ITR: A Global Grid-Enabled Collaboratory for Scientific Research

We propose to develop, prototype, test and deploy the ***first Grid-Enabled Collaboratory for Scientific Research (GECSR)*** on a global scale. A distinguishing feature of this proposal is the tight integration between the science of collaboratories, a globally scalable working environment built on the foundation of a powerful fully functional set of working collaborative tools, and an agent-based monitoring and decision-support system that will allow collaborating scientists to perform data intensive analysis tasks efficiently. Assessment of the methodology of scientific collaborations and the iterative evaluation of the tools by a team independent of the developers will be a critical element ensuring the success of the proposed work. The assessment will focus on the ability of the target communities (High Energy and Nuclear Physics, other sciences and eventually other fields of research) to collaborate one-on-one and in groups on a variety of scales across long distance networks; it will abstract requirements and barriers to effective communication and shared work; and it will aim to develop standardized, broadly applicable guidelines and tools for effective collaboration in a variety of working contexts.

The initial targeted early-adopter community will be the major collaborations of experimental High Energy and Nuclear Physics (HENP), who face unprecedented challenges in accessing, processing, and sharing Petabyte-scale data, and who have successfully developed some of the most ambitious Grid systems[1,2,3,4] in cooperation with leading computer scientists in the US and Europe, as well as developing scalable videoconferencing tools and the early components of a Grid-enabled Analysis Environment (GAE[5]).

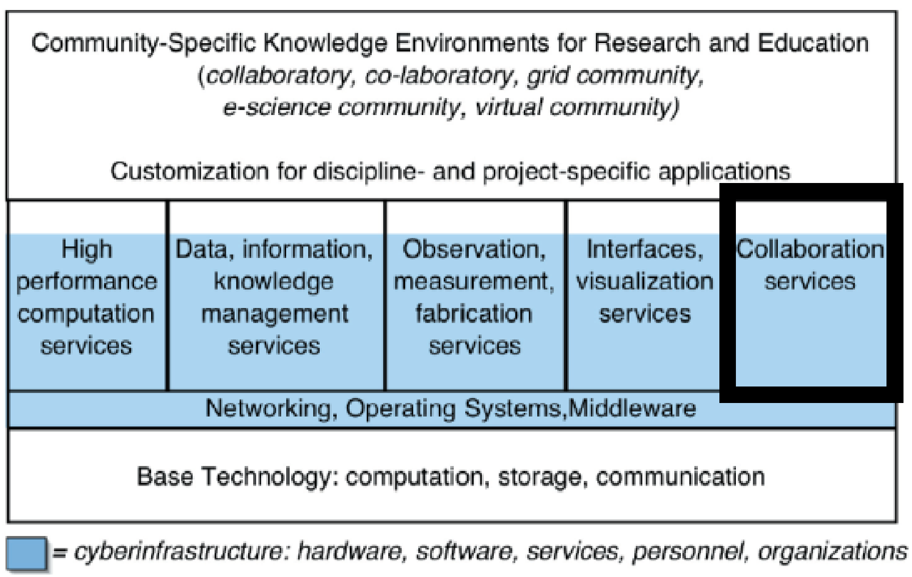


Figure 1: Integrated cyberinfrastructure services to enable new knowledge environments [6]

The recent report of the NSF Blue-Ribbon Advisor Panel on Cyberinfrastructure[6] identifies five key service categories that will provide a foundation for the comprehensive knowledge environments that will enable individuals, teams and organizations to revolutionize scientific practice (see Figure 1 above). The HENP community has addressed four of these services, based largely on common underlying middleware, in the widespread deployment of grid-enabled high performance computing resources, through a number of data grid projects that will facilitate management of data, information and knowledge[7], through instruments that can easily be monitored remotely through eLogs and other interfaces, and through the developing GAE[5]. The proposed Grid-Enabled Collaboratory for Scientific Research will provide the collaboration services cyberinfrastructure that is required for the HENP community to fully realize a functionally complete research environment that can revolutionize how and what physicists can do and who can participate

HENP's Data Intensive Challenges

The major HENP experiments of the next twenty years will break new ground in our understanding of the fundamental interactions, structures and symmetries that govern the nature of matter and spacetime in our universe. Among the principal goals at the high energy frontier are to find the mechanism responsible for mass in the universe, and the "Higgs" particles associated with mass generation, as well as the fundamental mechanism that led to the predominance of matter over antimatter in the observable cosmos.
The largest collaborations today, such as CMS[8] and ATLAS[9] who are building experiments for CERN's[10] Large Hadron Collider (LHC; [11]) program, each encompass 2000 physicists from 150 institutions in more than 30 countries. Each of these collaborations include 300-400 physicists in the US, from more than 30 universities as well as the major US HEP laboratories. The current generation of experiments now in operation and taking data at SLAC[12] and Fermilab (D0[13] and CDF[14]) are similar in scale to the US contingent of the next-generation experiments. Each of these experiments faces unprecedented challenges in terms of:

- The data-intensiveness of the work, where the data volume to be processed, distributed, accessed and analyzed by a major experiment are in the Petabyte (10^{15} Bytes) range now, and are expected to rise to the Exabyte (10^{18} Bytes) range within the next ten years.
- The complexity of the data, particularly at the LHC where the physics discovery potential is related to the very high intensity (luminosity) as well as the high energy of the collisions, such that ~20 interactions accompany the particle interaction of interest
- The global extent and multi-level organization of the physics Collaboration, leading to the need to collaborate and share data-intensive work in fundamentally new ways.

Addressing this last point forms the basis of this proposal. The new paradigm of "Grids" and grid-computing[15] is thought to hold the key to addressing the computing and data-management needs of HENP. There are significant efforts underway that are exploring and developing the grid toolkits and middleware which will be required for success in HENP, yet the hardest problems are not connecting and enabling resources like networks, computers and storage, but rather in effectively and efficiently connecting and enabling physicists to do their science with these new capabilities.

HENP physicists already perform experiments and analyses in tightly coupled cooperating groups. The collaborations can comprise thousands of people, with day-to-day research conducted in smaller teams that work closely together and then share their results with the larger collaboration for verification and further analysis. These teams range in size from very small groups of 1-5 physicists up to groups containing hundreds of physicists. These teams are collaborative and often competitive with one another. Although in the past they, and the larger collaboration, maintained contact through weekly meetings at the experiment site, the worldwide scope and shear size of newer HENP collaborations such as CMS and ATLAS make weekly face-to-face meetings unrealistic.

The team of physicists working together on an analysis, calculation, or simulation problem needs to be able to securely share and discuss work processes, data, results, and observations. Physics research is not an 8-5 endeavor. Experiments run on a 24x7 schedule and the data is distributed to centers throughout the world for analysis as it is generated. Although physicists work closely together, they also spend a significant portion of their time working separately. Thus a suitable collaborative environment must support both connected and disconnected work.

In addition to the HENP community's leading role in the social and technical infrastructures for collaboration, the LHC experiments' specific focus on secure and timely flow of large volumes of data presents challenges for collaboratory design and use. HENP scientists have successfully developed some of the most ambitious Grid systems[1,2,3,4] in cooperation with leading computer scientists in the US as well as Europe, as well as developing scalable videoconferencing tools[16] and the early components of a Grid-enabled Analysis Environment (GAE[5]). Solving the problem of how to support distributed collaborations which use massive datasets is critical for high-energy physicists, but meeting this challenge is emerging as a universal issue. For instance, recent expert reports from NSF[6] and from NIH[17] stress that scientific progress in a variety of fields will demand tools and capabilities, often termed "cyberinfrastructure", that can accommodate the production and manipulation of unprecedented amounts of data. While this proposal emphasizes the needs of CMS and ATLAS, there is a deep conviction that experiences with these communities will generalize to other domains. However, to effectively generalize from experiences in the HENP context will require an ongoing effort to systematically capture lessons learned, including both successes and failures. Therefore, an important element of the proposed project is the continuing evaluation of the Collaboratory to advance a general understanding of the impact of collaboratory tools on scientific practice, with specific attention to the role of collaboratories in supporting collaborative data-intensive research. Indeed, in the largest international scientific enterprises, Collaboration is probably the most vital part of an effective cyberinfrastucture for research.

The CMS and ATLAS collaborations are an ideal target for developing and evaluating a collaborative environment due to the way the HENP community is fundamentally organized, the presence of a high performance networking infrastructure and the data-centric focus of collaborative activity. The LHC collaborations present a unique opportunity for studying collaboration and collaboration tools as they are the first to face the challenges presented in pursuing global-scale computationally-intense science, leading the way for many disciplines to follow.

State of the Art of Collaboratory Tools

A broad range of collaboratory tools and environments have been developed over the last decade to support human-to-human interactions and remote scientific investigation. These tools include videoconferencing capabilities [18,19], messaging facilities[20], remote instrument access[21], and Grid computing efforts[22] among others. Some of the tools that have been developed recently to support scientific collaboration are described below. These tools are some of the leading collaborative tools available today and most have been developed by or with the help of members of the team of collaboratory developers involved in this proposal.

A CHEF Collaborative Framework: The CHEF project [23] is an active and long-term project at the University of Michigan, where it is used as the enterprise-wide learning management system and collaborative framework. CHEF (CompreHensive collaborative Framework) is used as the portal framework for the NSF-funded NEESGrid[22] project. CHEF makes use of a number of Apache/Jakarta tools including: Apache Web Server[24], Tomcat Servlet Container[25], Jetspeed portal framework[26], and Turbine services framework[27].
To these base technologies, CHEF has added a number of capabilities:

- Support within Jetspeed for groups of users in addition to individual users.
- A Portlet development methodology which decomposes Portlets into their presentation components (Teamlets) supported by persistent services accessed via a standardized API.
- Extended Jetspeed login authentication to use the Grid[28] as an authentication provider. The proxy credential is also stored within Jetspeed to allow other (non-Teamlet) Portlets to access the user's proxy credentials in order to perform Grid operations.

In addition to this Grid-enabled portal framework NEESGrid provides a number of pre-integrated Teamlets and Portlets which are automatically available for every group which is created within CHEF including: an announcement capability, persistent chat, a shared calendar, role-based access control, a threaded discussion system and a number of tools for interacting with Grid resources (e.g. GridFTP, LDAP browser). The CHEF framework development efforts are currently funded at Michigan in the NEESGrid project and the CHEF project will continue to grow and incorporate a number of tools and capabilities being developed in that context.

VRVS[18] – The VRVS system is already well developed, and relatively mature. It provides scalable desktop videoconferencing and interface to H.323 commercial products.

WLAP[29] – A system for the recording and web playback of audio, video and PowerPoint slides. It is based on the Synco-mat application developed by Charles Severance and extended by the Atlas Collaboratory Project at the University of Michigan. It's use requires only a web browser and a free version of RealPlayer. It has been tested over a period of several years in recording the CERN Summer Student Lectures, and is now the principal tool used in recording Atlas software training for its complex detector description and software architecture applications. It currently represents one the highest quality, open source web lecture archiving tools available. The deployment in ATLAS is made possible through support of the NSF Physics Division and US ATLAS..

Access Grid[19]- The access grid (AG) is a relatively new development that supports group-to-group interaction via videoconferencing tools. A primary feature of the AG is the concentration on providing excellent audio hardware and large projection screens. This environment supports natural conversations with all the videos projected in a many-to-many conference. VRVS interoperates with the Access Grid and provides scalable access to an AG along with interfacing of H.323 hardware.

VNC [30] – The Virtual Network Computer system for sharing windows and applications.

Pervasive Collaborative Computing Environment[31] – The Pervasive Collaborative Computing Environment (PCCE), at LBNL, is building collaboration tools that support connecting people to work together on an ad hoc or continuous basis. Tools that support the day-to-day connectivity and underlying needs of a group of collaborators are important for providing light-weight, non-intrusive, and flexible ways to stay in touch and work together. Tools available to date include a secure presence and messaging tool and collaborative computational workflow tools.

Peer-to-peer file sharing system[32] - In a typical scientific collaboration, there are many different locations where data would naturally be stored. The Scalable and Secure Peer-to-Peer Information Sharing Tool Project at LBNL is developing a lightweight file-sharing system that makes it easy for collaborators to find and use the data they need. This system is easy-to-use, easy-to-administer, and secure. It allows collaborating groups to form ad hoc and share files from local systems and archives. An XML-based Resource Discovery Messaging Framework (RDMF) based on a reliable and secure group communication protocol (www-itg.lbl.gov/CIF/GroupComm) provides resource discovery.

MonaLisa[33] **-** A monitoring framework and basis for a state-of-the-art collaborative system, reflective of the future system architectural design: a multithreaded, auto-discovering "services architecture".

Approach

The design of software systems in general, and of collaborative systems in particular, is most effective when carried out in a series of incremental stages[34, 35, 36, 37, 38]. To enable us to achieve the most effective environment, given the resources, we plan to pursue an iterative process for the design and deployment of the integrated knowledge environment. These stages consist of requirements analysis, tool development, tool deployment, and evaluation – with the typical evolution of a system moving iteratively through the stages several times. The process of progressing through each iteration focuses on five key activities.

To determine the most appropriate tools for the HENP collaboration environment, the initial step required is a study of the collaboration patterns of the physicists. This study will allow us to determine the most critical tools required in the physicist's collaboration environment. We expect that the final collaboration environment will be a persistent space which allows participants to locate each other, use asynchronous and synchronous messaging, share documents, share analyses and results, share applications, search for relevant information, and hold videoconferences. It will leverage existing and developing tools such as Grid services[38], the Pervasive Collaborative Computing Environment (PCCE), the Virtual Rooms Videoconferencing System (VRVS), shared physics analysis tools such as GAE, and possibly electronic notebooks.

The environment will be deployed incrementally with the high impact tools, as determined by the study, deployed first. Throughout the lifetime of the project, continued evaluations will be performed to assess the impact of deployed tools and the needs of the collaborators as they evolve.

The following sections detail the concepts, methods and deliverables that will shape each of these steps over the course of the project. We believe that this type of iterative process will be critical to the delivery of a useful integrated collaborative knowledge environment due to the anticipated changes in the nature of work that an LHC physicist engages in over the course of the project.

Assessment

In order to ensure that a useful environment is produced, it will be critical to work closely with end-users of the system to identify their needs. Researchers from the Collaboratory for Research on Electronic Work (CREW) at the University of Michigan School of Information will carry out this work. While physicists ultimately have the most complete knowledge of the functionality required in a collaboration environment for physics, it is necessary to rely on a team of researchers with expertise in social science methods and requirements analysis who are experienced in distilling common themes from diverse and complicated feedback provided by end users. In addition to working closely with users in order to understand the functionality that the system must provide, the assessment team plays a critical role in the adoption of the environment by providing project visibility to target users very early in the design process, managing expectations, and keeping the end users engaged. We propose to conduct this assessment process over the entire length of the project. Due to the size of the collaborations involved as user communities, it will be necessary to focus data collection effort on different elements of the collaboration in different years. Additionally, the requirements will change dramatically in response to each generation of the collaborative environment. The ongoing assessment effort will produce a high-level requirements document summarizing findings and a functional specification, aimed at and written with the development team, each year. We anticipate that findings from this work will also be disseminated in peer-reviewed journals and conferences.

Several methods will be employed to assess the community's collaborative tool needs. Surveys, likely web-based, provide the assessment team with data from a large cross-section of the community on a fixed set of questions that can be analyzed using quantitative methods. We anticipate developing a questionnaire to elicit information about scientific practice, including patterns of communication and collaboration, as well as information about the technical environment from a large sample of physicists and graduate students involved in the ATLAS, CMS, D0, and CDF collaborations. Interviews and site visits, while more labor intensive, provide much richer data than can be gained in surveys. We will visit a number of U.S. institutions involved in the CMS and ATLAS collaborations in order to observe first-hand the working environment and to be able to interact face-to-face with physicists and students. We will visit a representative sample of institutions in the various collaborations, including key research universities, top national laboratories, and teaching institutions, in order to assess the needs of all members of the community, from researchers to undergraduate physics students. We will also rely on collected data regarding current and future tool use to gain additional insight about attitudes towards collaborative tools and to identify factors that facilitate adoption. The assessment effort subsequent to the first release will focus heavily on understanding the use of the deployed tools. Relying on several data sources will allow us to identify the needs of the HEP community and understand which needs are most urgent.

This assessment phase is critical to the success of the collaborative environment by integrating social scientists into the development process. Several efforts to develop collaborative tools have failed to gain wide adoption because of a lack of attention to the underlying social science that influences use (e.g. NCSA Habanero[39] and NCSA Collage[40]). Without a thorough assessment of user requirements and a careful and ongoing analysis of the social and human factors of the collaboration environment, this project faces the risk of developing into a technology-driven environment that fails because it does not take decades of social science research into account.

Evaluation

The elements of the iterative process described so far – assessment, development and deployment – will give us a clear picture of how successful we were in developing an integrated collaborative knowledge environment for the HEP community. Additional work is needed, however, to assess the impact of the environment on the community and on the way science is done in the community. This type of analysis is critical to developing a better understanding of how collaborative environments can transform how science is done, who can participate and what types of problems can be solved.

We propose a four-year program of evaluation to identify the effects of deploying a collaboration environment in a scientific community. This effort will be lead by the Collaboratory for Research on Electronic Work (CREW) at the University of Michigan School of Information and will work closely with the assessment team to streamline data collection and understand the nature of scientific work being conducted in the community. The evaluation team will be responsible for developing the instruments and protocols necessary to gather the data to assess how the collaborative tools impact performance. The assessment team, as part of their data collection effort, will administer surveys and interviews that include measures for the evaluation effort in order to limit duplication of effort, while analysis of this evaluation data will be the responsibility of the evaluation team. Additionally, the evaluation team will rely on publication-based metrics of performance, such as co-authorship and citation impact, correlated with various measures of collaboratory use, to assess the scientific significance of collaboratory tools. The exact set of performance data needed for this kind of evaluation is one of the research goals of the NSF ITR-funded Science of Collaboratories (SOC)[41] project at the School of Information and we anticipate leveraging expertise from the SOC team in the design of the evaluation for the LHC collaboratory project. We propose that this evaluation take place over the entire four years of the project in order to obtain pre-tool baseline data and to assess the changes with regard to that baseline with each package release.

The results of this evaluation will be shared with the larger scientific community and will inform the development of future tools, environments and collaborations in other science, engineering and education domains. The findings of this team will be distributed as technical reports and in peer-reviewed journals and publications.

Development

Delivering a Global, Grid-Enabled Collaboratory for Scientific Research

We have identified a number of applications, development efforts, and a framework (discussed previously) that already exist, and are being actively used for collaboration. These tools will be integrated and form the basis of our intended first year package. Capabilities that have been identified as important for this release include a presence indication and directory service, text messaging, shared windows, integrated videoconferencing, and a shared file system. The CHEF framework will be integrated with these components. Some specific enhancements to these tools and development efforts planned under this proposal include the following:

VRVS [18] – The VRVS system is already well developed, and relatively mature. The main thrust of additional work in this area would be to:

- Work with the CHEF/NEESGrid team to integrate VRVS with the CHEF system, thereby enabling a convenient single GUI for the user who wishes to engage in collaborative activities.
- Add the capability to integrate VO Management Tools in VRVS within the CHEF framework, so that users can create both ad-hoc peer groups and associate them with transient or persistent Virtual Rooms.
- Enable an option such that a transient Virtual Room can be converted to a Persistent Virtual Room, which would imply that a certain level of archiving be applied to the various data sources present in the room. Persistent Rooms would remain "open for visits" indefinitely. The contents of the Persistent Rooms would be session data such as indexed MPEG audio/video streams etc.
- Provide mobile handheld (e.g. PDA) access to the VRVS system. A first version of a handheld client will be available within the first Project year.
- Improve and customize VNC (or some derivative) as part of the VRVS system, which would allow users to share specific application windows on their desktops, or their entire desktop. Thus, a text editor such as vi or Word could be shared among any number of collaborators.
- Add instant messaging facilities, or integrate existing IM facilities, or integrate tools like Trillian (which aggregates IRC, AOL, ICQ, MSN etc.) into VRVS

Quadtree-based Replica Management Service– An adaptive quadtree approach is proposed to support replica management service in a dynamic file sharing system within a VO or across VOs. It is anticipated that Grid-based collaborative LHC analysis will require frequent data replication and distribution based on dynamic data access pattern, continuing generation of new data, and diverse scopes of human-to-human collaborations. Current Grid middleware (e.g., GDMP) does not support regional and global optimization of dynamically distributing, replicating, and sharing large LHC datasets on the Grid. A preliminary study[42] has demonstrated that a quadtree-based algorithm can adaptively decompose and share datasets on the Grid based on their characteristics such as frequency of file access. The quadtree approach will be integrated with VO-based group communication protocols to provide collaborators with a shared file system.

Authorization Tools – These will be built up using existing VO Management Tools [44] to enable an effective ***Language of Access.*** More specifically, these tools would enable creation of ad-hoc groups of physicists, individually authenticated using X.509 certificates. We already have Java-based tools that implement much of this functionality. The new work would be to integrate these tools in the Collaborative GUI and extend the scope to encompass groupware concepts.

MonaLisa [33] – The MonaLisa monitoring system will be customized and extended to support monitoring and control of the collaborative system. This will include real time monitoring of the VRVS servers, CHEF, networks that interconnect peer groups, and so on. Another possible use of the MonaLisa service in this context is as a distributed task-monitoring tool, which could be extended to allow primitive distributed debugging capability.

Shared Workspaces – Successful collaborations require users to have the ability to share desktop windows. Traditional whiteboard (as in VRVS), window and desktop sharing (VNC), and file synchronization tools will be evaluated and integrated into the CHEF framework as appropriate. Early development of specialized modules, for example java applets, optimized for quality of service under dynamically changing bandwidths will be explored and developed.

There are several system-wide goals that will guide the development efforts throughout this project. The new working environment we propose will have the following novel features to provide a new, integrated, easy to use working environment:

- ***Persistent Collaboration:*** on the desktop, in small and large conference rooms, in halls and in Virtual Control Rooms; enabled by very cost-effective prototype setups[1]
- A flexible, ***extensible structure of hierarchical (role-based), openly persistent (forever-open, café-room), and ad hoc peer groups***, implemented by extending and generally the concepts in existing Virtual Organization management tools [43] and the virtual space concepts in VRVS as well as other tools
- A ***"Language of Access"*** and Control***:*** to people, meetings, casual conversations; scheduled and "on demand". Enabling access on demand to encourage spontaneous collaboration while limiting the ability to interrupt. Including concepts of collaboration such the "partly open door", the "cost of interrupting" [opening that door] and an initial "exchange of value" [using up credits by opening the door].
- ***Human-System-Human as well as Human-Human interactions for Collaborative Work.*** To enable effective sharing of the work with complex distributed systems; to make the ensemble of human-to-human interactions, and joint interactions between humans and the system consistent, and tolerable.
- Evaluation, Evolution and Optimization of the System using proven, effective, iterative evaluation methods
- Optimization of the environment for performance: using Clarens [lightweight client-server interactions], MonALISA monitoring, limits on processes to avoid bottlenecks
- ***Collaborative agent-based decision support***. For data-intensive transactions; system and humans decide what to do "together".

Object persistence is a critical attribute of a collaborative environment. Without the ability to archive and make available the documents, whiteboards, conversations and interactions that occur in collaborative environments we lose one of the most transforming aspects such environments could provide. Persisting documents and interactions in collaborative sessions create a form of group memory, keeping otherwise transient events or results available, not only for those who participated in the original session, but any others who may need to find such results.

We need to provide the users of the system with a ***language of access*** to allow each user to define the circumstances under which various collaborators, groups or agents can interrupt them and potentially at various levels of interruption. The language should be rich enough to allow defining automated ways of responding under specified circumstances and must encompass dynamically defining new groups or interactions. This language should also extend to allowing authorization decisions for persistent objects and the context sensitive use of collaborative tools. Such a capability is required to assure the users that they would be able to use the system for their benefit, rather than the system using them for its benefit.

A customizable dynamic environment that archives interactions and allows dynamic creation of groups could quickly produce so much data and meta-data than a typical user will be hard pressed to keep up. Providing ***agent-assisted collaboration*** as an integral part of our environment will be critical to giving users the capabilities they will need to find information, join groups and interact will colleagues. Agents that can act on the behalf of the user and be invested with increasing degrees of intelligence and capability will be required for effective use of a powerful collaborative environment.

While some aspects of these attributes will be present in our initial release, we are intending a staged implementation throughout the life of the proposal using the results of our assessment and evaluation to emphasize the most important, effective concepts at each stage.

First Year: "Delivering New Collaborative Capability"

The First Year Package will integrate a selected set of components in the CHEF environment. Each tool we identify for the release will be enhanced as needed to provide security, persistence, language of access capabilities and an agent interface. They will also each be integrated into the CHEF framework.

This framework immediately offers us a rich set of such capabilities together with integrated Grid authentication and an open-source portal framework with well-defined developer APIs and documentation. One of us (Severance) is the lead developer in the portal component of the NEESGrid project

[1] Funded by the participating groups and experiments, outside of this proposal

This will allow us to focus our resources on the other key components of the collaborative system, as we can directly begin integration and adaptation of the selected components in this base infrastructure. The choice of which components to include is largely governed by what features the initial environment is required to have.

The environment we propose can be best understood by providing a set of user capabilities and how we intend to provide them given existing tools or capabilities.

- Users or groups could create *Sessions which* would be a virtual environment within CHEF that assembles itself based on defined characteristics and the history of previous *Sessions* of that group or project. The *Session* begins by calling up previous metadata and tools and ends by archiving the products of itself for either renewed synchronous work or for asynchronous login by one of the parties. (Michigan, LBNL, Caltech, UTA)
- Audio and video capabilities will be available for any *Session*. These capabilities will be provided by VRVS. (Caltech)
- Upon demand, any participant can call up audio recording clipping for archived reminders/notes/memos available for text transcription, dispatch as email, or printing. All audio recording clipping must be acceptable to all parties in a *Session*. This will involve aspects of our shared file system, VRVS, CHEF and use of language of access/authorization features. (Michigan, Caltech, Iowa)
- Both remote parties will be able to simultaneously view and control a common application: including editors, graphics viewers and data analysis tools. We will enable this via VNC integration into CHEF/VRVS. (Michigan, Caltech)
- Automatic versioning will occur with incremented file names and/or metadata captured and catalogued for future retrieval. This will be automatic, or triggered by activity, or an explicit instruction. Layers of versions will be saved for each item produced.. (UTA, Iowa, LBNL)
- For asynchronous use, either party should be able to access the set of all previous *Sessions*. It should be apparent through off-line notification or an explicit presence tool when activity in the *Session* archive has occurred. (LBNL,UTA, Caltech, Michigan)
- A part of out-of-*Session* activity will include IM capability in order to move sequential dialog away from email. IM tools and rendering should have scientific/Greek character representation and directly couple to archived *Session* materials, where relevant. IM sessions should be archivable to a chosen *Session* archive after the "conversation". (Caltech, LBNL)
- All products of the *Session* will be automatically archived: this could be all versions of the evolution of manipulations of a shared postscript file, any documents prepared, audio memos created, or scripts applied or written during interactive computing. Transcripts of changes to documents will be available for later scrutiny and possible conversion to macro scripts.(All)

We believe that the creation of an initial system with these capabilities would be readily accepted by early adopters and lead to enhancements that will make measurable positive difference in collaborative interactions, and do so immediately. Because evolution to such tools will be a significant change, the early adopters should readily see the advantages for their personal and project productivity and imagine possible capabilities which go beyond those of the in-person interaction. As a research effort, abstraction of this frequent one-one/few-few activity into a collaboratory environment for early use and evaluation would serve a variety of important purposes for this project. First, because the one-one/few-few interaction is a heavily practiced activity, if the new tools are used, the result will be many encounters and hence statistically significant opportunities for evaluation and considerable experience and feedback to the developers. Second, both the technology and sociology of collaboratories cover new ground so such heavy use will provide considerable new and useful data for study in both arenas.

Second Year: "Augmenting Dynamic Collaboration Capability"

The focus of the development effort in Year 2 will shift from the integration effort of the first package release to the development of new tools that leverage capabilities that will exist in the framework after the initial release and add completely new capabilities to the environment. The final development priorities for year 2 of the project will be based on the assessment conducted in the first year to determine the users' needs. Although we don't know the final result of this process, we can based on our experience with collaboratories and the scenarios presented above, describe our current projection of the paths we expect year 2 development efforts to follow.

In Year two, we will have results from the requirements analysis of the first year. We will use the results of that analysis to tune our second release at the end of Year 2.

The tools described below are likely to comprise the release:

Electronic Scientific Notebook - We plan to customize one of the existing notebooks and integrate it in our collaboratory environment. The best candidate is a product from the HEPBook project. The project is currently funded through a DOE SBIR grant and the development team is working in collaboration with members of the Fermilab Computing Division. Its major components are the following..

- **The KnowledgeBook/ HEPBook** [44] **-** provides a graphical user interface (GUI) to enter multi-media entries, extensive search capabilities, access control, authorization, and protocols to connect to remote repositories. The **KnowledgeBook** also has a set of software agents to store, communicate, authenticate, control access, and to manage the configuration. The **HEPBook** has entry types and software agents with HEP specific knowledge in addition to the functionality of the knowledge book
- **An Agent Framework within the KnowledgeBook -** Expertise is the key ingredient to develop the agents. Eventually the experts, in our case the physicists, will want to modify and add new agents themselves, without having to come to us, the developers, to request a change. To that end we will provide an application programmer interface (API) to define new entry types and their agents. It will allow the end user to use languages other than Java and the Java Expert System Shell (JESS), which we will use for some agents. The basis for the agent platform will be FIPA-OS (Foundation for Intelligent Physical Agents)[45].

Collaborative, Grid-Enabled Dialogs – Collaborative work-sharing will be enhanced by decision support from the monitoring and other Grid-management agents, who will interact with the collaborators in a given session to determine their needs, match the needs to Grid system conditions, and begin a simple dialog in case there are difficult choices to be made. The effectiveness of these dialogs, and their development, will require in-depth evaluation, using the methodology summarized above.

Location-aware Monitoring. We could use a geographic-information-systems-based Grid information broker (GIS-GIB)[46]. This uses Grid information services to provide the physicist with a location-aware mechanism for monitoring virtual-organization-based collaborations, discovering interesting collaborations, and analyzing the spatial and temporal characteristics of collaborations. This GIS-GIB will help high-energy physicists dynamically identify where and when interesting collaborations are taking place. In addition, The GIS-GIB can be used to organize and analyze the collaborative efforts that are geographically distributed and substantially variant within a temporal scope. This GIS-GIB is an existing component of a Grid portal project that is conceptually based on OGSA.

Grid Computational Portal Toolkit – There is currently an effort (U Michigan, Indiana, U Texas, ANL, NCSA) to develop a complete computational grid toolkit based on CHEF/Jetspeed [47]. This will provide basic job submission, job monitoring, workflow, and other grid computational operating through the Jetspeed environment. The GridFTP [48] and LDAP portlets which are already in use in the NEESGrid are products of this collaboration. These basic Grid computational tools will be an excellent compliment to the sophisticated Grid-based data analysis tools as proposed by GriPhyN and the other Physics data grid projects[1, 2, 3, 4].

Web Services Definitions for all CHEF Services – All of the CHEF services will be adapted so as to support WSDL bindings using the OGSA services and Grid security [39]. This will allow collaborative capabilities to be added to any application that is capable of acting as an OGSA client. One can envision a number of JAVA-based "DesktopLets" which fully interoperate with the browser based teamlets and portlets.

Third and Fourth Year: "Extending Dynamic Collaboration"

In years three and four, our focus will be based very much on the results of the requirements analysis for years one and two as well as the analysis of the usage of the year one software release. As such, the items in this section may or may not be included in the final release. In a way, this section is input to the requirements analysis team as ideas that we recommend to be examined closely.

Collaborative Framework - CHEF will continue to evolve in years three and four. This list should be considered possibilities that will depend on funding, user requirements, and many other factors.

Federated Collaborative Servers – The ability for a set of CHEF servers to "join a club" using a single federated signon (i.e. Grid credential). Within the club, servers will exchange group membership information so when a user visits any of the club sites, they see all of the groups they belong to across the entire club regardless of which server the group resides on. The user will be switched from server to server transparently as they move between their groups.

Progressive Automation of Grid-Enabled Dialogs – As the requirements and desire elements in the decision support tree are better understood, some human-agent dialogs will be automated, and/or abstracted into efficient tools on the desktop to facilitate choices.

Clarens [49] – The Clarens Web Services/Portal dataserver is a system for delivering application specific data from scientific data repositories. The Clarens server is able to receive client requests in various formats, which may include demands for significant data processing on the server. Clients are authenticated using X.509 certificates. For the Collaboratory, we would use Clarens as a portal mechanism for the storage and (query-based) retrieval of persistent Virtual Room archive data, as well as a vehicle for the reception and handling of VO requests. In the later stages of the project, we would use Clarens servers in a Peer to Peer configuration. In this architecture, Peer Groups of physicists are formed on the basis of a shared interest (e.g. in an analysis channel). They become a VO and appear in a Virtual Room. They share potentially large collections of object data (collections of event objects). These collections are streamed between individuals in the Peer Group. The data requests are brokered by the Clarens servers, which sit at the next level up in the hierarchy (typically at Tier2/3 centres), in a "Super" Peer group. The bursty, high-throughput, data traffic flows across optimized networks between the Clarens servers, and then on to the client machines. This form of dynamic, high throughput data exchange between individual physicists could become a significant feature of physics analysis at the LHC. It may be one of the distinguishing aspects of future HEP collaboratories.

Learning Management Components – One of the other applications of CHEF is in a structured learning environment. By the end of year two, there are a number of planned tools that will be made available to the collaborations as they are completed. A partial list of tools include: Links to libraries of information, assignment management, gradebook, lecture capture and playback, online quizzes, voting/instant poll capability, roster management, workshop registration tool, and others. This learning management system will support a number of standards including: IMS Question and Testing [50], SCORM [51], and OKI [52]. These tools may or may not have relevance to the collaboration, but we can explore their use for the training and outreach components of the collaboration.

Deployment and Support

In order to develop and deliver the most suitable and efficient GRID-enabled collaboratory environment to the scientific research community, this collaboration must include very strong participation of the end-users - physicists from the High Energy Physics community. Most of the collaborators are members of US ATLAS and US CMS. Among this group, many are also members of the currently running experiments, CDF and D0 at the Tevatron [53]. There are some collaborators from the US LHC machine group. They all need a GRID-enabled collaboratory environment to perform activities, such as physics analysis, detector commissioning and operation, and accelerator commissioning and operation.

We plan to deploy prototype collaboratory environments frequently for the internal users in this collaboration. The users perform their activities in the prototype environment and feedback their response to the assessment team and the development team. The deployment for the internal user group will be done as needed. Once any individual tool is significantly improved or new functionality is added to the environment, the deployment team will test the new integrated environment and then deploy it for the internal users.

We plan to deploy the collaboratory environment once a year for the external users. The candidate external user groups include, but are not limited to, CDF and D0 at the Tevatron and ATLAS and CMS at the LHC. As a part of our effort on Education and Outreach, we will also deploy the collaboraatory environment for the QuarkNet [72] participants.

We propose to have one dedicated computer professional for deployment and support. Since the person needs to support a large number of internal and external user groups, he/she will reside at Fermilab and utilize the infrastructure of the US CMS Tier-1 center for the most efficient deployment and support of the user community.

The person will also collect user's response to the deployed collaboratory environment through daily interaction with users, and feedback the information to the development team for further improvement. Therefore he/she is also required to be capable to understand the technical details of the collaboratory environment in order to provide appropriate information to the development team.

Broader Impacts

Though HENP will be the targeted community for this proposal, the tools, systems, and the integrated working environment, proposed here will be designed and constructed to be broadly applicable, so that the GECSR could pave the way for new modes of collaborative work in many fields of science and engineering, education, and industrial research. This development is ***imperative*** for the health and success of current and future global physics collaborations. The collaborations enabled and fostered by the work proposed here, with the support of governments and funding agencies around the world, will fulfill the ***requirement that scientists from all regions are allowed, encouraged, and given the means to function as full partners in the scientific process of search and discovery.*** This proposal is the first attempt to make this possible, with a focus on a new generation of collaborative tools to be employed, integrated with Grid-enabled and other analysis tools that make data-intensive analysis by a global collaboration, using worldwide-distributed resources, possible.

The broader implications of this proposal, and its mission to grant full partnership to groups in all world regions are driven by (1) The right of groups and governments that have contributed to the global Collaborations to share and collaborate in the data analysis, and thus in the process of search and discovery that is at the root of all basic research. (2) The need to justify global projects, and global Collaborations as a valid and vital means of conducting future leading-edge scientific research and other endeavors, without exploitation of the poorer nations and world regions and (3 The need to develop effective means collaboration and cooperation among sub-groups with vastly different academic and cultural backgrounds, and especially to allow effective collaboration by university-based students in the native cultural milieu of their home countries.

Background of our Groups

This proposal brings together several groups with complementary and extensive knowledge and experience in several areas of collaborative scientific research:

Caltech: Caltech's expertise is in videoconferencing software, virtual rooms, Virtual Room Videoconferencing System, agent based monitoring, virtual organization management, high performance wide area network implementation, management and operations, Grid systems and Grid-Enabled Analysis development, self-learning distributed systems optimization. The Caltech group, led by the US CMS Collaboration Board Chair, originated and has had responsibility for transatlantic networking for HEP since 1982. It currently co-manages the operation and development of the "LHCNet" links between CERN and the Starlight in Chicago [54]. It led the MONARC project [55] that developed the worldwide distributed computing model for LHC data analysis, and originated the Data Grid hierarchy concept that has been adopted by the LHC experiments. It shares leadership in the Particle Physics Data Grid [2] and iVDGL[56]. It developed VRVS [16] which went into production in 1997 and currently runs on 14, 900 registered hosts in 64 countries.

The Caltech CMS group in HEP has worked in close collaboration with the Center for Advanced Computing Research (CACR) since 1996. This collaboration began with the GIOD (Globally Interconnected Object Databases) [57] project funded by Hewlett Packard, and the first accurate evaluation of the backgrounds to Higgs particle searches using the diphoton signature, using pre-TeraGrid prototypes that took advantage of the large-scale computational resources at CACR and other NPACI sites. This collaboration led to the development of the first prototype Tier2 center at CACR and SDSC, by the CMS groups at Caltech and UCSD. Caltech and UCSD are two of the four TeraGrid [58] sites (along with Argonne and NCSA).

Caltech, and the CMS group in particular, has been very active in fostering collaborations for science and education on a global scale. The group has provided assistance in the design and implementation of Tier2 centers in developing countries since 1999, including India, Pakistan, China, Korea and Brazil among others. Recent events organized by this group include a lecture series on Grids, Networks and national IT infrastructure at the renowned Nathiagali Summer College in Islamabad Pakistan (founded by Nobel Laureate Abdus Salam in 1974), a session on Grids at UERJ in Rio (and made available to the world scientific community using VRVS and AMPATH) in

February 2002, and a session on the "Role of New Technologies in Developing the Information Society" in Bucharest in November 2002 [75].

Fermilab: Fermilab is the US center of CMS for detector construction, commissioning and operations, and physics analysis. It is also the US CMS Tier-1 center for the computing. It is also one of the US institutions working on the LHC machine construction and operation. Many of the Fermilab team members are also members of CDF or D0. Their participation in this proposal is as a user group: to help define requirements and evaluate the prototype collaboratory environment. In addition, the members from the Computing Division (CD) will coordinate work on this proposal and other computing projects at Fermilab, which include the US CMS computing project and the HEPbook development project in CD.

Florida International University is ranked by the Carnegie Foundation for the Advancement of Teaching as a Doctoral/Research University-Extensive, the highest ranking in its classification system. FIU has more than 32,686 students, 1,100 full-time faculty, and 90,000 alumni, making it the largest university in South Florida and placing it among the nation's 25 largest colleges and universities. It is unique in that it has the highest proportion of international students and faculty of any major university in the country. Its mission includes being the principal educational and research interface between the State universities and South and Central America and the Caribbean. AMPATH provides opportunities for Hispanic minority students at the University to work on the project in technology and administrative positions. The FIU nuclear physics group was established in 1995 when 8 positions (6 experimental and 2 theoretical) were created in partnership with the Thomas Jefferson National Accelerator Facility (Jefferson Lab). The Jefferson Lab-based group does experiments using the 6 GeV electron accelerator, CEBAF. They are just beginning a High Energy Physics group with membership on the CMS experiment.

Lawrence Berkeley National Laboratory: The members of the Lawrence Berkeley National Laboratory team bring to the proposal expertise in physics collaborations and development, deployment and use of a broad range of collaborative tools including a pervasive collaborative computing environment, secure and reliable group communication protocols [59], secure chat, Grid computing, peer-to-peer shared file systems, and the Access Grid videoconferencing systems [60].

Maryland: The members of the University of Maryland team have extensive experience in coordinating many groups with a large number of people both in CMS and D0. They worked as coordinators of the CMS hadron calorimeter construction project, the CMS Jet and Missing Et physics group, the D0 physics group and its subgroups (top, new phenomena), the D0 Computing group, the D0 simulation group, etc. The Maryland team brings to the proposal close ties to the user community. Also, the team takes a responsibility for deploying the developed collaboratory environment and user support..

University of Michigan: The University of Michigan is an international leader in the design and assessment of collaboratories, in the development of open-source environments for Web-based classroom and research collaboration, and in the capture and storage of lecture and presentation content for replay over the Web. In terms of collaboratories, faculty at the School of Information (SI) developed and deployed the NSF-funded Space Physics and Aeronomy Research Collaboratory (SPARC) in 1993, which remains the world's oldest, continuously operating collaboratory. Over the past decade, SI faculty have led several other important NSF and NIH sponsored collaboratory projects, including the currently funded George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES)[61] – where Dr. Finholt is a co-PI on NEESgrid [22], the grid-based collaboratory component of NEES. In addition, SI faculty are leading an effort to develop general principles for collaboratory design and use through the NSF ITR-funded Science of Collaboratories [41] project, where Dr. Finholt is also a co-PI. In terms of environments for Web-based collaboration, faculty and staff at the University of Michigan's Media Union (MU) have produced some of the world's largest academic systems for sharing course and research content, Coursetools[62] and Worktools [63] – now used daily by over 35,000 students, faculty, and staff at the University of Michigan. A team led at the University of Michigan Media Union is developing CHEF, which will provide an open-source environment for the next generation of Course/Worktools. Dr. Severance is leading the effort to produce grid-enabled versions of CHEF, such as for the NEESgrid project. Finally, in terms of capture and replay of lecture content over the Web, faculty and staff at the MU and the Department of Physics produced the first applications for archiving and replaying CERN seminars and tutorials. Specifically, Professor Neal, Dr. McKee, and Dr. Severance have directed and shaped the Web Lecture Archive Project [29], which in an expanded form will provide a key mechanism for education and outreach in this proposal. Overarching these independent activities is the Michigan Grid Research and Infrastructure Development (MGRID) center[64] – an interdisciplinary effort funded by the University of Michigan administration to develop an internal grid testbed (mgrid.org). MGRID center staff will work with key projects, such as the research proposed here, to enhance evolution of data and computing grid applications at Michigan and elsewhere. Drs. Neal and Finholt are appointed to the Executive Board of MGRID, and Drs. McKee and Severance are key members of the MGRID technical advisory team.

UTA: Close interactions between the HEP researchers and CS researchers at UTA have led to recent acquisitions from the NSF MRI program and joint supervision of graduate students. The HEP group at UTA played a leading role in developing tools (Grid Application Toolkit GRAT[65]) for large scale Monte Carlo data generation using Data Grid middleware like Globus [66] and Condor [67]. Thousands of cpu days are made available for specific studies over short periods of 1-2 weeks at 3 national laboratories and 7 universities in the U.S. The collaboratory tools proposed here will allow collaborations to share and manage the task of data production. The CS collaborators at UTA have extensive experience in multimedia (author of the MtreC MPEG4 Toolkit [68]) and database research.

University of Iowa: Expertise in Grid research, education, and outreach in HENP through the collaboration between GROW (Grid Research and educatiOn group @ IoWa) and university CMS community, collaborative visualization (e.g., NSF funded Laboratory for Immersive Visualization for the Environment [71]), and geographical-information-systems-based Grid monitoring.

University of Oklahoma: The University of Oklahoma group, as member of the U.S. ATLAS Grid test bed team, has contributed to automated production of simulated events using grid resources. The group will contribute to early adoption of collaborative tools by D0 physics analysis groups.

Education and Outreach

The Education and Outreach (E&O) program associated with this Medium ITR will enable a collaborative learning community engaging non-traditional students in our studies of particle physics at the energy frontier. In the collaborative learning environments that will result for the development components of the proposal, students will become empowered to do science projects and/or help develop the learning environment itself, as a student peer group. Our efforts will involve high school teachers and individuals engaged in education research. We will deploy collaboratory tools into local schools to disseminate high-energy physics knowledge into the classroom, enhance physics classroom activities, and explore how the tools adapt to the school community. In addition, QuarkNet-affiliated [70] teachers will utilize the collaboratory tools through summer high-energy research programs. QuarkNet is a successful national program of particle physics education and outreach, supported by NSF that partners high school teachers with physicist mentors. Teachers work as researchers and collaborators in forefront physics projects. The program, now in its fifth year, has over 50 participating university and national laboratory centers, f ive of which are represented in this ITR proposal. The E&O team envisage a competition, modeled after the ThinkQuest [76] concept of forming 5 person teams with a coach to develop a project to be shared. While ThinkQuest projects are website lessons, we could develop live lesson plans that would involve remote teams linked together into the collaborative learning environment. Moreover teams could and should consist of geographical separated members with like interests. By providing access to our facilities and research activities, we will build personal fulfillment and science excitement in students as well as open their eyes to new career opportunities. Teachers will be reinvigorated through greater exposure to pedagogical innovations that will translate into successful physics classroom experiences.

Our first goal is to disseminate the collaboratory tools into schools located near our partners. The tools will provide a 'real window' into the exciting world of particle physics at remote accelerator laboratories for many K-12 students and teachers. For education and outreach involving K-12 teachers, we will draw from the pool of active and highly regarded QuarkNet Teachers affiliated with QuarkNet Centers associated with our participating groups. Teachers and students will utilize VRVS teleconferencing capability and archived web lectures to support and strengthen relationships and interactions between students, teachers, and physicists. Through these tools, students and teachers will interact with physicists at universities and laboratories to bring high energy physics into the classroom. Classroom-classroom interactions will also be a vital part of the proposal to foster teacher mentoring as well as to build a community among students. This will maintain and build the ties between the many QuarkNet Centers – offering the possibility of physics and analysis meetings, seminars, sharing of teaching strategies, and development of classroom transfer applications. This will enhance the effectiveness of the QuarkNet "Collaboratory" at the same time as the physicists and computer specialists are developing tools for the effective operation of particle physics collaboratories. We will also provide internships for students to actively work on developing, prototyping, testing and using the collaborative tools. The Education and Outreach Area Coordinator will establish a group that will discuss communication and development issues related to education. These discussions will be aimed at the goals of developing collaborative learning sessions referencing both QuarkNet and ThinkQuest team lesson development model mentioned above, with a new dimension of live interactive participation, and collaborative class and research projects, among undergraduates and high school students.

QuarkNet experiences clearly demonstrate that once engaged in research, many teachers are eager to continue research activities for periods of 4-8 weeks during summer breaks if opportunities are available. Likewise, physicist mentors who have made a significant investment of personal time and effort in developing teacher-researchers at their Center, are keen to capitalize on that investment by continuing the working and research relationships in subsequent years if resources are available. This proposal extends the methods for supporting and interacting with QuarkNet teachers, directly benefiting both teachers and QuarkNet.

Specifically, funding for this proposal would provide research opportunities for up to eight teachers per year for eight weeks of summer research each. Three groups will be supported: (1) **University of Texas at Arlington** (and affiliated groups at the **University of Oklahoma** and **Langston University**); (2) the **University of Iowa**; and (3) **Florida International University (FIU)**. Of these, UTA, OU and UI are three of the twelve original QuarkNet Centers funded by NSF since 1999. They have well established relationships between teachers and physicists and have reached tens of thousands of high school students through the programs. FIU is a new QuarkNet center. The groups also cover significant geographic regions and have the potential to engage underrepresented groups in HENP. UTA/OU/Langston cover North Central Texas and Central Oklahoma and Langston is Oklahoma's only Historically Black College. UI reaches rural, middle America. FIU is one of the largest minority universities in the continental US and the largest Hispanic serving minority university with over 32,000 students, (55% Hispanic and 20% African American). The three E&O groups will naturally form a mini-collaboratory focused on providing outreach and access for undergraduate and high school minority students to the world of particle physics.

UTA/OU/Langston will also concentrate on minority recruiting programs. Since we are delivering a new collaborative environment in this proposal, it would be appropriate to focus on how this new paradigm can help traditionally under-represented students in physics. Langston and OU collaborate in running a D0 Monte Carlo simulation farm. Langston HEP has a Linux cluster with D0 software installed and this cluster's condor pool is shared with OU's condor pool. The collaborative tools to be developed through this proposal, with active student involvement at Langston, can open doors for traditionally underrepresented students to participate in the greatest scientific discoveries at U.S. and international laboratories. A recent initiative between UTA and local schools is the TECOSE (TExas Cosmic Observations by School Experimenters) [71] project. This is part of an international initiative to have high school students participate directly in research on the physics and origins of ultra high energy cosmic rays. A network of detectors will be mounted on the roofs of high schools. We are working on a pilot project with three Fort Worth, Texas, high schools; two schools are predominantly minority Hispanic schools, and the third is in a rural area south of Fort Worth. The collaborative tools proposed in the medium ITR will be deployed in the TECOSE project and play an invaluable role in opening the world of physics for these K-12 students.

In the area of education research, we will develop direct and collaborative links with colleagues engaged in studying the effectiveness of collaboratory tools from an educational perspective at the undergraduate level as well as in the education of scientists joining large-scale research projects. To do this, we provide half-time support for a graduate student doing dissertation research in the School of Education at Michigan on the effectiveness of information delivery using the WLAP technology. Undergraduates participating in a summer Research Experience for Undergraduates (REU) and general physics undergraduates will be used as subjects in research on design principles for the WLAP technology and in an attempt to study the role of individual difference in the design of instructional materials for same [73, 74]. The results of this research will be immediately applied to the research and development efforts of the ITR that relate to human-computer interface and will be incorporated into the other aspects of this educational and outreach proposal. Through this research, physicists will gain insight from professional educators into the nature and effectiveness of information delivery using the WLAP technology. Simultaneously, educators will gain insight into the nature and operation of science experimentation in large collaborations, which may spawn further research.

The University of Iowa team will leverage four existing programs to conduct education and outreach for this ITR: Iowa QuarkNet, Science Education Center of School of Education, GROW, and the Studio of Academic Technologies for Education.

Florida International University will introduce the tools to local high schools and evaluate their impact. Use will be made of CHEPREO (Interregional grid-enabled center for high-energy physics research and education outreach), an integrated program of high energy physics research at CMS, education and outreach, and networking infrastructure development. FIU has forged a strong relationship between physics and education faculty that will support this project. FIU will also lead a coordinated program for E&O by creating a community of practice among the various collaborating institutions. This will be enabled through interactive meetings and mini-workshops using the collaboratory tools to insure the integration of E&O project goals which are to engage minorities, improve physics classroom curriculum directed by teacher/student contribution and response, add a new dimension to existing successful pedagogical approaches such as QuarkNet and ThinkQuest, and employ education research techniques to monitor and approve these approaches.

Results from Prior NSF Support

Harvey Newman: John Hopkins University Subaward Agreement No.08002-48195 under NSF Grant No. PHY-9980044 (9/15/1999 - 8/31/2003; $1,051,771), "Accessing Large Data Archives in Astronomy and Particle Physics" Results: The sheer amount of data that are used in the fields of astronomy and particle physics has forced both communities to move from flat files to highly organized databases. This project contributed to 1) The organization of data for efficient access, 2) Storage of data in widely distributed locations, 3) Efficient handling of entire major archives. Other NSF support: GriPhyN (Ongoing KDI- and GriPhyN-related developments to be applied in the iVDGL. Development of ODBMS-based scalable reconstruction and analysis prototypes working seamlessly over WANs; Grid Data Management Pilot (together with EU DataGrid); Grid-optimized client-server data analysis prototype development (Steenberg et al.), MONARC simulation and Self Organizing Neural Nets (Legrand et al.); modeling CMS Grid workloads (Holtman, Bunn et al.); optimized bit-sliced TAGs (Stockinger et al.); TeraGrid prototype for seamless data production (Litvin et al.; with Livny at Wisconsin and Koranda at NCSA).
Julian Bunn: John Hopkins University Subaward Agreement No.08002-48195 under NSF Grant No. PHY-9980044 (9/15/1999 - 8/31/2003; $1,051,771), "Accessing Large Data Archives in Astronomy and Particle Physics" Results: The sheer amount of data that are used in the fields of astronomy and particle physics has forced both communities to move from flat files to highly organized databases. This project contributed to 1) The organization of data for efficient access, 2) Storage of data in widely distributed locations, 3) Efficient handling of entire major archives.
Tom Finholt: CMS-0117853 (8/01/2001 – 9/30/2004; $1,814,026), "NEESgrid (Network For Earthquake Engineering Simulation): A Distributed Virtual Laboratory For Advanced Earthquake Experimentation and Simulation". Results: This project will contribute to the system integration effort for the George E. Brown, Jr. Network for Earthquake Engineering Simulation. It will define the user requirements for the NEESgrid and contribute several software components including collaboration tools through the CompreHensive CollaborativE Framework (CHEF). Representative Papers: Severance, C.R. (2003). Integrating the grid into CHEF. NEESgrid Technical Report TR-2003-1 and Finholt, T.A., Wierba, E.E., Birnholtz, J.P. & Hofer, E. (2002). NEESgrid user requirements document. NEESgrid Technical Report.

Shuichi Kunori, Shawn McKee: No prior support

Conclusion: Transition to a New Era of Collaboration in Scientific Research and Education

We are proposing to launch a next-generation collaborative framework for scientific research, harnessing multi-institutional expertise, and existing and new tools across a broad range of technologies, to begin the transition to a new era of collaboration in the sciences. Just as there is an ongoing transition to grid computing, enabling transparent access between people and computing resources, we foresee a transition to collaborative environments that enable new modes of transparent, persistent and spontaneous interaction between scientists located at sites around the globe. The Collaboratory we envision will provide secure, intelligent, easy to use collaborative tools, integrated with a discipline-specific toolset and environment for each scientific field. This makes it a key required element for the success of today's global scientific collaborations, and gives it the potential of becoming a cornerstone of more tightly-knit worldwide collaborations of the future. Extension of the GEC system and concepts to the formation of Grid-enabled Learning Environments will provide educators and students alike with richer and easier modes of access and interaction with scientists located in all world regions; making them, from their classrooms and homes, part of the discovery process of frontier science. The proposal's research outcomes, from the development of the Collaboratory to a deeper understanding of effective modes of global communication, could have a profound and broad impact on research and education, as well as the way multi-institution organizations operate, by enhancing our ability to interact and share work collectively, across global networks.

Management Plan

The PIs, Co-PIs and other senior personnel have extensive experience with the management and execution of large collaborative projects. In addition, this project will be part of the LHC research program, where thousands of physicists are engaged in an ongoing managed international collaborative effort, in which the core of the project team is already embedded as collaborators in their physics experiment.

The PI and co-PIs will provide leadership to the project. The PI is Professor Harvey Newman (California Institute of Technology), who is chair of the US CMS collaboration board and represents the US collaboration in the CMS experiment. Professor Newman has in-depth experience in leading such research collaborations, as well as large scale data analysis efforts in Collider experiments. He is PI of the DOE SciDAC Particle Physics Data Grid Collaboratory project, a co-PI of the International Virtual Data Grid Laboratory, a managing partner of the "wavelength triangle" advanced network development program between the US and Europe, and initiated and directs the VRVS effort at Caltech.

The Project Coordinator is Shuichi Kunori of University of Maryland, who has long experience in leadership of physics analysis groups. The Technical Coordinator is Shawn McKee of University of Michigan, who has significant technical expertise in grids, collaborative systems and networks.

The experience of our management team will enable us to meet the significant management challenges of the project: the broad scope of the problem to be addressed, the diverse multi-disciplinary teams, and the diverse external constituencies including national labs and international organizations.

The proposal has four work areas with dedicated teams, namely Assessment and Evaluation, Support and Deployment, Development, Education and Outreach. Each of these teams will have an Area Coordinator. Together with the Project Coordinator and the Technical Coordinator they form the Management Team, that will be responsible for running the project, will meet regularly, track the progress of the project and administer funds.

In addition, a User Group will be formed with the responsibility of feeding back requirements and results of using the delivered tools to the rest of the project. Interested and early-adopter members of the research community can discuss and give feedback on issues and suggestions for: more effective modes of communication during work sessions; required features or new paradigms, and future development work. A Web site will be maintained, to keep the scientific community up to date with development plans and issues, and encourage online feedback.

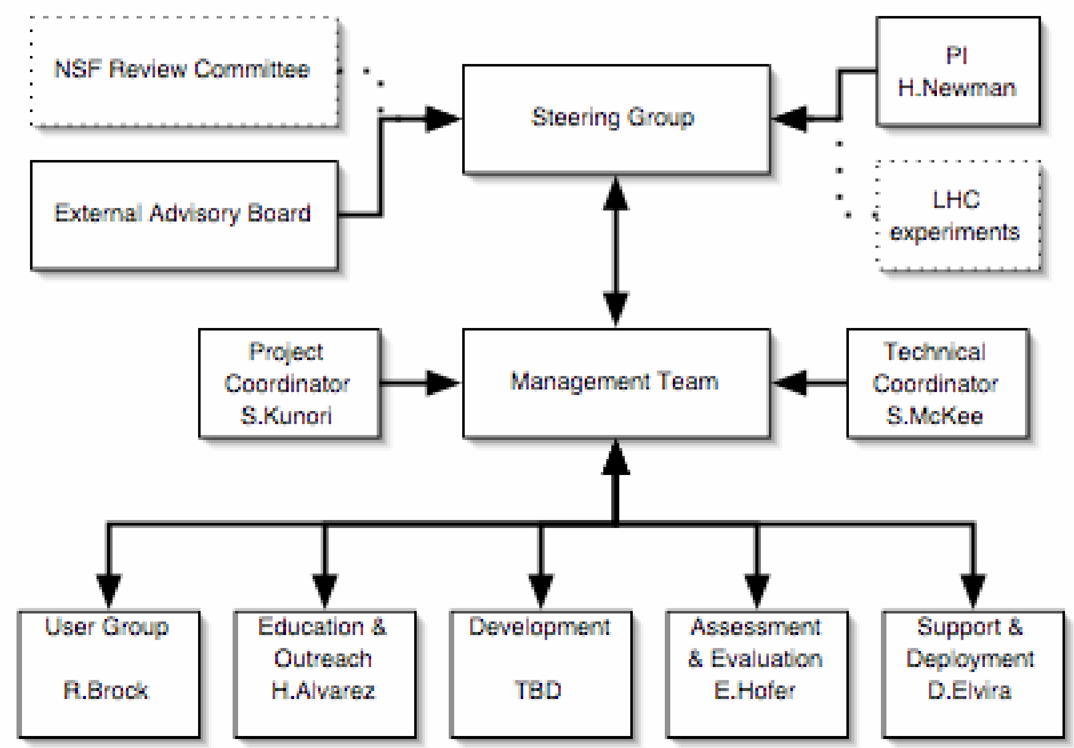
The Education and Outreach Area Coordinator will establish a group that will discuss communication and development issues related to education. These discussions will be aimed at the goals of developing collaborative learning sessions, and collaborative class and research projects, among undergraduates and high school students.

The Technical Coordinator has overall responsibility for the day-to-day coordination of the technical work of the project, for identifying needed inter-team coordination and work, and working closely with the project coordinator on their joint responsibilities.

The Project Coordinator has overall responsibility for project coordination and deliverables, as well as for calling meetings, writing reports, dissemination, and the use of the delivered tools in the physics collaborations, as well as working closely with the technical coordinator on their joint responsibilities. The Project Coordinator will report to the Steering Group for the Management Team.

A Steering Group will be established consisting of the project PIs, the Project Coordinator and the Technical Coordinator, and US Atlas and US CMS research program managers or designee. The Steering Group is responsible for the technical and scientific direction of the project.

In addition an External Advisory Committee will engage both user community and research community to advise the Steering Group. Members will include people from the physics experiments, from Grid projects, CS researchers and IT experts.



Our experience with other projects of comparable scope leads us to define a management approach that emphasizes regular integration goals and capability milestones; frequent meetings and consultation both internally and with key external stakeholders; and extensive use of advanced collaboration technologies, already used widely by most participants.

Communities of LHC researchers will test capabilities and tools delivered through 'testbed' activities and Data Challenges, feeding back results to the development teams in an iterative fashion. In particular, direct collaboration with the Virtual Control Room and Physics Analysis Center efforts at Fermilab (including development of a "Tier-1 Prototype") will facilitate this valuable feedback. The US LHC Research Program, which is jointly funded by the NSF and the DOE, will provide testbed facilities out of existing resources and will provide matching effort for deployment and integration for Data Challenges and Virtual Control Room activities. Other testbed facilities will be provided by participating institutions, including Caltech and CACR.