

GA-A25539

**COLLABORATIVE TECHNOLOGIES FOR
DISTRIBUTED SCIENCE: FUSION ENERGY AND
HIGH-ENERGY PHYSICS**

by

D.P. SCHISSEL, E. GOTTSCHALK, and G. ABLA

AUGUST 2006



DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

COLLABORATIVE TECHNOLOGIES FOR DISTRIBUTED SCIENCE: FUSION ENERGY AND HIGH-ENERGY PHYSICS

by

D.P. SCHISSEL, E. GOTTSCHALK,* and G. ABLA

This is a preprint of a paper to be presented at the 24th Symposium on Fusion Technology, September 11–15, 2006, Warsaw, Poland, and to be published in *Fusion Engineering and Design*.

*Fermi National Accelerator Laboratory, Batavia, Illinois

Work supported by
the U.S. Department of Energy
under DE-FC02-01ER25455

GENERAL ATOMICS PROJECT 30106
AUGUST 2006



ABSTRACT

This paper outlines a strategy to significantly enhance scientific collaborations in both Fusion Energy Sciences and in High-Energy Physics through the development and deployment of new tools and technologies into working environments. This strategy is divided into two main elements, collaborative workspaces and secure computational services. Experimental and theory/computational programs will greatly benefit through the provision of a flexible, standards-based collaboration space, which includes advanced tools for ad hoc and structured communications, shared applications and displays, enhanced interactivity for remote data access applications, high performance computational services and an improved security environment. The technologies developed should be prototyped and tested on the current generation of experiments and numerical simulation projects. At the same time, such work should maintain a strong focus on the needs of the next generation of mega-projects, ITER and the ILC. Such an effort needs to leverage existing computer science technology and take full advantage of commercial software wherever possible. This paper compares the requirements of FES and HEP, discusses today's solutions, examines areas where more functionality is required, and discusses those areas with sufficient overlap in requirements that joint research into collaborative technologies will increase the benefit to both.

1. INTRODUCTION AND VISION

The large-scale experiments needed for fusion energy sciences (FES) and high-energy physics (HEP) research are staffed by correspondingly large, globally dispersed teams. At the same time, theoretical work has come to rely increasingly on complex numerical simulations developed by distributed teams of scientists and applied mathematicians and run on massively parallel computers. These trends will only accelerate. Operation of the most powerful accelerator ever built, the Large Hadron Collider (LHC) [1] at CERN, will begin next year and will dominate experimental high-energy physics. The experimental program will require collaborative efforts on a heretofore-unprecedented scale, with two LHC experiments each involving approximately 2000 physicists. A future global HEP project, the International Linear Collider (ILC) [2], is expected to place even more demands on work performed in collaborative environments. The fusion program will be increasingly oriented toward the International Thermonuclear Energy Reactor (ITER) [3] where even now, a decade before operation begins, a large portion of national program efforts are organized around coordinated efforts to develop and qualify promising operational scenarios. Substantial efforts to develop integrated plasma modeling codes are also underway in the U.S., Europe and Japan. While both FES and HEP have a significant track record for developing and exploiting remote collaborations, with such large investments at stake, there is a clear need to improve the integration and reach of available tools. We believe that there is sufficient overlap in the requirements between FES and HEP, that joint research into collaborative technologies to significantly improve the collaborative scientific environment will increase the benefit to both.

Common requirements amongst FES and HEP include the need to support remote operations of experimental facilities; distributed code development, computing and visualization; and a wide range of planning and coordination activities. All of these activities will benefit from sharable applications and displays, and improvements in interpersonal and group communications integrated with extensible data services. Implementations will build on emerging technologies and standards and the convergence of telecommunication and computing technologies. A service-oriented approach would be extended to high performance computation and data management, yielding interactive large-scale data access over the wide area networks. The latter will be especially important for supporting simulations and long-pulse or continuous experiments, which can generate very large data stores. In all cases, a crucial element in the approach should be to test technologies on existing experiments and in existing collaborations as part of the design and development process. Finally, the entire implementation should depend on responsiveness to user needs, interoperability in heterogeneous environments, robustness, and ease of use. The envisioned work, for the purposes of this paper, is divided into two main elements that combine to deploy a distributed science environment (Fig. 1).

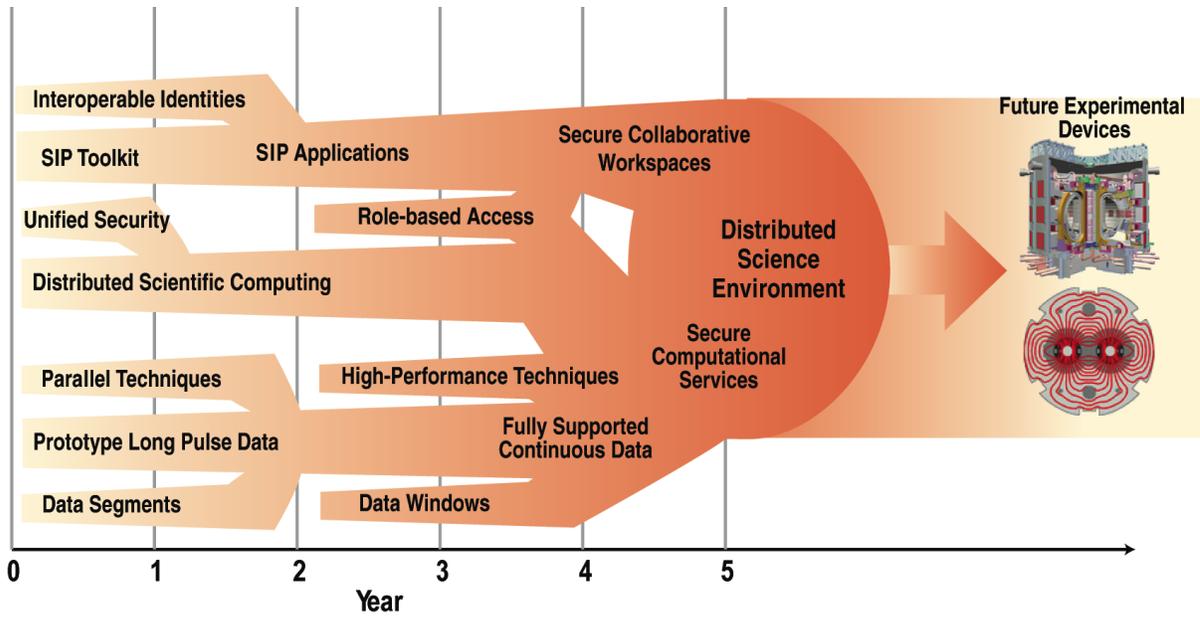


Fig. 1. Separate efforts can build towards secure collaborative workspaces and secure computational services, which will be integrated into a distributed science environment for FES and HEP.

2. COLLABORATIVE WORK SPACE

Our vision is to create a flexible and extensible “collaboration fabric” that will support diverse ad hoc and structured interactions. Effective remote participation on the scale envisioned requires the provisioning of a working environment where personnel at different sites can engage in dialogues enriched by data flows and where the “friction” and barriers to communication are significantly reduced. We anticipate the convergence of physical and logical communications channels so that phone, audio, video, email, messaging, and data can be integrated into a common framework. Web pages, which already support integrated and shared workspaces, such as electronic logbooks, code and experimental run management, records of presentations and publications, personnel databases, and physical site maps can be ‘communications enabled’, so that just as currently there are ‘mailto:’ links we will be able to have ‘speak to:’, ‘instant message to:’, ‘video to:’, and ‘share with’ links. Streaming of video and audio from remote control rooms would be integrated with ad hoc voice and video channels. Advanced interactive directory services need to be provided, allowing people and data streams to be identified, located, scheduled and connected to. In addition, real time presence information, role-based “dialing,” and dynamic call forwarding should become available. Presence mechanisms would be used to convey information on experimental status, data analysis or code runs.

Based on its flexibility and extensibility, Session Initiation Protocol (SIP) [4] appears to be a solid basis for this new environment. SIP is very likely to become a critical part of the communications infrastructure that we all use, and tools which leverage a common infrastructure are more likely to succeed than stand-alone systems. In order to support and attract users who are using existing applications like Virtual Rooms Videoconferencing System/Enabling Virtual Organization (VRVS/EVO) [5] and the AccessGrid (AG) [6] there needs to be a method to allow users of these tools to participate, within the scope of their resources, in the communications fabric that would be built. This is in line with our overall philosophy that attempts to accommodate the capabilities of available user agents, providing the richest environment that can reasonably be deployed on each.

Since the ability to share complex visualizations and applications among remote participants is a necessary adjunct for scientific dialogues, interpersonal communications media should be enriched via remote sharing of displays and applications among researchers. Distributed shared displays will be an important element supporting remote control rooms, but development is required to control and consolidate network transfers in a way that overcome bottlenecks arising from the inherent latency of long distances. Network bandwidth optimization by compression mechanisms, data segmentation and intelligent caching techniques need to be employed. Tools for implementing access and privacy policies should be implemented and integrated with authentication and authorization mechanisms.

3. SECURE COMPUTATIONAL SERVICES

The National Fusion Collaboratory project (NFC) [7] has deployed a service oriented computational and data management infrastructure, that emphasizes accessibility rather than portability. Tasks are carried out as loosely coupled services running on a distributed set of servers. The interfaces are simple and generic, describing the tasks to be accomplished, but do not prescribe how they are to be done. Implementation details and complexities are hidden whenever possible. Building on the success of the NFC, additional physics codes should be placed on high performance computers as services and added to the FusionGrid (a U.S. FES data and computational grid) [8]. FusionGrid's monitoring and distributed authorization tools should then be integrated into the collaboration workspace described above. The interoperability of FusionGrid with international grids needs to be realized. Extensions to MDSplus [9,10] should be implemented to support long-pulse or continuous data sources, and evaluated to provide remote access to mission-critical HEP controls data. This will improve the interactivity of data-intensive applications and should apply equally well to long-pulse and continuous experiments and simulations. Additionally, this will require concurrent writing, reading, analysis and visualization of data segments from extremely large records and databases. More challenging is the integration across time scales. The data sets may encompass more than a factor of 10^9 in significant time scales, leading to requirements for efficient browsing of very long data records and the ability to describe and locate specific data records accurately from within very long time series. Support for record identification, logging and tracking will be needed by users trying to correlate physical phenomena across the large set of data signals and the wide dynamic range of the time base.

The activities described above need to be carried out in an environment that allows appropriate management of resources and ensures the integrity of data, the security of facilities and the privacy of individual users. It will become necessary to develop and deploy a User, Credential, and Authorization Manager (UCAM). This new security management tool should combine the present capabilities of the FusionGrid authorization management system and credential management system into a single unified system and make it available to the HEP community. UCAM would support varied authentication techniques to facilitate single sign-on, including static passwords, one-time password (OTP), and X.509 credentials [11]. In this way, users and sites can participate regardless of their authentication techniques. Single sign-on should be provided through issuance of proxies for users and support renewal of proxies from authorized services. To ease the burden on end users, UCAM needs to support the creation, renewal, and revocation of X.509 credentials for users, hosts, and services. To allow all stakeholders to control their own resources, UCAM should provide a web interface for requesting new accounts, managing user information, and for requesting, granting, and revoking permissions. Since each experiment may have different personnel assigned to key

roles (e.g. session leader or engineering operator) on each run day, it is a requirement that permissions change on a daily basis through dynamic application of role-based authorization. Taken as a whole, these improvements on computational services will positively impact today's experimental and simulation science as well as show a path forward for offering supercomputer class codes in a services-oriented model.

4. DISCUSSION AND SUMMARY

Although this paper focuses on ways to improve scientific collaboration, the communities have considerable experience in placing remote collaboration tools in the hands of real users. For FES, the ability to remotely view operations and to control selected instrumentation and analysis tasks was demonstrated as early as 1992 [12]. Full remote operation of an entire tokamak experiment was tested in 1996 [13,14]. Today's experiments invariably involve a mix of local and remote researchers, with sessions routinely led by scientists at remote institutions. For example, over the last 2 years, 5-10% of experimental time on the Alcator C-Mod tokamak was devoted to runs with remote session leaders. For HEP, a 1990 proposal for a project that came to be known as the "World Wide Web" was aimed at providing easy access to "information such as reports, notes, databases, computer documentation and on-line systems help." Subsequent development of other tools, such as VRVS, continued the trend of exploiting the web to improve communication and collaboration. The advantage of these capabilities to experimental facilities is substantial: they can regain operational flexibility that is lost when the experimental schedule is complicated by travel requirements of collaborators. The benefit to collaborators is the ability to take their run time when machine conditions, subsystem and personnel availability is optimum for their experiments thereby enhancing scientific productivity. The history of collaborative technology adoption in these scientific fields indicates that as fast as new technology matures they would be put to real world usage to improve the science.

The goal of the work outlined in this paper is to significantly enhance scientific collaborations. This will be accomplished through the development and deployment of new tools and technologies into the working environments of FES and HEP. Experimental and theory/computational programs will greatly benefit through the provision of a flexible, standards-based collaboration space, which includes advanced tools for ad hoc and structured communications, shared applications and displays, enhanced interactivity for remote data access applications, controlled resource usage, computational services and an improved security environment. The technologies developed should be prototyped and tested on the current generation of experiments and numerical simulation projects. The work also needs to maintain a strong focus on the needs of the next generation of mega-projects, ITER and the ILC.

To carry out the proposed tasks, the work would leverage existing computer science technology and take full advantage of commercial software wherever possible. However, it is clear that not all of the requirements for scientific collaboration can be met without new development. For both new development and integration of existing tools, the overall approach will be one of rapid prototyping, deployment, testing and iteration.

To accomplish the goals, we envision the work to be divided into the two categories of, collaborative workspaces and secure computational services. The first aims at developing a set of tools or components that exploit the convergence of communications and computing technologies. Modules need to be created to support ad hoc and structured interpersonal communications, persistent collaboration environments along with shared displays and applications. Commercial user agents should be deployed, along with those developed for use in the science community, which would be adapted to work in this framework. Advanced directory services need to be provided that allow people and data streams to be identified, located, scheduled and connected into a flexible collaboration fabric. The second element will build on the service-oriented model developed by the National Fusion Collaboratory (NFC). As part of this project, several new advanced physics codes should be made available as network-available services, freeing researchers from implementation details and allowing a sharper focus on physics issues. Computational service deployments need to be extended to high performance parallel computers. MDSplus a widely used, service-oriented data management system, should be extended to support long-pulse or continuous data streams, meeting the requirements of existing and future experiments like LDX, MDE, KSTAR, EAST, and ITER, as well as the needs of numerical simulations. To provide the necessary security infrastructure, a federated web portal should be built, integrating tools for credential management, authorization and resource management.

At the conclusion of the work outlined in this paper, we envision an enhanced collaborative research environment. Together, the tools outlined will link the geographically diverse scientific communities assisting present day FES and HEP research, and show a clear path for successful collaboration into the future on ITER and ILC. For FES, this environment can be used for production research on present day facilities. For ITER, the path will be clear, showing how a grid-based application services model can support worldwide research even for supercomputer class codes and showing the components required for a successful collaboration framework. For HEP this environment can be used in a remote operations center that provides secure access to LHC and CMS control rooms at CERN as well as serving as a prototype control room and test facility for design efforts that are underway for a future International Linear Collider.

REFERENCES

- [1] <http://lhc.web.cern.ch/lhc/>
- [2] <http://www.linearcollider.org/>
- [3] F.W. Perkins, “ITER physics basis: overview and summary.” *Nucl. Fusion* **39** (1999) 2137–74.
- [4] J. Rosenberg, H. Schulzrinne, et al., “SIP: session initiation protocol,” Internet Engineering Task Force (IETF), RFC 3261, June 2002.
- [5] D. Adarnczyk, et al., “Global platform for rich media conferencing and collaboration,” Proc. Conf. for Computing in High Energy and Nuclear Physics, La Jolla, California, 2003.
- [6] L. Childers, T. Disz, R. Olson, M.E. Papka, R. Stevens, T. Udeshi, “Access grid: immersive group-to-group collaborative visualization,” Proc. 4th Intl. Immersive Projection Technology Workshop, Ames, Iowa, 2000.
- [7] D.P. Schissel, et al., “Grid computing and collaboration technology in support of fusion energy sciences,” *Phys. Plasmas* **12** (2005) 058104.
- [8] D.P. Schissel, et al., “Building the U.S. national fusion grid: results from the national fusion laboratory project,” *Fusion Engin. Design* **71** (2004) 245–250.
- [9] T. Fredian, J. Stillerman, “MDSplus remote collaboration support–internet and world wide web,” *Fusion Engin. Design* **43** (1999) 327.
- [10] T. Fredian, J. Stillerman, “MDSplus. Current developments and future directions,” *Fusion Engin. Design* **60** (2002) 229–233.
- [11] R. Housley, et al., “Internet X.509 public key infrastructure certificate and crl profile,” RFC 2459, <http://www.ietf.org/rfc/rfc3280.txt> (2002).
- [12] R. Fonck, et al., “Remote operation of the ITER BES experiment from an off–site location,” *Rev. Sci. Instrum.* **63** (1992).
- [13] S. Horne, M. Greenwald, et al., “Remote control of Alcator C–Mod from LLNL,” *Fusion Technol.* **32** (1997) 52.
- [14] B.B. McHarg, T.A. Casper, S. Davis, D. Greenwood, “Tools for remote collaboration on the DIII–D National Fusion Facility.” *Fusion Engin. Design* **43** (1999) 343.

ACKNOWLEDGMENT

This work was supported in part by the U.S. Department of Energy under DE-FC02-01ER25455.