U.S. Department of Energy

Scientific Discovery through Advanced Computing: National Collaboratories & High Performance Network
Summary Report

Proposal Number: 82694
A National Collaboratory to Advance the Science of High Temperature Plasma Physics for Magnetic Fusion

Organization Name: GENERAL ATOMICS
Principal Investigator: Schissel, David

---

1. Scientific and/or technical merit of the project.
When answering this question consider the following elements as appropriate:
a) The potential of the proposed project to make a significant impact on the effectiveness of SciDAC applications researchers.
b) The degree to which an application area can benefit from collaborative technology.
c) The extent to which the project will test important collaborative technologies.
d) The extent to which the results of the project are extensible to other program or discipline areas.

Reviewer 3 Comments:
This proposal will create and deploy collaborative software tools throughout the national magnetic fusion research community. The work focuses on linking all of the major experimental research facilities, relevant computing facilities, and other resources, in a manner that will permit remote scientists to more quickly and effectively use all of these tools. The collaboratory components will focus on three major areas of activity: security, remote and distributed computing, and scientific visualization. The work appears highly synergistic in that both the applications area and the collaboratory CS efforts should benefit substantially from this partnership. The proposal provides a very focused plan of activity, including both research and advanced development, in addition to the expected deployment. In particular, the security work builds on the latest work on user authorization, with the plan here to integrate it with the latest security technology. The complexity of the authentication policies required for this environment will be an excellent test of the capabilities of the Akenti tools. The work on the certificate authority implementation and the issue of credential delegation also pose interesting and useful topics that should enhance the general capabilities of collaboratory environments. Much of the work on remote and distributed computing will focus on deploying much of the existing technology to the experimental centers – as should be expected. However, there are a few new wrinkles that must be addressed, including end-to-end resource management, managing a diverse set of policies. Again here, this very practical environment should help to improve and harden the Grid tools. The remote visualization topic has been covered in a lot of prior work, but the issues here are a different aspect of the problem. This issue here is dealing with large remote display walls to help convey the status of various experiments and to share large volumes of real-time data quickly. This is a very focused, specific problem that again will contribute to many other grid research efforts. From the application standpoint, this type of research environment should provide a dramatic new approach for carrying out their research in the future and help to accelerate results while making far more effective use of their existing and planned facilities.

Reviewer 4 Comments:
The PI proposes a National Fusion Collaboratory prototype that will use new developments in security, distributed computing and visualization to significantly impact design of fusion energy source. Computer Science areas are well defined and well understood; integration of the technology with actual requirements of the fusion scientists is well described, including several scenarios specific to the fusion community.

Success of the proposed project benefits fusion researchers as follows: first, by allowing real-time data analysis and communication among geographically dispersed teams of scientists; second, by allowing more efficient utilization of scarce lab facilities, allowing more throughput; third, by allowing comparison of experimental and simulation data "on the fly"; fourth, raising scientists' productivity by making data, analysis codes, and other tools available through a single interface that separates the scientist from the details of location and logons. The collaboratory will facilitate the communication between theoretical and simulation fusion research communities; and that large-scale visualization will make the experience equally engaging in all locations. In particular, there is a 15-minute window between plasma pluses; the more experimentalists can understand in that time window, the better able they are to configure the next pulse. Also, access to higher performance facilities can reduce plasma nonlinear dynamics analysis time from days to minutes. The project proposes advances in key collaborative technologies. For security, development of fair use policies and automated conveyance of those policies to certificate management; integration with MDSplus and SQL server systems; and extends the Grid/Akenti architecture. For distributed computing, Globus will be extended to include end-to-ed resource management with advance resource reservation. For visualization, workstations will be integrated with the larger screen Access Grid. New software will be written to create an apparent single point of contact for display resources. New software will be created to enable visual comparison of simulated and experimental data and to better represent regions of uncertainty. Distributed security will be extended to apply to SQL server security. All grid tools are widely employed in other disciplines and the functionality of the proposed new modules has parallels in other disciplines. Thus, results should be extensible to other fields.

Reviewer 5 Comments:
Proposal 82694 (Dr. D.P. Schissel, et al) addresses the creation of a national collaboratory in fusion research. The team brings a history of work in this community, building toward increasing sophistication in their collaborative tools. Computer Science research is already very clearly focused, on; a collaborative control room, collaborative meeting room, and visualization tools. Innovatively, the proposal addresses community needs for code development, security architecture sufficient to encompass COTS databases, pre-emptive scheduling, data visualization tools incorporating solution uncertainty. The choice of the three main fusion labs appears well thought out, particularly in view of their prior history of meaningful collaboration in their science work. Additionally, the proposal recognizes the centrality of the effectiveness of this desire for meaningful collaboration. Solid provisions have been made for an oversight committee from the fusion community. Although the project will ultimately be quite ambitious, initial plans indicate an understanding of the major issues involved, and of opportunities for cost savings and usage community broadening via reuse of extant tools and equipment. Work-up testing will using extant equipment, and the effort will leverage DOE Science Grid work, and several COTS and research tools. Details of the proposed work could have been significantly more complete.

Reviewer 6 Comments:
This proposal represents a strong, complete plan for the investigation and deployment of collaboration technologies enabling the sharing of information between experimental laboratories. The proposed project implements an end-to-end infrastructure allowing the secure and flexible distribution of magnetic fusion data across a wide range of participant sites. The expected results of this effort will result in vast reduction in time and complexity of the analysis of ongoing experiments and simulations. As such, a successful completion of this work will significantly increase the pace with which such research is conducted and results disseminated. A key consideration in the design of this infrastructure is the ability of researchers to efficiently transfer the vast data-sets between local and remote resources. The high volume of data in conjunction with the quality of service demands of synchronous communication will place significant burdens on collaborative systems and supporting infrastructure. Historically, performance has not frequently been a focus in the investigation of collaborative systems. Hence, the ability to meet the unique needs of the target environments will be a key ingredient to its success. While not emphasized in this proposal, the success of this work is predicated on the discovery of techniques advancing performance in collaborative systems. However, if found, these are likely to be applicable across a wide range of problem domains.

Reviewer 7 Comments:
The problem area here is magnetic fusion. One of the unique features of this application area is that the

collaboration task has deadlines - there are pulses to be tracked here and decisions have to be made regarding a pulse before the next pulse is emitted about 15 minutes later. Tasks with deadlines are those that benefit most from CSCW tools, so seems like a good application area. Some interesting scenarios are presented here. The components of the proposal are the usual suspects: authentication, distributed processing, distributed collaborative visualization, etc. Some distinguishing features are tiled displays, visualization comparison with uncertainty. However, no new idea was presented in this proposal. In fact, not even sure some of the problems mentioned here are open - in particular automatic propagation of security credentials seems to be solved by Foster's system (earth system grid). Perhaps the twists here are to extend previous security work to commercial databases and to integrate large displays and fusion tools with grid tools. I wish previous research in distributed/real-time scheduling was mentioned to convince the reader that some new ideas will be pursued here

2. Appropriateness of the proposed method or approach.

When answering this question consider the following elements.

a) The degree to which the project adheres to the management philosophy of incorporating collaboration into the project execution.

b) The quality of the plan for ensuring interoperability and integration with software produced by other SciDAC efforts.

c) The extent to which the project incorporates broad community (industry/academia/other federal programs) interaction.

d) Quality and clarity of proposed work schedule and deliverables.

e) Knowledge of and coupling to previous efforts for collaborative technologies such as DOE 2000.

Reviewer 3 Comments:

For all three CS concentrations, the proposal gives significant detail regarding the issues and the plans for addressing these issues. The plans appear to be quite sound, building on all of the latest technologies, using them where appropriate, and giving a strategy for how to extend it where it is currently insufficient. All of these plans appear solid. I am impressed with the careful connection of application needs to specific tools that will be put into the collaboratory. The PI's have decided to keep there attention of a few key areas of work that they need most and not stray off into a wide range of topics that may lose time and energy.

Reviewer 4 Comments:

The overall project manager is a technical expert from a large fusion science lab. The project management team is composed of computer scientists and technical experts. The fusion community will participate through an oversight committee to advise on requirements and provide feedback. Existing SciDac software will be extended. The community consists of General Atomics (nuclear engineer/lab manager), ANL (computer science/parallel computing/grid; scientific programming; virtual reality/grid), LBNL (on-line storage/security), MIT (instrumentation/database; fusion scientist), Princeton U (computer science/visualization; operating systems/parallel computing; computational plasma physics), and U. Utah (visualization/parallel computing; computational steering). The work schedule and deliverables are specified in detail. Project PI's are leaders in the fusion, collaboratory/Grid and visualization communities. They have current funding in DOE Science Grid, DOE Certificate Authority, and other DOE 2000 and SciDAC funding.

Reviewer 5 Comments:

The proposal only very briefly mentions "live user tests, both for offline analysis and computation". If the project goal is to produce a _working_ community collaboratory, these tests deserve considerable elevation in importance. Metrics and scientific studies to be used in project validation are only minimally discussed. Particularly, the claim that display walls enable "more effective" exploration by scientists begs formal confirmation and quantification. Language is equivocal on the "feasibility" of Access Grid deployment. Deployment can be presumed, but "effective use" needs to be formally investigated in this community's context. In pre-emptive scheduling, the host site must agree to give up their resource on (your) demand. Will they? Why should they? Who will bear the costs of the disruption to their work already in progress? Working out such questions may well be harder than implementing the job submission code itself. The proposal, already expensive, does not include display walls at the premier 3 main control rooms, which

would seem to preclude many possible utilization effectiveness experiments. The proposal is generally weak in the degree of detail associated with construction of all the network services that will be required. However, the proposed price tag and the expertise of the researchers chosen appear appropriate for the work. A more detailed implementation plan is, however, generally desirable. The proposal seems to co-mingle "collaboration" with "communication". Collaboration spans a longer time base, and includes phases without communication. Collaboration encompasses work practices and pressures of communities, and thus is not completely approachable by upgraded video bandwidth. It is a long haul from "develop the infrastructure necessary for..." to "create and operate a working collaboratory in this domain". The goals (task schedule) are not very aggressive as stated, in that (with the exception of the Scientific Visualization group) they frequently use words like "demonstration" and "create", and thus tend to stop short of promising the full-deployment stages that will prove-out the system.

Reviewer 6 Comments:
The proposed project is centered around the integration of collaborative technologies appropriate for widely dispersed experimental environments. This project is clearly grounded in the previous and ongoing DOE projects, and cites many examples of other projects that may make use of the proposed environment. The proposal tasks of security, distributed systems, and visualization are built on well-accepted technologies with proven track records. These systems are extended where the facilities are not perfectly matched with the project requirements. The aggressive schedule clearly outlines a plan for the development and integration of a number of technologies. However, there is not inexplicit task and (associated deliverable) for the investigation of security and resource allocation policies. Defining policies representing desirable behavior in collaborative endeavors is a complex and time consuming task. Specific coalition models are currently under investigation at a number of institutions. Reporting the processes and results of policy investigation may be warranted.

Reviewer 7 Comments:
Fine on all fronts.

3. Competency of applicant's personnel and adequacy of proposed resources.

Reviewer 3 Comments:
Overall, the team has an excellent track record for research and development relevant to this project. The CS team includes many of the recognized leaders in security, grid computing, and visualization. They are the right people to do the work.

Reviewer 4 Comments:
The PI and co investigators are good computer scientists, some with a great deal of experience in fusion energy research. They have numerous publications that jointly cover areas of expertise described. Research resources appear to be adequate for the proposed research.

Reviewer 5 Comments:
In the creation of a working collaboratory involving instruments of such great cost and limited national availability, involving research colleagues from the Sociology, Communications, Ethnography, Computer Supported Cooperative Work, and/or Computer Mediated Communications fields is strongly encouraged. Metadata, data archival, and intellectual property (rights management) issues would also benefit from the inclusion of researchers from the Digital Library community.

Reviewer 6 Comments:
The investigators of this project represent some of the leaders in the fields of visualization, collaborative technologies, and security in computational grids. As such, it seems likely that they will be able to successfully carry out the proposed research.

Reviewer 7 Comments:
Only proposal with researchers with publications is decent places. But the work proposed here seems to be all system development. Not clear these researchers have anything useful to contribute to this project.

4. Reasonableness and appropriateness of the proposed budget.

Reviewer 3 Comments:
Overall, the budget looks reasonable for the amount of work described. I have a few minor concerns about pieces. For example, one of the two separate proposals from ANL (the one covering visualization) has no one on the project at more than 25% time –that does not seem to me like they are substantially committed to doing the work. I am also concerned that Stevens is listed on this proposal at 10% time, when he is a senior manager at ANL and also listed on about six other proposals. The other observation is that most of the budgets do appear to be targeted at approximately $ 250 K per year. I find it remarkable that the workload is so balanced among the various institutions that they need exactly equal efforts. My suspicion is that the budgets could be cut a bit, but it is very hard to tell.

Reviewer 4 Comments:
The budget is primarily for personnel, with minor amounts for travel, materials, and computer support. This is consistent with the project description.

Reviewer 5 Comments:
The budget requested appears reasonable for a major collaboratory spanning the key sites of a whole science specialty. However, proposals of this magnitude should have more detail provided concerning the tasks involved.

Reviewer 6 Comments:
As the budget is realistic in that the vast majority of expenses are based on human costs. The use of COTS and open source software, and the commodity hardware will greatly enhance the accessibility of its solutions.

Reviewer 7 Comments:
Individual institutions are receiving modest amount of money but there are too many institutions.