutions. To this end, DOE researchers are

working to create what an *Earth System Grid* (ESG): a virtual collaborative environment that links distributed centers, users, models, and data. This ESG will provide scientists with virtual proximity to the distributed data and resources that they require to perform their research. Note that the primary goal is not interpersonal collaboration, but rather seamless and high-performance access to data and compute resources.

The ESG group, which involves researchers at ANL and LBNL as well as NCAR, USC/ISI, and LLNL, has created replica management and data transfer tools, and integrated these tools into climate model data analysis systems to enable automatic selection of the “best” copy of data. Experience with the use of these tools has emphasized the importance of persistent Grid infrastructure services for production use, so that ESG users can authenticate themselves, determine availability of Grid resources, and then access and/or transfer required datasets. For example, much of the climate data of current interest to ESG users resides at NERSC. However, in the absence of a production Certificate Authority and a GSI-enabled GridFTP server on NERSC storage systems (i.e., basic Grid infrastructure services), it is very difficult to provide the sort of automated access that users need. The realization of the goals of the DOE Science Grid project will hence be of immediate benefit to the climate community.

**Supernova Factory Collaboratory:** Over the past several years, astronomers and astrophysicists have been conducting in-depth sky searches with the goal of identifying supernovae in their earliest evolutionary stages in order to measure their changing magnitude and spectra during the two to four weeks of their most "explosive" activity. Computational analyses of these early experiments have demonstrated that the expansion of the universe is accelerating, apparently driven by an unknown new force - now called dark energy. While these experiments have been daunting tasks in terms of both the number and volume of observations required, they have achieved notable success, being named Science magazine's Breakthrough of the Year for 1998. (Science, 18 December 1998). This early success has spurred development of a more ambitious search program—the Supernova Factory—currently under development at LBNL by Saul Perlmutter and the Supernova Cosmology Project (<http://www-supernova.lbl.gov/>). This is an earth-based observation program utilizing observational instruments at Haleakala and Mauna Kea, Hawaii and Mt. Polomar, California. When fully implemented, this search program will also utilize instruments at observatories in Chile and the Canary Islands. This program also serves as a development testbed for the next generation search program, the space-based Supernova Acceleration Probe (SNAP, <http://snap.lbl.gov/>).

The future Supernova Factory computing environment will use Grid resources to support the data acquisition, archiving and analysis efforts. Major elements of this environment include the analysis tools based on a core of Grid-based services, development of workflow management tools for control and monitoring of acquisition, archiving and analysis tasks, and integration of collaborative tools to allow efficient and timely control of observing operations by members of the collaboration. The Supernova Factory will require a domain-specific workflow framework built on Grid capabilities.

## 1.2.2 Prototyping Grid Infrastructures

Project participants, particularly at ANL and LBNL, have been exploring the creation of Grid infrastructures for several years: indeed, much of the information summarized in Section 1.1.2 was gained as a result of research and development activities conducted by the PIs or their collaborators. We summarize here some of the more notable and recent activities.

**GUSTO.** Globus project participants at ANL and ISI led the creation of the “Globus Ubiquitous Supercomputing Testbed Organization” during 1997-99, which deployed early versions of Globus services across some 60+ sites worldwide. GUSTO served as a testbed for early Grid Information Service, Grid Security Infrastructure services, resource monitoring, and resource management technologies. At the SC’97 conference, GUSTO provided access to several hundred computers with over 3000 processors. Success stories included a record-setting synthetic forces simulation involving >1000 processors across 7 DOD and university sites [47]. GUSTO experiences have motivated significant refinements to core services, concerned for example with GIS scalability.

**National Technology Grid and Information Power Grid.** Support from NSF enabled the creation of the Globus-based National Technology Grid [48], a Grid infrastructure based at NCSA and SDSC and designed to support the university community; NASA researchers, under the direction of Bill Johnston, are creating the Globus-based Information Power Grid [5, 49]. These projects include multiple production CAs (one of these being operated for NCSA by ANL), production MDS services, and production deployment of resource access services. They have

motivated extensions to core services, concerned for example with support for multiple CAs and improved packaging.

**Supernova Factory.** A preliminary version of a Globus environment at LBNL has been used to support the Supernova Factory program with Globus managed computer and data resources. This experience has led to considerable refinement of the operational model, and in fact, led directly to the model depicted in Figure 3.

## 1.3 Research Design and Methods

---

The primary goal of the DOE Science Grid Collaboratory Pilot is to deploy, evaluate, refine, evolve, and support a set of core Grid services across DOE labs that will make an initial set of computing and data resources uniformly available.

This deployment activity has as its three-year goal the creation of a persistent, production infrastructure that is supported and available in the same way as other DOE production services, such as networks, domain name services, supercomputers, and beamlines. During the prototype phase of the Science Grid (the first two years) we will involve a “friendly” set of DOE applications that can provide real world requirements, testing, and evaluation; evaluate interoperability with other Grid infrastructures; and create an infrastructure that supports experimental activities as well as prototype production use. In the third year, we will expand the DOE Science Grid to additional sites and shift from a prototype to a production infrastructure for use by all parts of the SciDAC programs as well as other DOE researchers.

In the following, we first expand upon this general statement of goals, providing technical details as follows:

- o We describe our plans for central directory and security services.
- o We explain how we will develop and multi-lab DOE Science Grid prototype, and then expand this prototype into a production DOE Science Grid.
- o We present technical work focused on the creation of Grid monitoring and auditing infrastructure.
- o We describe our plans for application outreach and user services.

Then, we address organizational issues, providing details on:

- o How the DOE Science Grid project will address SciDAC needs for the definition and adoption of standard Collaboratory architectures, protocols, interface, and services.
- o Our management structure.
- o How we will use collaboratory technologies within the DOE Science Grid project itself.
- o Our tasks and milestones.
- o How we expect the work performed in this project to be transferred to the larger community.
- o How the proposed work relates to, leverages, and supports other work within and outside DOE.

We conclude this section with additional material relating to the SciDAC Collaboratory Pilot proposal call’s evaluation criteria.

### 1.3.1 Central Directory and Security Services

We will collaborate with DOE’s Energy Science Network (ESnet, [www.es.net](http://www.es.net)) to create, operate, and evaluate a scalable and global directory service and certificate authorities, thus providing the essential services required to enable resource discovery and authentication across all SciDAC applications, users, and resources. These services will allow Science Grid-CSE users to authenticate once, using public key technology, and then access any Science Grid resource (to which access is authorized) without further authentication. It will also provide the mechanisms needed for secure verification of user and resource identity and for the certification of attributes that might be used in authorization decisions.

Work with ESnet on the directory service will address naming and indexing issues that arise when multiple virtual organizations must be supported concurrently, with resources and participants in common; scaling issues with respect to performance and reliability when directories encompass thousands of resources at hundreds of locations; support for general cataloging services (e.g., data replica catalogs) within the Science Grid directory infrastructure and name space; and maintenance of the directory service. This is a critical issue that is not being addressed elsewhere in the Grid community.

Work with ESnet on certificate authorities will address the definition of a DOE Science Grid security model; the creation of a “Root” Certification Authority (to sign the institutional CA certificates that facilitate cross-institutional identities); the creation of institutional and other independent CAs to issue certificates to associated groups of users; the creation of a Science Grid-wide Certificate Revocation List “repository;” and maintenance of the CA.

### **1.3.2 Creation of a Multi-Lab Grid Prototype for a Global-Scale Science Grid**

The DOE Science Grid has two goals: Provide a near term infrastructure for DOE collaborative science, and help establish the design parameters for services needed to support a global scale science Grid.

To support DOE laboratories, we will deploy, evaluate, and (as needed) enhance the services needed to “Grid-enable” key compute and storage resources at ANL, LBNL, NERSC, ORNL, and PNNL, thus creating a multi-site Grid suitable for use by a wide range of application projects. Accredited users of this Grid will be able to allocate and use compute, storage, network, and other resources in an on-demand fashion, using standard interfaces, and to coordinate the operation of these resources across multiple sites. We list the resources to be integrated into the initial Science Grid in Section 4.

This effort will include integration of GridFTP remote access mechanisms with mass storage systems, hence enabling high-performance access to remote storage systems via uniform mechanisms [31]; integration of Globus server-side software with computer systems, hence providing uniform mechanisms for reservation, allocation, and submission to compute resources, and subsequent management of computation [24, 50]; deployment of monitoring and auditing tools; enhancements to local system administration procedures to address Grid operation; evaluation of user experience with Grid software; and identification and correction of problems associated with missing functionality and scaling problems in current implementations.

A model for configuration and management of the several different types of user systems (as opposed to resource platforms) that are involved in Grids will also be developed.

The multi-lab Grid will evolve over the three-year period from an initial testbed, encompassing mid-range computing and storage resources, and network cache systems at each laboratory, to a production Grid, encompassing the flagship, topical, and experimental facilities in SciDAC as well as other major facilities.

As DOE’s flagship computing facility, the National Energy Research Scientific Computing Center (NERSC, [www.nersc.gov](http://www.nersc.gov)) participation in the prototype Science Grid is very important. It will allow investigation of the issues that need to be addressed before using Grid technology to provide access to NERSC production facilities. NERSC staff will work with the rest of the Science Grid team to integrate Grid software with non-production NERSC systems, each of which is a prototype for a NERSC production system. NERSC evaluation of, and feedback about, the Grid software will consider a range of issues important for a production environment, including stability of the Grid middleware, and efficacy and operational issues of integrating the Grid services with such systems. This particularly relates to integrating Grid security and authentication mechanisms with the NERSC authentication and authorization system; integration with the batch schedulers; and use, configuration, and management of the Grid Information Service. Once a prototype deployment is accomplished, then the use of Grid services by applications to access these NERSC resources will be evaluated.

Several NERSC systems will participate in the Science Grid. Initially, these will be a high-performance Linux cluster and the PROBE, HPSS mass storage testbed system. The cluster will serve to introduce NERSC personnel to Grid software. Subsequently, Grid software will be deployed in “limited production mode” on one of the major NERSC compute platforms and one storage platform. This is targeted for end of Q2FY03 (March, 2003). The most likely systems would be the IBM SP - seaborg, the production HPSS, and possibly the physics data processing system (PDSF).

Application projects will be engaged at each stage in the deployment, evaluation, and refinement process. Applications programmers will also have access to collections of user-developed Grid programming tools deployed in Science Grid, e.g., MPICH-G2 and Condor-G.

Establishing the parameters and initial infrastructure for a global-scale science Grid involves, among other things, addressing the issues identified in the preceding section.

The following work areas are all part of deploying a prototype large-scale Grid.

### **1.3.2.1 Grid Information Services (GIS)**

The GIS is a fundamental Grid service, providing the basic capability to locate and characterize resources. The GIS is also one of the two areas in which the project will collaborate closely with ESnet in establishing the design for a global Grid.

The GIS must be configured in a global namespace that will be consistent with a higher-level name construct. In fact, from the global Grid point of view, the DOE Science Grid is a virtual organization that exists in a consistent namespace.

The information about the resources that are the leaf nodes of the GIS must be managed in order to ensure accuracy. There must also be agreement about how new information objects (e.g., accounting or allocation management information) will be introduced, and that this is done in a way consistent with the recommendations and draft standards coming out of the Global Grid Forum's Grid Information Service Working Group.

The architecture and deployment of GISs must establish a structure within an organization that provides both appropriate administrative scoping and performance.

Once GISs have been established, then the information pertaining to them must be propagated to all of the other Science Grid participants in order to establish an N-way federation of sites and/or facilities. As work on the Global GIS progresses, and such services are established, the Science Grid GISs will register with higher-level nodes (e.g., a DOE Science Grid node) in order to obviate the need for N-way site configuration. Global GIS issues that will be addressed include flexible name spaces (that include, e.g., aliasing), user defined branches of the name space, use of commercial metadirectories for query management and results caching, etc.

We will work closely with SciDAC and other DOE projects to determine the types of information required in the Grid Information Service and to ensure that the information delivered meets applications requirements for accuracy and timeliness. We will also participate in the GIS working group of the Global Grid Forum.

### **1.3.2.2 Certification Authority**

CAs issue X.509 identity certificates for all Grid users. These certificates are the basis of the Grid single sign-on and for global auditing (e.g., for allocation management). The "global" nature of the identity has two aspects. One is entity naming so that there will not be namespace collisions, and the other is policy agreement on how strong a binding there is between the name in the certificate and the individual to whom the certificate was issued, and on what uses are permitted for the certificates. A third issue for globalizing the use of certificates is a trusted and reliable certificate revocation mechanism. This is another service that will be designed in partnership with ESnet.

Administrative scoping is not as important for CAs as with the GIS. However, if identity establishment is tied to corporate personnel records (as is frequently the case) and/or if a physical presence is required to establish identity, then having a CA at each site is important. It is possible for an off-site CA to issue certificates, and this will be important for accommodating non-Lab (e.g., university) collaborative participants. Policy and procedures for remote certification will be established working with ESnet.

The Science Grid will initially be established with an N-way configuration of CAs, with the goal of moving toward a global approach: a root CA (at ESnet) signing site CA certificates and incorporating certificate chain evaluation into the Grid authentication software.

Again, we will work closely with SciDAC and other DOE projects to determine application requirements for authentication and authorization, and to ensure that CA services are sufficiently easy to use. We will also work closely with the SciDAC Collaboratory project "Security and Policy for Group Collaboration" and in the Security working group of the Global Grid Forum. We also anticipate developing a Grid-wide repository of Certificate Revocation Lists, or pointers to those lists, for Science Grid CAs.

### **1.3.2.3 Deploy Globus**

The Globus Toolkit will provide the basic, initial Science Grid services. In particular, Globus provides for querying the GIS, a uniform job submission language, co-scheduling mechanism (when the underlying resources support this), security services and utilities, Grid user management on end systems, and a uniform data access mechanism.

The Globus software must be configured and tested on each of the computing platforms that participate in the Science Grid. Many of the configuration issues come from the GIS and CA configurations noted above. The software must also be of compatible versions in areas like the security services components. As noted below, the Science Grid may establish its own Globus distribution that incorporates the required configuration, and as the Globus software is modularized, different configurations will be appropriate for different types of Science Grid participant. For example, we expect to develop specialized distributions appropriate for resource sites, instrument sites, and data providers, and for different application classes. We also plan to develop Science Grid configuration validation test suites; and tools for both static and dynamic reliability and sensitivity analysis for the Science Grid.

### **1.3.2.4 Grid Tertiary Storage**

Access to tertiary storage is a critically important for laboratories. Many, if not most, laboratories are data driven at some point in their workflow, either from standard modeling input data, from experiment data collection, or from data that represents knowledge accumulation that must be shared among collaborators.

There are many aspects of managing data in Grids—see, for example, the GriPhyN ([www.griphyn.org](http://www.griphyn.org)) and PPDG ([www.ppdg.net](http://www.ppdg.net)) projects. However, in the Science Grid we will focus on providing a uniform, versatile, high performance, and adaptable, network interface to tertiary storage. The approach is called GridFTP, and this service will be installed at tertiary storage sites, Science Grid platforms, and on end-user/application/client systems where data otherwise originates or is stored.

The initial tertiary storage systems will be the PROBE testbed component of NERSC's HPSS storage system and an EMASS based system at PNNL. In both cases, GridFTP needs to be installed, Grid user accounts established, and high-speed network connectivity established to the storage system platforms. A particular focus of work in this area will be to understand how to balance user requirements for flexible, high-performance access with resource provider requirements for security and ease of operation.

### **1.3.2.5 Security Infrastructure**

There are several aspects to Grid security. The Science Grid, like many laboratories, will be a relatively “open” environment in the sense that many pieces of the Grid infrastructure will reside on systems directly connected to the global Internet. This means that there must be a model and services that provide secure access (authenticated and authorized) to Grid resources, there must be mechanisms for protecting user data at an appropriate level of confidentiality, and there must be a codified model that can explain to site security personnel how the Grid will interact with local security policy.

Once the security model is established, then the requirements of that model must be implemented (e.g., in system configurations and software design), and monitoring and enforcement mechanisms must be defined and deployed in order to ensure that the model is adhered to.

An important issue for building the Science Grid will be dealing with site firewalls. The goal is to have a sufficiently well developed and enforced security model that the ports needed by Grid services will be made available. If this is not possible—especially initially—we will develop a secure and authenticated Grid forwarding service that uses, e.g., the ssh tunneling mechanism. This will be an aspect of the Grid security R&D.

### **1.3.2.6 System Services/Science Grid Issues**

Though the philosophy of the Grid software is to keep participating systems as independent as possible, there are some “community” issues that will have to be addressed so that users to perceive the DOE Science Grid as a “system.” For example, there are issues in tracking problems through a collection of machines that may be scattered all over the country, and beyond.

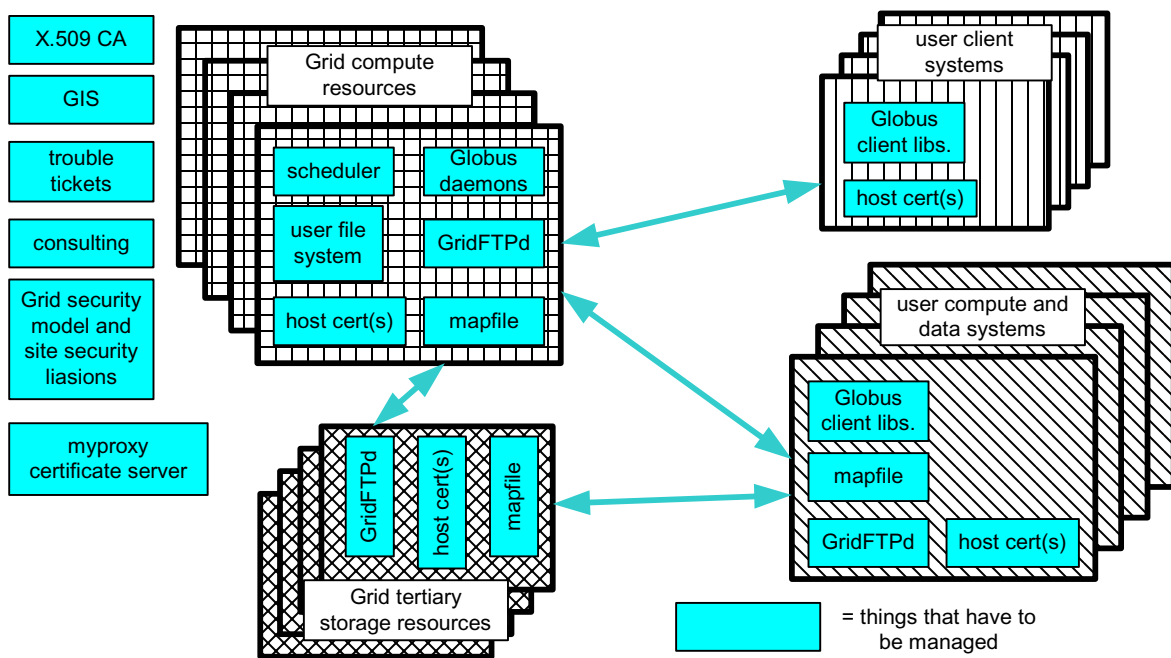
A DOE Science Grid Deployment Working Group will be established and meet weekly. This group will consist of the people who install and configure Grid software. This is where cross-site issues will be identified and resolved, and Grid system administration information exchanged. Another task of this group will be to identify the Grid system administration techniques and tools that are needed to keep the Grid functioning as a system.

Network connectivity is generally not a problem in the ESnet community, but university and other collaboratory partners may need assistance with high-speed connectivity. The Grid is highly network communication oriented, so reliable and high-speed connectivity is essential.

Resource owner allocation agreements will have to be worked out so that the Science Grid will become a long-term resource sharing mechanism. The technology for this is addressed in section 1.3.3.

The security model implementation will be dealt with by the deployment WG, especially where site firewalls are involved that need to allow Grid services communication.

As noted above, configuration management of user systems that are included in the Grid is an open issue, and appropriate models and techniques must be developed. (E.g., as in Figure 3.)



**Figure 3. Grid and user system configuration model elements**

Finally, a critical issue will be evaluating and refining Grid resources for use in DOE's production supercomputing environment. This will entail addressing, for example:

- o What is the process to decide that Grid software is near enough to production, including some level of security analysis, that it is ready to go on a production system?
- o NERSC has promised levels of service and utilization and have developed a lot of code to get the job flow to run well on the systems. If the Grid software disrupts the ability to fully schedule the system, we will impact the size of the allocation provided to DOE researchers. NERSC will need to validate that the Grid software at least "does no harm" to a fully loaded system.

- o How will Grid software be supported and what is the service commitment by the producers of the software if something is not working correctly.

These issues will be addressed through NERSC participation in the Science Grid.

### **1.3.3 R&D Tasks: Extending the Grid Technology Base**

#### **1.3.3.1 Monitoring and Auditing Infrastructure**

Use of high-performance applications across the DOE Science Grid will require access to accurate, up-to-date information on the structure and state of available resources. For example, are all the resources required by the application available at this moment? Are the necessary network QoS reservations in place? Are the resources free for the entire period of the run or are there conflicts downstream? The DOE Science Grid information services and scheduling tools will be available to SciDAC researchers for determining the best configuration based on the information available at the start of a job. *However, the present Grid infrastructure is missing the monitors and tools to dynamically assess its resources and respond.* The resources across the DOE Science Grid are constantly changing—nodes fail in a compute resource, a drive fails in a storage archive, the power goes out in California, etc. Wide-spread acceptance of the DOE Science Grid as a viable computational resource requires the creation and deployment of appropriate monitoring and auditing tools.

We will develop, deploy, and operate the DOE Science Grid-wide monitoring and auditing infrastructure required to identify “faults” (whether failures or performance bottlenecks) in various resources (networks, computing, storage). By logging selective information from these monitors we will also be able to account for resource usage by SciDAC users. The third part of this task is to develop easy to use tools so users can access their allocation status and monitor their DOE Science Grid applications.

The first step in this task is to examine the mechanisms for fault detection and auditing inside the DOE computer centers today, because this is what computational scientists are accustomed to, then deploy an infrastructure that is at least (if not more) informative across the Science Grid. While allocation and auditing tools are fairly mature in the computer centers, they tend to be site specific—created in an ad hoc fashion over decades. Fault detection inside computer centers is almost non-existent.

Allocation is the money of computational scientists. All production computer centers distribute allocations to given groups or individuals. In contrast most (all) Grid testbeds installed to date do not consider allocation in their usage models. For the Science Grid to transition into a production environment in year three of this project, allocation must be incorporated into the authentication and authorization models. During the first two years we will evaluate options and develop the monitoring and auditing tools required to enforce the different allocation schemes deployed at NERC, ORNL, ANL, PNNL, and LBL. Issues we will address include evaluation of available resources based on a job request. Resources that have insufficient allocation remaining will be eliminated from consideration.

One option that will be evaluated is to develop interfaces to local allocation and user databases at each site so that usage across the Science Grid can be monitored directly and authorization can be determined by a query to the local site. Another option is to install monitor daemons at each site that periodically update the central directory services. In this case, authorization would be determined by a query to the central directory services. There are security and social issues that may prevent sites from sharing all their information with a central site. Such considerations will be part of the evaluation. We will also follow the work of the Grid forum Account Management Working Group and be compliant with their recommendations. The SciDAC Scalable Systems Software ETC proposal could be another opportunity for collaboration. This ETC is focused on the software inside a large computer center, which includes the local accounting components.

The Science Grid presents some interesting challenges in allocation management. Will users be allowed to trade allocations? Who would determine the exchange rate? If future allocations are given at the Science Grid level rather than the local sites, how will the relative worth of the resources be determined? These are policy decisions and not a part of this proposal (although we may consider the technical feasibilities during the third year). Our approach will be to supply tools that allow SciDAC scientists to manage their allocations on any individual Science Grid resources as easily as if they were at a local site.

### **1.3.3.2 Fault Tolerance**

Fault tolerance is another area where existing Grid technology needs to be improved in order to be accepted in production environments. Sites do not want to take on the liability that a failure at another site may cause failures at their own site. Very straightforward behavior by the Grid environment could lead to unexpected results. For example, a large site in Science Grid drops off line due to a network fault. This resource becomes unavailable and jobs start getting directed to the remaining sites. These sites begin to see data pouring into their systems for staging, their network bogs down, and the average number of jobs doubles. Resources become strained; some start to fail. Managing failure is key part of Science Grid infrastructure. And failure does not have to be catastrophic. If a job has given a requirement of 100 MB/s minimum bandwidth between the data archive site and the computing site, what happens if the measured rate drops to 80 MB/s half way through the job, or to 20 MB/s?

The first step in fault tolerance is monitoring. In the first part of this task we monitor allocation and usage. For fault detection we will be monitoring the network (bandwidth, latency, and Science Grid usage), storage (available GB for staging), and computers (availability, load, node failure). There are already monitors that measure network bandwidth and latency, but we will also factor in how these metrics are being impacted by use of the Science Grid. For example, we will try to determine what percentage of the observed bandwidth is being consumed by Science Grid jobs versus other jobs.

Our initial approach will be to detect faults or dips below requested resources and report these problems to the user and the Science Grid log. (E.g., see [17].) This monitoring information will also be fed back into the Grid scheduling components.

Once we have experience detecting problems, we will explore approaches to automatically adapt to faults by restarting job(s) on other resources or rescheduling on the same resources. We would like to ensure that if a user submits a job through the Science Grid that it will be run to completion. More dynamic fault detection and automatic recovery will be investigated in the out years if additional funds are available.

### **1.3.3.3 Grid Security**

Grid security has many characteristics in common with system security, but also has significant differences. Grids are built from resources at multiple sites that are primarily dedicated to the Science Grid, together with a loose federation of resources at other sites that are incorporated into the Science Grid by virtue of these sites instantiating Grid services and (possibly) providing reciprocal user allocations. As such, it is expected that there will be different security approaches at the various sites that participate in the Science Grid.

When developing a security model one has to take into account the assets to be protected, the environment in which those assets exist, the vulnerabilities and associated risks, and the consequences of compromise of the assets operating in Grid environments, and finally, the adversaries and their motivation. Threats arise from motivated adversaries plus vulnerabilities, consequences arise from actualized risks together with the value and/or importance of the assets and scope/extent of the systems being attacked.

The assets associated with Grids are:

- o Grid resource use and/or access is valuable: computing systems, data management and mass storage systems, scientific / engineering instruments, collaboration services, and communications systems
- o Intellectual property is potentially valuable and/or proprietary (due to scientific, commercial, and national interests): source code and data, collaborative interaction, the knowledge that is built up in multi-disciplinary, problem solving frameworks
- o Cross-site trust is valuable: access to resources that can be shared is an important Grid capability

Where Grids differ from single or enterprise systems is in their environment – with its mix of users and institutions - and the fact that large multi-disciplinary collaborations may represent a more tempting target to some adversaries. This is because of the high level of information and knowledge that can be represented by the structured frameworks that tie together a lot disparate resources to solve large-scale problems.

The main sources of vulnerability in Grids are the underlying systems, Grid services and applications, communications infrastructure, users, and the diverse security models for local systems and their servers.

These assets and vulnerabilities, together with the access authorization needs of some user communities, leads to a set high level security requirements that will be further explored and refined over the course of this project:

- o Cyber risk mitigation and cross-site integrity
- o Infrastructure assurance
- o Application control channel integrity and confidentiality
- o Data integrity and optionally confidentiality: in transit, in middleware, and in storage
- o Identity management, authentication, and single identity sign-on without clear text passwords
- o Authorization via policy-based access control for individual, group, and role
- o Non-repudiation

The scope of the Grid security R&D work will be to refine this analysis, to clarify the requirements, and to develop a security model that addresses the requirements. Once a model is derived, the existing Grid and system security tools and approaches will be evaluated against the model and refined and augmented as necessary.

One specific area that will be addressed will be the design of a Grid firewall proxy service that is capable of proxying all Grid services through a single port. We believe that this is feasible, but will likely to require some changes in the Grid information service to distinguish local and remote resources. Such a service is expected to be quite important at sites where the Grid users may have little or no influence on the site boundary security policy.

### 1.3.4 User Support Services

The Science Grid brings powerful new tools for distributed computing which can be built into applications and problem-solving environments. However, Grid services are a relatively new capability, largely unfamiliar to application, collaboratory and tool developers – a new approach that needs to be learned and a new infrastructure that must be supported. Therefore user support is a key element of the proposed work. There are four aspects; each is a pilot scale effort, which will support the initial users and develop scalable support approaches.

- o User Education
- o Applications Integration Support
- o Helpdesk Support
- o User Feedback

*User Education:* We will provide tutorials on using Grid capabilities targeted at SciDAC application, tool and collaboratory developers (including best practices described in 1.3.5). Two venues will be used: (a) in depth tutorials at regular intervals for any interested SciDAC participant, with locations rotating among the four pilot project sites, and (b) shorter tutorials at scientific community meetings held by SciDAC projects and through the Access Grid [51]. The objective is to equip SciDAC developers with the knowledge they need to develop Grid applications. Periodically, we will also hold tutorials for system administrators charged with installing and locally supporting Grid software.

*Applications Integration Support:* In addition to user education, it is important to follow through with help on specific applications problems. To this end, we will provide staff time to directly assist SciDAC developers. The intent is to assign a single point of contact for a particular SciDAC team, who will become familiar with the objectives of the team and provide specific advice on Grid services usage. At this time, it is not possible to know how many teams the SciDAC Program will fund. However, the adoption of Grid technologies is likely to be spread out in time, requiring only a modest level of effort to provide integration support.

*Helpdesk Support:* Troubleshooting services are vital to providing the reliability essential to SciDAC applications. The aims of the proposed efforts in this area are (a) to develop first level troubleshooting capabilities at the pilot Grid sites, and (b) to deploy an integrated trouble ticket system that enables the DOE Science Grid project team to review and respond to any second level problems. We will use EHSQ, a state-of-the-art trouble ticket system developed by PNNL. EHSQ is a production web-accessible problem reporting and tracking tool currently used by 18 software and hardware support teams at PNNL.

*User Feedback:* User support services not only provide assistance to those employing Grid capabilities, they also provide a conduit for essential feedback to the developers of Grid technologies. Developer and user feedback will be captured and analyzed via feedback surveys at tutorials, comments captured from applications developers, and information on trouble tickets, respectively. In addition, we will conduct periodic web surveys of Grid applications/collaboratory/tool developers and users.

Taken together, we believe that these user services provide a complete package, supporting outreach and use of the Science Grid, and that they also capture the essential feedback needed to improve and extend Grid services and operations, making them more useful for the SciDAC program. These suppositions will be tested, evaluated, and refined in the course of this project. The end result will be a scalable operations and support plan for wide deployment of the DOE Science Grid in DOE.

### **1.3.5 Standards Advocacy and Enablement**

An important role for the DOE Science Grid project is to act as an advocate and enabler of “standards” and hence code sharing, infrastructure sharing, and interoperability across SciDAC and other DOE and non-DOE projects. The SciDAC call notes that:

*“Integration of work efforts across all projects funded under this notice will occur following the awards, to preclude duplication of effort and to maximize leveraging and coordination. Projects are expected to work closely with other SciDAC teams, where identified during this integration. Coordination through a participatory management process will continue for the life of the projects.”*

We believe that the DOE Science Grid project team is well positioned to facilitate this coordination and participatory management process, due to its focus on the delivery of common mechanisms for key collaborative problems, its distributed, multi-laboratory structure, and its strong ties with other DOE and non-DOE projects. To maximize our impact in this area, we propose the following specific activities:

- o Definition, documentation, and community advocacy of “best common practices” for key collaborative problems addressed by Science Grid-CSE technologies, such as authentication, resource discovery, and resource access. Our experience to date shows that collaborative developers are typically only too happy to adopt such technologies, if someone will only take the time to explain how they are used and if they are broadly deployed.
- o Convene regular DOE Science Grid User Meetings that will bring together Science Grid-CSE developers, Science Grid-CSE users, and resource providers to discuss deployment issues, future directions, and so forth.
- o Bring forward at GGF and IETF relevant standards relevant to DOE Science Grid operation.

### **1.3.6 Use of Collaboratory Technologies**

Project management will make extensive use of collaborative technologies, as well as face-to-face meetings. Important elements of the DOE Science Grid “Collaboratory” will be:

- o Secure shared repositories for code and documentation.
- o Extensive use of archived email lists for Science Grid-CSE developers, resource providers, and users.
- o Use of Access Grid technologies for multi-site coordination meetings (M-to-M interactions), as well as for user tutorials (1-to-N interactions).
- o Use of real time collaboration tools (e.g., video and shared screens) for problem resolution.

### **1.3.7 Tasks and Milestones**

The Gantt chart on the next page summarizes the tasks and milestones that we anticipate for the project.

ID	WBS	Task Name	4Q01			1Q02			2Q02			3Q02			4Q02			1Q03			2Q03			3Q03			4Q03			1Q04			2Q04			3Q04												
			J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M
1	a	<b>Application / User Milestones</b>																																														
2	a.1	project begin																																														
3	a.2	Grid jobs on local systems at LBNL, ANL, ORNL, PNNL																																														
4	a.3	Grid jobs on NERSC non-prod. Resources																																														
5	a.4	Grid tertiary storage access at NERSC																																														
6	a.5	Grid tertiary storage access at PNNL																																														
7	a.6	Jobs across the federated Science Grid																																														
8	a.7	Jobs across the scalable Science Grid																																														
9	a.8	Proto user: Cosmology supernova search																																														
10	a.9	Proto user: Regional Air Quality																																														
11	a.10	Collaboratory: Earth Sciences Grid																																														
12	a.11	Collaboratory: STAR																																														
13	a.12	Build virtual orgs.																																														
14	a.13	Fault tolerant job execution																																														
15	a.14	project end																																														
16																																																
17	b	<b>Deployment and Development</b>																																														
18	b.1	<b>Grid Information Services</b>																																														
19	b.1.1	GIS installed at each Lab																																														
20	b.1.2	GIS installed at NERSC																																														
21	b.1.3	GIS cross Lab federation																																														
22	b.1.4	Sci Grid integration with ESNet global GIS (LBNL/ESNet)																																														
23	b.2	<b>Certification Authority</b>																																														
24	b.2.1	CAs installed at each Lab																																														
25	b.2.2	CA installed at NERSC or NERSC uses ESNet																																														
26	b.2.3	inter-Lab X.509 certificate policy																																														
27	b.2.4	scalable Grid CA arch. (LBNL/ESNet)																																														
28	b.2.5	CA: cross Lab federation																																														
29	b.2.6	integration with ESNet root CA (LBNL/ESNet)																																														
30	b.3	<b>Deploy Globus</b>																																														
31	b.3.1	Globus installed at each Lab																																														
32	b.3.2	Globus installed at NERSC																																														
33	b.3.3	Globus cross Lab operation																																														
34	b.3.4	Globus in a scalable Science Grid (all)																																														
35	b.3.1	<b>Incorporate computing resources</b>																																														
36	b.3.1.1	Sun SMPs at LBNL																																														
37	b.3.1.2	Aix at PNNL																																														
38	b.3.1.3	Linux Cluster at PNNL																																														
39	b.3.1.4	XX at ORNL																																														
40	b.3.1.5	small IPM SP at NERSC																																														
41	b.3.1.6	Linux cluster at NERSC																																														
42	b.4	<b>Grid Tertiary Storage</b>																																														
43	b.4.1	Probe/HPSS@NERSC + GridFTP																																														
44	b.4.2	NERSC Probe configured for data replication																																														
45	b.4.3	EMASS@PNNL + GridFTP																																														
46	b.5	<b>Security Infrastructure</b>																																														
47	b.5.1	develop Grid security model (LBNL)																																														
48	b.5.2	design and implement security monitoring																																														
49	b.5.3	ongoing security monitoring																																														
50	b.5.4	implement Grid security model (all)																																														
51	b.6	<b>Auditing and Fault Monitoring (ORNL)</b>																																														
52	b.6.1	job monitoring techniques																																														
53	b.6.2	resource utilization auditing techniques																																														
54	b.6.3	deploy on Sci Grid resources																																														
55	b.6.4	resource usage data mgmt and accounting																																														
56	b.6.5	allocation mgmt																																														
57	b.6.6	fault notification and response																																														
58	b.7	<b>User Services (PNNL)</b>																																														
59	b.7.1	develop user system config. model																																														
60	b.7.2	deploy Sci Grid install pkg. for config. model																																														
61	b.7.3	Application Integration Support																																														
62	b.7.4	<b>Helpdesk support</b>																																														
63	b.7.4.1	first level troubleshooting capability																																														
64	b.7.4.2	deploy integrated trouble ticket system																																														
65	b.7.4.3	User Education																																														
66	b.8	<b>System services /Science Grid issues (all)</b>																																														
67	b.8.1	Science Grid deployment Working Group																																														
68	b.8.2	resource owner allocation agreements																																														
69	b.8.3	Grid sys. admin techniques and tools																																														
70	b.9	<b>Applications</b>																																														
71	b.9.1	Cosmology job mgmt integrated																																														
72	b.9.2	Regional air quality job mgmt integrated																																														
73																																																
74	c	<b>Integrate R&amp;D from other projects</b>																																														
75	c.1	CoG tools and Web interface																																														
76	c.2	Data replica management																																														
77	c.3	STACS tertiary storage management																																														
78	c.4	Cache management																																														
79	c.5	Workflow management																																														
80	c.6	Grid Mon. Arch. Network monitoring																																														
81	c.7	CPU Resource reservation																																														
82	c.8	Certificate based access control																																														
83	c.9	Restricted delegation																																														

### 1.3.8 Technology Transfer and Application

A primary goal of the Science Grid is to provide persistent services to application projects that have committed to using these technologies, such as the Particle Physics Data Grid, Earth Systems Grid, and Center for Collaborative Problem Solving, and enable new application projects to build on these technologies. We build on a solid base of experience within NSF and NASA projects in which Foster and Johnston have been involved, with Johnston in particular serving as Project Manager for the NASA Information Power Grid project, and on large-scale DOE projects at all four participating laboratories led by Bair, Foster, Geist, and Johnston.

We plan to work directly with a number of SciDAC Collaboratory Pilots and application projects to both obtain additional input on requirements and enable large-scale evaluation in practical settings. The Co-investigators already have close collaborations with each of these projects. This will further help to ensure technology transfer.

We will continue to work within the Global Grid Forum, in which Foster and Johnston play leadership roles, with the goal of aligning our work with that of other agencies and industrial groups.

We anticipate that this work will be institutionalized in several different ways.

First, the explicit involvement of NERSC is intended to eventually lead to incorporating the NERSC production facilities into the DOE Science Grid. This should have several benefits:

- o Users will more easily move between local/other computing facilities and NERSC.
- o It should be easier to incorporate NERSC systems into workflow frameworks, where only part of the work is performed on supercomputers with, e.g., the rest being involved with data collection and processing on other systems. The fact that this cannot currently be easily done is already proving to be a significant obstacle to projects in climate and chemistry.

Cross-center experiments with the other national supercomputer centers will become possible as many of the other centers are also deploying Grid software. Such experiments will be performed as part of this project. The investigators have close ties with both NSF PACI Centers and with the Pittsburgh Supercomputer Center.

Second, we anticipate that mid-range computing centers of various sorts – Lab and/or department computing systems, topical centers, etc., will find the Grid a convenient way to provide a common access mechanism that allows for the possibility of sharing and load leveling across such facilities. The leadership provided by NERSC will encourage and facilitate deployment at these sites.

Third, it is already clear that several data intensive collaborations including, e.g., high energy physics, will be adopting Grid technology to manage their data and collaborative resources. The DOE Science Grid's work with ESnet is intended to define and set up operational Grid Information System components and X.509 Certification Authority components that will support global scale collaborations.

In order to further our overall goal of achieving broad adoption of open standards/open source-based Grid infrastructure, we are investigating the feasibility of establishing a **Consortium for Open Grid Software (COGS)**, a body dedicated to the creation, documentation, and distribution of high-quality open source Grid software. Details are not finalized but we will continue to pursue this goal within the context of the DOE Science Grid.

### 1.3.9 Connections

The proposed DOE Science Grid Collaboratory Pilot leverages and supports a collection of projects and proposals that will together develop, deploy, apply, and evaluate the DOE-CSE. Several collaboratory pilots are very interested in using the Science Grid infrastructure. Additionally, a collection of Grid R&D projects will utilize, enhance, and evolve the Science Grid.

#### 1.3.9.1 DOE Collaboratory Pilots and Other Application Projects

The following is a *partial list, intended to be illustrative rather than complete*, of Collaboratory Pilots and other DOE application projects that have expressed strong support for the goals of the DOE Science Grid project and a strong interest in working with us to establish and use the DOE Science Grid.

The **Particle Physics Data Grid** (PPDG) project is applying Grid technologies to enable the distributed management and analysis of extremely large data sets produced by major physics experiments.

The **Earth System Grid** (ESG) project is applying Grid technologies to enable the distributed access to and analysis of extremely large data sets produced by climate simulations.

The **Center for Collaborative Problem Solving in the Earth Sciences Community** focuses frameworks and workflow for collaborative problem solving, with application to the climate modeling and assessment arena.

The proposed **National Collaboratory to Advance the Science of High Temperature Plasma Physics for Magnetic Fusion** will exploit Grid technologies as it creates a collaboratory for fusion simulation.

The **Supernova Search Factory**, observational cosmology project will use Grid services, and will use the Science Grid if it is funded.

### **1.3.9.2 DOE Collaboratory Middleware and Networking Technology Projects**

A number of DOE Collaboratory and Networking projects and proposals will inject technologies into the DOE Science Grid effort, and benefit from the core services and deployment activities of the DOE Science Grid project. We have referred to many of these in the text of the proposal.

### **1.3.9.3 DOE SciDAC Enabling Technology Centers**

The work of the DOE Science Grid project also complements work in several proposed SciDAC Enabling Technology Centers, including the following.

The proposed **Scalable Data Management ETC** will provide high-speed network access to high-performance storage systems, and looks to the Science Grid-CSE effort for integration of these storage systems into a wide area collaboratory setting.

The **Visualization and Data Understanding ETC** will provide high-speed network access to high-performance visualization systems, and expects to leverage Science Grid-CSE services when it integrates these systems into a wide area collaboratory setting.

The **Common Component Architecture ETC** will provide sophisticated component-based mechanisms for development of advanced scientific applications. This project will look to the Science Grid-CSE for the tools required to enable composition of geographically distributed components.

The **Scalable Systems Software ETC** will provide an integrated suite of systems software for the effective management and utilization of terascale computational resources. Meta-computing component interfaces will be defined that enable them to link to the Science Grid.

### **1.3.9.4 Other Grid Efforts**

The DOE Science Grid project will work closely with, and complement the activities of, other Grid activities within the U.S. and worldwide, with a view to creating interoperable Grid infrastructures in support of multi-institutional and multi-agency collaboratory projects. In particular, we note the NSF's National Technology Grid, NASA's Information Power Grid, the EU's European Data Grid, and ASCI's DISCOM projects as activities with which we already have close ties and with whom we plan to coordinate closely. This coordination will be achieved via a combination of personal contacts and the Global Grid Forum.

## **1.3.10 Evaluation Criteria**

We address specifically the special evaluation criteria described in the call.

*Potential to make a significant impact in the effectiveness of SciDAC applications researchers.* We address this issue in Sections 1.1.3, 1.3.1, 1.3.2, 1.3.4, and 1.3.9, and point also to the strong letters of support from various application groups.

*The degree to which an application area can benefit from collaborative technology.* This topic is addressed in Section 1.3.6. While our primary goal is creation of collaborative technologies, we also expect to make considerable use of collaborative technologies in the project.

*The extent to which the project will test important collaborative technologies.* As explained in Section 1.1.3, a major goal of this project is to take existing collaborative technologies such as Globus and deploy them on a large scale and in a DOE-oriented production setting. In so doing, the project will certainly test these technologies in new ways and expose deficiencies that will motivate improvements and further research and development.

*Extent to which the results of the project are extensible to other program or discipline areas.* The entire goal of this project is to deploy and make available general-purpose technologies of relevance to a wide range of program and discipline areas. The letters of support indicate the breadth of relevance: see also Section 1.3.8.

*Degree to which the project adheres to the management philosophy of incorporating collaboration into the project execution.* As we explain in Sections 1.3.6 and **Error! Reference source not found.**, the DOE Science Grid project is highly distributed and collaborative in nature, and its management structure and mechanisms are designed with this in mind. The use of a management council, and an Engineering Working group that holds weekly telecons to discuss the issues and state of the deployment has proven effective in other multi-institutional projects.

*The quality of the plan for ensuring interoperability and integration with software produced by other SciDAC efforts.* As we explain in Section 1.3.5, we believe that the DOE Science Grid project will have a significant positive impact on interoperability and integration within and across SciDAC projects.

*The extent to which the project incorporates broad community (industry/academia/other federal programs) interaction.* As we explain in Section 1.3.8, we have strong connections with other federal and international Grid efforts, via collaborations that we have established and nurtured over several years. In addition, the Global Grid Forum that Foster and Johnston co-founded provides a wonderful vehicle for community interaction. Industrial connections are growing, with strong interest from Sun, IBM, Microsoft, and others, although this interest has not yet translated into direct support for our efforts.

*Quality and clarity of proposed work schedule and deliverables.* Section 1.3.7 provides a detailed work schedule and set of deliverables.

*Knowledge of and coupling to previous efforts for collaborative technologies such as DOE 2000* The LBNL Distributed Security Research Group grew out of, and participated in several DOE 2000 projects. Several of the people who worked on those projects will be working on the Science Grid, and the knowledge and expertise gained in the DOE 2000 security projects will be applied to the Science Grid. DOE2000 has also supported early work on DOE applications of the Globus Toolkit/.

*Relevance of the proposed research to the terms of the announcement.* We believe this is clear: we will develop and deploy collaborative middleware services of direct relevance to a range of DOE-relevant applications, and in so doing with advance our knowledge of our how to construct and operate these services.

*Uniqueness of the proposer's capabilities.* We do not believe that any other team could come anywhere close to fielding a comparable combination of expertise in Grid technologies, collaborative applications, and creation of Grid infrastructures. This is indicated in the Preliminary Studies section (1.2) and in the bios of the investigators.

*Demonstrated usefulness of the research for proposals in other DOE Program Offices as evidenced by a history of programmatic support directly related to the proposed work.* OBER and HENP have provided support for the Earth System Grid and Particle Physics Data Grid projects, and continue to express strong support for further development of Grid technologies.

## 2 Literature Cited

---

- [1] *The Grid: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, eds. 1998, Morgan Kaufmann. [http://www.mkp.com/books\\_catalog/1-55860-475-8.asp](http://www.mkp.com/books_catalog/1-55860-475-8.asp)
- [2] *National Collaboratories - Applying Information Technology for Scientific Research*, Committee on a National Collaboratory - National Research Council. 1993, Washington, D. C: National Academy Press. <http://books.nap.edu/catalog/2109.html>
- [3] "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", I. Foster, C. Kesselman and S. Tuecke. Intl. J. Supercomputer Applications, 2001. **(to appear)**. <http://www.globus.org/research/papers/anatomy.pdf>
- [4] The NSF PACIs are the Alliance/NCSA (<http://www.ncsa.uiuc.edu/>) and NPACI/SDSC (<http://www.npaci.edu/>) "PACI," PACI.
- [5] "NASA's Information Power Grid," IPG. <http://www.ipg.nasa.gov>
- [6] DataGrid. 2001. <http://www.cern.ch/grid>
- [7] NEESgrid is a virtual laboratory for the earthquake engineering community. NEESgrid is funded by the NSF NEES project to develop a system design for integrating experimental and computational facilities for use by the earthquake engineering community. NEESgrid. <http://www.neesgrid.org/>
- [8] "Distance Computing and Distributed Computing (DisCom2) Program," DISCOM. <http://www.cs.sandia.gov/discom>
- [9] The Global Grid Forum ([www.Gridforum.org](http://www.Gridforum.org)) is an informal consortium of institutions and individuals working on wide area computing and computational Grids: the technologies that underlie such activities as the NCSA Alliance's National Technology Grid, NPACI's Metasystems efforts, NASA's Information Power Grid, DOE ASCI's DISCOM program, and other activities worldwide. Grid Forum.
- [10] "Real-Time Generation and Cataloguing of Large Data-Objects in Widely Distributed Environments", W. Johnston, G. Jin, C. Larsen, J. Lee, G. Hoo, M. Thompson, B. Tierney and J. Terdiman. International Journal of Digital Libraries - Special Issue on Digital Libraries in Medicine, 1998. <http://www-itg.lbl.gov/~johnston/papers.html>
- [11] "Visual Servoing for Online Facilities", B. Parvin, J. Taylor, D. E. Callahan, W. Johnston and U. Dahmen. IEEE Computer. <http://www-itg.lbl.gov/~johnston/papers.html>
- [12] "High-Speed Distributed Data Handling for On-Line Instrumentation Systems," W. Johnston, W. Greiman, G. Hoo, J. Lee, B. Tierney, C. Tull and D. Olson. In *ACM/IEEE SC97: High Performance Networking and Computing*. 1997. <http://www-itg.lbl.gov/~johnston/papers.html>
- [13] "Collaboratories: Building Electronic Scientific Communities," R. Bair, in *Impact of Advances in Computing and Communications Technologies on Chemical Science and Technology*. 1999, National Academy Press.
- [14] "Development and Use of a Virtual NMR Facility", K. A. Keating, J. D. Myers, R. A. Bair and P. D. Ellis. Journal of Magnetic Resonance, 2000. **143**: p. 172-183.
- [15] "Condor," M. Livny and e. al. <http://www.cs.wisc.edu/condor>
- [16] "QoS as middleware: Bandwidth broker system design", G. Hoo, W. Johnston, I. Foster and A. Roy. 1999.
- [17] "A Quality of Service Architecture that Combines Resource Reservation and Application Adaptation," I. Foster, A. Roy and V. Sander. In *Proc. 8th International Workshop on Quality of Service*. 2000.
- [18] "Globus: A Metacomputing Infrastructure Toolkit", I. Foster and C. Kesselman. Intl J. Supercomputing Applications, 1997. <http://www.globus.org/research/papers.html>
- [19] "Certificate-based Access Control for Widely Distributed Resources," M. Thompson, W. Johnston, S. Mudumbai, G. Hoo, K. Jackson and A. Essiari. In *Eighth Usenix Security Symposium*. 1999. <http://www-itg.lbl.gov/Akenti/papers.html>

- [20] "Generic Authorization and Access control API (GAA API)," GAA. [http://www.isi.edu/gost/info/gaa\\_api.html](http://www.isi.edu/gost/info/gaa_api.html)
- [21] "Distributed Security Architecture," M. Thompson. Submitted SciDAC proposal, March, 2001, Lawrence Berkeley National Laboratory.
- [22] "A Directory Service for Configuring High-Performance Distributed Computations," S. Fitzgerald, I. Foster, C. Kesselman, G. v. Laszewski, W. Smith and S. Tuecke. In *6th IEEE Symp. on High-Performance Distributed Computing*. 1997. <http://www.globus.org/documentation/papers.html> An updated view of the Grid Information Service is in "Globus Grid Tutorial Part 2: Information Services," at <http://www.globus.org/tutorial>
- [23] "Forecasting Network Performance to Support Dynamic Scheduling Using the Network Weather Service," R. Wolski, in *Proc. 6th IEEE Symp. on High Performance Distributed Computing*. 1997: Portland, Oregon.
- [24] "A Resource Management Architecture for Metacomputing Systems," K. Czajkowski, I. Foster, N. Karonis, C. Kesselman, S. Martin, W. Smith and S. Tuecke, in *The 4th Workshop on Job Scheduling Strategies for Parallel Processing*. 1998. p. 62--82.
- [25] "Secure, Efficient Data Transport and Replica Management for High-Performance Data-Intensive Computing," B. Allcock, J. Bester, J. Bresnahan, A. L. Chervenak, I. Foster, C. Kesselman, S. Meder, V. Nefedova, D. Quesnel and S. Tuecke. In *Mass Storage Conference*. 2001.
- [26] "A Wide-Area Implementation of the Message Passing Interface", I. Foster, J. Geisler, W. Gropp, N. Karonis, E. Lusk, G. Thiruvathukal and S. Tuecke. *Parallel Computing*, 1998. **24**(12): p. 1735--1749.
- [27] "Condor-G," M. Livny. <http://www.globus.org/retreat00/presentations/miron-08-00/>
- [28] "NetSolve: A Network Server for Solving Computational Science Problems", H. Casanova and J. Dongarra. *International Journal of Supercomputer Applications and High Performance Computing*, 1997. **11**(3): p. 212-223.
- [29] "The AppLeS Parameter Sweep Template: User-Level Middleware for the Grid," H. Casanova, G. Obertelli, F. Berman and R. Wolski. In *Proc. SC'2000*. 2000.
- [30] "Application-Level Scheduling on Distributed Heterogeneous Networks," F. Berman, R. Wolski, S. Figueira, J. Schopf and G. Shao, in *Proc. Supercomputing '96*. 1996.
- [31] "The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets", W. Allcock, A. Chervenak, I. Foster, C. Kesselman, C. Salisbury and S. Tuecke. *Journal of Network and Computer Applications*, 2001. <http://www.globus.org/research/papers.html#DG1>
- [32] "PACI Grid Portal Lets Computational Science Community Access High-Performance Computing Resources Across Country," PACI. <http://www.ncsa.uiuc.edu/access/Headlines/00Headlines/001107.PACIGrid.html>
- [33] "Common Component Architecture Toolkit," D. Gannon. <http://www.extreme.indiana.edu/ccat/about.html>
- [34] "A Java Commodity Grid Kit", G. v. Laszewski, I. Foster, J. Gawor and P. Lane. *Concurrency: Experience and Practice*, 2001. <http://www.globus.org/cog/documentation/papers/index.html>
- [35] "CoG Kits: A Bridge Between Commodity Distributed Computing and High-Performance Grids," G. v. Laszewski, I. Foster and J. Gawor. In *ACM 2000 Java Grande Conference*. 2000. San Francisco. <http://www.globus.org/cog/documentation/papers/index.html>
- [36] "CoG Kits: Enabling Middleware for Designing Science Appl," K. Jackson and G. v. Laszewski. Submitted SciDAC proposal, March, 2001, Lawrence Berkeley National Laboratory and Argonne National Laboratory.
- [37] "Pervasive Collaborative Computing Environment," D. Agarwal. Submitted SciDAC proposal, March, 2001, Lawrence Berkeley National Laboratory.
- [38] The Globus project is developing fundamental technologies needed to build computational grids. Grids are persistent environments that enable software applications to integrate instruments, displays, computational and information resources that are managed by diverse organizations in widespread locations. "The Globus Project," Globus Project. [www.globus.org](http://www.globus.org)

- [39] "Data-Intensive Computing," R. Moore, C. Baru, R. Marciano, A. Rajasekar and M. Wan, in *The Grid: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, Editors. 1999, Morgan Kaufmann. p. 105-129.
- [40] "The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Data Sets", A. Chervenak, I. Foster, C. Kesselman, C. Salisbury and S. Tuecke. *J. Network and Computer Applications*, 2001.
- [41] "Particle Physics Data Grid", PPDG. 2000. <http://www.cacr.caltech.edu/ppdg/>
- [42] "GridFTP: Universal Data Transfer for the Grid", Globus Project. 2001. <http://www.globus.org/datagrid/>
- [43] "Extensible Computational Chemistry Environment (Ecce) Data-Centered Framework for Scientific Research," D. R. Jones, T. L. Keller, K. L. Schuchardt, H. L. Taylor and D. K. Gracio, in. *Domain-Specific Application Frameworks: Manufacturing, Networking, Distributed Systems, and Software Development*, M. Fayad and R. E. Johnson, Editors. 1999, John Wiley & Sons: New York.
- [44] "Bringing TelPresence Microscopy and Science Collaboratories into the Class Room", J. Bonkalski, R. Anderson, S. Jones and N. Zaluzec. *TeleConference Magazine*, 1998. **17**(9). <http://www.ornl.gov/doe2k/Bottom.html>
- [45] "CUMULVS: Providing Fault-Tolerance, Visualization and Steering of Parallel Applications," G. A. Geist, J. A. Kohl and P. M. Papadopoulos. In *Parallel Scientific Computing Workshop*. 1996. Domaine de Faverges-de-la-Tour, Lyon, France.
- [46] "Wide-Area ATM Networking for Large-Scale MPPS," P. M. Papadopoulos and G. A. Geist, in *SIAM conference on Parallel Processing and Scientific Computing*. 1997.
- [47] "Application Experiences with the Globus Toolkit," S. Brunett, K. Czajkowski, S. Fitzgerald, I. Foster, A. Johnson, C. Kesselman, J. Leigh and S. Tuecke. In *Proc. 7th IEEE Symp. on High Performance Distributed Computing*. 1998: IEEE Press.
- [48] "From the I-WAY to the National Technology Grid", R. Stevens, P. Woodward, T. DeFanti and C. Catlett. *Communications of the ACM*, 1997. **40**(11): p. 50-61.
- [49] "Grids as Production Computing Environments: The Engineering Aspects of NASA's Information Power Grid," W. E. Johnston, D. Gannon and B. Nitzberg. In *Proc. 8th IEEE Symposium on High Performance Distributed Computing*. 1999: IEEE Press.
- [50] "A Distributed Resource Management Architecture that Supports Advance Reservations and Co-Allocation," I. Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt and A. Roy. In *Proc. International Workshop on Quality of Service*. 1999.
- [51] "Access Grid: Immersive Group-to-Group Collaborative Visualization," L. Childers, T. Disz, R. Olson, M. E. Papka, R. Stevens and T. Udeshi, in *Proc. 4th International Immersive Projection Technology Workshop*. 2000. [www-fp.mcs.anl.gov/fl/Accessgrid](http://www-fp.mcs.anl.gov/fl/Accessgrid)
- [52] "The NetLogger Methodology for High Performance Distributed Systems Performance Analysis," B. Tierney, W. Johnston, B. Crowley, G. Hoo, C. Brooks and D. Gunter. In *Proc. 7th IEEE Symp. on High Performance Distributed Computing*. 1998. <http://www-didc.lbl.gov/NetLogger/>



## 3 Biographical Sketches

---

### 3.1 William Johnston

---

William E. Johnston is a Senior Scientist and head of the Distributed Systems Dept. in the National Energy Research Scientific Computing Division of Lawrence Berkeley National Laboratory (<http://www-itg.lbl.gov/~wej/>), and the project technical manager for the Information Power Grid at NASA Ames Research Center ([www.ipg.nasa.gov](http://www.ipg.nasa.gov)).

Research interests include high-speed, widely distributed computational and data "Grids" and wide area network-based distributed systems; Public-Key security architectures and authorization systems, and; use of the global Internet to enable remote access to scientific, analytical, and medical instrumentation. Other professional activities include Principal Investigator for several US Dept. of Energy, Office of Energy Research and DARPA projects related to these topics, and executive committee co-chair for the Grid Forum ([www.gridforum.org](http://www.gridforum.org)).

Mr. Johnston has worked in the computing field for more than 30 years, and has taught computer science at the under graduate and graduate levels. He has an M.A. in Mathematics and Physics from San Francisco State University.

Selected publications. (See <http://www-itg.lbl.gov/~johnston/papers.html>)

"High-Speed, Wide Area, Data Intensive Computing: A Ten Year Retrospective," William E. Johnston. 7th IEEE Symposium on High Performance Distributed Computing July 29-31, 1998, Chicago, Ill.)

"The NetLogger Methodology for High Performance Distributed Systems Performance Analysis," (See [52].)

"Real-Time Widely Distributed Instrumentation Systems," William E. Johnston. In *The Grid: Blueprint for a New Computing Infrastructure*. Edited by Ian Foster and Carl Kesselman. Morgan Kaufmann, Pubs. August 1998.

"Authorization and Attribute Certificates for Widely Distributed Access Control." (See [19].)

"Real-Time Generation and Cataloguing of Large Data-Objects in Widely Distributed Environments." (See [10].)

"Rationale and Strategy for a 21st Century Scientific Computing Architecture: The Case for Using Commercial Symmetric Multiprocessors as Supercomputers," William E. Johnston. International Journal of High Speed Computing. June, 1998.

"Visual Servoing for Online Facilities." (See [11].)

"Research & Development Priorities for Communications and Information Infrastructure Assurance," William J. Huntman (Los Alamos National Laboratory), Sharon E. Jacobsen (Oak Ridge National Laboratory), William E. Johnston (Lawrence Berkeley National Laboratory), Douglass L. Mansur and Kathleen C. Bailey (Lawrence Livermore National Laboratory). Presidential Commission on Critical Infrastructure (PCCIP) taskforce report. June, 1997.

"High-Speed Distributed Data Handling for On-Line Instrumentation Systems." (See [12].)

### 3.2 Ray Bair

---

Dr. Raymond Bair is the Associate Director of the Computational Sciences and Mathematics Department (CSMD) at Pacific Northwest National Laboratory (PNNL). CSMD brings together multidisciplinary teams to develop tera-scale applications, develops technologies and systems for collaborative problem-solving environments, and conducts research in scalable technologies critical to intensive computing (programming models, solvers, data analysis, etc.). Dr. Bair is the DOE National Collaboratories Program Coordinator, working with DOE Program Manager Mary Anne Scott to promote collaboration among DOE's collaboratory projects and between these projects and other DOE and Federal agency efforts. He also leads the Computer Science and Enabling Technologies thrust area in PNNL's Computational Sciences and Engineering Initiative.

Recently, Dr. Bair completed 10 years of service on the DOE ESnet Steering Committee, representing DOE Basic Energy Sciences Programs in development of DOE's national and international network backbone, and leading development of the most recent ESnet Strategic Plan.

From 1990 to 1994, he lead the team responsible for creating a new national supercomputer facility, the Molecular Science Computing Facility (MSCF) at PNNL, a major component of a new DOE national scientific user facility, the Environmental Molecular Sciences Laboratory (EMSL). In 1995 he was promoted to lead the organization charged with development of the computing infrastructure and instrument development facilities for EMSL: Computing & Information Sciences (C&IS).

Dr. Bair and his C&IS organization built a prominent collaborative effort in EMSL, spanning research, development, deployment and operations. One result is EMSL's Virtual Nuclear Magnetic Resonance Facility, where over 25% of the many external users of EMSL's high field NMR instruments do their work remotely and collaboratory though the capabilities of the VNMRF.

From 1987 to 1990, Dr. Bair was Director of New Product Development at BioDesign, Inc. (currently Molecular Simulations, Inc.). He managed software product development for the pharmaceutical, polymer and materials industries; took an active role in developing interactive molecular modeling software; and established the software product development department and its processes.

Dr. Bair holds a Ph.D. in Computational Chemistry from the California Institute of Technology, and a B.S. in Chemistry and Mathematics from Westminster College, PA.

#### Recent Awards

*Federal Laboratory Consortium Award for Excellence in Technology Transfer*, in recognition of the development of a revolutionary collection of computational chemistry software, Molecular Science Software Suite, and it's transfer to the private sector (2000).

#### Selected Publications

"Development and Use of a Virtual NMR Facility," K. A. Keating, J.D. Myers, R. A. Bair, P. D. Ellis, J. G. Pelton, and D. E. Wemmer, *Journal of Magnetic Resonance*, **143**, 172-183 (2000).

"Collaboratories: Building Electronic Scientific Communities" (invited), R. A. Bair, *Proceedings of National Research Council Chemical Sciences Roundtable Workshop on Impact of Advances in Computing and Communications Technologies on Chemical Science and Technology*, National Academy Press, Washington, D.C., 125-140 (1999).

"Collaboratorium" (invited), R. A. Bair, contributor and section editor, *National Magnetic Resonance Collaboratorium, a Report by the Committee for High Field NMR: A New Millennium Resource*, National High Field Magnet Laboratory, 24-27 (1998).

### 3.3 Ian Foster

---

Mathematics and Computer Science Division  
Argonne National Laboratory  
9700 South Cass Avenue  
Argonne, IL 60439

+1 630 252 4619  
+1 630 252 3378 fax  
[foster@mcs.anl.gov](mailto:foster@mcs.anl.gov)  
[www.mcs.anl.gov/~foster](http://www.mcs.anl.gov/~foster)

---

#### Professional Preparation

University of Canterbury, New Zealand	Computer Science	BS (Hons I), 1977-1979
Imperial College, London, England	Computer Science	PhD, 1986-1988
Argonne National Laboratory (ANL)	Computer Science	Postdoc, 1989-1990

#### Appointments

2000-present Professor, Computer Science, UC

1998-present Senior Scientist, Mathematics and Computer Science Division (MCS), ANL  
 1999-present Executive Committee, Computation Institute, UC  
 1999-present Associate Director, MCS, ANL  
 1999-present Enabling Technologies Steering Committee, National Computational Science Alliance  
 1999-present Executive Committee, Global Grid Forum  
 1996-1999 Associate Professor, Computer Science, UC  
 1992-1998 Scientist, MCS, ANL  
 1990-1992 Assistant Scientist, MCS, ANL

### **Related Publications**

- [1] I. Foster, C. Kesselman, S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," *Intl J. Supercomputer Applications*, 2001 (to appear).
- [2] I. Foster, C. Kesselman, "Globus: A Metacomputing Infrastructure Toolkit," *Intl J. Supercomputer Applications*, 11(2):115-118, 1997.
- [3] I. Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt, A. Roy, "A Distributed Resource Management Architecture that Supports Advance Reservations and Co-Allocation," *Proc. Intl. Workshop on Quality of Service*, 27-36, 1999.
- [4] I. Foster, C. Kesselman, G. Tsudik, S. Tuecke, "A Security Architecture for Computational Grids," *ACM Conference on Computers and Security*, 83-91, 1998.
- [5] I. Foster, J. Insley, C. Kesselman, G. von Laszewski, M. Thiebaut, "Distance Visualization: Data Exploration on the Grid," *IEEE Computer*, December 1999.

### **Other Significant Publications**

- [1] K. Czajkowski, I. Foster, C. Kesselman, "Resource Co-Allocation in Computational Grids," *Proc. 8th IEEE Symp. on High Performance Distributed Computing*, IEEE, 1999.
- [2] I. Foster and N. Karonis, "A Grid-Enabled MPI: Message Passing in Heterogeneous Distributed Computing Systems," *Proc. SC'98*, 1998 (CD ROM).
- [3] I. Foster, J. Geisler, W. Nickless, W. Smith, S. Tuecke, "Software Infrastructure for the I-WAY Metacomputing Experiment," *Concurrency: Practice and Experience*, 10(7):567--581, 1998.
- [4] I. Foster, P. Worley, "Parallel Algorithms for the Spectral Transform Method," *SIAM J. Scientific Computing*, 18(3), 1997.
- [5] I. Foster, "Compositional Parallel Programming Languages," *ACM Trans. Prog. Lang. Syst.*, 18(4):454--476, 1996.

### **Synergistic Activities**

- **Pedagogical**: Development of widely used texts: *Strand: New Concepts in Parallel Programming* (Prentice Hall, 1990), *Designing and Building Parallel Programs* (Addison Wesley, 1995: [www.mcs.anl.gov/dbpp](http://www.mcs.anl.gov/dbpp)), and *The Grid: Blueprint for a Future Computing Infrastructure* (Morgan Kaufmann, 1999: [www.mkp.com/grid](http://www.mkp.com/grid)); also, teaching of numerous related tutorials.
- **Research tools**: Development of software tools and systems that have seen extensive use in research and teaching, including: Program Composition Notation compiler and runtime system, Parallel Spectral Transform Shallow Water Model, Nexus communication library ([www.mcs.anl.gov/nexus](http://www.mcs.anl.gov/nexus)), Globus distributed computing toolkit ([www.globus.org](http://www.globus.org)).
- **Service**: Including: numerous program committees and review committees; chair of numerous technical workshops and conferences, including IEEE HPDC '98, IEEE Frontiers '99; Software Architect, 1995 I-WAY, ACM/IEEE SC'95; SC'XY steering committee, 1999-present; Editorial Board, IEEE Trans. on Parallel and Distributed Systems, 1997-present.
- **Leadership**: Convenor and member of the Executive Committee, Global Grid Forum, an international organization focused on standards and best practices in "Grid" computing ([www.gridforum.org](http://www.gridforum.org)).
- **Awards and Honors**: British Computer Society Award for Technical Innovation, 1989; Best paper award, IEEE/ACM Supercomputing '95 Conference, 1995; GII Next Generation Award, 1997.

### **Collaborators**

David Abramson (Monash); Deb Agarwal (LBNL); Richard Alkire (UIUC); Bruce Allen (Wisc-Milwaukee); John

Anderson (ILM); Paul Avery (Florida); Ruth Ayt (UIUC); Ray Bair (PNNL); Steve Barnard (NASA Ames); Fran Berman (UCSD); Robert Biswas (NASA Ames); Maxine Brown (UIC); Randy Butler (NCSA); Charlie Catlett (ANL/Chicago); David Ceperley (NCSA); Mani Chandy (Caltech); Ann Chervenak (USC); Andrew Chien (UCSD); Alok Choudhary (NWU); Karl Czajkowski (USC); H. Dachsel (PNNL); Terrence Disz (ANL); Jack Dongarra (Tennessee); John Drake (ORNL); Steve Fitzgerald (USC); Dennis Gannon (Indiana); Jonathan Geisler (NWU); Bill Gropp (ANL); Steve Hammond (NCAR); Robert Harrison (PNNL); Bill Hibbard (Wisc-Madison); Bob Hollebeck (Pennsylvania); Rob Jacob (Chicago); Chris Johnson (Utah); Andrew Johnson (UIC); Lennart Johnsson (Houston); Nancy Johnston (LBNL); Nick Karonis (NIU); Ricky Kendall (Ames); Ken Kennedy (Rice); Steven Kent (Chicago); Carl Kesselman (USC); Dave Kohr (SGI); Rakesh Krishnaiyer (Intel); Tim Kuhfuss (ANL); Stephen Lau (LBNL); Craig Lee (Aerospace); Jason Leigh (UIC); Kai Li (Princeton); Bob Lindell (USC); Miron Livny (Wisc-Madison); Stu Loken (LBNL); Bob Lucas (LBNL); Rusty Lusk (ANL/Chicago); Andrea Malagoli (Chicago); Joe Mambretti (Northwestern); Rick McMullen (Indiana); Michael McRobbie (Indiana); Reagan Moore (SDSC); Richard Mount (SLAC); Klara Nahrstedt (UIUC); Cliff Neuman (USC); Harvey Newman (Caltech); Jarek Nieplocha (PNNL); Jason Novotny (NCSA); Mike Papka (ANL/Chicago); Manish Parashar (Rutgers); Larry Price (ANL); Dan Reed (UIUC); John Reynders (LANL); Bob Rosner (Chicago); Alain Roy (Chicago); Subhash Saini (NASA Ames); Chuck Salisbury (ANL); Volker Sander (Juelich); Ed Seidel (MPI); Larry Smarr (NCSA); John Shalf (NCSA); Arie Shoshani (LBNL); Rok Sosic (Active Tools); Paul Stelling (Aerospace); Rick Stevens (ANL/Chicago); Wai-Mo Suen (WUStL); Valerie Taylor (NWU); George Thiruvathukal (Loyola); Brian Tierney (LBNL); Michael Tobis (Wisc-Madison); Brian Toonen (ANL); Gene Tsudik (Irvine); Steven Tuecke (ANL); Dean Williams (LLNL); Rich Wolski (Tennessee); Patrick Worley (ORNL); Rob Van der Wijngaart (NASA Ames); Gregor von Laszewski (ANL); Maurice Yarrow (NASA Ames); Lou Zechter (NASA Ames).

**Graduate Advisor:**

Keith Clark (Imperial College, London)

**Thesis Advisor & Postgraduate-Scholar Sponsor for: 2 PhDs, 6 Postdocs**

Rakesh Krishnaiyer (Intel), Ravi Nanjundiah (Indian Institute of Science), Juan Restrepo (U. Arizona), Michael Tobis (U. Wisconsin), Warren Smith (NASA Ames), George Thiruvathukal (Loyola), Gregor von Laszewski (Argonne), Ming Xu (Platform Computing).

### 3.4 AI Geist

---

Section Leader, Oak Ridge National Laboratory, <http://www.csm.ornl.gov/~geist>

AI Geist is a senior research staff member at Oak Ridge National Laboratory and leader of the Distributed Computing Section in the Computer Science and Mathematics Division. AI is one of the original developers of PVM (Parallel Virtual Machine) which became a world-wide de facto standard. AI continues to be the technical manager of the Heterogeneous Distributed Computing project at ORNL. He was actively involved in both the MPI-1 and the MPI-2 design teams. He is a member of the IEEE Taskforce on Cluster Computing and is presently involved in the DOE 2000 Common Component Architecture (CCA) forum. The CCA is an attempt to define a standard for interoperability between high-performance (possibly parallel) software components ranging from math routines to computational steering modules. AI is a co-PI on the DOE 2000 Electronic Notebook project and his notebook software is in use by hundreds of research groups in industry, labs, universities, and medical research facilities around the world. In his 15 years at Oak Ridge National Laboratory, AI has published two books and over 150 papers in areas ranging from heterogeneous distributed computing, numerical linear algebra, parallel computing, collaboration technologies, solar energy, and solid state physics. AI is a member of Phi Kappa Phi, Tau Beta Pi, SIAM, and served on numerous conference program committees and has been chairman of four conferences including the Joint PC Cluster Computing Conference (JPC4-5) and the Fifth Distributed Supercomputing Conference.

AI has won numerous awards in high-performance and distributed computing including: the Gordon Bell Prize (1990), the IBM Excellence in Supercomputing Award (1990), an R&D100 Award (1994), two DOE Energy 100

awards (2001), the American Museum of Science and Energy Award (1997), and the Heterogeneous computing challenge several times (1992, 1993, 1995, and 1996). He joined Oak Ridge National Laboratory in 1983 after receiving a BS in Mechanical Engineering from North Carolina State University.

R. Armstrong, D. Gannon, Al Geist, K. Keahey, S. Kohn, L. McInnes, S. Parke,. "Toward a Common Component Architecture for High Performance Scientific Computing." Proceedings of HPDC'99, August 1999

J. Kohl, Al Geist, "Monitoring and Steering of Large-Scale Distributed Simulations." Proceedings of Applied Modeling and Simulation Conference (AMS'99), September 1999

P. Gray, Al Geist, S. Scott, V. Sunderam, "Bringing Cross-Cluster Functionality to Processes Through the Merging and Splitting of Virtual Environments." Journal of Parallel and Distributed Computing, 1999

Al Geist, J. Kohl, S. Scott, P. Papalopoulos. "HARNES: Adaptable Virtual Machine Environment for Heterogeneous Clusters." Parallel Processing Letters Vol. 9 No. 2 253-273, 1999

S. Scott, B. Luethke, Al Geist, Ray Flanery, "Cluster Command & Control (C3) Tools Suite." Proceedings of Third Distributed and Parallel Systems Conference (DPYS2000) June 2000

Al Geist, "PVM and MPI: What else is needed for Cluster Computing?" Proceedings of EuroPVM-MPI 2000. Springer LNCS, September 2000

Al Geist, J. Schwidder, B. Luethke, S. Scott, "M3C: a web-based management tool for federated clusters." Proceedings of IEEE International Conference on Cluster Computing, December 2000

### **3.5 William Kramer**

---

William Kramer is the Head of the NERSC High Performance Computing Department and the Deputy Director of the NERSC Division. His primary responsibilities in these roles is as Principle Investigator of the NERSC Program, which provides DOE's premier supercomputer facility NERSC is unique and quite possibly the most powerful unclassified facility in the U.S. NERSC's paradigm establishes a new balance of intellectual services and traditional computer cycles and storage and also generated shift for DOE computational science from traditional vector computing to massively parallel.

Mr. Kramer is responsible for all aspects of the NERSC facility consisting a 2,500 processor IBM SP, 700 processor Cray T3E, three Cray SV-1, a large cluster of workstations, a HPSS Mass storage system with over a petabytes of archive storage, high performance networks and over 100 other systems. The current computational power of the entire facility is almost 5 teraflops of peak performance, with a 1 Petabyte archive. A technical staff of 65 provides 24 by 7 support to 2,400 clients throughout the United States. As one of the first employees of NERSC at LBNL, Mr. Kramer led the re-implementation of NERSC, including reinstallation of all systems, and hiring over 70 staff members. He established an integrated service and systems architecture. Mr. Kramer led the installation, test, and introduction of the largest unclassified IBM SP system, first large T3E, the largest I/O configuration on a T3E, early J-90s, HPSS, first use of checkpoint/restart in a MPP production environment, and the first to demonstrate the ability to manage very large MPPs with utilization over 95% and many other technical innovations. He also was responsible for the creation of a new 20,000sf computing facility with OC-12 connections.

Mr. Kramer has worked in high performance computing for almost 20+ years, playing a key role in the creating and implementing of NASA's largest supercomputer facility - the Numerical Aerodynamic Simulation Facility. He has a breadth of experience from scientific computing, operating system and networking. While at NASA he helped start and became the program manager for a \$400M research program to improve the Air Traffic Control system. He has played high level leadership roles in SC conferences, DECUS, CUG and other organizations. His most recently was the Information Architect for SC 2000 and led the SCinet 2000 activity - which set networking records in a number of areas. Mr. Kramer has advanced degrees from Purdue University and University of Delaware and is currently attending UC Berkeley.

Selected Publications and Presentations

Wong, Adrian T., Leonid Oliker, William T. C. Kramer, Teresa L. Kaltz, and David H. Bailey, "Evaluating System Effectiveness in High Performance Computing Systems",. Proceedings of SC2000, November 2000.

Wong, Adrian T., Leonid Oliker, William T. C. Kramer, Teresa L. Kaltz, and David H. Bailey, "System Utilization Benchmark on the Cray T3E and IBM SP", the 5th Workshop on Job Scheduling Strategies for Parallel Processing, May 2000, Cancun Mexico.

Kramer William, Francesca Verdier, Keith Fitzgerald, James Craw, and Tammy Welcome, "High Performance Computing Facilities for the Next Millennium", presented at SC99, Portland, OR, and published as part of the Tutorials Program, November 1999.

Wong, Adrian T., Leonid Oliker, William T. C. Kramer, Teresa L. Kaltz, and David H. Bailey, "Evaluating System Effectiveness in High Performance Computing Systems",. LBNL Technical report LBNL 44542, November 1999

Simon, Horst D., William T. C. Kramer, and Robert F. Lucas , "Building the Teraflops/Petabytes Production Supercomputing Center". Presented at EuroPar '99 in Toulouse, France, September 1999, and published in the Proceedings.

Shoshani, Arie, Craig Tull, Brian Tierney, Harvard Holmes, Robert Lucas and William T.C. Kramer, "Large Scale, Data Intensive Computing", referred, full day tutorial at the SC '98 Conference, November 8, 1998, Orlando, FL.

Kramer, William T.C., "Computational Climate Activities at NERSC", invited presentation at 1998 Computing in Atmospheric Sciences Workshop, Imperial Palace Annecy France, July 2, 1998 .

Simon, Horst, C. William McCurdy, William T.C. Kramer and Alexander X. Merola, "The New NERSC Program for FY 1997 - An Overview of Current and Proposed Efforts", December 18 1996, revised April 3, 1997.

Kramer, William T.C., Horst Simon, C. William McCurdy, "Reinventing the Supercomputer Center", referred tutorial at the Supercomputing '96 Conference, November 1996, Pittsburgh, PA.

Kramer, William T.C., "NASA's Advanced Air Transportation Technologies Program and Free Flight", November 5, 1995, an invited presentation at Digital Avionics Systems Conference, Cambridge MA.

Kramer, William T.C., "NASA's Advanced Air Transportation Technologies Program and Free Flight", October 21, 1995, an invited presentation at Airline Dispatcher Federation Annual Symposium, Dallas Texas.

Castagnera, Karen, et al, "NAS Experiences with a Prototype Cluster of Workstations", referred, paper presented at Supercomputing 94, Washington, D.C., November 15, 1994

Kramer, William T.C., "Large Memory Applications", invited presentation at the Digital Equipment Computer User Society Symposia, Anaheim, CA, December 1992.

Walter, Howard, Robert VanCleaf, and William T.C. Kramer, "System Management in the UNIX Supercomputer Environment", referred, full day tutorial presented at Supercomputing 90, New York, NY, November 1990.

Craw, James, and William T.C. Kramer, "Computational Services in the UNIX Supercomputer Environment", referred paper presented at Supercomputing 89, Reno, NV, November, 1989.

Kramer, William T.C., "Digital Signal Processing with Microprocessors", a graduate thesis at the University of Delaware.

Kramer, William T.C., "Techniques of Protocol Validation", May 1985, Proceedings of the Digital Equipment Users Society Vol. 11 No. 1.

## 4 Description of Facilities and Resources

With the exception of the NERSC production facilities, which will be phased in based on evaluation and refinement of the Grid software as described in the proposal, the other facilities will be available early in the project as indicated in the task plan.

### Compute Systems Integrated in Pilot Phase

Site	System Name	Processors	Operating System	Memory	Theoretical Performance (Teraflops)
PNNL	Jupiter	104	AIX	80 GBy	0.25
PNNL	Colony	192	Linux	54 GBy	0.1
LBNL	portnoyc	4, SMP	Solaris	.5 GBy	
LBNL	fluffy	6, SMP	Solaris	2 GBy	
LBNL	diesel	8, SMP	Solaris	4 GBy	
LBNL	slappy	8, SMP	Solaris	1 GBy	
LBNL	clipper	8, SMP	Solaris	1 GBy	
NERSC	alvarez	256	Linux		
NERSC production	seaborg	158 x 16	AIX	158 x 12 GBy	
ANL	Chiba City	574	Linux	128 GBy	0.25
ANL	Denali	128	IRIX	32 GBy	0.064
ANL	Quad	80	AIX	40 GBy	
ORNL	Colt	64	True64	32 GBy	0.085
ORNL	Eagle	724	AIX	372 GBy	1.080
ORNL	Falcon	256	True64	128 GBy	0.342

### Data Storage Resources Integrated in Pilot Phase

Site	System Name	Storage System	Operating System	Rotating Storage (Terabytes)	Archival Storage (Terabytes)
PNNL	Nwarchive1	Custom/ E-mass	Solaris/ IRIX	0.4	20
LBNL	Sun cluster	NFS	Solaris	0.75	0
NERSC	Probe	HPSS	AIX	1.5	?
NERSC production	HPSS	HPSS	AIX	7.5	1300
ORNL	Probe	HPSS	AIX	1.5	20
ORNL	Eagle		AIX	9.2	
ORNL	Falcon		True64	5.5	
ANL	IBM 3495 Tape Library	ADSM			60
ANL	Chiba City		Linux	8.4	
ANL	Denali	XFS Filesystem	IRIX	2	
ANL	Quad	SSA & FC Arrays	AIX	1	