C.1. The Challenge: Data-Intensive Science and Virtual Data

An unprecedented requirement for geographically dispersed extraction of complex scientific information from very large collections of measured data provides the impetus for this broad, coordinated effort to expand the boundaries of Information Technology (IT). The requirement is posed by four of the most far-reaching physics experiments of the next decade—ATLAS, CMS, LIGO, and SDSS—who partner here with leading IT researchers in the Grid Physics Network (**GriPhyN**) project. The **GriPhyN** research agenda aims at IT advances that will enable groups of scientists distributed worldwide to harness Petascale processing, communication, and data resources to transform raw experimental data into scientific discoveries. We refer to the computational environment in which these scientists will operate as Petascale Virtual Data Grids (PVDGs). The goals of the **GriPhyN** project are to achieve the fundamental IT advances required to realize PVDGs and to demonstrate, evaluate, and transfer these research results via the creation of a Virtual Data Toolkit to be used by the four major physics experiments and other projects. The synergy with the experiments, which depend on PVDGs to maximize their ability to extract scientific discoveries from massive quantities of data, will help **GriPhyN** achieve its goals within the next few years.

This research proposal is structured as follows. In Section C.1, we review the four physics experiments, the virtual data grid concept, the associated IT research challenges, and the Virtual Data Toolkit. In Section C.2, we outline our IT research activities, application experiments, and toolkit development work. We conclude with a discussion of related work, management issues, and our technology transfer, education, and outreach plans.

C.1.a. The Frontier Experiments: Motivating Factors and Principal Clients

The immediate motivation and the primary testing ground for the virtual data grid technologies to be developed within **GriPhyN** are four frontier experiments that are exploring fundamental forces of nature and the structure of the universe: the CMS[1] and ATLAS[2] experiments at the LHC (Large Hadron Collider) at CERN, Geneva, LIGO[3,4,5] (Laser Interferometer Gravitational-wave Observatory), and SDSS[6] (Sloan Digital Sky Survey). For the next two decades, the LHC will probe the TeV frontier of particle energies to search for new phenomena. LIGO will detect and analyze, over a similar span of time, nature's most energetic events sending gravitational waves across the cosmos. SDSS is the first of several planned surveys that will systematically scan the sky to provide the most comprehensive catalog of astronomical data ever recorded. The National Science Foundation has made heavy investments in LHC, LIGO, and SDSS. (US LHC scientists also receive support from DOE and SDSS is also supported by the Alfred P. Sloan Foundation.)

The IT authors of this proposal choose to collaborate with these experiments not only because of their tremendous scientific importance but also because together they span an extremely challenging space of data-intensive applications in terms of timeframe, data volume, data types, and computational requirements: see Table 1.

Table 1: Characteristics of the four physics experiments targeted by the GriPhyN project

Application	First Data	Data Rate MB/s	Data Volume (TB/yr)	User Community	Data Access Pattern	Compute Rate (Gflops)	Type of data
SDSS	1999	8	10	100s	Object access and streaming	1 to 50	Catalogs, image files
LIGO	2002	10	250	100s	Random, 100 MB streaming	50 to 10,000	Multiple channel time series, Fourier transformations
ATLAS/ CMS	2005	100	5000	1000s	Streaming, 1 MB object access	120,000	Events, 100 GB/sec simultaneous access

The four experiments also share challenging requirements, including: (1) collaborative data analysis by large communities; (2) tremendous complexity in their software infrastructure and processing requirements: separating out the rare "signals" in the data that point the way to scientific discoveries will be hard to manage, extremely complex, and computationally demanding; and (3) decadal timescales in terms of design, construction and operation, with a consequent need for great flexibility in computing, software, data, and network management.

The four experiments are each committed to developing a *highly distributed* data analysis infrastructure to meet these requirements. These infrastructures are distributed both for technical reasons (e.g., to place computational and data resources near to demand) and for strategic reasons (e.g., to leverage existing technology investments). ATLAS and CMS are the most ambitious, anticipating a need for aggregate data rates of ~100 Gbytes/sec, around 60 TeraOps of fully-utilized computing power, and the fastest feasible wide area networks, including several OC-48 links into CERN. Their hierarchical worldwide Data Grid system is organized in "Tiers," where Tier0 is the central facility at CERN, Tier1 is a national center, Tier2 a center covering one region of a large country such as the US or a smaller country, Tier3 a workgroup server, and Tier4 the (thousands of) desktops[7].

C.1.b. Virtual Data Grids as a Unifying Concept

The computational and data management problems encountered in the four physics experiments just discussed differ fundamentally in the following respects from problems addressed in previous work [8,9]:

- *Computation-intensive as well as data-intensive*: Analysis tasks are compute-intensive *and* data-intensive and can involve thousands of computer, data handling, and network resources. The central problem is coordinated management of computation and data, not simply data movement.
- *Need for large-scale coordination without centralized control*: Stringent performance goals require coordinated management of numerous resources, yet these resources are, for both technical and strategic reasons, highly distributed and not amenable to tight centralized control.
- *Large dynamic range in user demands and resource capabilities*: These systems must be able to support and arbitrate among a complex task mix of experiment-wide, group-oriented, and (thousands of) individual activities—using I/O channels, local area networks, and wide area networks that span several distance scales.

These considerations motivate the study of the virtual data grid technology that will be critical to future data-intensive computing not only in the four physics experiments, but in the many areas of science and commerce in which sophisticated software must harness large amounts of computing, communication and storage resources to extract information form measured data.

We introduce the *virtual data grid* as a unifying concept to describe the new technologies required to support such next-generation data-intensive applications. We use this term to capture the following unique characteristics:

- A virtual data grid has *large extent*—national or worldwide—and *scale*, incorporating large numbers of resources on multiple distance scales.
- A virtual data grid is more than a network: it layers sophisticated *new services* on top of local policies, mechanisms, and interfaces, so that geographically remote resources can be used in a coordinated fashion.
- A virtual data grid provides a new degree of *transparency* in how data-handling and processing capabilities are integrated to deliver data products to end-user applications, so that requests for such products are easily mapped into computation and/or data access at multiple locations. (This transparency is needed to enable optimization across diverse, distributed resources, and to keep application development manageable.)

These characteristics combine to enable the definition and delivery of a potentially unlimited virtual space of data products derived from other data. In this virtual space, requests can be satisfied via direct retrieval of materialized products and/or computation, with local and global resource management, policy, and security constraints determining the strategy used. The concept of *virtual data* recognizes that all except irreproducible raw experimental data need 'exist' physically only as the specification for how they may be derived. The grid may instantiate zero, one, or many copies of derivable data depending on probable demand and the relative costs of computation, storage, and transport. In high-energy physics today, over 90% of data access is to derived data. On a much smaller scale, this dynamic processing, construction, and delivery of data is precisely the strategy used to generate much, if not most, of the web content delivered in response to queries today.

Figure 1 illustrates what the virtual data grid concept means in practice. Consider an astronomer using SDSS to investigate correlations in galaxy orientation due to lensing effects by intergalactic dark matter[10,11,12]. A large number of galaxies—some 10^7— must be analyzed to get good statistics, with careful filtering to avoid bias. For each galaxy, the astronomer must first obtain an image, a few pixels on a side; process it in a computationally intensive analysis; and store the results. Execution of this request involves virtual data catalog accesses to determine whether the required analyses have been previously constructed. If they have not, the catalog must be accessed again to locate the applications needed to perform the transformation and to determine whether the required raw data is located in network cache, remote disk systems, or deep archive. Appropriate computer, network, and storage resources must be located and applied to access and transfer raw data and images, produce the missing images, and construct the desired result. The execution of this single request may involve thousands of processors and the movement of terabytes of data among archives, disk caches, and computer systems nationwide.

Virtual data grid technologies will be of immediate benefit to numerous other scientific and engineering application areas. For example, NSF and NIH fund scores of X-ray crystallography labs that together are generating Petabytes of molecular structure data each year. Only a small fraction of this data is being shared via existing publication mechanisms. Similar observations can be made concerning long-term seismic data generated by geologists, data synthesized from studies of the human genome database, brain imaging data, output from long-duration, high-resolution climate model simulations, and data produced by NASA's Earth Observing System.

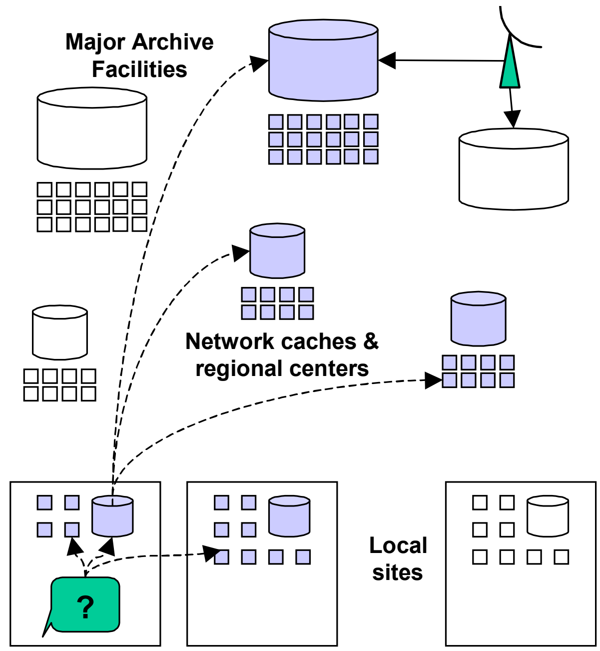


Figure 1: A request from a PVDG user may result in accesses to local caches, network caches, or remote archives; and also in computation at sites across the grid. In this example, the request is satisfied by data from a major archive facility, data from one regional center and computation on data at a second regional center, plus both data and computation from the local site and neighbor.

C.1.c. Information Technology Challenges Associated with Virtual Data

We summarize as follows the fundamental IT challenge that we address: *scientific communities of thousands, distributed globally, and served by networks with bandwidths varying by orders of magnitude, need to extract small signals from enormous backgrounds via computationally demanding (Teraflops-Petaflops) analyses of datasets that will grow by at least three orders of magnitude over the next decade: from the 100 Terabyte to the 100 Petabyte scale*. Furthermore, the data storage systems and compute resources are themselves distributed and only loosely, if at all, under central control.

Overcoming this challenge and realizing the Virtual Data concept requires advances in three major areas, which we summarize here and detail in Section C.2.a. These questions form the core of our IT research program; investigations of related issues such as high-performance I/O and networking, databases, security, and agent technologies will be addressed as needed to advance this overall goal.

Virtual data technologies. Advances are required in *information models* if we are to represent both physical data structures and data structure manipulations (virtual data), hence integrating procedural and data models. We require new methods for cataloging, characterizing, validating, and archiving the *software components* that will implement virtual data manipulations integrated with existing information models and transport protocols (e.g., those created by the NSF Digital Library Initiative). These methods must apply to an environment in which software components, data elements, and processing capabilities are distributed, under local control, and subject to update.

Policy-driven request planning and resource scheduling. PVDG users, whether individual scientists or large collaborations, need to be able to project and plan the course of complex analyses involving virtual data. As a single data request may involve large amounts of computation and data access, and the computing needs of a collaboration may involve thousands of requests, this *request planning* task is well beyond the current state of the art. New techniques are required for representing complex requests, for constructing request representations, for estimating the resource requirements of individual strategies, for representing and evaluating large numbers of alternative evaluation strategies, and for dealing with uncertainty in resource properties and with failures. The PVDG proper needs to be able to allocate storage, computer, and network resources to requests in a fashion that satisfies global and local constraints. Global constraints include community-wide policies governing how resources dedicated to a particular collaboration should be prioritized and allocated; local constraints include site-specific policies governing when external users can use local resources. We require mechanisms for representing and enforcing constraints and new policy-aware resource discovery techniques.

Execution management within national-scale and worldwide virtual organizations. Once a request execution plan has been developed, we require mechanisms for *execution management* within a high-performance, distributed, multi-site PVDG environment—a heterogeneous "virtual organization" encompassing many different resources and policies—to meet user requirements for performance, reliability, and cost while satisfying global and local policies. *Agent computing* will be important as a means of giving the grid the autonomy to balance user requirements, grid throughput, and grid or local policy when deciding where each task or subtask should execute; new approaches will certainly be required to achieve fault tolerance (e.g., via checkpointing) at a reasonable cost in these extremely large and high-performance systems. *Simulation* of Grid behavior will be an important evaluation tool.

C.1.d. The Virtual Data Toolkit: Vehicle for Deliverables and Technology Transfer

The success of the **GriPhyN** project depends on our ability not only to address the research challenges just listed but also to translate research advances into computational tools of direct utility to application scientists.

We will achieve this latter objective via the development of an application-independent Virtual Data Toolkit: a *suite of generic virtual data services and tools* designed to support a wide range of virtual data grid applications. These services and tools will constitute a major **GriPhyN** deliverable and will enable far more than just our four chosen physics applications; in fact, they will open up the entire field of world-distributed Petascale data access and analysis.

Figure 2 illustrates the layered architecture that we plan to realize in the Virtual Data Toolkit. PVDG applications are constructed by using various *virtual data tools*. For example, a request-planning tool would implement a method for translating a high-level request into requests to specific storage and compute services; an agent-based diagnostic tool might implement agents designed to detect and respond to performance problems[13].

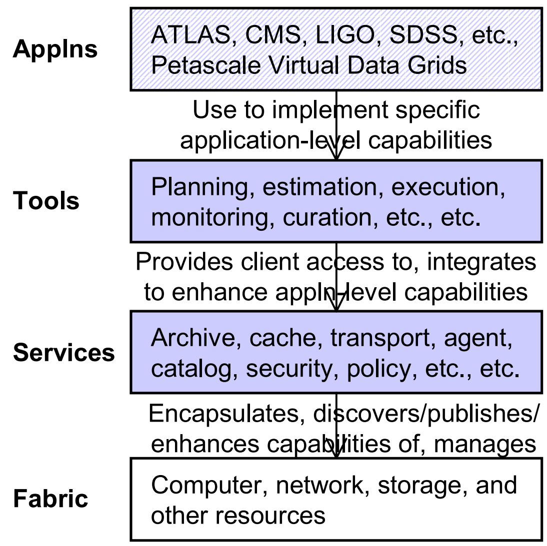


Figure 2: Applications built on the Virtual Data Toolkit exploit lower-level services that realize the virtual data abstractions needed to produce a PVDG

The implementations of these virtual data tools rely upon a variety of *virtual data services*, which both encapsulate the low-level hardware fabric from which a PVDG is ultimately constructed and enhance their capabilities[14]. Examples of such services include archival storage services, network cache services, metadata catalogs, transformation catalogs, component repositories, and information services. PVDG services are distinguished by a focus on highly distributed implementation, explicit representation of policy, integration with system-wide directory and security services, performance estimation, trouble-shooting services, and high performance. For example, a PVDG storage service would likely support, in addition to functions for reading and writing a dataset, methods for determining key properties of the storage service such as its performance and access control policies, and would be designed to cooperate with system-wide services such as failure detection, resource discovery, agent execution, and performance estimation. We emphasize that the layering discussed here reflects logical distinctions and not strict hierarchical interfaces: performance considerations will motivate flexible combinations of functionality.

The PVDG architecture that we propose to pioneer builds substantially on the experience of **GriPhyN** personnel and others in developing other large infrastructures, in the Internet, Grid[15,16,17], and distributed computing contexts[18]. We also leverage considerable existing technology: for example, the uniform security and information infrastructure that has already been deployed on a global scale by the Grid community. These technology elements and new approaches to be pioneered in this project will enable the establishment of the first coherently managed distributed system of its kind, where national and regional facilities are able to interwork effectively over a global ensemble of networks, to meet the needs of a geographically and culturally diverse scientific community.

C.1.e. The Need for a Large, Integrated Project; Substantial Matching Support

The scientific importance and broad relevance of these diverse research activities, and the strong synergy among the four physics experiments and between physics and IT goals (see Sec. C.1.a.), together justify the large budget requested for this project. This project represents a tremendous opportunity for both IT and physics. While IT researchers can investigate fundamental research issues in isolation, only a collaborative effort such as that proposed here can provide the opportunities for integration, experimentation, and evaluation *at scale* that will ensure long-term relevance and reusability in other application areas. And none of the physics experiments can, in isolation, develop the Virtual Data Grid technology that will enhance their ability to extract new knowledge and scientific discoveries from the deluge of data that is about to inundate them.

A five-year duration is needed because of the complexity of the problems to be tackled and the long time scale of the physics experiments that serve as testbeds. For example, ATLAS and CMS are developing *now*, using simulated Terascale data, the systems that will analyze data that will flow in Petabyte-scale quantities in 2005. A five-year scope allows the **GriPhyN** project to engage with these experiments as a full partner throughout this vital period.

The participating institutions recognize the broad relevance and exciting goals of this project and have hence committed to a total of $1,486,000 in institutional funds over the 5 years of the project, plus substantial effort from staff and scientists not funded by the project. These contributions and other aspects of the budget are summarized as an overall budget justification included with the University of Florida budget justification. In addition, we have received strong expressions of interest from potential industrial partners, which we expect to see translate into substantial equipment and manpower support.

C.2. Research Activities

We structure our research project as three distinct but tightly integrated activities (see Figure 3):

- ***IT research***: groundbreaking IT research motivated and guided by challenges encountered by domain scientists in meeting their computational and data management needs.
- ***Application experiments***: prototyping new information technology and interfacing it with scientific applications in real-life test-bed environments defined by CMS, ATLAS, LIGO, and SDSS requirements.
- ***Tool development***: turn "winning" prototypes into production quality tools to be used by the scientific community.

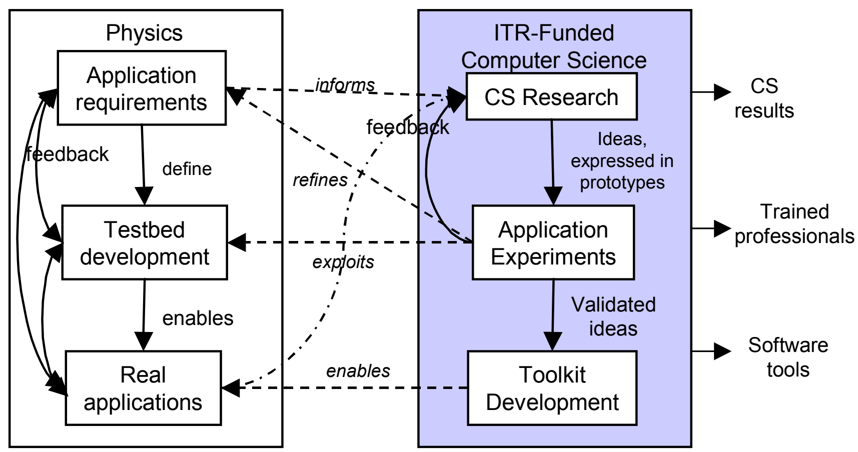


Figure 3: Relationship between GriPhyN ITR research components (on the right) and the four Physics project

The 26 researchers—12 computer scientists and 14 experimental physicists—that we have brought together to attack these problems have been intrigued by the research challenges, the possible impact, and the scope of this project. These researchers represent a wide inter- *and* intra-disciplinary spectrum of research interests, talent, and experience. The computer scientists are organized in seven groups—Berkeley, Chicago, Florida, Indiana, San Diego, USC, and UW Madison—each managed by local **GriPhyN** personnel and each contributing a unique capability to the project and responsible for one major research activity. The physicists are distributed across a comparable number of institutions, chosen with a view to IT expertise and connections with the physics experiments.

The rich collection of technology and software already developed by project participants means that we can start all three of the phases just listed in parallel. The result will be a steady stream of IT research results, application experiences, and production quality software.

Table 2 summarizes the high-level goals of each project component. We plan to coordinate these diverse efforts via the definition and frequent review of a *PVDG architecture* that defines the interfaces between key components, and the scheduling of frequent *integration events* in which components developed by different groups are brought together for interoperability testing. Figure 4 illustrates PDVG architecture elements, from a user's perspective.

C.2.a. Information Technology Research

The primary IT research goals of the project are outlined above in C.1.c. Here we expand on the specific technical issues to be addressed in each research area and organize them into research efforts. We also indicate who among the senior personnel leads each of these efforts and which research groups are involved. This allocation of responsibilities and research foci should not be viewed as drawing hard boundaries between the different IT research groups: we expect considerable cross-fertilization of ideas among groups.

Table 2: Timeline for the IT research, Virtual Data Toolkit, and Physics Experiment Components

	Year 1	Year 2	Year 3	Year 4	Year 5
Virtual Data	Preliminary data models for virtual data.	Procedural / declarative representations	Info. model for catalog planning & scheduling	Integrate perf. estimates into info models	Integrate policy constraints into info models
Request Planning	Model for policy driven request planning	Integration of virtual data and resource models	Request planning algorithms	Integration of local / global policies	Optimization across local & global policies
Request Execution	Language for distributed service	Integration with request planning model	Integration of fault tolerance mechanisms	Support for dynamic replanning	Simulation of Grid behavior
Virtual Data Toolkit	Package basic services	Centralized virtual data services	Distributed virtual data services	Scale virtual data services to Petascale	Enhance usability, performance, etc.
SDSS	Catalog model, data model	Distributed statistics	Distributed virtual data presentation	Distributed generation of merged sky maps	Distributed creation of derived catalogs
LIGO	Catalog, information model	Data streaming through caches	Opportunistic computing of derived products	Event-based derived products	Optimization of product derivation
LHC CMS / ATLAS	Proto Tier-2 system	Replication, caching, and distributed databases	Production prototype for PVDG	Production quality PVDG	Full distributed generation of derived data products

C.2.a.1. Virtual Data Technologies

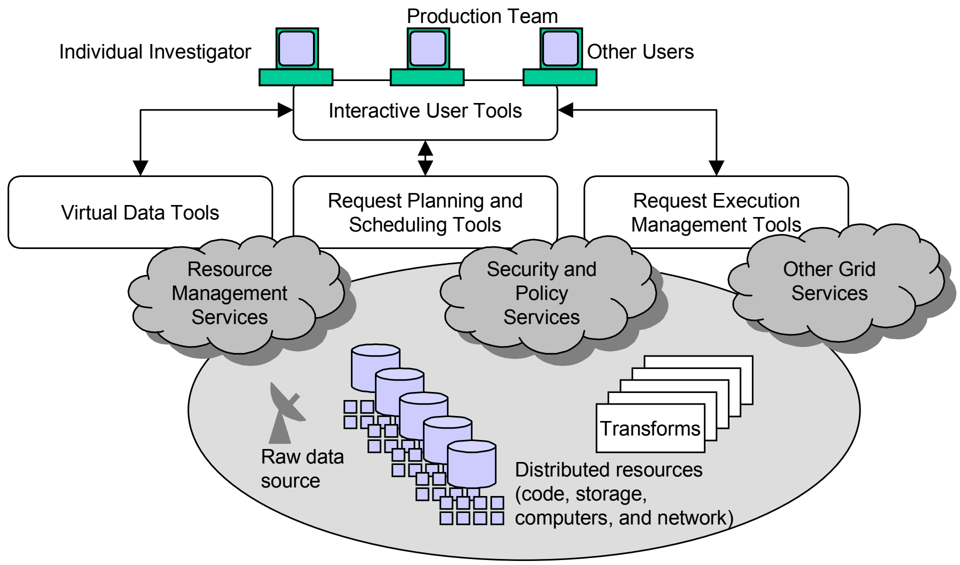
A *virtual data catalog* provides a single (logical) source for the many types of information required to create an effective virtual data grid. These information types include *metadata* (the attributes, etc., that applications use to locate data and procedures of interest); *procedures* (definitions for the procedures used to create virtual data, together with associated information such as performance characteristics); *replicas* (information concerning the physical location of instantiated data and procedures); and *grid information* (information concerning the resource fabric of the virtual data grid, including physical characteristics, instantaneous performance, and future availability).

These different information types have rather different characteristics and so in practice may be maintained in different catalog structures. However, uniform representations and access methods simplify the task of the application seeking to make efficient computation/data access tradeoffs when planning a request execution. The catalog can also store plans of both previous and currently executing requests.

Virtual data services must include the curation functions used to publish data or procedures and associated metadata; replication functions for creating duplicates; instrumentation functions for obtaining accurate performance data; and trend analysis functions for monitoring system behavior and generating predictions of future state[74,75]. These services must deal with information that is highly distributed and maintained by local organizations.

The major research challenges that we propose to address in this area relate to (1) the nature of the information models used to integrate declarative and procedural representations of data and (2) the services used to initiate and manage the execution of virtual data procedures when a virtual data catalog access results in computation. This effort is performed by the San Diego and Berkeley teams and is led by Moore.

Figure 4: User view of PVDG architecture, showing the diverse users that compete for resources, the tools used to formulate and execute requests, the services used to coordinate use of Grid resources, and the resources themselves.



Information Models. We will start our investigations of information models by studying existing data flow models[19] and object-oriented active databases[20]. These models typically assume centralized specifications and implementations; we must consider how to extend them for a distributed and highly heterogeneous environment. One challenge will be to design curation procedures and repository structures able to deal with procedural representations of data, so that users can easily define new derived data types by defining the required input data, output data, and transformation functions. Interesting security and version control problems arise. We will leverage ongoing work by **GriPhyN** and other researchers on extensible data formats[21,22].

Services. It will be important to understand how to decompose functionality between the virtual data catalog, planning, and execution components of a PVDG. Initial work in experiment management has shown that it can be advantageous to integrate execution components directly into the data catalog[23]. Alternatively, one can separate the planning and control functions from the cataloging function, providing interfaces for accessing and manipulating information maintained in the virtual data catalog. We will investigate the use of a service-oriented virtual data catalog that provides a collection of orthogonal, distributed data cataloging services such as code curation, distributed metadata storage, and data discovery. In addressing these issues, previous work within the digital library community[102] and tertiary storage for High Energy Physics applications[24] is relevant, as is work on extensible metadata representations within SRB/MCAT[25] and indexing in STACS[26].

C.2.a.2. Policy-Driven, Virtual Data Request Planning and Scheduling

The second major challenge that we will address relates to the planning and scheduling of data analyses in the virtual data grid environment. The factors that guide the development of a plan include user requirements, global and local policy, and overall state. User requirements may include not only the virtual data request but also optimization criteria such as completion in the shortest time or usage of the fewest computing resources. Any plan is necessarily constrained by resource availability, and consequently, we must obtain all available state information. This complicates planning, as the global system state can be large and determining future system state can be difficult. We will explore the use of advanced reservations[27,28] and of plans with alternative branches.

The complex interrelationships among different data representations (procedural vs. declarative), data locations (archived vs. cached), policies (local vs. global), and computations (different user queries, background tasks, etc.) make planning and scheduling a challenging and rewarding problem. We will approach this problem by (1) developing an architecture and mechanisms for policy-driven request planning and scheduling, (2) instantiating these mechanisms in a variety of alternative request planning algorithms, and (3) evaluating these alternatives via both simulation studies and empirical evaluation in the context of our motivating physics problems. We will address hierarchical planning methods, global and local policies, optimization criteria, and dynamic replica management. For simulation studies, we intend to exploit the extensible simulation toolset developed within the MONARC project[29,30] for simulation[31] and optimization of distributed LHC analysis systems. We will also investigate scalable network simulation tools[32,33,34], the Ptolemy II heterogeneous modeling and design system[35], and Chien's MicroGrid technology. Foster leads this effort, which includes U. Chicago, Florida, USC/ISI, and Indiana University.

Hierarchical planning and scheduling methods. As noted above, request planning and scheduling in virtual data grids is extremely demanding. New techniques are required for representing complex requests, for constructing request representations via the composition of representations for virtual data components, for representing and evaluating large numbers of alternative evaluation strategies, and for dealing with uncertainty in resource properties.

The heterogeneity of the underlying system, the difficulty of obtaining and maintaining global state information[36], and the complexity of the overall task requires a hierarchical approach to planning and scheduling. Hence, we intend to study hierarchical methods in which specialized agents are responsible for planning and scheduling subplans, which can be assembled into an overall solution for the virtual data query. Subplanners may be responsible for specific sets of resources or may be optimized for specific types of requests. For example, a storage planner with specialized knowledge of data movement and replication strategies might know how to co-reserve space in storage systems and networks and incorporate knowledge about the interchangeability of alternative data sources to construct alternative plans that are failure resilient.

In our research, we will develop methodologies for building both subplanners and the coordinating global planners. These methodologies will include methods by which subplanners can determine what resources are available, what guarantees are available for these resources, and the quality of the guarantee[37]. We plan to develop negotiation protocols by which subplanners may interact with global planners and with each another. Specific subplanners will be developed, with an initial focus on replica management (see below). We will evaluate alternative planning strategies using both simulations and application experiments. We also plan to explore alternative approaches to the execution of planning algorithms, first investigating planners that generate complete plans (potentially with alternatives) before any execution takes place and subsequently considering incremental planning algorithms in which planning and execution phases are more tightly coupled. Florida (Rajasekaran) will also investigate randomized techniques for achieving better performance.

Local and global policies. A virtual data grid must be able to allocate storage, computer, and network resources to requests in a fashion that satisfies global and local policies. Global policy includes community-wide policies governing how resources dedicated to a particular collaboration should be prioritized and allocated. Local policies are site-specific constraints governing when and how external users can use local resources and the conditions under which local use has priority over remote use. The execution of a plan will fail if it violates either global or local policy. Hence we require mechanisms for representing policies and new resource discovery techniques that can take into account policy information. We also need mechanisms for enforcing policy[38]. Policy must also be expressed in a format that is interpretable by PVDG virtual data catalog and request planning components—and by system users.

Research in policy-driven resource allocation has focused on simple symmetric matching[39], attribute certificates[40] and policy independent enforcement mechanisms[41]. Efficient implementation of virtual data in a large-scale, multi-institutional setting will certainly require more general mechanisms for policy expressions and compliance checking, such as the use of knowledge representation schemes. We will investigate more sophisticated policy matching strategies and the integration of policy into planning activities. Hierarchical policy specifications may simplify the integration of policy concerns into planning. For example, the LHC community might wish to make processing of new data a top priority. Remaining resources may then be devoted to CMS or ATLAS analyses, constrained by the amount of time any one resource owner is willing to commit. We will develop policy representations, query, and enforcement mechanisms that facilitate the specification of such policy.

Optimization Criteria. The purpose of planning and scheduling is to optimize the response to a query for virtual data given global and local policy constraints. Different optimization criteria may be applied to a PVDG request: minimize execution time, maximize reliability, minimize use of a particular resource, etc. For a given metric, optimization is driven by resource characteristics and availability. In Nile[42], for example, it was assumed that the bottleneck was data sources first, wide-area network bandwidth second, and free processors third. In this context, it was shown that the scheduling problem could be cast in terms of max-flow, allowing an exact solution. Rearrange these resource availabilities, and the nature of the scheduling problems changes greatly[71,43]. Policy and advanced reservation further complicate the optimization process. Initially, we will investigate how policy constraints interacts with existing optimization criteria, such as a generalized version of the greedy approach taken by Nile. We will also investigate the use of alternative optimization criteria, such as least cost solutions and solutions that balance cost and time. Later, we will develop a unified framework for policy-constrained optimization designed to balance the requirements of user, community, and resource owner.

One important class of resource policies maps resource allocation decisions into monetary exchange[44]. Such approaches have the significant advantage of uniformity and simplicity but may provide too little control. We plan to investigate and experiment with such methods in realistic PVDG environments.

We can draw here upon a substantial body of experience and technology obtained in Condor, Nile, and other projects. In Condor, matchmaking technology links tasks with the "best" available resource. Some recent work has also addressed the co-allocation of multiple resources. The Nile approach allows upper bounds on completion time.

Dynamic replica creation. Explicit application-level request planning can be complemented by system-wide strategies designed to optimize the use of scarce resources. For example, replicas of large datasets can be created within network caches to reduce load on networks and central archives. We will investigate both application-level replica creation strategies (explicit prefetching) and automated methods based on dynamic monitoring of access patterns and analysis of past behavior. Work on semantic caching[45] and the DBIS-Toolkit[46] is relevant.

C.2.a.3. Execution Management in National-Scale and World-Wide Virtual Organizations

Once an execution plan for a virtual data request has been developed, we require mechanisms for managing the execution of the request within a high-performance, distributed, multi-site PVDG environment. We require specialized strategies for fault tolerance and security when using heterogeneous, dedicated and opportunistically obtained resources. Request execution is driven by a plan, constructed using techniques described above

Elements of the request execution environment may interact with the virtual data catalog, especially if active database technology is pursued. Consistent with our toolkit-based approach, we intent to construct the PVDG execution environment in terms of layered services, building on the extensive research done in Condor, SRB, Globus, and other systems. We take these results in new directions by focusing on four main topics: distributed execution agents, co-allocation agents, reliable event services, and interactions between execution and planning. The Florida, UW Madison, and San Diego (Marzullo) teams carry the main responsibility for this effort, led by Livny.

Distributed Execution Agents. Request execution requires control structures able to sequence the subtasks identified in the request execution plan. When components are pipelined, a control structure must initiate co-allocation of resources; when tasks are sequenced, a control structure must detect termination of one subtask and initiate the execution of the next. In addition, the control structure should detect possible failure and decide when alternatives should be pursued. Control of request execution is complicated by the fact that components may be executed across heterogeneous, distributed resources. Previous work on request instantiation has targeted simple task graphs, is contained within a single organization, and is driven from a centralized point[22]. Such a centralized approach is not appropriate for our purposes, because it is not scalable and does not work well in an environment that can partition. We plan instead to base our control structure on mobile agents[47,48,49]. For example, each subtask can have associated with it a mobile agent to monitor the progress of its subtask, modify the subtask's behavior to accommodate changes in the environment, and synchronize with other subtasks in the execution plan. We will build on the existing work in mobile agents and their runtime environments to construct a decentralized, scalable and fault-tolerant control structure. We expect that this concrete application of mobile agents will lead to new and practical protocols for mobile agent communications, agent planning, and agent fault-tolerance.

Co-allocation Agents. Execution of a request may proceed in stages. However, all but the simplest stages consist of a pipeline spanning multiple resources. Consequently, co-allocation of resources is a basic service that will be required by execution agents. Co-allocation is complicated by lack of availability, loss of reservation, and failures. To improve the probability of success, co-reservation should start with a set of alternative resources, and define a strategy by which these alternatives are tried. We will develop methods for optimizing co-allocation specifications based on desired cost and reliability, and develop strategies for mapping these specifications into specific resource sets. Initial strategies will include depth-first (backtrack on failure), breadth first (select first set of appropriate resources), and bounded breadth first, with both reserved and best-effort resources being considered.

Reliable Event Service. An essential element of the agent-based execution environment is the ability to detect events such as the completion or failure of a task, the availability of a new resource, or a subtask changing state. It is critical that there be a uniform, consistent, and reliable means for propagating events across the PVDG. To date, a wide area event notification infrastructure does not exist. (The Grid Forum's performance group is discussing the issue, and this research will feed into that process.) Constructing such a system is complicated by the diversity of events that one may need to generate, the need to propagate events across administrative domains, and the need to deliver events reliably. We will investigate how epidemic communication[50,51] and fault tolerance[52,53,54] techniques can be combined with event delivery systems to provide a general Grid event service.

Interaction with Planning The preceding discussion assumed that the request execution component operates from a fixed plan, with contingencies represented by alternatives. One can also operate request execution and planning concurrently, planning the next phase of request execution when the current phase is complete. This approach can be more robust to failure and may be more appropriate when request characteristics such as the volume of generated data cannot be determined *a priori*. Furthermore, the approach can adapt to changes in the environment, offering the possibility of a better solution. We intend to investigate mechanisms that will enable an execution agent and a planning agent to communicate and will develop concurrent planning and execution algorithms.

C.2.b. Application Experiments and Virtual Data Grid Prototyping

Our application experiments are designed both to stress test software systems developed within the IT research and Virtual Data Toolkit efforts and to improve understanding of how PVDG concepts can be used effectively within a particular scientific discipline. These experiments will rely on substantial testbeds to be constructed in partnership with the experiments and the NSF PACIs. These testbeds are described in detail under "Facilities."

C.2.b.1. SDSS

The SDSS project has been generating data since 1999 and already has produced 3 TB of data. By 2005, it will have detected over 200 million objects on the sky, each with a few hundred physical attributes. The availability of this data enables a wide variety of sophisticated statistical analyses: for example, (1) repeated correlation analyses over 100 million galaxies as a function of different cuts by color, type, etc, created by dynamical database queries on the Science Archive (requires the generation of Monte-Carlo catalogs hundred times larger in size and performing the same calculations)[55]; (2) maximum likelihood determination of the cosmological parameters over pixelized two/three dimensional maps of the galaxy distribution[56,57,58]; and (3) the statistical analysis of gravitational lenses involving re-processing of 10 million images to measure local shearing of the images[10,11,12]. Small-scale analyses of this sort have been performed; however, large-scale analyses are extremely complex and expensive in terms of computational and storage resources and so will benefit greatly from PVDG capabilities for their fullest realization.

SDSS is an important partner for the **GriPhyN** project for several reasons. Because its data will be available early, it will serve as an early testbed and driver for new tools. In addition, its computationally intensive analyses will serve to stress computational scheduling issues. **GriPhyN**-related SDSS activities are led by JHU.

C.2.b.2. LIGO

LIGO, like SDSS, will produce data well before the LHC experiments (by 2002). It represents a significant scaling up relative to SDSS in terms of scale and geographical distribution. LIGO data sources, computing facilities, and scientific users are widely distributed nationally and internationally. Optimal accessing and processing of the data demands a distributed grid of services, including tools to manage the storage and migration of over 100 Terabytes of data and the scheduling of Teraflop/s distributed computing resources. LIGO also introduces additional concerns that will serve to stress other aspects of virtual data grid technology. LIGO provides a continuous stream of data that should be replicated remotely. Given a network faster than the LIGO data rate, this "data diffusion" can be done in a non-continuous fashion, providing a good test case for grid policies and network-monitoring services.

Opportunistic computing pushes the planning and request execution services. LIGO data may contain very faint but regular signals from rotating compact objects such as neutron stars. Searching for these signals when the source sky position is unknown can use an essentially infinite amount of computing resources[59,60]. We plan to use **GriPhyN** services to run such jobs in deep background over many machines.

LIGO virtual data will result from requests such as "gravitational strain for 2 minutes around each of 200 gamma-ray bursts over the last year" (an attempt to find gravitational-wave signatures of such astrophysical events). Others include data streams that have been reduced, decimated, or whitened, and a fusion of data streams from multiple interferometers, including other gravitational-wave detectors. LIGO's participation in **GriPhyN** includes data analysis teams at UW Milwaukee and UT Brownsville, and is led by Caltech.

C.2.b.3. ATLAS and CMS

ATLAS and CMS, the two largest of the four LHC experiments, will begin to accumulate stored data at the rate of several Petabytes/year in 2005, increasing each year thereafter. Simulation data will be generated well before that data. The CMS and ATLAS Data Grid prototypes, described in C.1.a, will be used as testbeds to enable large-scale evaluation of critical issues associated with the implementation and use of the virtual data toolkit, such as: co-allocation of CPU, data handling and network resources; the nature and feasible extent of job self-description in terms of the required resources, to enable policy-driven prioritization; load-balancing within and across wide area networked sites; and compact dataset extraction (Terabytes from Petabytes), and transport where needed. The production tasks for processing and reprocessing data (experiment-wide and group-oriented) require high resource utilization and throughput, while the thousands of jobs per day supporting physicists' individual analyses require rapid turnaround, in minutes to a few hours. These production tasks provide an excellent opportunity to study usage patterns; verify/optimize the efficiency of access and the success in responding to bottlenecks and long transaction-times; and study the ability of the system to respond to a shifting workload.

ATLAS and CMS need planned **GriPhyN** capabilities and will find extracting physics from the data significantly more difficult and less productive without them. These experiments will work closely with **GriPhyN** and related projects and move quickly to incorporate successful results of this research into their distributed analysis systems. ATLAS and CMS participation in **GriPhyN** are led by Indiana and Florida, respectively.

C.2.c. Development of Virtual Data Toolkit

As discussed in Section C.1.d, a primary **GriPhyN** deliverable will be a suite of *virtual data services* and *virtual data tools* designed to support a wide range of applications. The development of this *Virtual Data Toolkit* (VDT) will enable the real-life experimentation needed to evaluate **GriPhyN** technologies. The VDT will also serve as a primary technology transfer mechanism to the four physics experiments and to the broader scientific community.

VDT development will be an ongoing activity throughout the five-year duration of the project, with major new Toolkit versions being released approximately yearly. (Interim releases will provide minor enhancements and correct problems.) Each VDT release will integrate new technologies developed by **GriPhyN** researchers with "best practice" technologies from other sources to achieve significant improvements in functionality, scalability, reliability, manageability, and performance. Technologies that we expect to import from external sources include high-speed storage management[24,26,61], parallel I/O[62], high-speed data movement, matchmaking[39], policy representation[40,41], digital library access[9], authentication/authorization[68,69], and scalable object database technologies.

We note that **GriPhyN** project participants have considerable experience with the development and support of sophisticated research software systems with large user communities: Condor[63], MCAT/SRB[25], and the Globus Toolkit[17] are just three examples of widely used packages with user communities in the 100s.

The following release plan summarizes our initial vision of the main features delivered each year.

VDT-1 (Basic Grid Services) provides an *initial set of grid enabling services and tools*, including security, information, metadata, CPU scheduling, and data transport. VDT-1 will support efficient operation on O(10 TB) datasets, O(100) CPUs, and O(100 MB/s) wide area networks and will build extensively on existing technology.

VDT-2 (Centralized Virtual Data Services) provides a *first set of virtual data services and tools*, including support for a centralized virtual data catalog, centralized request estimation, centralized request planning, network caching, and a simple suite of distributed execution mechanisms. Representation and exchange of local policies will be supported for network caches.

VDT-3 (Distributed Virtual Data Services) supports *decentralized and fault tolerant execution and management* of virtual data grid operation, via integration of distributed execution mechanisms able to select alternatives in the event of faults, agent-based estimation and monitoring mechanisms, and iterative request planning methods. This version will support O(100) TB datasets, O(10 TB) network caches, O(1000) CPUs, and O(400 MB/s) networks.

VDT-4 (Scalable Virtual Data Services) *scales virtual data grid operation to realistic magnitudes*, supporting applications involving widely distributed O(1 PB) datasets, O(100 TB) network caches, and O(10,000) CPUs.

VDT-5 (Enhanced Services) enhances VDT functionality and performance as a result of application experiences.

VDT development will be led by Livny, under the direction of the **GriPhyN** Technical Committee (see below).

C.3. Integration and Coordination Strategy; Management Plan

The participants have worked together extensively in the past and have, over many years, gained considerable experience with the successful management of projects of comparable scope to what is proposed here. The IT researchers have a strong history of past collaboration; most have also worked closely with the physics participants. The PIs and other senior personnel have extensive experience with the management and execution of large collaborative scientific projects, as a result of their leadership roles in the Nile[64] National Challenge project, Globus project, DOE PPDG project, and NSF PACIs, as well as major physics experiments.

In managing the **GriPhyN** project, we will need to take into account the number of institutions collaborating and the related but distinct goals in IT research and in capabilities for physics and astrophysics experiments. We recognize that we must achieve significant accomplishments in each area. Particular attention must be paid to managing interactions between IT and physics research and to coordinating interactions with the motivating physics applications. We discuss here our proposed management approach.

Project Structure: While the project involves many institutions and personnel from both IT and physics, we have structured subprojects so that coordination costs are manageable. IT research efforts are located at U.Chicago, USC/ISI, UW Madison, UCSD, and UCB. (Indiana and Florida also participate, thanks to matching funds.) Each IT subtask involves a subset of these institutions. Software development efforts are further concentrated at U.Chicago, USC/ISI, and UW Madison, with one postdoc also located at Caltech. The physicists include leading members of computing groups from four experimental collaborations. Again, effort is focused at a small number of institutions; others participate as unfunded collaborators.

Management Structure: Project management will operate as follows. PI Avery will serve as the primary point of contact for NSF and external relations and will provide overall scientific leadership. Co-PI Foster will chair a Technical Committee, charged with coordinating IT research, software development, application experimentation, and testbed activities. A full-time Project Coordinator will be hired to coordinate day-to-day operations.

The Technical Committee (TC) will comprise at least one person per collaborating site, capable of representing the site's principal scientific and technical contributions to the project. The TC is responsible for establishing and monitoring progress on regular integration goals and capability milestones, as well as coordinating the exchange of information on scientific progress. TC members act as local project managers and the TC will identify task forces for the many multi-site components of the project. The TC will meet monthly, mostly by teleconference or videoconference, as well as communicating via email and maintaining an active project Web site.

Internal Advisory Mechanism: We also propose to create a formal internal advisory mechanism charged with advising the PIs on long-term strategy relating to resource allocations, relationships with other projects, and other matters of strategic importance. This Collaboration Board (CB) will normally consist of the local project PI from each collaborating site. The CB will meet 2 or 3 times per year, mostly by videoconference.

External Advisory Committee: We plan to create an External Advisory Committee consisting of senior IT and physics researchers and managers. The following list indicates the types of people that we expect to enlist; we have not yet contacted them concerning their participation: Bill Johnston, LBNL; Paul Messina, Caltech; Dan Reed, UIUC; Joel Butler, Fermilab; Horst Simon, LBNL; Manuel Delfino, CERN; Ben Segal, CERN; Douglas van Houweling, UCAID; Jim Gray, Microsoft; Fran Berman, UCSD; Tom Coleman, Cornell.

C.4. Relationship to Other Projects

The crosscutting nature of the **GriPhyN** project means that our proposed research necessarily impinges on the concerns of many other groups. We discuss here how the proposed **GriPhyN** effort relates to, and proposes to leverage, other work, looking in particular at the IT research community, the NSF PACI program, the U.S. physics community, industry, and international projects. *We note that an exciting aspect of the **GriPhyN** effort is that the leadership role played by **GriPhyN** personnel in various communities means that it is feasible, essentially for the first time, to bring these projects together in a productive partnership*.

C.4.a. Information Technology Research Community

The ambitious goals proposed for **GriPhyN** are possible only because we are able to leverage and build upon a large body of prior work in data-intensive computing, distributed computing, networking, resource management, scheduling, agent technologies, and data grids. This necessarily brief review of this prior work explains both how we propose to leverage it and how our work will achieve advances in the state of the art.

The problem of managing large amounts of (scientific) data has received considerable attention[65,8,66], with **GriPhyN** participants centrally involved. This work provides us with advanced technologies for tertiary data management[24,26,61], metadata management[9], and high-speed data movement. **GriPhyN** is distinguished by its focus on the integrated management of large amounts of computation and data, in order to support complex physics analyses and to realize the concept of virtual data.

The research and prototyping activities proposed here complement R&D activities being performed under the general rubric of "Grids[67]." These Grid activities are developing basic mechanisms in such areas as security[68,69], remote computing[70], scheduling[71,72], data access[9,73], and instrumentation[74,75]; they provide building blocks that will facilitate virtual data grid R&D. Relevant development efforts include the NASA Information Power Grid[15] and projects concerned with data visualization corridors[76,77]. **GriPhyN** participants working in these areas will ensure that connections are established with **GriPhyN** research.

There are also commonalities of interest with research on networking, distributed databases, digital libraries, data mining, distributed operating systems[78], content distribution[79], and agent technologies. Research on distributed database and digital library focuses on enabling the federation of semantically diverse data sources and interoperability of services[80,81,82,83], while work on distributed data mining[84] emphasizes the ability to analyze massive datasets. **GriPhyN** can exploit this work but focuses on the organization of computationally intensive operations on distributed data and emphasizes the policy-driven management of those resources.

Research on mobile agents has recently focused on agent services such as security, communications, and fault-tolerance. However, research has been hampered by a lack of applications that clearly need the technology. PVDG management and the control structures of execution plans appear to be excellent candidates. **GriPhyN** should provide a long-needed testbed for mobile agent platforms.

C.4.b. NSF PACI Program

The two NSF Partnerships for Advanced Computational Infrastructure (PACIs) are developing a PACI Technology Grid[85], a national-scale network of computing and storage resources supporting standard Grid services. PACI efforts have provided early proofs of principle for several technologies to be exploited here and contribute software that **GriPhyN** can leverage. They can, in addition, provide an early operational testbed for the new technologies and application concepts to be developed within **GriPhyN**. We plan to establish a close collaboration with both PACIs. As stated in the letters of support, both PACIs are enthusiastic about such a relationship and will consider adopting **GriPhyN** as a (no-cost) application technology team to facilitate communication and cooperation.

C.4.c. U.S. Physics Community

The **GriPhyN** project includes senior representatives from the ATLAS, CMS, LIGO, and SDSS projects, which are committed to working with **GriPhyN** on the exploration and development of virtual data grid technologies. The support provided by these projects in terms of use case analysis, testbed construction and access, and experimental studies will be vital for the success of **GriPhyN**.

Existing Information Technology funding within the U.S. Physics community is focused on short-term data management issues (e.g., see recent work within the BABAR experiment[86]); this pragmatic focus complements the more ambitious goals of **GriPhyN**. For example, the Particle Physics Data Grid[87] (PPDG) project funded initially for one year by the DOE NGI program (follow-on funding has been requested from DOE High-Energy and Nuclear Physics and DOE Mathematical Information and Computational Sciences) is focused on applying existing grid middleware services to high-energy and nuclear physics problems. The co-PIs of the PPDG (Mount and Newman) and many PPDG senior personnel are members of the **GriPhyN** collaboration.

C.4.d. Industrial Collaborations

While the problems to be tackled in **GriPhyN** are motivated by unique scientific requirements and are well beyond the current state of the art, **GriPhyN** technologies should be of considerable relevance to U.S. industry. We have discussed our goals with U.S. companies and received encouraging responses. The letters of support from Cisco, Compaq, HP, IBM, SGI, and Sun suggest that we can expect fruitful collaborative relationships with industry.

C.4.e. International Connections

Each physics experiment within **GriPhyN** is a large international collaboration and **GriPhyN** participants have been centrally involved in the development of national and global networks in support of scientific research for the last 15 years. The technologies to be developed here will have a strong positive impact on the abilities of CMS, ATLAS, LIGO, and SDSS scientists to participate in and contribute to these collaborations. In addition, strong interest from international collaborators indicates that we can anticipate many opportunities for collaborative work on PVDG technologies. The links required to ensure coordination of R&D activities are already being formed: for example, Co-PI Foster has been asked to join the Board of a budding CERN "Data Grid" effort.

We also expect to see strong connections established between the **GriPhyN** project and NSF's international networking program, via for examples connections with the STAR-TAP group at UIC/EVL[88].

C.5. Broader Impact of Proposed Research

The development of virtual data grid technology is likely to have a significant impact on research and education well beyond the bounds of this specific project and its collaborators. The compelling nature of the problems to be solved and the technology to be developed generates enthusiastic responses from students and researchers alike. In addition, virtual data grid technology is inherently democratic. While a student or faculty member at a small school cannot easily travel to CERN, they can, with virtual data grid technology, access, analyze, and publish from LHC data. Similarly, information technologistis from diverse backgrounds can contribute in many different ways to the construction of virtual data grid testbeds and the execution of virtual data grid experiments.

C.5.a. Advancing Knowledge in Computer and Computational Science Disciplines

We view our ability to transfer **GriPhyN** knowledge and technology to the computer and computational science communities as being a critical measure of success. This transfer will occur via the development of a Virtual Data Toolkit, the prototyping of operational Petascale Virtual Data Grids, and the collaborations in which **GriPhyN** personnel already participate: e.g., the various physics experiments, the NSF PACIs, and the Grid community.

The **Virtual Data Toolkit** will provide an easily accessible snapshot of current "best practice" in virtual data grid technology. Its utility will be enhanced by its integration of external technologies (e.g., metadata, transport, security, etc.) and by a modular, service-oriented design that allows users to select only required functionality.

The deployment of operational **Petascale Virtual Data Grids** by the ATLAS, CMS, LIGO, and SDSS experiments provides a tremendous opportunity for successful technology transfer to these large communities. Furthermore, success in these contexts will provide a powerful advertisement for **GriPhyN** technology, hence encouraging adoption by other communities with similarly demanding requirements.

GriPhyN PIs are major contributors to the NSF PACIs, DOE Science Grid effort, and NASA's Information Power Grid project, and play leadership roles in the Grid Forum[89]. We will work closely with these projects to ensure technology transfer in both directions. We anticipate mutually beneficial interactions and the establishment of formal agreements that will allow collaboration on research, technology development, and testbeds.

C.5.b. Educational Merit: Advancing Discovery, Promoting Teaching and Learning

The **GriPhyN** project provides an outstanding opportunity to train the next generation of scientists who must necessarily work at the intersection of computer science and computational science. We intend to exploit this opportunity to the fullest. The postdoctoral associates and graduate students for whom funding is requested will all receive considerable training. Many subprojects are accessible to students from a wide range of backgrounds and with a wide range of skills. Subprojects can range from development of user-level interfaces and GUIs, building local resource management agents, exploring heuristics and optimization strategies for Grid utilization, up to the creation of architectures for virtual data. Experience and education research have proven the effectiveness of incorporating K-12, undergrad, and beginning graduate students into labs with an overarching project goal, in which they are given tasks that clearly fit into the larger picture. The **GriPhyN** project provides precisely the kind of exciting environment that can provide an accessible entry to research at the frontiers of computational science.

We also plan an ongoing program of *external* education and outreach designed to expose faculty and students at other institutions to **GriPhyN** research. *This program will engage all **GriPhyN** senior personnel, with each committing to lecturing and mentoring activities at other (particularly minority-serving) institutions.* To facilitate and coordinate these various activities, we propose to employ an outreach/education coordinator at UT Brownsville, responsible for (1) serving as an initial point of contact for other groups and universities interested in participating in **GriPhyN**; (2) organizing "how to" workshops/tutorials at various local and national conferences; (3) providing technical support (e.g., installation instructions, user's manuals, CVS repositories, etc.) for the **GriPhyN** software/toolkits; and (4) serving as an interface between one or more of the physics applications (e.g., LIGO) and **GriPhyN** IT researchers. We will also exploit existing outreach programs within the high energy physics and gravitational wave communities, such as QuarkNet[90] and the LIGO outreach program based at the Louisiana facility.

UT Brownsville is well suited for such an individual because it is a minority serving institution that already has close ties with LIGO, first as a member of the LIGO Scientific Collaboration and second as a recipient of an NSF grant (PHY-9981795) to integrate research and education between UT Brownsville and LIGO. By Spring 2001 UT Brownsville will have a 64-node Beowulf cluster for LIGO data analysis that can also serve as a **GriPhyN** testbed.

C.6. Results from Prior NSF Support

We summarize briefly significant results from prior NSF support obtained by selected **GriPhyN** senior personnel.

Bruce Allen (NSF 9728704, UWM: Data analysis tools, techniques and algorithms for gravitational wave detectors) completed and published the first end-to-end analysis of data from the Caltech 40-m prototype interferometric gravitational wave detector, searching for signals emitted by coalescing binary stars in our galaxy[91]. This is the first observational gravitational-wave astrophysical limit for such sources and a model for upcoming LIGO analysis. The UWM group wrote the code[92] and constructed the 48-node Beowulf cluster used for analysis[93].

Paul Avery (PHY-9318151: Nile National Challenge project) developed a scalable framework for distributing data-intensive applications, including a Java-2 based control system which uses CORBA.

Ian Foster (CCR-8809615: Center for Research in Parallel Computation) conducted research on parallel languages[94], the Nexus runtime system[95], dynamic resource management mechanisms, and multimethod communication[96]. CRPC-supported technologies were applied within the I-WAY networking experiment[97] and contributed to the Globus toolkit[17], which provides mechanisms for communication, resource management[98], security[69], data access[73], and information[99] in high-performance distributed environments. Globus is used extensively within the NSF PACI National Technology Grid and by many NSF-supported research projects.

Michael Franklin (IRI-95-01353 CAREER: Broadcast Disks: Data Management for Asymmetric Communications Environments) developed and extended the Broadcast Disks paradigm for combining broadcast scheduling and client caching to efficiently deliver data to large populations of clients, via dynamic client prefetching, update dissemination, and the integration of pull requests through the use of a client backchannel. The Dissemination-Based Information System (DBIS) toolkit[46] allows the construction of networks of "Information Brokers" that simplify the construction of large-scale dissemination-based services.

Carl Kesselman (CCR-8899615) (ASC 96-19020 Center for Research in Parallel Computation) conducted research in parallel programming languages[100] and the Nexus runtime system[95]. This funding also supported early work on Globus (see details under Foster).

Albert Lazzarini (NSF PHY-9210038: LIGO Laboratory data analysis tools, distributed computing and data archive) The LIGO Data and Computing Group is building a data analysis analysis system using a distributed parallel computing and data management paradigm. Database updates are generated at 3 geographically remote sites and are consolidated using a WAN. Job scheduling for computationally intense jobs is performed through a middle-layer software agent or "manager API" which sets job priorities and is aware of system resources.

Miron Livny (CDA-9726385: MetaNEOS) developed a framework for solving very large optimization problems with distributively owned opportunistic resources. The framework has been implemented as a runtime support library and used to solve open optimization problems.

Keith Marzullo (ASC-9318151: Nile HPCC Grand Challenge project) developed and deployed a distributed environment for the execution of highly data-parallelized HEP/CLEO jobs[42,101], developing novel scheduling methods and making significant progress on integration techniques of existing software tools for HEP. He also developed novel techniques for retrofitting fault-tolerance into a CORBA-based distributed application[53,54].

Reagan Moore (NSF ASC 96-19020: National Partnership for Advanced Computational Infrastructure - NPACI, 10/1/97 - 9/30/02) developed infrastructure to support data intensive computing by integrating archival storage, data handling systems[102], collections management[103], digital libraries[104], and data grids[105]. SRB[25] provides attribute-based access to data through a metadata catalog and supports execution of remote proxies through integration with the University of Maryland DataCutter system[106].

Alex Szalay (KDI-9980044: Large Distributed Archives in Phyics and Astronomy) in a collaborative effort between JHU (Szalay, Vishniac, Pevsner in Physics and Goodrich, CS) with Caltech (**Newman**, **Bunn** and Martin), Fermilab (Nash, Kasemann, Pordes) and Microsoft (Jim Gray) seeks to create a distributed database environment with intelligent query agents. The project is exploring different data organization to speed up certain types of distributed queries[107,108,109], comparing object-oriented and relational databases[110,111], and building intelligent query agents[112].

REFERENCES:

[1] M. Della Negra, Spokesperson, The CMS Experiment, A Compact Muon Solenoid, cmsinfo.cern.ch/Welcome.html.

[2] P. Jenni, Spokesperson, The ATLAS Experiment, A Toroidal LHC ApparatuS, atlasinfo.cern.ch/Atlas/Welcome.html.

[3] Abramovici A, Althouse WE, Drever RWP, Gursel Y, Kawamura S, Raab FJ, Shoemaker D, Sievers L, Spero RE, Thorne KS, Vogt RE, Weiss R, Whitcomb SE, Zucker ME, LIGO - The Laser-Interferometer-Gravitational-Wave-Observatory, *Science* 256: (5055) 325 (1992).

[4] B.C. Barish and R. Weiss, LIGO and the Detection of Gravitational Waves, *Phys. Today*, 52: (10), 44 (1999).

[5] The LIGO Experiment, http://www.ligo.caltech.edu/

[6] Szalay, A.S.: The Sloan Digital Sky Survey, *CiSE*, Vol 1, No 2, 54 (1999); see also http://www.sdss.org/.

[7] L. Barone et al. (The MONARC Architecture Group), *Regional Centers for LHC Computing*, MONARC-DOCS 1999-03, monarc.web.cern.ch/MONARC/docs/monarc_docs/1999-03.html (1999).

[8] Davis, F., W. Farrell, J. Gray, C. R. Mechoso, R. W. Moore, S. Sides, M. Stonebraker, EOSDIS Alternative Architecture, Final Report submitted to HAIS, Sept 6, 1994.

[9] Reagan Moore, Chaitanya Baru, Richard Marciano, Arcot Rajasekar, Michael Wan, Data-Intensive Computing. In *The Grid: Blueprint for a Future Computing Infrastructure*, pages 105-129. Morgan Kaufmann Publishers, 1999.

[10] Fischer, P., McKay, T.A., Sheldon, E., Connolly, A., Stebbins, A. and the SDSS collaboration: Weak Lensing with SDSS Commissioning Data: The Galaxy-Mass Correlation Function To 1/h Mpc, *Astron.J.* in press, 2000.

[11] Luppino, G.A. and Kaiser, N., *Ap.J.* 475, 20, 1997.

[12] Tyson, J.A., Kochanek, C. and Dell'Antonio, I.P., *Ap.J.*, 498, L107, 1998.

[13] Nelson Minar, Kwindla Hultman Kramer, and Pattie Maes, Cooperating Mobile Agents for Mapping Networks, *Proceedings of the First Hungarian National Conference on Agent Based Computation*, 1999.

[14] R. Aiken, J. Strassner, Brian Carpenter, I. Foster, C. Lynch, J. Mambretti, R. Moore, B. Teitelbaum, Terminology for describing middleware for network policy and services, Internet RFC, 1999.

[15] W. Johnston, D. Gannon, B. Nitzberg, Grids as Production Computing Environments: The Engineering Aspects of NASA's Information Power Grid, *Proc. 8th High Performance Distributed Computing Symposium*, IEEE Press, 1999; see also www.nas.nasa.gov/IPG

[16] A. S. Grimshaw, W. A. Wulf, The Legion Vision of a Worldwide Virtual Computer, *Communications of the ACM*, 40(1), 1997.

[17] I. Foster and C. Kesselman, The Globus Project: A Status Report, *Proceedings of the Heterogeneous Computing Workshop*, IEEE Press, 4-18, 1998; also see www.globus.org

[18] CORBA, www.omg.org

[19] C. Bauzer Medeiros, G. Vossen, M. Weske, WASA: A Workflow-Based Architecture to Support Scientific Database Applications (extended abstract), *Proc. 6th Database and Expert System Applications*, pp 574-583, 1997.

[20] J. Wiener and Y. Ioannidis, A Moose and a Fox Can Aid Scientists with Data Management Problems, *Proc. 4th International Workshop on Database Programming Languages*, New York, NY, August 1993, pp. 376-398.

[21] K Blackburn, A Lazzarini, T Prince R Williams, XSIL: Extensible Scientific Interchange Language, *Proc. HPCN99 conference*, Amsterdam, April 1999.

[22] G Aloisio, G Milillo, R Williams, An XML Architecture for High-Performance Web-Based Analysis of Remote-Sensing Archives, *Future Generation Computing Systems* 15 (1999) 687-698.

[23] Y. Ioannidis, M. Livny, S. Gupta, and N. Ponnekanti, ZOO: A Desktop Experiment Management Environment, *Proc. 22nd International VLDB Conference*, Bombay, India, September 1996, pp. 274-285.

[24] L. M. Bernardo, A. Shoshani, A. Sim,H. Nordberg, Access Coordination of Tertiary Storage for High Energy Physics Application, *MSS 2000*.

[25] Baru, C., R, Moore, A. Rajasekar, M. Wan,The SDSC Storage Resource Broker, Proc. CASCON'98 Conference, Nov.30-Dec.3, 1998, Toronto, Canada; also www.npaci.edu/DICE/SRB

[26] A. Shoshani, L. M. Bernardo,H. Nordberg, D. Rotem, and A. Sim, Multidimensional Indexing and Query Coordination for Tertiary Storage Management, *SSDBM* 1999.

[27] Roch Guerin and Henning Schulzrinne, Network Quality of Service. In *The Grid: Blueprint for a Future Computing Infrastructure*, pages 479-503. Morgan Kaufmann Publishers, 1999.

[28] I. Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt, A. Roy, A Distributed Resource Management Architecture that Supports Advance Reservations and Co-Allocation. *Intl Workshop on Quality of Service*, 1999.

[29] H. Newman, Worldwide Distributed Analysis for the Next Generations of HENP Experiments, *CHEP2000*, Padua, Italy, February 2000. http://chep2000.pd.infn.it/, paper number 385.

[30] Models of Networked Analysis at Regional Centers for LHC Experiments: www.cern.ch/MONARC

[31] H. Newman and I. LeGrand, The MONARC Toolset for Simulating Large Distributed Processing Systems. To appear in the *Proceedings of the 2000 Winter Simulation Conference, Orlando* (http://www.wintersim.org).

[32] Bajaj, et. al., Improving Simulation for Network Research, USC Computer Science Technical Report 99-702b, March 1999, University of Southern California, Los Angeles, California.

[33] Fujimoto, et. al., PDNS—Parallel/Distributed NS, http://www.cc.gatech.edu/computing/pads/pdns.html

[34] Prakash and R. Bagrodia, Using Parallel Simulation to Evaluate MPI Programs, *Proceedings of the 1998 Winter Simulation Conference*, WSC '98.

[35] J. Davis, R. Galicia, M. Goel, C. Hylands, E.A. Lee, J. Liu, X. Liu, L. Muliadi, S. Neuendorffer, J. Reekie, N. Smyth, J. Tsay and Y. Xiong, Ptolemy II: Heterogeneous Concurrent Modeling and Design in Java, Technical Report UCB/ERL No. M99/40, University of California, Berkeley, CA 94720, July 19, 1999. See http://ptolemy.eecs.berkeley.edu/ptolemyII/.

[36] S. Chen, K. Nahrstedt, Distributed Routing with Imprecise State Information, *IEEE Seventh International Conference on Computer Communications and Networks*, 1998.

[37] C. Lai, G. Medvinsky, B. C. Neuman, Endorsements, Licensing, and Insurance for Distributed System Services, *Proc. 2nd ACM Conference on Computer and Communication Security*, 1994.

[38] Matt Blaze, Joan Feigenbaum, and Martin Strauss, Compliance Checking in the PolicyMaker Trust Management System, *Proc. FC'98*, 1998.

[39] Rajesh Raman, Miron Livny, Marvin Solomon, Matchmaking: Distributed Resource Management for High Throughput Computing, *Proceedings of the Seventh IEEE International Symposium on High Performance Distributed Computing*, July 28-31, 1998, Chicago, IL.

[40] W. Johnston, C. Larsen, A Use-Condition Centered Approach to Authenticated Global Capabilities: Security Architectures for Large-Scale Distributed Collaboratory Environments, LBNL Techical Report #3885, 1996.

[41] T. Ryutov and C. Neuman, Representation and Evaluation of Security Policies for Distributed System Services, *Proc. of the DARPA Information Survivability Conference*, January, 2000.

[42] A. Ricciardi, M. Ogg and K. Marzullo. The Nile System Architecture. *Proceedings of the Eleventh International Conference on Systems Engineering*, 9-11 July 1996, Las Vegas, NV, USA, pp. 414-419.

[43] F. Berman and R. Wolski and S. Figueira, J. Schopf, G. Shao, Application-Level Scheduling on Distributed Heterogeneous Networks, *Proc. Supercomputing 96*, November, 1996.

[44] T. Malone, J. Yates, R. Benjamin, Electronic Markets and Electronic Hierarchies, *Communications of the ACM*, 30(6):484-497, 1987.

[45] S. Dar, M. Franklin, B. Jonsson, D. Srivastava, and M. Tan. Semantic Data Caching and Replacement, Proc. of the Twenty-Second International Conference on Very Large Data Bases, Bombay, India, September, 1996, pp 330-341.

[46] M. Altinel, D. Aksoy, T. Baby, M. Franklin, W. Shapiro and S. Zdonik, DBIS-Toolkit: Adaptive Middleware for Large Scale Data Delivery, (Demonstration Description), *Proc. ACM SIGMOD Int'l Conf. on Management of Data*, Philadelphia, PA, June, 1999, pp 544-546.

[47] Dejan Milojicic, Frederick Douglis, Richard Wheeler (Eds), *Mobility: Processes, Computers, and Agents*, ACM Press, 1999.

[48] Y Yang, O F Rana, D W Walker, C Georgousopoulos, R Williams, Mobile Agents on the SARA Digital Library, *Proc. IEEE Advances in Digital Libraries 2000*, Washington DC, May 2000.

[49] Brian Tierney, William Johnston, Brian Crowley, Gary Hoo, Chris Brooks, Dan Gunter, The NetLogger Methodology for High Performance Distributed Systems Performance Analysis, *HPDC'99*, 1999.

[50] K. P. Birman, M. Hayden, O. Ozkasap, X. Zhen et. al. Bimodal multicast. *ACM Transactions on Computer Systems*, May 1999,17(2):41-88.

[51] M. J. Lin and K. Marzullo. Directional gossip: Gossip in a wide area network. *Proceedings of the Third European Dependability Computing Conference* (EDCC-3), Prague, Czech Republic, 15-17 September 1999, Springer-Verlag LNCS 1667, pp. 364-379.

[52] S. Rao, L. Alvisi and H. M. Vin. Egida: an extensible toolkit for low-overhead fault-tolerance. *Proceedings of the 29th Annual International Symposium on Fault-Tolerant Computing*, Madison, WI, USA, 15-18 June 1999), pp. 48-55.

[53] Dag Johansen, Keith Marzullo, F. B. Schneider, Kjetil Jacobsen, and Dmitrii Zagorodnov. NAP: practical fault-tolerance for itinerant computations. In *Proceedings 19th IEEE International Conference on Distributed Computing Systems*, Austin, TX, USA, 31 May-4 June 1999, pp. 180-189.

[54] Matt Clegg and Keith Marzullo. Predicting physical processes in the presence of faulty sensor readings. In *Proceedings of IEEE 27th International Symposium on Fault Tolerant Computing*, Seattle, WA, USA, 24-27 June 1997, pp. 373-378.

[55] Kerscher,M., Szapudi,I. and Szalay,A.S.: An objective comparison of estimators for the two--point correlation function: dispelling the myths, *Astrophys. J. Lett.*, , in press (2000).

[56] Vogeley,M.S. and Szalay,A.S.: Eigenmode Analysis of Galaxy Redshift Surveys, Theory and Methods, *Astrophys. J.*, 465, 34 (1996).

[57] Tegmark,M., Hamilton,A.J.S., Strauss,M.A., Vogeley,M.S. and Szalay,A.S.: Measuring the Galaxy Power Spectrum with Future Redshift Surveys, *Astrophys. J.*, 499, 555 (1998).

[58] Matsubara,T., Szalay,A.S. and Landy,S.D.: Cosmological Parameters from the Eigenmode Analysis of the Las Campanas Redshift Survey, *Astrophys. J.*, in press (2000).

[59] P.R. Brady and T. Creighton, *Phys.Rev.* D61 (2000) 082001

[60] P.R. Brady, T. Creighton, C. Cutler, B.F. Schutz, *Phys.Rev*. D57 (1998) 2101-2116

[61] A. Shoshani, A. Sim, L. M. Bernardo, H. Nordberg, Coordinating Simultaneous Caching of File Bundles from Tertiary Storage, *Twelfth International Conference on Scientific and Statistical Database Management*, SSDBM 2000.

[62] Remzi H. Arpaci-Dusseau, Eric Anderson, Noah Treuhaft, David E. Culler, Joseph M. Hellerstein, David A. Patterson, Katherine Yelick, Cluster I/O with River: Making the Fast Case Common, *IOPADS '99*, 1999.

[63] The Condor Project, High Throughput Computing, www.cs.wisc.edu/condor.

[64] D. Cassel, PI, The Nile Project, Scalable Solution for Distributed Processing of Independent Data, www.Nile.cornell.edu.

[65] Moore, R., J. Lopez, C. Lofton, W. Schroeder, G. Kremenek, M. Gleicher, Configuring and Tuning Archival Storage System, *Proceedings of the 16th IEEE Symposium on Mass Storage Systems*, March 1999.

[66] Ann Chervenak, Ian Foster, Carl Kesselman, Charles Salisbury, Steven Tuecke, The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Data Sets, *Proc. NetStore*, 1999.

[67] I. Foster and C. Kesselman, editors. *The Grid: Blueprint for a Future Computing Infrastructure*. Morgan Kaufmann Publishers, 1999.

[68] Cliff Neuman, Security. In *The Grid: Blueprint for a Future Computing Infrastructure*. Morgan Kaufmann Publishers, 1999.

[69] I. Foster and C. Kesselman and G. Tsudik and S. Tuecke, A Security Architecture for Computational Grids, *ACM Conference on Computers and Security*, 83-91, ACM Press, 1998.

[70] Miron Livny and Rajesh Raman, High Throughput Resource Management. In *The Grid: Blueprint for a Future Computing Infrastructure*, pages 311-337, Morgan Kaufmann Publishers, 1999.

[71] Fran Berman, High-Performance Schedulers. In *The Grid: Blueprint for a Future Computing Infrastructure*, pages 279-309. Morgan Kaufmann Publishers, 1999.

[72] K. Czajkowski, I. Foster, and C. Kesselman, Co-Allocation Services for Computational Grids. Proceedings of the IEEE Symposium on High-Performance Distributed Computing, 1999.

[73] Joseph Bester, Ian Foster, Carl Kesselman, Jean Tedesco, Steven Tuecke, GASS: A Data Movement and Access Service for Wide Area Computing Systems, Proc. IOPADS'99, ACM Press, 1999.

[74] R. Wolski, Forecasting Network Performance to Support Dynamic Scheduling Using the Network Weather Service, Proc. 5th IEEE Symposium on High-Performance Distributed Computing, IEEE Press, 1996.

[75] Tony DeWitt, Thomas Gross, Bruce Lowekamp, Nancy Miller, Peter Steenkiste, Jaspal Subhlok, ReMoS: A Resource Monitoring System for Network Aware Applications, Carnegie Mellon University, CMU-CS-97-194, 1997.

[76] Smith, P. H., J. van Rosendale, Data and Visualization Corridors, Report on the 1998 DVC Workshop Series, http://www.cacr.caltech.edu/Publications/DVC/

[77] I. Foster, J. Insley, C. Kesselman, G. von Laszewski, M. Thiebaux, Distance Visualization: Data Exploration on the Grid, *IEEE Computer*, December 1999.

[78] Amin Vahdat, Eshwar Belani, Paul Eastham, Chad Yoshikawa, Thomas Anderson, David Culler, Michael Dahlin, WebOS: Operating System Services For Wide Area Applications, *7th Symposium on High Performance Distributed Computing*, 1998.

[79] M. Beck and T. Moore, The Internet2 Distributed Storage Infrastructure Project: An Architecture for Internet Content Channels, *Computer Networking and ISDN Systems*, 30(22-23):2141-2148, 1998.

[80] Sergey Melnik, Hector Garcia-Molina, Andreas Paepcke, A Mediation Infrastructure for Digital Library Services, http://www-diglib.stanford.edu/cgi-bin/get/SIDL-WP-1999-0126

[81] C. Baru, R. Frost, R. Marciano, R. Moore, A. Rajasekar, M. Wan , Metadata to Support Information-Based Computing Environments, 2nd IEEE Conference on Metadata (METADATA'97), September, 1997, Greenbelt, MD

[82] Chaitanya Baru and Arcot Rajasekar, A Hierarchical Access Control Scheme for Digital Libraries, Proc. 3rd ACM Conf. on Digital Libraries, June, 23-25, 1998, Pittsburgh, PA

[83] Arms, W., C. Blanchi, E. Overly, An Architecture for Information in Digital Libraries, D-LIB magazine, Feb 1997, http://www.dlib.org/dlib/february97/cnri/02arms1.html

[84] E.g., National Scalable Cluster Project, http://nscp.upenn.edu/data_mining.htm

[85] Stevens, R., Woodward, P., DeFanti, T., & Catlett, C., From the I-Way to the National Technology Grid. Communications of the ACM, 40(11), 51-60, (1997).

[86] BaBar Collaboration, www.slac.Stanford.edu/BFRoot/www/Computing/offline/Databases/www/homepage.html

[87] H. Newman, R. Mount, co-PI's, The Particle Physics Data Grid, www.cacr.caltech.edu/ppdg

[88] STAR-TAP, www.startap.net

[89] Grid Forum, www.gridforum.org

[90] QuarkNet, http://quarknet.fnal.gov

[91] B. Allen, K. Blackburn, P. Brady, J. Creighton, T. Creighton, S. Droz, A. Gillespie, S. Hughes, S. Kawamura, T. Lyons, J. Mason, B.J. Owen, F. Raab, M. Regehr, B. Sathyaprakash, R.L. Savage, Jr., S. Whitcomb, A. Wiseman, Observational Limit on Gravitational Waves from Binary Neutron Stars in the Galaxy, Phys. Rev. Lett. 83 (1999) 1498. Also available at xxx.lan.gov/abs/gr-qc/9903108.

[92] B. Allen and contributors, Gravitational Radiation Analysis and Simulation Package (GRASP), a software package for analysis of interferometric data, available at www.lsc-group.phys.uwm.edu/~ballen/grasp-distribution/. Version 1.9.7 is dated 1999.

[93] The UWM beowulf system is currently composed of 48 alpha 21164 machines with 9 Gbytes of distributed RAM and a 29 Gflop aggregate peak throughput. It is documented at www.lsc-group.phys.uwm.edu/~www/docs/beowulf.

[94] I. Foster, Compositional Parallel Programming Languages, *ACM Trans. Prog. Lang. Syst.*, 18(4):454--476, 1996.

[95] I. Foster, C. Kesselman, S. Tuecke, The Nexus Approach to Integrating Multithreading and Communication, *J. Parallel and Distributed Computing*, 37:70--82, 1996.

[96] I. Foster, J. Geisler, C. Kesselman, S. Tuecke, Managing Multiple Communication Methods in High-performance Networked Computing Systems, *Journal of Parallel and Distributed Computing*, 40, 35-48, 1997.

[97] 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

[98] K. Czajkowski, I. Foster, N. Karonis, C. Kesselman, S. Martin, W. Smith, and S. Tuecke. A resource management architecture for metacomputing systems. In The 4th Workshop on Job Scheduling Strategies for Parallel Processing, pages 62-82. Springer-Verlag LNCS 1459, 1998.

[99] S. Fitzgerald, I. Foster, C. Kesselman, G. von Laszewski, W. Smith, and S. Tuecke, A Directory Service for Configuring High-performance Distributed Computations, *Proc. 6th IEEE Symposium on High-Performance Distributed Computing*, 365-375, IEEE Press, 1997

[100] K.M. Chandy, C.F. Kesselman, C++: A Declarative Concurrent Object Oriented Programming Notation, *Research Directions in Object Oriented Programming*, MIT Press, 1993.

[101] Alessandro Amoroso, Keith Marzullo and Aleta Ricciardi. Wide-area Nile: a case study of a wide-area data-parallel application. In *Proceedings of 18th International Conference on Distributed Computing Systems*, Amsterdam, Netherlands, 26-29 May 1998, pp. 506-515

[102] Rajasekar, A., R. Marciano, R. Moore, Collection Based Persistent Archives, *Proceedings of the 16th IEEE Symposium on Mass Storage Systems*, March 1999.

[103] Chaitanya Baru, Managing Very Large Scientific Data Collections, poster session, *5th International Conference on High Performance Computing HiPC'98*, Dec. 17-20, 1998, Chennai, India.

[104] C. Baru, A. Gupta, V. Chu, B. Ludäscher, R. Marciano, Y. Papakonstantinou, P. Velikhov, XML-Based Information Mediation for Digital Libraries, demonstration session, *Digital Libraries '99*, http://www.npaci.edu/DICE/Pubs/

[105] B. Ludäscher, Y. Papakonstantinou, P. Velikhov, Navigation-Driven Evaluation of Virtual Mediated Views, *Intl. Conference on Extending Database Technology* (EDBT), Konstanz, Germany, © LNCS, Springer, 2000.

[106] Michael D. Beynon, Renato Ferreira, Tahsin Kurc, Alan Sussman, Joel Saltz, DataCutter: Middleware for Filtering Very Large Scientific Datasets on Archival Storage Systems, In *Proceedings of the Eighth Goddard Conference on Mass Storage Systems and Technologies*, College Park, MD, March 27-30, 2000.

[107] Chaudhary,A., Duncan,C., Goodrich,M. and Szalay,A.S.: Generalized Balanced Aspect Ratio Trees in R^2, *4th CGC Workshop on Computational Geometry*, (1999).

[108] Szalay, A.S., Kunszt, P.Z., Thakar, A., Gray, J., and Slutz, D.: The Sloan Digital Sky Survey and its Archive, *Proc. ADASS 99*, in press (1999).

[109] Kunszt, P.Z., Szalay, A.S., Csabai, I., and Thakar, A.: The Indexing of the SDSS Science Archive, *Proc. ADASS 99*, in press (1999).

[110] Szalay, A.S., Kunszt, P. Thakar, A., and Gray, J.: Designing and Mining Multi-Terabyte Astronomy Archives: The Sloan Digital Sky Survey, Microsoft Technical Report, MS-TR-99-30, (1999).

[111] Szalay, A.S., Kunszt, P. Thakar, A., Gray, J., Slutz, D., and Brunner, R.J.: Designing and Mining Multi-Terabyte Astronomy Archives: The Sloan Digital Sky Survey, *SIGMOD 2000 Conference*, (2000).

[112] Thakar, A., Kunszt, P.Z. and Szalay, A.S.: Multi-threaded Query Agent and Engine for a Very Large Astronomical Database, *Proc. ADASS 99*, in press (1999).