TeraGrid Planning Process Report: June 2007 Workshop for Science Gateways

Katherine A. Lawrence
<kathla@umich.edu>
Ann Zimmerman
<asz@umich.edu>
Collaboratory for Research on Electronic Work
School of Information
University of Michigan
Ann Arbor, MI 48109-2112
http://www.crew.umich.edu

September 5, 2007

Table of Contents

Executive Summary

TeraGrid is a national, comprehensive, distributed infrastructure integrating multiple resources at nine resource provider facilities. In late spring 2007, the NSF awarded a grant to the University of Michigan's School of Information (UM-SI) to facilitate a community-driven, participatory planning process for the future of TeraGrid. This process is intended to anticipate the changes that are already occurring in the use of HPC services and resources. To gain an understanding of the requirements of current and potential future users of TeraGrid, UM-SI conducted a series of three workshops. The first workshop, which focused on the needs of those developing TeraGrid Science Gateways and the needs of Gateway users, was held on June 7, 2007, immediately following the TeraGrid '07 Conference in Madison, Wisconsin. All twenty-one projects currently designated as TeraGrid Science Gateways were invited, and representatives from seventeen gateways attended the workshop.

The guiding question for the workshop was "What would TeraGrid be if it met the needs of your science gateway perfectly?" Answering this question required a better understanding of the future goals of each project and the specific role that high-end computing resources and services (such as those provided by TeraGrid) would play in meeting the scientific needs of the gateway user communities. The workshop was organized to encourage small group discussion of these topics and cross-fertilization of ideas. The first exercise focused on the question "What role does TeraGrid play in meeting the scientific needs of your gateway user communities?" and generated a list of six important topics:

1. Creating a *community environment* with appropriate templates and shared code (including the approach of horizontal integration with other infrastructures, virtual platforms, virtual organizations, and common authentication)
2. Managing *scheduling flexibility and allocations*, such as throughput on demand or opportunistic scheduling
3. Sharing of *best practices through education and training* for gateway creations (including educating faculty and graduate students)
4. Identifying *appropriate performance* (in response to issues such as high throughput vs. high performance or the need for handling large amounts of data)
5. Supporting *ease of use and reliability* of software
6. Offering a *system that is consistent across machines*, thereby creating a real grid

The second exercise was based on these six topics, asking "Regarding [this topic X], what do our science gateways need from TeraGrid?" The group generated many ideas and solutions, and then prioritized solutions according to three criteria:

- Most important to end users
- Most important to science gateway developers (i.e., makes development easier)
- Impact per resources spent (money, people, etc.); broad impact in terms of publications, students, classes, etc.

Two overarching themes came out of this exercise. The first was supporting interaction and cross-fertilization among the science gateway development communities. Specifically, developers wanted to share code and successful solutions, and they hoped for greater financial

and professional support for developing gateways. The second was reducing the hurdles to getting on the TeraGrid. For example, developers were concerned with the reliability and tracking of upgrades, the length of the software development cycle, and the bureaucracy associated with using TeraGrid.

In sum, gateway developers are excited at the potential of TeraGrid to make HPC available to end users and communities who would otherwise be unable to conduct their research as effectively or efficiently. In addition, they are excited at the opportunity for distributed communities to work together on common solutions. Meanwhile, they are eager to move TeraGrid toward a collaborative mindset that enables the developers to focus on the unique needs of their gateway communities. At present, they find that too much energy is focused on re-creating custom solutions when standardized systems or a TeraGrid-hosted gateway layer would suffice. The prominent cross-topic patterns identified from the workshop outcomes advocate for:

- Basic services that gateways can use instead of creating or hosting their own.
- Templates and standardized systems to save developers the time of recreating things that others have already built.
- Standardization that would make TeraGrid a *real* grid that could support the effective use of allocations and meta-scheduling.
- Operating more effectively as a community in order to better support the education and development needs of gateway developers.

Introduction

TeraGrid is a national, comprehensive, distributed infrastructure integrating multiple resources at nine resource provider facilities. The project is funded by the National Science Foundation (NSF) and provides access to computational resources, primarily in the form of supercomputers, large amounts of storage space, visualization services, fast networks, and software. Following a 5-year construction phase, TeraGrid became operational in late 2004. TeraGrid's resources currently include more than 150 teraflops of computing capability and more than 15 petabytes of online and archival data storage. High-performance networks provide rapid access and retrieval to data. TeraGrid supports a variety of use cases ranging from exploiting a single TeraGrid resource to combining resources across sites. The latter capability opens up new possibilities for conducting scientific work. Since TeraGrid became operational, many changes have occurred. For example, TeraGrid resources have been opened up to new user communities through the TeraGrid Science Gateways program, the number of resources providers has grown, and a user portal has been developed to make it easier to obtain information about accounts and resources and to simplify access to TeraGrid services. Further changes are certain to come. For instance, by the time the current grants expire in 2010, a petascale resource will be on the horizon, the user community will increase and diversify, and new policies and services are likely to be needed to meet the needs of users and the expanding pool of high performance computing (HPC) resources.

In late spring 2007, the NSF awarded a grant to the University of Michigan's School of Information (UM-SI) to facilitate a planning process. The planning process itself is led by a steering committee consisting of individuals representing key stakeholder communities and various areas of expertise. The outcome of this process will be a report that discusses options for the future delivery of TeraGrid services and resources based on the diverse needs of science and engineering communities. This report will be written by the steering committee for the audience of stakeholders. The process is intended to anticipate the changes that are already occurring in the use of HPC services and resources. By relying on a community-driven, participatory process, the steering committee and facilitation team hope to achieve wide acceptance of the planning outcomes.

One objective of the first phase of the process is to gain an understanding of the requirements of current and potential future users of TeraGrid. During summer 2007, as the steering committee was being established, researchers and staff at the UM-SI organized and conducted a series of three workshops to elicit preliminary information on user requirements that could be used to inform subsequent phases of the planning process. This report describes the results of the first workshop, which focused on the needs of those developing TeraGrid Science Gateways and the needs of Gateway users. The workshop was held on June 7, 2007, immediately following the TeraGrid '07 Conference in Madison, Wisconsin. All twenty-one projects currently designated as TeraGrid Science Gateways were invited, and representatives from seventeen gateways attended the workshop.

The remainder of this report is organized as follows. We begin with a brief description of the TeraGrid Science Gateways program. Next, we discuss the workshop's guiding question, the participants, and the results of a pre-workshop survey. From here we summarize the activities

that were used to address the guiding question and to acquaint attendees with each other and their projects. The main part of the report analyzes the results of the workshop discussions and activities. We conclude with a summary of the key findings from the workshop and the results of the meeting evaluation survey.

TeraGrid Science Gateways

Early in its history, TeraGrid conceived the idea for what has become the TeraGrid Science Gateway program. Recognizing that many disciplinary communities were building elements of their own cyberinfrastrucuture, TeraGrid set out to form partnerships that would provide TeraGrid resources and services to user communities through tools and environments they were already using (Catlett, Beckman, Skow, & Foster, 2006). Essentially, the gateway concept recognizes that many of today's scientists use desktop computing applications and web browsers to conduct their work. The program also recognizes that different communities have different needs.

TeraGrid's role is as a back-end service provider with the gateways serving as the front end to the user. At the time of the workshop, there were 21 projects with an allocation on the TeraGrid, which is the criterion used to designate a project as a TeraGrid Science Gateway. A wide range of disciplines are represented including astronomy, biology, chemistry, computer science, earth science, engineering, materials science, and physics. These projects are similar in that they have external funding to build a community-specific cyberinfrastructure. For more information about the TeraGrid Science Gateways, see Wilkins-Diehr (2006) or the gateways section of the TeraGrid web page (www.teragrid.org/programs/sci_gateways/).

Workshop Purpose and Participation

The workshop was designed to assess how TeraGrid could meet the needs of Science Gateways, specifically the end users of the gateways and the gateway developers. The guiding question for the workshop was "What would TeraGrid be if it met the needs of your science gateway perfectly?" Answering this question required a better understanding of the future goals of each project and the specific role that high-end computing resources and services (such as those provided by TeraGrid) would play in meeting the scientific needs of the gateway user communities. Thus, the workshop activities were organized with these topics in mind. Secondary goals of the workshop were to solicit ideas for engaging others in the planning process and to provide an opportunity for gateway developers to meet and interact with their colleagues.

We contacted the lead Principal Investigator (PI) of each gateway project, inviting them to attend the workshop or send a qualified gateway development team member in their place. (See Appendix A for a copy of the invitation and Appendix B for the list of participating gateways and participants.) Although a few gateways were unable to participate due to scheduling conflicts, illness, or other reasons, the following 17 Science Gateways were represented at the workshop, representing a broad range of TeraGrid's partners:

Biomedical Informatics Research Network (BIRN)

Caltech Science Gateways (Grid Analysis Environment)

Computational Science and Engineering Online (CSE-Online)

GEOsciences Network (GEON)

GISolve (Geographic Information Science Gateway)

GridChem (Computational Chemistry Grid)

Linked Environments for Atmospheric Discovery (LEAD)

Massive Pulsar Surveys using the Arecibo L-band Feed Array (ALFA)

nanoHUB

National Biomedical Computation Resource (NBCR)

NESSSI (National Virtual Observatory or NVO)

Neutron Science Instrument Gateway

Open Science Grid Consortium and Project (OSG)

Southern California Earthquake Center Earthworks Project

Telescience Project

TeraGrid Visualization Gateway

Virtual Laboratory for Earth and Planetary Materials (VLAB)

Results of Pre-Workshop Survey

Before the workshop, we asked participants to answer questions in a brief online survey. The goal of this survey was to gather some general information about the projects and to spur attendees (and their Gateway colleagues) to think in advance about the topics to be discussed at workshop. We also asked the participants to create an introductory slide about their gateway projects using a template we provided. (A full list of survey questions is in Appendix C. Slides are available on the planning process website: www.teragridfuture.org.)

The survey results showed that participants represented a range of roles on science gateway projects, including project directors, researchers, software developers/engineers, graduate students, and project managers. The science domains represented by their gateway projects were also diverse and included earth sciences, chemistry, biology, physics, astronomy, and nanotechnology, as well as projects open to all domains. Almost all of the projects were funded by NSF—ten exclusively by NSF—and some received additional funding from the Department of Energy and National Institutes of Health as well as more specialized funding sources. Gateways are primarily designed to support researchers, educators, and university-level students, but seven of the respondents indicated support for students in grades K–12 (see figure 1). Three of the gateways are open to the general public and two expect use by policy makers. More than half (ten respondents) consider their gateway development process to devote an even balance of effort between research and development; most of the remainder have a stronger bias towards development (see figure 2). All of the gateways have made some development progress; about two-thirds have at least 50% of their gateway in operation (see figure 3).

Figure 1: Expected User Types

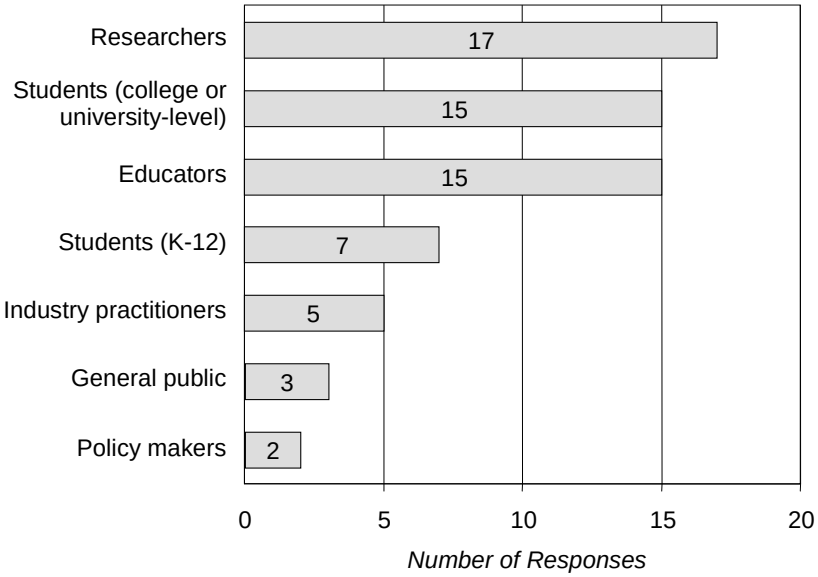


Figure 2: Focus on Research (as Opposed to Development)

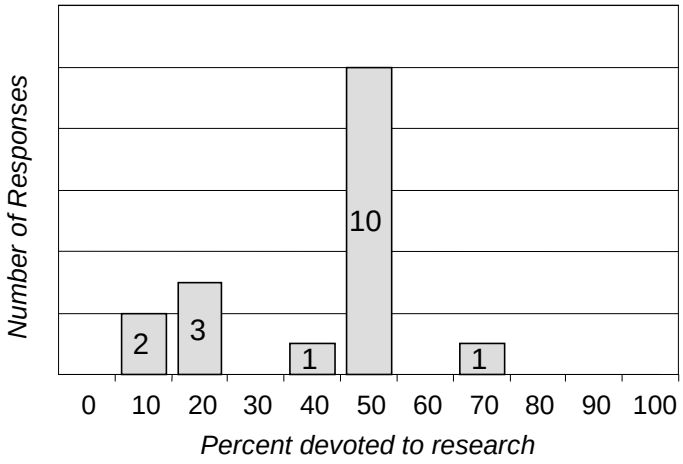
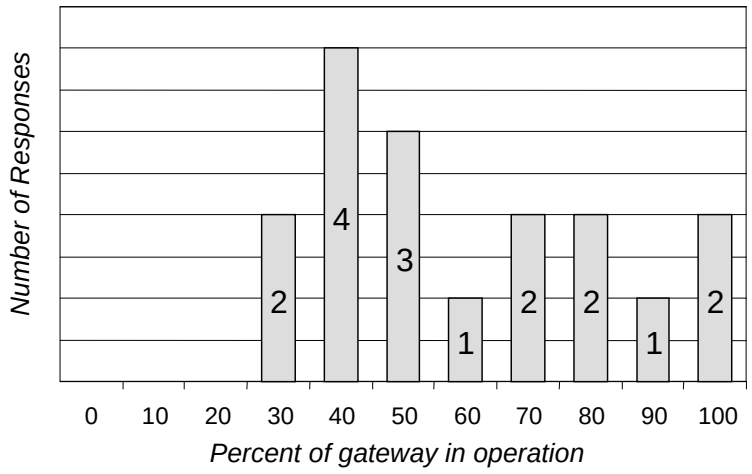


Figure 3: Development Completed



The survey included several open-ended questions. Participants indicated a range of current and future user needs that are driving the services and capabilities of their gateways (many of which were raised again during the workshop). Not surprisingly, TeraGrid primarily supports the storage, management, manipulation, and transfer of large amounts of data. Gateways also depend on visualization and simulation capabilities which may not always use large amounts of data but which cannot be performed on local, desktop computers or servers. TeraGrid is a valuable resource because the high-end speed and capacity are beyond what end users could afford, and the resources are critical to rapid advancement of scientific discovery.

Other open-ended questions encouraged attendees to consider the current state of TeraGrid, TeraGrid's potential, and how their projects can work with TeraGrid to their mutual benefit. Responses indicated that the main ways that TeraGrid could exploit its potential reach were through outreach, communication, and technological clarity. For example, one person observed that to improve understanding of TeraGrid, education needs to be twofold: inform senior researchers about the capabilities of the system and teach the hands-on users (postdocs, students, staff) how to work with TeraGrid. Another recognized the importance of balancing the different needs of power users and regular users on the same system or gateway. To promote grassroots adoption, TeraGrid might consider how to ease access through graduated security and other simplified mechanisms. TeraGrid could also seek out ways of working with other NSF cyberinfrastructure projects and national networks.

Participants also saw potential ways in which TeraGrid could better explain what it has to offer. TeraGrid management needs to clarify its unique relationship to the gateways, specifying the services that will be available, how requests for allocations and services will be handled, and in turn, what gateways must provide TeraGrid. One participant also emphasized the need for better real-time information about the current resources so as to help schedule and match user needs with available services. Several participants identified the key role of Globus and the Globus Toolkit to help make the resources more portable, reliable, and easier to use. Others remarked on the need for scalable tools and services, made available to gateway developers. Such advances would help make TeraGrid attractive to new users and communities.

Workshop Structure and Activities

The workshop was organized to encourage small group discussion and cross-fertilization of ideas. Appendix D contains a copy of the workshop agenda. The majority of the time was devoted to generating issues, ideas, and solutions related to supporting the TeraGrid requirements of gateway end users and developers.

The workshop began with an overview of the TeraGrid planning process and a summary of the pre-survey (as described above). Next, each of the participants introduced his or her gateway project, following the slide template we had provided in advance. Their introductions included their own role on the gateway, the field/discipline of the end user audience, their project start date, the project's goals, the current and expected number of users, computational requirements, their measures of success, and an interesting tidbit. As noted above, these slides are available on the planning process web site.

Next, to warm up the participants for their small group discussions, we began with a brief physical exercise that was intended to help attendees get to know each other and to illustrate the similarities and differences between the gateway projects. For example, in the open space at the center of the room, the attendees arranged themselves in their geographic locations across the US. Next, they organized themselves in order based on the number of people working on their gateway projects. Curiously, the majority of projects have fewer than a dozen people working on them (though some may have interpreted the instructions to mean "number of people at their location," not across the entire project). Finally, the participants stood in a two-dimensional space that represented the number of different kinds of end users versus the level of comfort of their end users with using high performance computing services. One participant pointed out that the real goal of the gateways should be to make HPC accessible to all, regardless of ability. The remainder of the workshop was divided into two main exercises.

Exercise 1: World Café

The World Café is an effective but simple method for generating small group conversations that benefit from the insights of people throughout the larger group. The conversations begin at small tables of four people each, and each table is outfitted with a paper tablecloth and colored markers. The small group members are encouraged to share in the task of taking notes on the tablecloths, while making a point of listening carefully, connecting ideas, and noticing deeper patterns. After twenty minutes, one person stays at each table as the "host" while the remaining participants rotate to other tables with a different mix of people. The hosts share the essence of their prior conversation with the newcomers (although they do not formally facilitate), and the newcomers add connections and ideas from their prior tables. At the end of this second, twenty-minute round, people return to their original tables and report back, collecting ideas so that they can present their key themes to the other groups. Ultimately, these conversations build on each other as people cross-pollinate and link their ideas.

The overarching question discussed during the World Café exercise was "What role does TeraGrid play in meeting the scientific needs of your gateway user communities?" We asked the participants to listen for common themes during the small group presentations. While the larger group identified and discussed these common themes, we took notes of the topics that emerged. (All ideas and notes are summarized in the next section.) From this discussion, the participants identified six topics that formed the basis for the second exercise of the workshop.

Exercise 2: Wandering Flip Charts

The Wandering Flip Charts exercise is an efficient means of engaging a large number of people in the discussion and solution of many issues in a short period of time. During a break, the facilitators wrote each of the six topics at the top of a flip chart, and the flip charts were distributed around the room. For this second exercise, participants were instructed to wander to a flip chart with a topic that interested them the most, discuss the topic with the others there, and list the issues on the flip chart. The overarching question was, "Regarding [this topic X], what do our science gateways need from TeraGrid?"

Although this exercise is designed for participants to meet in small groups of three or four people for five minute intervals at a flip chart, the participants had discussed many of the issues as a larger group before the break and so they instead wandered around at will, writing issues and

ideas that were important to them and discussing them with their colleagues. After sufficient time generating ideas, the flip chart sheets were removed and posted alongside the charts, and then the group repeated their circulation among the charts, except this time the goal was to generate solutions to the issues that were listed for each topic.

Reconvening after another fifteen minutes, the group discussed what criteria should be considered when prioritizing the solutions that had been generated. For example, solutions that would be most important to end users might be different from those that would be most viable. Each participant received three colored sticker dots for each criterion (described below) to vote for their favorite solutions.

After voting, we concluded with a general group discussion of the results. We asked if anyone felt that something was missing or underrepresented. We also discussed what their role and TeraGrid's role might be in informing each other about requirements and opportunities. To close, we asked each participant to make a final comment, if desired. The facilitators took notes during this final portion of the workshop and also handed out a workshop evaluation form.

Findings

Key Themes Generated

The discussion during the "world café" exercise was lively and wide ranging. After the small groups reported to each other, the broad discussion of the question "What role does TeraGrid play in meeting the scientific needs of your gateway user communities?" revealed more than a dozen common themes, some of which were related to each other. The groups raised many issues and questions which were captured in the subsequent idea and solution generation exercise, so we have grouped and synthesized them in that section, but a synopsis of the small group reports can be found in Appendix E.

The themes were distilled into the following six main topics (in order of those garnering the most votes in the subsequent prioritization stage of the activity):

1. Creating a *community environment* with appropriate templates and shared code (including the approach of horizontal integration with other infrastructures, virtual platforms, virtual organizations, and common authentication)
2. Managing *scheduling flexibility and allocations*, such as throughput on demand or opportunistic scheduling
3. Sharing of *best practices through education and training* for gateway creations (including educating faculty and graduate students)
4. Identifying *appropriate performance* (in response to issues such as high throughput vs. high performance or the need for handling large amounts of data)
5. Supporting *ease of use and reliability* of software
6. Offering a *system that is consistent across machines*, thereby creating a real grid

Additional themes that were not included in the second exercise were:

- Managed persistence of capabilities and interfaces, information and data

- Web availability of non-web services
- Roles—e.g., on demand
- Software service provider support - service role
- Database support
- Balancing of cost, throughput, and time—implied by this is meta-scheduling
- Equity-barter

With the six themes listed above, the group began the "wandering flip charts" exercise. The results of this exercise are described in the next section, combined with prioritization outcomes.

Prioritized Solutions

After completing the issue-generation and solution generation parts of the second exercise, the group discussed as a whole what criteria would be important for voting on and prioritizing solutions. The group suggested a number of criteria, and the top three were:

- Most important to end users
- Most important to science gateway developers (i.e., makes development easier)
- Impact per resources spent (money, people, etc.); broad impact in terms of publications, students, classes, etc.

Other criteria which were not used for voting included most viable or likely to work, most sustainable or scalable, most innovative, and most transformative.

In the following subsections, we review the issues and the prioritized solutions, identifying the different ways that these ideas are important to the future development of TeraGrid. We also include a table itemizing the solutions and the votes they earned for each of the three criteria above.

1. Creating a community environment with appropriate templates and shared code (including the approach of horizontal integration with other infrastructures, virtual platforms, virtual organizations, and common authentication)

The issues associated with this theme arose as part of the concern that a great deal of effort is expended developing duplicate code and systems that have already been created successfully elsewhere. Because each gateway has limited resources, they would like to see some aspects of the systems they create be made available as standard or shared services and tools. For example, new gateway communities would be better served by tapping into pre-existing, generic templates for creating a hub, as well as by accessing web services such as authorization, job submission, data distribution, data access, and resource discovery. Likewise, centralized TeraGrid support for such services and virtual organizations would make this standardization more effective. In fact, one person commented that if TeraGrid were to provide a gateway framework as a hosted service, new gateway developers could readily develop domain science applications for their community without having to deploy and maintain an entire framework. A related issue is the need for greater standardization across TeraGrid sites. For example, users would like to be able to run the same code on different machines at different sites. A TeraGrid-hosted framework could reduce much of the overhead associated with using varied resources. (This issue was also expressed in the context of creating a "real grid"—see subsection 6 below.) In support of this

cross-fertilization, participants were eager to have high bandwidth between the greatest possible number of resources.

Thus, the solutions considered to be the most important for supporting these concerns, particularly to ease the development of science gateways, were (1) the development of gateway framework templates built upon toolkits which may already exist, (2) partnering with National Lambda Rail or Internet2 in order to provide greater bandwidth between sites, and (3) examining what is happening in other grid communities with the aim of learning best practices from them. Although each of these solutions is most beneficial for gateway developers, some participants also felt they represented a good use of resources. A solution that would particularly benefit end users (but received votes for the other two criteria as well) would be to have common scheduling of jobs across different TeraGrid sites.

Users [a]	Gateways [b]	Resources [c]	TOTAL	SOLUTIONS
	5	3	**8**	Develop gateway framework templates built upon toolkits which may already exist
3	4	1	**8**	Peering with NLR (National Lambda Rail)/Internet2, etc.
4	1	2	**7**	Have common scheduling of jobs across different TeraGrid sites
	4	1	**5**	Examine what's going on in other grid communities (OSG, eScience, etc.)
3		2	**5**	Hosting web services
	1		**1**	Integrate applications and other resources for resource discovery → uniform software stack for different kind of machines (Note: this may be difficult in practice, but common interfaces at least are essential)
	1		**1**	Keep "data" open with instant acknowledgment of the generator
				WSDL publication should be maintained (Note: WSDL=Web Services Definition Language)
				Documentation (someone drew an arrow from WSDL... to here)
3	2	9	**32**	**Subtotals and Totals**

[a] Most important to end users
[b] Most important to science gateway developers (i.e., makes development easier)
[c] Impact per resources spent (money, people, etc.); broad impact in terms of publications, students, classes, etc.

2. Managing scheduling flexibility and allocations, such as throughput on demand or opportunistic scheduling

The issues associated with this theme reflect difficulties that already exist for making effective use of TeraGrid resources through scheduling and allocations. Participants would like to see

scheduling issues addressed by being able to send jobs to available sites, which would require meta- or co-scheduling systems that share a common interface at the very least. One person noted that a common file system would be useful as well, although this might be a stretch). Meta- or co-scheduling could be accomplished with an automated system providing an application code interface that could send jobs or data to the most responsive resources. Allocations thus need to be parceled out in ways that are more sensitive to the variety of user accounts—people, communities, subgroups, and shared accounts. As the popularity of gateways (and the number of users more generally) grows, allocation processes will need to be reconsidered.

The solutions generated for these issues were primarily of value to end users, but the most important solution—seriously pursuing meta-scheduling by funding development for it—is very valuable for end users and gateways and a good use of funding resources. In fact, meta-scheduling may be critical for the effective use and growth of cyberinfrastructure going forward, as it is already inhibiting growth today. Other important solutions included allocating separate resources for on-demand computing, standardizing file system interfaces and metadata for working across nodes and tools, and improving the availability and quality of information about the status of resources.

Users	Gateways	Resources	TOTAL	SOLUTIONS
3	3	5	**11**	Take meta-scheduling seriously, not as a future dream! Allocate funding for development
5			**5**	Allocate separate resources for on-demand computing
3	1		**4**	Uniform file system interface • Needs to be reliable/efficient
	3	1	**4**	Improve the availability and quality of information services
1	2		**3**	Standard/common metadata for sharing job information across nodes & tools, i.e. maybe an "ontology"-based resource (both hardware & software) description language
12	9	6	**27**	**Subtotals and Totals**

3. Sharing of best practices through education and training for gateway creations (including educating faculty and graduate students)

This theme focuses on issues of making it easier for gateways to use and develop within TeraGrid, but it also speaks to more broadly spreading awareness of cyberinfrastructure among those that would likely use it. Some of the barriers that exist for gateway developers include finding out what factors will contribute to their success, creating many basic services from scratch, updating gateways to work with new releases of TeraGrid, and developing and deploying desktop applications.

The primary solutions for these issues are seen as providing multiple levels of education to both end users and developers. For example, graduate students and postdocs need technical education

while senior researchers who may not be hands-on users of TeraGrid will need to know what they can do with it. Workshops for gateway developers should be targeted at beginning, intermediate, and advanced levels, identifying interoperation between gateways at the higher levels. Here, too, the idea of providing common, basic services to avoid the duplication of effort was suggested.

Users	Gateways	Resources	TOTAL	SOLUTIONS
2	1	3	6	Tiered education • Grad students, postdocs – technical education • More senior researchers need to know what *can* be done with TeraGrid
1	1	2	4	Workshops • Beginner • Intermediate • Advanced (inter-operate between gateways)
1	2	1	4	Generalize and provide basic services – to avoid duplication of effort– it is possible?
3			3	Make sure your users get the training
	1	1	2	Documentation – keep it up
	2		2	Have standard software stacks for types of nodes • Compute node software stack • Date node software stack • Pop node software stack
		2	2	Program of work driven by science users – desktop
7	7	9	23	**Subtotals and Totals**

4. Identifying appropriate performance (in response to issues such as high throughput vs. high performance or the need for handling large amounts of data)

This topic reflected a particularly acute concern among participants that HPC is focused too much on building single, giant machines—the tension of capability versus capacity. They would like to see greater balance in the attention towards throughput and towards performance, particularly to meet the end-user expectation that jobs would run in seconds or minutes—not typically in days. Some believe that high throughput needs may be outpacing demand for traditional HPC and should be considered as a significant change as TeraGrid goes forward. Many opportunities exist for TeraGrid to serve its users in new ways. For example, TeraGrid could serve as a library and resource of community data sets, database expertise, and storage. The use of TeraGrid allocations could also be more dynamic, such as partitioning resources to let some portion focus on throughput while other parts can be geared toward HPC or allowing users to schedule a large number of nodes to accommodate a large number of data sets. Likewise, TeraGrid could offer support for on-demand capability and using cycles opportunistically. Finally, disk storage is currently not commensurate with CPU performance, but might be made available with that in mind.

This topic prompted the solution that, by far, garnered the most votes: Participants would like to see the next generation focus perhaps only half of the funding on single, high-capacity machines while the remainder would serve towards developing "content" for the distributed environment that science gateways provide. Content includes everything from middleware (including robust, backward-compatible releases), gateway interfaces (such as portals and management services, including new gateways), and science applications (running on TeraGrid systems and accessible through science gateways) with accompanying tutorials and outreach material. Also, some felt that offering clustered, lower-cost machines might be helpful toward this end, and many were also eager to have a reliably performing global file system with a fast local input/output. The group felt that this division of funding between high-end and high-throughput would provide the most impact for resources spent and it would serve the end users well. We asked participants if they wished to use their votes for a "solution" that was a foregone conclusion, buy they were adamant that they wished to support this proposal made and championed by one of the participants.

Users	Gateways	Resources	TOTAL	SOLUTIONS
4	1	10	15	Do not invest $200M into a single machine • $100M in a capacity machine • $100M in developing the distributed software and integration of the distributed environment of the Science Gateways (portals, interfaces, information, data and job management and their integration)
4	2	1	7	Reliable, performance global file system with local high-throughput I/O (reliably performing global file system with a fast input/output locally)
				Common software for data collections, data archives, retrieval, and "curation"
				Applications support for wide set
8	3	11	22	**Subtotals and Totals**

5. Supporting ease of use and reliability of software

This issue centers on simplifying access to and between TeraGrid resources, possibly by making some of the systems generic across sites. Participants felt gateways need:

- Technology that allows codes to be ported easily to every TeraGrid computer, including support for that portability (installation, compilation, etc.) (See the NSF Middleware Initiative (NMI) Build and Test facility as an example of a support center.)
- Generic services for gateways
- Support of virtual machines and services
- A simple process to acquire resources
- A quick results/queuing policy
- Mechanisms to validate results across multiple sites
- Accuracy and precision of results

Another concern was making it easier for gateway developers to create applications. Participants suggested that facilities support automatic testing of application software. They also saw an opportunity for increasing application development by making systems that help graduate students and non-web developers to create applications rapidly on their own.

To these ends, participants supported the solution of providing guidelines and a framework for building web interfaces and services. They were also in favor of funding integrative testing of software stacks.

Users	Gateways	Resources	TOTAL	SOLUTIONS
2	3	1	6	Provide guidelines and framework for building web interfaces and services
	3	2	5	Fund integrative testing of software stack
2	1		3	Simplify the resource allocation process and fast turn around time of proposal response
	2	1	3	Utilize SQA best practices for software development
		3	3	Supposed gateways that run applications "as is" no web enabling
	2		2	Application developers should provide test data to compare versus successive builds
				Adopt commercial solutions • Amazon • SunGrid
4	11	7	22	**Subtotals and Totals**

6. Offering a system that is consistent across machines, thereby creating a real grid

The topic of creating a true "grid" system garnered the least overall votes for solutions, perhaps only because the need for standardization and consistency across such a grid was expressed in so many other topics. For example, participants envisioned certain standard capabilities that could produce reliable outcomes across the grid, such as executing functions, authenticating once, moving data, monitoring progress, and locating data (via Uniform Resource Identifiers). Another desirable feature would be a connection to or interoperability with university computing systems. Like the electricity grid, they also suggested a system that would automatically provision nodes or storage, rather than allocating in advance. (Automatic provisioning, of course, relies on having a uniform stack that would allow software to work across the TeraGrid.) Instead, users could be sent periodic bills for their usage.

In fact, of all the solutions receiving votes, this topic included single solutions with the third and fourth largest number of votes—solutions that were considered to be important to end users as well as for gateway developers or for producing impact. One person noted, though, that the liability of this approach is that if gateways are too tightly linked to TeraGrid's approach, they may be less able to make use of other grid environments.

Users	Gateways	Resources	TOTAL	SOLUTIONS
4	1	5	10	Standardize certificate based authentication/authorization – see GAMA (Grid Account Management Architecture, http://grid-devel.sdsc.edu/gridsphere/gridsphere?cid=gama) for example
3	5	1	9	End-to-end support for Virtual Organizations
				Virtual machines
7	6	6	19	**Subtotals and Totals**

Closing Discussion

In their closing comments, participants emphasized again certain key avenues of support that would improve both the development of gateways and the experience for the end users of those gateways. Nearly all these comments focused on two main themes: (1) supporting interaction and cross-fertilization among the science gateway development communities and (2) reducing the hurdles to getting on the TeraGrid.

First of all, interaction among developers of TeraGrid Science Gateways as a way of exchanging ideas and successful solutions would save people from reproducing existing code or pursuing unproductive efforts. Developers would also like to know what code or features could be generalized across gateways or would have wide appeal among different types of users. By sharing how certain features have been made available, providing some standard features as web services, and even sharing how they manage their software development as a project, gateway developers would feel better supported. This could be accomplished through an online forum or listserv venue as well as with regular, in person meetings. (Many of the participants were unaware of the biweekly gateway teleconferences hosted by TeraGrid's Nancy Wilkins-Diehr, who subsequently invited all participants to join the group.) In addition to moral support, financial support in the form of funding incentives to those who propose and develop gateways and professional support in the form of academic credit for building gateways would be welcome. A more active "developers' gateway" could make this exchange more successful, allowing gateways to share with TeraGrid the burden of communicating best practices.

The second key issue was the variety of hurdles that make using and building on TeraGrid so difficult. A chief hurdle is simply the lack of reliability of the TeraGrid, particularly with regard to upgrades. Some developers find that middleware is not tested and deployed in such a way as to establish that it is truly useful, reliable, and usable. Moreover, gateways find it difficult to keep up with new versions of things, making it more likely that gateway systems will break when upgrades are made. In addition, the length of time it takes to build middleware and other software elements is too slow to keep up with the evolution of requirements. A different type of hurdle is the process for obtaining a TeraGrid allocation and the time-consuming aspects of interacting with TeraGrid bureaucracy. (Few of the gateways at the workshop actually get funding from TeraGrid, although some attendees were under the impression that most gateways, other than theirs, received financial support from TeraGrid). As mentioned above, the provision of web services by TeraGrid could make the construction of and access to gateways considerably easier. Similarly, the ability to program a user application that runs directly from the TeraGrid

could simplify access. Some wondered whether, in the future, something like TeraGrid should come from industry as a way of bypassing these difficulties.

Participants indicated that, to an extent, these two large issues could be alleviated with better education and communication between TeraGrid and the gateway developers and among the developers themselves. They also felt that progress could be made by extending themselves further towards connecting with other gateways and resources. Finally, by making it possible for users to help themselves, they might scale the system more organically.

Conclusions

In conclusion, it is clear that gateway developers are excited at the potential of TeraGrid to make HPC available to end users and communities who would otherwise be unable to conduct their research as effectively or efficiently. In addition, they are excited at the opportunity for distributed communities to work together on common solutions. Meanwhile, they are eager to move TeraGrid toward a collaborative mindset that enables the developers to focus on the unique needs of their gateway communities. At present, they find that too much energy is focused on re-creating custom solutions when standardized systems or a TeraGrid-hosted gateway layer would suffice. The prominent cross-topic patterns observed in the exercises above advocate for:

- Basic services that gateways can use instead of creating or hosting their own.
- Templates and standardized systems to save developers the time of recreating things that others have already built.
- Standardization that would make TeraGrid a *real* grid that could support the effective use of allocations and meta-scheduling.
- Operating more effectively as a community in order to better support the education and development needs of gateway developers.

Workshop Evaluation

At the end of the workshop, we left time for participants to complete a survey that asked them to evaluate the information received prior to the meeting, the clarity of the workshop goals, the quality of the presentations, instructions, and workshop activities, and their overall impressions of the event. Based on this feedback, the workshop was a success. The evaluations of the workshop activities averaged greater than 4 on a scale of 1 to 5. Although many participants expressed a preference for one exercise over another, both exercises were singled out as helpful and effective in participants' feedback comments. For future workshops, participants suggested:

- Extending the presentations about each of the gateways
- Promoting a discussion forum
- Bringing real cases from users and discussing practice aspects
- Focusing less on technical problems, solutions, and objectives, instead considering aspects such as funding, organization, etc.
- More workshops leading to deployed solutions
- Continuing to involve gateways as part of an active planning and tracking process

They also look forward to seeing their suggestions implemented so that the gateways are used more extensively, their development solutions and technologies are shared with each other, and

users are more satisfied. They also hope that future NSF solicitations incorporate the suggestions that they generated in this workshop.

We observed that the workshop was very successful in providing an opportunity for individuals developing gateways to come together and interact. We also noted, however, that it was challenging to get attendees to maintain a consistent post-2010 focus on their needs and those of their users. In retrospect, this did not seem surprising. When individuals with common concerns first have the chance to meet and discuss their mutual challenges and goals, it is natural that they need time to focus on the issues of today before moving onto the future. This argues for providing ongoing face-to-face and online opportunities for interaction between gateway developers around the topics they themselves identified.

Acknowledgements

We acknowledge the following TeraGrid personnel who provided assistance that made the workshop possible: Kay Hunt, Nancy Wilkins-Diehr, and Dane Skow. Also, we thank the science gateway participants for their enthusiastic input and interest in the workshop. The detailed findings in this report were prepared with the assistance of Rebecca O'Brien. This workshop and report was supported by the National Science Foundation under Grant No. OCI-0724300 to the University of Michigan. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Catlett, C., Beckman, P., Skow, D., and Foster, I. (2006) Creating and operating national-scale cyberinfrastructure services. *CTWatch Quarterly, 2*(2): 2-10.

Wilkins-Diehr, N. (2006). Special issue: Science Gateways—common community interfaces to grid resources. *Concurrency and Computation: Practice and Experience*, *19*(6): 743-749.

Appendix A: Workshop Invitation

Dear (name of Gateway PI):

We would like to invite you or a member of your project to attend a workshop at the TeraGrid ’07 Conference in Madison, Wisconsin. This invitational workshop will be held on Thursday, June 7, from 12:30 p.m. until early evening. The purpose is to solicit information about the needs of developers of projects such as (name of gateway) that will be used to help guide future plans for the TeraGrid. Individuals working on projects like yours that integrate TeraGrid resources and services are important users of TeraGrid.

This workshop is the first in a series of activities associated with a planning process being supported by the National Science Foundation (NSF). The planning process will be led by a Steering Committee comprised of individuals from key stakeholder communities and facilitated by the University of Michigan’s School of Information (UM-SI). The workshop will be conducted by Ann Zimmerman and Katherine Lawrence of the UM-SI. Please see the end of this message for further information about the planning process.

The ideal participant in this workshop would be an individual working on (name of gateway) who understands the technical, social, and organizational issues related to the integration of TeraGrid resources into your project as well as the target users for (name of gateway). We are able to support costs related to attendance at the workshop. Alternately, if an appropriate person from (name of gateway) is already planning to attend to the TeraGrid ’07 Conference, and needs to extend their stay in order to participate in the workshop, we are able to subsidize some of the travel expenses associated with that person’s attendance.

We also hope that the timing of this workshop will coincide with plans you may have to attend—or will make it easier for you to attend—a Friday, June 8 morning staff meeting for TeraGrid Science Gateways developers, hosted by Nancy Wilkins-Diehr of TeraGrid.

Please reply to Ann Zimmerman (asz@umich.edu; 734-764-1865) by Wednesday, May 9 as to whether you or a colleague will be able to attend the June 7 workshop. Ann can also answer questions you might have about the workshop.

To make this short meeting as productive and engaging as possible, we will ask attendees to answer a brief survey in advance of the meeting. This survey will help give participants some advance consideration to the issues we will be discussing and allow them to efficiently share information with colleagues at the workshop.

Thank you for considering this invitation. We hope to see you in Madison!

Sincerely,

Ann Zimmerman and Katherine Lawrence, Workshop Co-Organizers
School of Information, University of Michigan

ABOUT THE TERAGRID PLANNING PROCESS

The NSF is providing support for a community-driven, participatory planning process whose goal is to provide information that will help guide the future evolution of TeraGrid. Current awards for the operation, user support, and enhancement of the TeraGrid facility will expire in 2010. By this date, a petascale computing resource will be on the horizon, the user community will have grown and diversified, and new policies and services are likely to be needed to meet the needs of users and the expanding pool of high-performance computing resources. In anticipation of these changes, the planning process is focusing on the needs of current and emerging user communities as a critical aspect in the development of a path forward for TeraGrid in 2010 and beyond.

Planning activities will be conducted over the space of approximately one year and will include a combination of face-to-face and online engagement designed to:

- gather information on user needs and priorities;
- compare user requirements; and
- develop options for the delivery of high-performance resources and services

The results of the planning process will be a report to the stakeholders that outlines options for the design of the next generation TeraGrid and for the delivery of high-end resources and services based on user requirements. The report will be written by the Steering Committee using the information and input gathered from stakeholders throughout the planning activities. The final version of the report is targeted for February 28, 2008.

The Steering Committee is currently being formed and a web site is in development. For further information, please contact Ann Zimmerman, School of Information, University of Michigan at asz@umich.edu or 734-764-1865.

Appendix B: List of Workshop Participants and Gateways

Adam	Brazier	ALFA Pulsar Survey
Julian	Bunn	Caltech Science Gateways (Grid Analysis Environment)
Xuiyi	Fan	CSE-Online
Ashraf	Memon	GEOsciences Network (GEON)
Yan	Liu	GISolve (Geographic Information Science Gateway)
Shaowen	Wang	GISolve (Geographic Information Science Gateway)
Sudhakar	Pamidighantam	GridChem (Computational Chemistry Grid)
Tom	Baltzer	Linked Environments for Atmospheric Discovery (LEAD)
Gerhard	Klimeck	nanoHUB
Wilfred	Li	National Biomedical Computation Resource (NBCR)
Steve	Meacham	National Science Foundation
Abani	Patra	National Science Foundation
Roy	Williams	NESSSI (National Virtual Observatory or NVO)
Steve	Miller	Neutron Science Instrument Gateway
Ruth	Pordes	Open Science Grid Consortium and Project (OSG)
Phil	Maechling	Southern California Earthquake Center Earthworks Project
Jeff	Grethe	Telescience Project and BIRN
Joseph	Insley	TeraGrid Visualization Gateway
Cesar	da Silva	Virtual Laboratory for Earth and Planetary Materials (VLAB)
Pedro	da Silveira	Virtual Laboratory for Earth and Planetary Materials (VLAB)

Gateways unable to participate included:

Bioportal (Biology and Biomedicine Science Gateway)
The Earth System Grid (ESG)
Network for Earthquake Engineering Simulation (NEES)
Open Life Sciences Gateway (OLSG)
Special PRiority and Urgent Computing Environment (SPRUCE)

Appendix C: Workshop Pre-Survey

Thank you for taking a few moments to prepare for the TeraGrid Planning Workshop to be held in Madison, WI on Thursday, June 7, 2007, from 12:30 - 6:00 pm.

The 14 questions in this survey are intended to help spur your thinking and to assist us in collecting data to seed the workshop. We realize that projects differ and not all questions will be relevant or easy to answer. Do the best you can. Skip questions that aren't relevant and leave us a comment in question 14 with any notes or other information you would like to convey.

Feel free to get input from others who are part of your gateway project. You can return to this survey to revise your answers by following the link in the e-mail we sent.

With the exception of name and contact information, data will be aggregated and individual responses will be anonymous.

Please complete the survey by Thursday, May 31.

Page 1 of 3: Information about You

1. Please provide your name and your gateway affiliation as you would like it to appear on your name badge.

2. We will be preparing a participant list to include with the workshop materials. Please provide your name and contact information in the space below. List as much information as you would like to make available. For example, affiliation, address, phone, fax, email, web page.

3. We will be serving lunch and a snack at the Workshop. Vegetarian options will be available. Do you have any other special dietary needs we should consider?

No
Yes (please specify)

Page 2 of 3: Basic information about your gateway

4. What is the name of your gateway?

5. What is the primary field/discipline of your gateway end-user audience? (e.g., chemistry, meteorology)

6. What is the funding source(s) for your project?

7. What types of people do you expect will use your Gateway? (check all that apply)

Researchers
Educators
Students (college or university-level)
Students (K-12)
Policy makers
General public
Industry practitioners
Other (please specify)

8. To what extent is your gateway a research project (as opposed to development)?

0 (0% research/100% development)
10
20
30
40
50 (50% research/50% development)
60
70
80
90
100 (100% research/0% development)

9. To what extent is your gateway in operation (as opposed to being in development)?

0 (0% operational/100% in development)
10
20
30
40
50 (50% operational/50% in development)
60
70
80
90
100 (100% operational/0% in development)

10. As part of the initial planning process, we will be conducting two more user workshops to include individuals with various levels of experience and knowledge regarding high-performance computing (HPC) and TeraGrid. We welcome recommendations of people, especially users of your gateway, who would be thoughtful about the topic and constructive in their participation. These people do not need lots of HPC experience or knowledge and could be “emerging users.” However, it would be most useful if they have a basic understanding of high-end resources and services and how they might be used in support of their own research. Graduate students, post docs, research assistants, and under-represented user groups (as well as PI-level nominees) are all welcome.

If you have recommendations, please provide their name(s) and institutional affiliation(s), along with a brief note as to why they would be appropriate.

Page 3 of 3: Open-ended questions

The following are open-ended questions to jump start your thought process... Polished essays are not required!

11. What current and future user needs are driving the services and capabilities that your gateway project provides or is planning to provide? (For example, are your users dealing with large amounts of data that they need to move around or visualize? Do they need to collaborate across disciplines or distance?)

12. Why will TeraGrid be a valuable resource to your project and end users?

13. How could TeraGrid become outstanding?

14. What other comments would you like to add? Are there explanations or clarifications that you would like to make? Use this space to ask questions, make comments, etc., on whatever topics you feel are relevant to the workshop and planning process.

Appendix D: Workshop Agenda

Time: Thursday, June 7, 2007, 12:30 to 6:00 pm

Location: Great Hall, 4th Floor Central Core, Wisconsin Union

Hosts: Ann Zimmerman (asz@umich.edu) and Katherine Lawrence (kathla@umich.edu)

Workshop Guiding Question

What would TeraGrid be if it met the needs of your science gateway perfectly? Answering this question requires a better understanding of the future goals of your project and the specific role that high-end computing resources and services (such as those provided by TeraGrid) play in meeting the scientific needs of your user communities. Thus, the workshop activities are organized with these topics in mind.

Workshop Goals

1. Assess the requirements of gateway developers. Specifically, how can TeraGrid improve the capabilities available to your gateway user communities? To do this, we will identify the common and different needs and priorities for your projects.
2. Solicit your ideas for engaging members of your project's user community in the planning process.
3. Provide you with an opportunity to interact with each other around topics of shared interest and give us ideas about how to support your continued participation and interaction throughout the planning process...and beyond.

Schedule

12:15–12:45 pm: Check-in and get lunch (Lunch will be served starting at 12:15, so we encourage you to arrive early)

12:45–1:30 pm: Introduction (and lunch)

- Welcome from workshop organizers and explanation of the goals and format of the workshop
- Presentation of overall participant survey data
- Participant introductions (accompanied by your project slides)—90 seconds each

1:30–1:35 pm: Brief bio-break

1:35–1:50 pm: Warm-up exercise to familiarize participants with the scope of each other's projects

1:50–2:55 pm: Idea generation activity from perspective of gateway end users

The driving question of this activity is: "What role does TeraGrid play in meeting the scientific needs of your gateway user communities?" Participants have the opportunity to explore this question in small groups.

2:55–3:35 pm: Large group discussion and prioritization of topics

- Small groups report the most important and interesting issues raised during the activity.
- Entire group discusses common themes and identifies most significant issues for use in the second part of the workshop.

3:35–3:55 pm: Break

3:55–4:45 pm: Idea & solution generation activity from perspective of gateway developers
In response to the key topics selected from prior activity, participants answer the question, "What do our science gateways need from TeraGrid?"

4:45–5:10 pm: Discussion of evaluation criteria and prioritization of solutions

5:10–6:00 pm: Reactions, feedback, and workshop evaluation survey
Participants will identify key issues to consider going forward. We will discuss how, in future workshops, events, and initiatives, we can most effectively engage the end users and developers of TeraGrid science gateways.

Workshop Outcomes

We will produce a report that will be posted on the planning web site shortly after the workshop. We will ask you to comment on the report before it is posted. The information collected during the workshop will carry forward in the planning process. It will be used in creating an overall picture of user needs and requirements. It will also be used to gain stakeholder input on the development of options for the future delivery of high-performance resources and services.

Appendix E: Ideas Generated in "World Café" Exercise

Group 1

Our group tried to maintain a looking forward perspective in our discussion—both the needs of the gateway user community and the roles of the TeraGrid. What would we like the roles of our community to be? What would we like TeraGrid to be?

User perspective: Software packages we don't have to purchase ourselves. Results are private unless we want them to be public. Storage and data management.

TeraGrid role: Software system integration. Ability to operate and administer these HPC systems. Check user credentials. Carry the culture of best practices. TeraGrid has visibility on all 18 of these projects, so they should provide feedback to the gateways about the experiences of other projects. TeraGrid should evaluate role of new technologies (ex: virtual machines). Also development of new parallel computing techniques. Not MPI forever. Perhaps developing relational databases. Evaluating and improving performance of our applications

Group 2

One person in our group who needs bandwidth. Most people at the table don't. Time and number of users – most people want to run in seconds or minutes. Very few running in days. Bringing people up the slope. Buying specific machines. User support. Gateway users shouldn't have to call user support a lot—it doesn't scale. Won't scale for ten thousands of users. Metascheduler on the TeraGrid. The user wants to minimize the turnaround time.

Group 3

What happens if TeraGrid goes away? It depends on the TeraGrid user community. For some it's the only way they can get the work done because they don't have it built into their local infrastructure. Connecting into TeraGrid. When you broaden the scope of the users coming in, how do you get users in? The allocation process will need different dynamics with large numbers of users coming in. More structured now to run production codes that run behind the gateway. Privacy. Transferring data back and forth. As the gateways grow, you're going to have a mix of many different types of users.

Group 4

We started by trying to figure out what is the TeraGrid. Too hard. Not just existing gateways, but how you bring new ones in. They don't want to have to develop lots of infrastructure. A way in—an existing package. The next layer was virtual machines. A virtual organization on top of virtual machines, so you could really have a community. Gateways facilitate running scientific computations in a faster way. Getting results their communities want to get without having to build infrastructure. Getting people more feedback on how their jobs are running. Gateways not always having to keep up with new versions of things. So, people don't have to worry about things breaking when upgrades are made. Community allocations are good, but people may also want to be able to use their own allocations.

Group 5

We started with developer needs and then went onto user needs. Data parallel resources. Communication interfaces—way to communicate with other grids. Making the submission of larger jobs easier on the TeraGrid. Rule-based authorization. As a user don't want to have to worry about authenticating to all the different grids—a common account. On-demand computing. 5-minute need vs. larger needs. A limited number of homogeneous resources. Harness for regression testing, so it's consistent across all the TeraGrid resources. Know code works on Tungsten, but I also want to run it on Big Red, so I want to make sure it runs efficiently.