TeraGrid User Workshop
Final Report

Ann Zimmerman
Thomas A. Finholt

July 2006

Collaboratory for Research on Electronic Work
School of Information
University of Michigan
Ann Arbor, MI 48109-2112
http://www.crew.umich.edu

Table of Contents

Introduction

TeraGrid is a national, comprehensive, and distributed infrastructure that integrates multiple resources at nine resource provider facilities.[1] Following a 5-year construction phase, TeraGrid became operational in late 2004. TeraGrid's resources include more than 102 teraflops of computing capability and greater 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.

TeraGrid is coordinated through the Grid Infrastructure Group (GIG) at the University of Chicago. The GIG works in partnership with the resource provider sites to carry out three interdependent sets of activities. First, TeraGrid aims to increase the size of the user community by an order of magnitude by working with entire communities to integrate TeraGrid capabilities with their infrastructure. This effort is referred to as “TeraGrid WIDE” and includes the development of a set of discipline-specific “Science Gateways.” A second major goal, known as “TeraGrid DEEP,” is to exploit the unique features of the current TeraGrid system to make possible cutting edge science. Third, "TeraGrid OPEN" involves provision of a persistent, reliable national cyberinfrastructure. Wherever possible, TeraGrid services are based on open standards and with the intention to achieve interoperation with peer grids operated by other U.S. agencies or by other countries.

TeraGrid is a key part of the nation's long-term cyberinfrastructure (CI) plan. Like other CI efforts, TeraGrid is an administratively complex partnership among multiple, geographically-distributed institutions. There is some risk that CI investments will not be optimized unless these projects have access to techniques that will help them identify, monitor and address potential problems. In late spring 2006, the NSF awarded a one-year grant to the University of Michigan’s School of Information (UM-SI) to conduct an external evaluation of TeraGrid. The primary goals of the evaluation are a) to provide specific information to TeraGrid managers that will increase the likelihood of TeraGrid success; and b) to give NSF leaders and policy makers general data that will assist them in making strategic decisions about future directions for cyberinfrastructure. In order to accomplish these objectives, the UM-SI evaluation study is assessing four aspects of the TeraGrid project: 1) progress in meeting user requirements; 2) impact on research outcomes; 3) quality and content of TeraGrid education, outreach, and training efforts; and 4) satisfaction among TeraGrid partners.

This report describes the results of the first major evaluation activity. On June 12, 2006 the UM-SI evaluation team conducted a workshop to begin to examine the relationship between TeraGrid’s development priorities and the needs of its users. The invitation-only workshop was funded by TeraGrid and was held at the University Place Conference Center in Indianapolis, Indiana. The *TeraGrid User Workshop Final Report* summarizes the data collected and

[1] The nine resource provider sites are Argonne National Laboratory, Indiana University, National Center for Atmospheric Research, National Center for Supercomputing Applications, Oak Ridge National Laboratory, Pittsburgh Supercomputing Center, Purdue University, San Diego Supercomputing Center, and the Texas Advanced Computing Center.

information gained during the workshop. It is a first step toward evaluating TeraGrid's progress in meeting user needs.

TeraGrid User Workshop

Workshop Design

The workshop was designed to assess user requirements from the perspective of how TeraGrid could enable the next big breakthrough in workshop attendees' lines of research. The workshop focused on the requirements of advanced users (i.e. TeraGrid DEEP). The goal of the event was to produce a prioritized roadmap based on the future needs of those present at the event.

Since the evaluation project officially started in late spring 2006, there was limited time available to arrange the workshop, especially as the organizers wished to hold the event in conjunction with the first TeraGrid Conference scheduled for June 12-15, 2006. The evaluation team selected individuals to invite with the assistance of TeraGrid personnel. An effort was made to achieve diversity in disciplinary areas and in the types of resources and services used. Invitees received an e-mail invitation from Charlie Catlett, TeraGrid Director, requesting their participation in the workshop. (See Appendix A for a copy of the invitation message.) In most cases, those who attended the workshop were advanced users with large allocations on the TeraGrid. Twelve individuals, representing the fields of biology, chemistry, geoscience, physics, and social and behavioral sciences, participated in the event.

Workshop Structure

The workshop was organized around a combination of large and small group activities. A copy of the workshop agenda is available in Appendix B. Due to the limited amount of time available to conduct the workshop, the organizers focused on generating as many ideas as possible about how TeraGrid could enable future discoveries in participants' lines of research. Most of this work was accomplished in two breakout groups using the nominal group technique.

The workshop began with a welcome from a member of the external evaluation team who described the afternoon’s agenda and explained the evaluation study and the relevance of the workshop to that investigation. This was followed by short introductions from all attendees. Individuals were asked to state their name and affiliation, to describe their general area of research, and to say what came to mind when they thought of TeraGrid. The responses to the latter show that perspectives of TeraGrid are very diverse. They range from views of TeraGrid as single machines at an individual supercomputing center to a cluster of supercomputers, which is sometimes accessible through a single desktop interface. The responses also show that some users are uncertain how to define TeraGrid while others see it as enabling them to do things that would not be possible otherwise. One person felt that TeraGrid was a problem seeking a solution since algorithms in his field are not advanced enough to capitalize on the TeraGrid. Following introductions by workshop attendees, Charlie Catlett presented an overview of TeraGrid with a special focus on the project's current status. The purpose of this presentation was to provide participants with a common understanding of the resources and capabilities available through the TeraGrid. The remainder of the workshop consisted of small group work followed by a large group discussion of the prioritized ideas generated in the workshop.

Nominal Group Exercise

The nominal group technique (NGT) is an effective method to generate many ideas from a group and to narrow those ideas down to a set of prioritized recommendations. The advantages of NGT are that it gives each individual an equal chance to participate, produces a greater number and more creative ideas than traditional small group exercises, and allows for private voting. Disadvantages of the technique are that it requires considerable advanced preparation, tends to be limited to a single topic, and requires participants to follow a structured approach. These drawbacks did not present difficulties in the case of the workshop described in this report. The evaluation team, with logistical support provided by TeraGrid personnel, were responsible for workshop planning; the topic was restricted to the subject of user needs; and workshop attendees cooperated fully with the planned activities.

The steps in NGT are: 1) generating ideas; 2) recording ideas; 3) discussing ideas; and 4) voting on ideas. The specific means by which these procedures were used in the TeraGrid User Workshop are described below.

Generating Ideas

Participants were assigned to one of two groups before the workshop began to save time and to provide a mix of disciplinary perspectives at each table. There were 6 users in each breakout group. To begin, individuals in each group were given a worksheet with the following two questions listed and were asked to write down their responses to them:

1. What do you think is the next big breakthrough to be made in your line of research that has a computational component?

2. What are the barriers that must be overcome in order to achieve this next big breakthrough? Please describe these in terms of the following types of challenges:
   - Social
   - Organizational
   - Technical

The purpose of these questions was to prepare participants to answer a third question, which was the main focus of the workshop and is described below. Time was not available to discuss the responses to the first two questions, but participants were asked to hand in this worksheet at the end of the workshop, and seven of the twelve individuals did so. The content of the worksheets are summarized at the end of this report.

Next, participants were given another worksheet and asked to write down their ideas in response to a third topic.

3. In light of the breakthrough that you described in question 1, and the barriers to achieving it that you noted in question 2, list at least 5 things (and up to as many as you want) that TeraGrid could do to enable the breakthrough.

Individuals were given approximately twenty minutes to record their ideas.

Recording and Discussing Ideas

The second main step in the NGT is to record the ideas generated in the previous step. Dane Skow, TeraGrid Deputy Director, and Noshir Contractor, Professor in the Department of Speech Communication at the University of Illinois, Urbana-Champaign, served as group recorders. Each recorder went around the table and asked each person in their group, one at a time, to give one idea from their response to topic 3. As the ideas were stated, the recorder numbered each consecutively and wrote them on a flipchart that was visible to all participants in the breakout group. The recorders continued to go around the table until all ideas were listed. During this process, the ideas presented were not discussed or debated as the goal was to get down as many ideas as quickly as possible. The complete list of ideas generated by each of the two breakout teams is available in Appendix C.

The third step in the activity provided the participants with the opportunity to review and clarify the meaning of each item that was recorded. The recorder facilitated this process by pointing to each idea on the list and asking if anyone had questions or comments about the item.

Voting: Round 1

The goal of the next step was to get the judgment of each person concerning the most important ideas on the list. Each breakout group member received a set of five index cards and was asked to select the five most important items from the list of ideas and to write one idea down on each card along with its corresponding idea number; the latter was recorded in the upper left-hand corner of the card. Participants were then instructed to review the set of cards and decide which one was the most important and to write the number "5" in the lower right-hand corner of that card. Following this, they were asked to look at the remaining four cards, select the one that was least important, and write the number "1" in the lower right-hand corner of that card. A similar process continued with the remaining three cards. This exercise resulted in each participant having a set of cards on which he gave a higher numerical score to the ideas he considered most important.

When the voting was completed, the recorders collected the cards from the members of their group and shuffled them together. Recorders were then instructed to make a tally sheet with numbers down the left-hand side corresponding to the ideas generated by his group. The number of points from each card was written next to the idea number on the tally sheet. At the end of this step, the recorder added up the number of points to obtain a total for each idea. Some ideas received no votes. The four ideas with the greatest number of points from each breakout group were selected as the highest priorities and were carried forward to the next exercise. The number of points that each idea received, including the number of people who voted for it, can be found in Appendix C.

Each breakout group shared their list with the other breakout group. This provided the opportunity for clarifications to be made to the ideas on the lists. The eight ideas that resulted from the small group work are listed below.

- All available TG sites with 50% capacity for simulations
- Cross-site runs 2-16 TB in core to operate – for qualitatively new science algorithms

- Education and outreach to create code that is parallelized – ramp up new users, graduate students
- Single point-of-contact to provide tech support in real-time for multi-site run
- Very large allocations (> 10^7 SUs)
- Large memory size (>10 TB)
- More CPUs (10X) – more of same kind
- High inter-machine communications bandwidth to processors (multiple machines)

Voting: Round 2

The eight ideas listed above were recorded on the outside of large envelopes; each envelope was labeled with one idea. The twelve workshop participants were each given $1,000 in "TeraGrid bucks" in denominations of $100 and instructed to "spend" the money however they wished among the eight ideas. For example, individuals could place $1,000 in the envelope with the idea they felt most strongly about, or they could give $400 to one idea and $600 to another, etc. The goal of this voting exercise was to identify ideas with high scores and those with low scores from many people.

User Priorities

Below are the results from the large group vote to prioritize the eight ideas that resulted from the two breakout groups. The ideas are listed in order from those that received the most TeraGrid bucks to those that garnered the least:

Education and outreach to create code that is parallelized	$2,800
Single point-of-contact to provide tech support in real-time for multi-site run	$1,800
Very large allocations (> 10^7 SUs)	$1,500
Large memory size	$1,400
High inter-machine communications bandwidth to processors	$1,200
Cross-site runs 2-16 TB in core to operate	$1,200
All available TG sites with 50% capacity for simulations	$800

With this second voting method, it is not possible to determine how many people voted for a particular idea. However, since each person was given $1,000 to spend, the results show that except for one item, all ideas received votes from at least two people. Further, the highest priority idea gained votes from a minimum of three people. One idea -- more CPUs -- received no votes.

Discussion of User Priorities

The last part of the workshop was a large group discussion focused on the results that emerged from the final voting exercise. This discussion was led by a member of the evaluation team.

Comments from participants indicated that education ranked high for several reasons. First, the rate of scientific progress is constrained by figuring out how to parallelize code. In addition, the techniques that are available now do not scale beyond what currently exists; fundamentally different algorithms and architectures will be needed to exploit future code. Second, it takes a long time to get up-to-speed in the use of high-performance computing (hpc) resources, which limits the numbers and kinds of people that principal investigators find it worthwhile to train. For instance, a workshop participant with a large TeraGrid allocation (i.e. a DEEP user) commented that some, but not all, doctoral students can learn and contribute meaningfully to a project in the time it takes to complete a degree, but the same is not true for a typical Masters level student. Thus, even though most of the workshop participants fall into the category of experienced, advanced users, they face education and training challenges. TeraGrid WIDE users, on the other hand, may not wish to become experts. They may even want TeraGrid, as one participant stated it, "to provide me with a solution, not teach me a solution." This raises the potential need for a service component in which TeraGrid offers services in response to needs that users have. It also points to a third education and training need, which is to help researchers see that there are better ways to do things, even if they do not carry out the work themselves. One person noted that the GIS community has done this effectively by holding workshops that target specific applications.

In addition to those noted above, various types of solutions were suggested to address the education and training needs that workshop participants identified. One approach focused on ways to energize the informal transfer of knowledge between users through the encapsulation of common knowledge. Start-up time for hpc use could be reduced through summer schools and workshops for students and for the larger professional community. These approaches are employed fruitfully in other areas. For example, the National Center for Atmospheric Research hosts summer institutes organized around an annual theme as a way to share and capture dispersed knowledge. Workshops, formal training, and re-playable online courses were considered applicable to a wide range of education and training needs and were thought to be models that TeraGrid could adopt.

The second most highly ranked idea to emerge from the voting was the need for a single point-of-contact to provide technical support in real-time for multi-site runs. However, the discussion indicated that participants saw this as a need for more than cross-site work. Attendees wished for an intermediary to serve as a consultant between a user and TeraGrid proper and for the option to apply for allocations of people just as one applies for allocations of computer cycles.[2] One participant noted that most solutions arise from people sitting down together and working through a problem. Thus, it is possible that the same improvements might result from ten people working together as are enabled by one bigger machine. Attendees thought it was important for hpc experts to understand their science goals, and they felt this was most likely to be achieved through a single point-of-contact working with them over time. This support is needed for all levels of users. The discussion leader raised the question of the scalability of the single point-of-contact model for all users. He proposed that the user community consider ways to help support each other. The open-source community might serve as a model in this respect.

Very large allocations were the third priority identified by the voting exercise, although it was evident in the discussion that this need was not equally valued by all participants. It was also

[2] TeraGrid has some aspects of such a support mechanism in place. The goal of the Advanced Support for TeraGrid Applications (ASTA) program is to provide sustained effort to maximize the effectiveness of application software and TeraGrid resources to support a project's scientific goals. ASTA provides one or more TeraGrid staff members at a minimum of 25% FTE time with a user's project. To date, TeraGrid has completed 7 ASTA projects, has 8 in progress, and five that have been proposed.

clear that attendees had different views of what was meant by "very large allocations." One user's work requires more degrees of freedom and thus more nodes in the process. His group wants to be able to conduct runs with more than 2,000 processors.[3] Another participant noted the difference between capacity and capability: "I need 200 processors. If there were more processors available, my run of 200 would get done faster." The tension that exists between different types of users was summed up by an attendee: "There are two types of people. One, there are people who want to do something that cannot be done. Two, there are people who want to get something done." Currently, TeraGrid is trying to serve both audiences, and over time this could grow to be difficult. TeraGrid needs to makes plans to deal with this issue.

Finally, a major factor in optimizing the work of users is the degree to which codes are able to make use of processors. Having the largest number of processors may not be a solution by itself if the code cannot keep up. Large-scale data transfer is still an open question in the minds of some of the users present at the workshop. TeraGrid needs to put more resources into solving this problem.

TeraGrid Impact on Science

The final part of the large group discussion was used to gain insights from participants regarding ways to measure the impact of the TeraGrid on their science. This is another area of focus in the external evaluation.

Participants suggested the following as markers of TeraGrid impact:

- Number of papers that cite TeraGrid as being used in their research
- Industry's interest in the results of work based on the use of TeraGrid as well as industry's interest in employing TeraGrid resources
- Media interest in research results that used TeraGrid resources. The science page of *The New York Times* has the single highest impact of any publication.
- Interdisciplinary research collaboration
- Proxies of impact. Identify markers of success in earlier projects and analyze current projects for those signs. For example, does every Nobel Prize winning discovery go through a particular stage?

Difficulties were noted with the measurements suggested, including the fact that usage does not indicate impact. Similarly, there is a difference between output and outcomes. It is easier to enumerate the former. Further, in many disciplines, it can take decades for impacts to be known.

Scientific Breakthroughs and Barriers

This section summarizes the seven responses received to questions one and two on the worksheet that participants completed at the beginning of the workshop. These answers provide additional insight into the priorities that resulted from the voting exercises.

The first question asked participants to describe the next big breakthrough in their line of research that depends on computational resources. Simulation and visualization of large and/or

[3] A few researchers are already using 2048 processors.

complex data sets were common themes that appeared in answers to this question. These breakthroughs are computationally intensive and will require large amounts of active memory. Other discoveries will arise from the integration of multiple types of data -- sometimes in real-time -- and through the use of tools to investigate these data. Finally, one participant noted the need for a cyberinfrastructure to help integrate tools that have overlapping and complementary features. This capability would enable users to create and share research workflows, enhance the analysis process, and facilitate knowledge sharing.

The primary social barrier to achieving new breakthroughs that participants identified concerned collaboration between multiple disciplines, especially hpc specialists and domain scientists. "Fear of the unknown" is a barrier to greater use of hpc resources by scientists. Education and outreach activities are needed to help scientists understand how grid resources can advance their research. Helping hpc experts to appreciate the needs of end users is important, too. Domain scientists and computer scientists should be encouraged to interact earlier in their training. Differences in cultures between disciplines such as biology, chemistry, physics, and engineering also present obstacles to new discoveries.

A common organizational barrier that was mentioned was a lack of funding and support for a sustained computational approach. A change from time-limited research efforts to longer term operational computational capabilities will require different types of staff. In addition, domain science organizations will need to provide more funds for information technology. Researchers and students at small college or universities will require access to grid computing accounts even though their institutions are not on the grid. Support will be needed to administer and manage access and credentials for these users.

Answers to the question of technical barriers illustrate the varying needs of the workshop participants. One set of responses related to data. Specifically, participants wished for a highly stable workflow and data management infrastructure to facilitate data storage, to manage data (metadata tools), and to move data to computational resources. A second theme that emerged was the need for developments in the areas of code and algorithm development. A final area concerned system capability and performance. Participants mentioned the need for access to 15,000,000 SUs in one year; access to more computational speed and memory, and three orders of magnitude improvement in time-to-result.

Conclusions

Several conclusions emerge from the information collected in the workshop and from the thirteen evaluation surveys that were returned at the conclusion of the event. First, it is clear that there are a range of technical needs even in a group of users who mostly have large allocations on the TeraGrid. As one individual noted on the workshop survey, "The voting system eliminated important issues that are of great concern to a small subgroup of the participants." While it is difficult to tell if this sentiment was felt strongly by most attendees, the workshop clearly showed that users have different demands for hpc capabilities. Second, although the technical requirements of users vary, the social and educational challenges they face are similar. From this standpoint, the priorities that emerged from the workshop appear to be ideas that were important to the group as a whole. Finally, responses to the workshop survey showed that participants welcomed the opportunity to meet and talk with other users. The large group discussion and the chance to interact with peers were mentioned by many as the aspect they enjoyed most about the workshop. Overall, attendees were very positive about all aspects of the workshop. Approximately 80% of the survey respondents gave the workshop as a whole a rating of very good -- the highest rating possible -- and the remainder classified it as good. The main area that survey responses noted for improvement was in the information they received prior to the workshop. From the perspective of the evaluation team, the workshop provided a valuable window into the needs of some members of the TeraGrid user community. The knowledge gained will be useful in the next stages of the evaluation study.

Acknowledgments

With the exception of travel expenses for the UM-SI evaluation team, the TeraGrid Grid Infrastructure Group provided support for the workshop.[4] The evaluation team gratefully acknowledges the following TeraGrid personnel who provided assistance that made the workshop possible: Kay Hunt, Diane Louise Jung, Scott Lathrop, and Carolyn Peters. Finally, we thank the attendees for their active participation and interest in the workshop.

[4] This report is based upon work supported by the National Science Foundation under Grant No. OCI 0603525 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.

Appendix A

Workshop Invitation

Dear TeraGrid User,

I would like to cordially invite you to attend a TeraGrid user workshop to be held on Monday, June 12, 2006 from noon-5 PM. This invitation-only workshop for advanced users of the TeraGrid will be held in Indianapolis in conjunction with the TeraGrid '06 conference (http://www.teragrid.org/events/2006conference/). TeraGrid will support up to $1,000 of your expenses to attend the workshop.

The purpose of the workshop is to investigate the impacts of the TeraGrid on your science and to develop a prioritized roadmap based on your future needs. The workshop is being conducted by the TeraGrid external evaluation team. Tom Finholt and Ann Zimmerman of the University of Michigan are evaluating TeraGrid's impact on research outcomes, progress in meeting user requirements, and quality and content of education, outreach, and training efforts. This workshop and other evaluation activities will provide important feedback that will help TeraGrid to better achieve its goals.

If you are not able to attend the workshop, we invite you to send one of your colleagues, who is an active user of the TeraGrid and who could address the workshop's themes. Please reply to Ann Zimmerman (asz@umich.edu; 734-764-1865) by Monday, May 24 whether or not you or a colleague will be able to attend. Ann can also answer any questions you might have about the workshop. Assistance with travel arrangements is available from Carolyn Peters (cpeters@mcs.anl.gov; 630-252-4909).

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

Charlie Catlett
Director of TeraGrid

Appendix B

Agenda

TeraGrid User Workshop
Monday, June 12, 2006
University Place Conference Center
Indianapolis, IN

Time & Location:

- noon - 5:00 p.m. ; Room 137

Workshop goals:

- Primary goal: To develop a prioritized roadmap for future enhancements to the TeraGrid based on the requirements of advanced users
- Secondary goal: To learn about the impacts of the TeraGrid on scientific outcomes and practice and to solicit ideas about how those impacts might be measured

Schedule

Noon - 1:20 p.m.: Welcome and Introductions

- *noon - 12:30 p.m.*: Welcome from external evaluation team and explanation of the goals and format of the workshop
- *12:30 - 12:50 p.m.*: Introduction of workshop participants
- *12:50 - 1:20 p.m.*: Welcome from Charlie Catlett, TeraGrid Director, and presentation on the TeraGrid

1:20 - 1:30 p.m.: Break

1:30 - 2:45 p.m.: Small Group Exercise

The purpose of the small group exercise will be to identify the ways that TeraGrid can help achieve future scientific breakthroughs in participants' lines of research.

2:45 - 3:15 p.m.: Group Reports and Voting

Small groups will report on the results of their work. Participants will prioritize the ideas that emerged from the groups.

3:15 - 3:30 p.m.: Break

3:30 - 4:50 p.m.: Large Group Discussion

Workshop participants will discuss the prioritized roadmap for TeraGrid that resulted from the workshop activities. There will also be a short session to solicit ideas from attendees on the impact of the TeraGrid on the outcomes and practice of their science and on ways that these impacts might be measured.

4:50 - 5:00 p.m.: Workshop Evaluation Survey

Appendix C

Ideas Generated in Breakout Groups

Group members: Bennett Bertenthal, Bruce Boghosian, Yuqing Deng, Hank Green, Leopold Grinberg, Ricky Robertson, and Duane Rosenberg
Recorder: Noshir Contractor

1) **All available TG sites with 50% capacity for simulations**: **9** (2)[5]
2) **Cross-site runs 2-16 TB in core to operate – for qualitatively new science algorithms**: **12** (3)
3) **Education and outreach to create code that is parallelized – ramp up new users, graduate students**: **18** (5)
4) Take 2 or more of the sites to dedicate to visualization (esp. progressive resolution multi-scale) – software will "soon" exist; hardware needed: **0**
5) Use uniform job specification to run similar jobs at multiple sites (perhaps use Globus?): **7**(3)
6) Real time signal processing for audio and video data "grand challenges": **0**
7) Get through the queue at **ALL** sites so processors to debug code – for an hour (rather than six weeks) -- minimize probability of one site holding it up: **5** (1)
8) Very large ensemble runs (small number of processors, but lots of them)—with some bandwidth: **2** (2)
9) Ability to pre-stage date (at least 50GB) independent of queue status: **4** (1)
10) Real-time monitoring of large data runs: **7** (3)
11) Real-time sharing of data where collaborative annotators use different data formats locally: **1** (1)
12) **Single point-of-contact to provide tech support in real-time for multi-site run**: **14** (4)
13) Web portals to allow collaborators to enter parameters and execute programs simultaneously – "auto generate job submission scripts": **5** (2)
14) IM chat among collaborators and with tech support: **0**
15) Familiarize help desk with cross-site runs – not clear if there is a source of help for tech issues related to cross-site runs: **2** (1)
16) Mid-level tech support to liase between new end users -- with TG tech support: **0**
17) Automated reservation system to handle queues based on real-time load balance – would need sites to have "load" data in machine-readable form (past and present data): **7** (3)
18) Contract with TG to get TG consultant at end user location (and who would pay?): **7** (2)
19) Respond to queue stoppage using "bid system" – allocation e-Bay: **0**
20) Software development kit for TeraGrid – for standard data structures, etc. – and user code repositories: **5** (2)
21) Parallel debugger MPICH-G2 (Added after the voting was completed)

---

[5] The number in bold is the total number of points the idea received in the breakout group voting. The number in parenthesis indicates the number of people who voted for that idea. The four ideas that received the highest number of votes are highlighted in bold.

Appendix C cont.

Group members: Roberto Dandi, Steve Dong, Christopher Gilpin, Axel Kohlmeyer, Phil Maechling, and P.K. Yeung
Recorder: Dane Skow

1) **Very large allocations (> 10^7 SUs)**: **9** (2)
2) Grid enabled workflow system: **5** (1)
3) Compilers which achieve closer to theoretical max performance: **5** (1)
4) Needs to know the goals of the science/scientists: **0**
5) **Large memory size (>10 TB)**: **6** (2)
6) **More CPUs (10X) – more of same kind**: **11** (3)
7) Guidance on how to create metadata: **3** (1)
8) Agreement on file formats for different modules of workflows: **3** (1)
9) Schedules which favor large jobs: **0**
10) Develop tools to integrate data: **0**
11) Visualization of very large data sets (> 5-10 TB): **3** (2)
12) Teach researchers from related fields how to collaborate (computationally): **3** (1)
13) Robust support for embarrassingly parallel jobs on clusters: **4** (1)
14) Metadata or catalog for retrieval/use of appropriate workflow modules: **4** (1)
15) More flexibility to swap allocations on specific machines: **1** (1)
16) Create individual relationships with support staff: **0**
17) **High inter-machine communications bandwidth to processors (multiple machines)**: **8**(2)
18) Algorithm development: **5** (3)
19) Hardware development: **0**
20) Fast turnaround on test jobs (zero queue wait time): **2** (1)
21) Fast (latency) and reliable (uptime) archive storage systems: **4** (2)
22) Portals (providing them): **1** (1)
23) Debugger for (Globus) jobs using multiple machines (integrate with total view): **3** (1)
24) (Random access) disk based storage shared across multiple sites: **1** (1)
25) Improved tools/capabilities for profiling large jobs (~1,000 CPUs): **0**
26) Real-time, interactive visualization: **5** (1)
27) More competent technical support (technology): **0**
28) Help accessing (faster) machines outside TeraGrid (workflow access): **0**
29) Work with commercial software vendors to parallelize code: **0**
30) Promote funding of computational sciences: **4** (1)

Appendix D

Participant List

TeraGrid User Workshop
University Place Conference Center
Indianapolis, IN
Monday, June 12, 2006

Bennett Bertenthal	University of Chicago
Bruce Boghosian	Tufts University
Charlie Catlett	TeraGrid Director
Noshir Contractor, G*roup recorder*	University of Illinois at Urbana-Champaign
Roberto Dandi	National Center for Supercomputing Applications
Yuqing Deng	Argonne National Laboratory
Steve Dong	Brown University
Tom Finholt, *Workshop co-organizer*	University of Michigan
Christopher Gilpin	University of Texas Southwestern Medical Center at Dallas
Hank Green	National Center for Supercomputing Applications
Leopold Grinberg	Brown University
Axel Kohlmeyer	University of Pennsylvania
Philip Maechling	University of Southern California
Ricky Robertson	University of Illinois at Urbana-Champaign
Duane Rosenberg	National Center for Atmospheric Research
Dane Skow, *Group recorder*	TeraGrid Deputy Director
P.K. Yeung	Georgia Institute of Technology
Ann Zimmerman, *Workshop co-organizer*	University of Michigan