Envisioning XD

TECHNICAL REPORT: TERAGRID EXTREME DIGITAL CAMPUS CYBERINFRASTRUCTURE AND CAMPUS BRIDGING REQUIREMENTS ELICITATION MEETING

XROADS:
Driving the Next Generation of Scientific Discovery

Technical Report: TeraGrid eXtreme Digital Campus Cyberinfrastructure and Campus Bridging Requirements Elicitation Meeting

Craig A. Stewart[1], Pervasive Technology Institute, Indiana University

D. Scott McCaulay[2], Pervasive Technology Institute, Indiana University

Susanne Jul[3], Amaryllis Consulting, LLC

Richard L. Moore[4], San Diego Supercomputer Center, University of California San Diego

XROADS is a collaborative effort that responded to the TeraGrid Phase III: eXtreme Digital solicitation (NSF solicitation 08-571 http://www.nsf.gov/pubs/2008/nsf08571/nsf08571.htm) under a planning grant from the National Science Foundation (award number 0928542). XROADS partners include University of California San Diego, University of California Berkeley, University of Chicago/Argonne National Laboratory, Clemson University, University of Delaware, Elizabeth City State University, University of Florida, Hispanic Association of Colleges & Universities, Indiana University, Internet2, Lawrence Berkeley National Laboratory, National Center for Atmospheric Research, Purdue University, Raytheon, and Redelfs LLC. In addition to funding support from the NSF, additional support for this report was provided by the Indiana University Pervasive Technology Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the Indiana University Pervasive Technology Institute, or any of the organizations that are partner to the XROADS proposal..

More information about the XROADS collaboration is available online at http://nsf.gov/awardsearch/showAward.do?AwardNumber=0928542.

Please cite as:

Stewart, C.A., D.S. McCaulay, S. Jul and R.L. Moore. *Technical Report: TeraGrid eXtreme Digital Campus Bridging Requirements Elicitation Meeting*. 2011. Available from: http://hdl.handle.net/2022/13026

[1] stewart@iu.edu, Franklin Hall 116, 601 East Kirkwood Avenue, Bloomington IN, 47405
[2] smccaula@indiana.edu, 2719 E 10th Street, Bloomington IN, 47408
[3] sjul@acm.org 171 Main St. #183, Los Altos CA 94022
[4] rlm@sdsc.edu, MC 0505, 9500 Gilman Drive, La Jolla CA, 92093-0505

Acknowledgments

We would like to thank several people who provided critical input into the execution of the resource elicitation meeting and this report. Dale Lantrip (Indiana University Pervasive Technology Institute) handled local logistics for the meeting. Susanne Jul led the discussion and the activities of the resource elicitation meeting. Thanks to Brian Davenport and Jack Kleinert (Raytheon) for note taking at the meeting and for their expertise in systems engineering – they were very helpful in rendering the output of the meeting into systems engineering terms. Thanks go to Malinda Lingwall (Indiana University Pervasive Technology Institute) for excellent editorial work in finalizing this document. Finally, and most particularly, we thank the (anonymous) participants who took valuable time out of their schedules to participate in this process and thus make this report possible.

This research was directly supported by grant award number 0928542 to the University of California San Diego and a subcontract of that award to Indiana University. This research was also supported by the Indiana University Pervasive Technology Institute and the Office of the Vice President for Information Technology. The Indiana University Pervasive Technology Institute is supported by several sources, and thanks particularly the Lilly Endowment for its support.

This report depends very much on the prior involvement of several XROADS partners in the TeraGrid, which has been funded in part by the NSF via the following grant awards:

- 0504086, 0503697, and 0742145 to the University of Chicago
- 0451237 and 0504075 to Indiana University
- 0122272, 0332113, 0451566, 0503944, 0910847 to the University of California San Diego

Any opinions expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation, the Indiana University Pervasive Technology Institute, or any of the organizations that are partner to the XROADS proposal.

Table of Contents

1 Executive summary

In 2008 the National Science Foundation (NSF) issued a solicitation entitled "TeraGrid Phase III: eXtreme Digital Resources for Science and Engineering (XD)," requesting proposals to design and operate the next phase of the TeraGrid [1]. This solicitation includes the statement,

> "… the TeraGrid has revolutionized the way in which members of the academic science and engineering community use leading-edge digital resources in their research."

The solicitation goes on to say:

> "The goal of this solicitation is to encourage innovation in the design and implementation of an effective, efficient, increasingly virtualized approach to the provision of high-end digital services – extreme digital services … For this reason, we refer to the next phase of the TeraGrid as 'eXtreme Digital', 'XD.'"

This solicitation calls specifically for a systems engineering approach to the design of the next phase of the TeraGrid.

In an effort to systematically investigate requirements for TeraGrid XD, the XROADS collaboration held during 2009 a series of requirements elicitation meetings (REM) with small groups of stakeholders. This report summarizes the conduct of and results from a requirements elicitation meeting on the topics of campus bridging and campus cyberinfrastructure. The meeting's goal was to develop a clearer and more functional definition of what the next phase of the TeraGrid should do to be a resource broadly useful to and used by university and college campuses throughout the US.

The question "What does campus bridging mean?" was discussed at some length. While the resource elicitation meeting did not reach a firm definition of campus bridging, the discussion was consistent with the definition subsequently adopted by the NSF Advisory Committee for Cyberinfrastructure Campus Bridging Task Force [2]:

> *Campus bridging* is the integrated use of cyberinfrastructure operated by a scientist or engineer, other cyberinfrastructure on the scientist's campus, at other campuses, the regional, national, and international levels in a seamless manner as if they were proximate to the scientist. A key aspect to this definition is the crossing of boundaries, both internal to and among organizations, as seamlessly as possible and when working within the context of a Virtual Organization (VO) making the 'virtual' aspect of the organization irrelevant (or helpful) to the work of the VO.

Consensus requirements for TeraGrid XD were identified at this requirements elicitation meeting. After the REM concluded, XROADS collaboration attendees and other XROADS staff created a formal summary of needs statements expressed in proper systems engineering format, and assigned criticality to those statements. These formal statements are listed below, each with an indicator of criticality ranging from critical (high) to desirable (lowest criticality ranking used).

- **TeraGrid XD overall**
    - N1. TeraGrid XD must create a suite of resources that fill the gap between resources that are feasibly provided at the campus level, and the very large Track I and II systems scheduled to be the vast majority of the TeraGrid by the middle of 2010. Criticality: Critical
    - N2. TeraGrid XD must help NSF occupy a leadership position in cyberinfrastructure within the US and world. Criticality: Highly Desirable

- N3. TeraGrid XD should create a reasonable way for campuses to donate compute resources and in so doing facilitate interoperability between local and national resources. Criticality: Critical

- N4. In order to provide a better, more easily usable system, TeraGrid XD should provide a consistent software environment, and in particular a high-quality scripting environment to support workflows and job management. Criticality: Critical

- N5. In order to provide a better, more easily usable system, TeraGrid XD should provide good mechanisms for interoperability of load between TeraGrid XD and commercial cloud providers. Criticality: Highly Desirable

- N6. TeraGrid XD should have a clear focus that is reasonably balanced between research, discovery, and learning. Criticality: Desirable

- N7. TeraGrid XD should create a newsletter that focuses specifically on medium and small schools, to convey information of practical use to them and to demonstrate real interest in broader engagement. Criticality: Critical

- **Resources for Training, Education, and Outreach Services (TEOS)**

  - N8. An Open Science Grid-style of contributed virtual clusters could provide a highly valuable resource for TeraGrid XD Training, Education, and Outreach Services (TEOS). Criticality: Highly Desirable

- **Networking**

  - N9. TeraGrid XD must work more actively with the applied networking community, particularly as represented at EDUCAUSE, and look for economically feasible ways to implement dedicated networking where needed by heavy-duty TeraGrid users. TeraGrid XD should have a networking structure fundamentally different than available today – perhaps making some use of dynamic lambda provisioning. The focus of the TeraGrid XD network design has to be on end-to-end solutions valuable to researchers starting at the desktop and going to and through the next generation of the TeraGrid. Criticality: Critical

- **Identity management**

  - N10. TeraGrid XD should depend upon InCommon (or perhaps, more generally, SAML) and Shibboleth (or its successors) for authentication. Criticality: Critical

- **Globally accessible file systems**

  - N11. TeraGrid XD must include a globally accessible file system and better tools for moving data within TeraGrid XD and between campus systems and TeraGrid XD while protecting as appropriate the security of data. Criticality: Critical

- **Advanced support**

  - N12. TeraGrid XD should consider embedding consultants (paid for via AUSS) within a very few, very large Virtual Organizations (VOs) – with co-investment of personnel provided by such VOs Criticality: Highly Desirable

- N13. TeraGrid XD should offer on-site training [at a major facility of TeraGrid XD]– perhaps with a variety of timing options – one week on site, once per month for 4 months perhaps. Criticality: Desirable

- N14. TeraGrid XD should establish a certification program, with different levels, including: User, Supporter, Educator, and Domain-Specific certification. Criticality: Highly Desirable

- N15. TeraGrid XD should provide good online training tools, to ease the challenges to small schools in particular. Criticality: Critical

- N16. TeraGrid XD should establish a "proposal shepherd" process to aid people through the allocations request process [and improve the allocations request process to boot]. Criticality: Desirable

In addition to the discussions that led to the statements of need above, there was a general and interesting discussion at this REM on the topic of Computational Science as a discipline. This finding was that Computational Science exists as a discipline, in addition to and distinct from computer science. The US science and engineering community needs a curriculum for computational sciences that includes problem decomposition, workflows, modeling, rapid prototyping, scalability, and validation and verification.

This report draws from the experience and expertise of a small focus group of US experts in cyberinfrastructure. It should be of value to the US science and engineering community and to the NSF as regards TeraGrid XD and its more general efforts related to the NSF Cyberinfrastructure Framework for 21st Century Science and Engineering (CF21) [3].

2 Introduction

The NSF solicitation NSF 08-571 "TeraGrid Phase III: eXtreme Digital Resources for Science and Engineering (XD)" [1], issued in 2008, includes the statement,

> "Over the past three years, the TeraGrid has revolutionized the way in which members of the academic science and engineering community use leading-edge digital resources in their research."

It then goes on to say:

> "The goal of this solicitation is to encourage innovation in the design and implementation of an effective, efficient, increasingly virtualized approach to the provision of high-end digital services – extreme digital services – while ensuring that the infrastructure continues to deliver high-quality access for the many researchers and educators that use it in their work. The integration of extreme-scale digital resources and services into a common framework that makes it easy for researchers to take advantage of multiple resources and services remains a challenge. Especially challenging is the desire of many researchers to be able to move between using local resources and national resources within a single, well integrated environment. New ideas and technologies have emerged that make it timely to revisit the architecture of the TeraGrid and to plan to address these challenges in the coming years. In preparation for the next operational phase of the TeraGrid …. The primary goal of the next phase of the TeraGrid is to enable major advances in science and engineering research, in the integration of research and education, and in broadening participation in science and engineering by under-represented groups, by providing researchers and educators with usable access to extreme-scale digital resources, beyond those typically available on a typical campus, together with the interfaces, consulting support and training necessary to facilitate their

use. For this reason, we refer to the next phase of the TeraGrid as 'eXtreme Digital', 'XD.'"

The solicitation text offers the following guidance regarding the preparation of proposals:

> "Key attributes of XD will be that … its design is clearly tied to the user requirements of the science and engineering research community."

In an effort to understand the stakeholder requirements and constraints for TeraGrid XD, the XROADS collaboration held during 2009 a series of seven stakeholder-oriented meetings with scientists, providers, and other future users of the cyberinfrastructure (CI). Each meeting aimed at understanding the needs and requirements of a different stakeholder community. The seven meetings were:

1. Power Users, Jul 27-28, San Diego CA
2. Training, Education and Outreach, Aug 4-5, Indianapolis IN
3. Wide Users, Aug 10-11, Chicago IL
4. Infrastructure Providers, Aug 17-18, San Diego CA
5. Data/Sensors/Instruments, Sep 15-16, San Diego CA
6. Clouds and Emerging Technologies, Sep 29-30, Chicago IL
7. Campus Cyberinfrastructure and Campus Bridging, Oct 15-16, Indianapolis IN

Reports from these meetings, plus an overall summary, will be made available from IU ScholarWorks [4].

Each meeting was unique in content, but the seven meetings were similar in format and were organized in keeping with system engineering approaches to requirements identification. This approach follows specific guidance in the NSF solicitation [1] that TeraGrid XD be "designed and implemented in a way that is consistent with sound system engineering principles." These meetings were conducted with a facilitator, in the form and format of what is called in systems engineering terms a *requirements elicitation meeting* (REM). General guides to the concept of resource elicitation are available [5, 6].

This report summarizes the conduct of and results from the Campus Cyberinfrastructure and Campus Bridging REM. This meeting's goal was to develop a clearer and more functional definition of what the next phase of the TeraGrid should do to effectively deliver services to campuses of US colleges and universities.

For the sake of clarity, we have in this report adopted the following conventions and definitions:

> *Cyberinfrastructure* (CI) consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible [7, 8].
>
> *TeraGrid* refers to the suite of services offered from 2004 through 2011 under a series of NSF grants (supplemented strongly by match contributions from awardees) including the awards 0504086, 0503697, and 0742145 to the University of Chicago to provide software, services, support, and organization; and a series of awards to Resource Partners that provided computing, data, storage, visualization, and interface facilities (including awards 0122272, 0332113, 0451566, 0503944, and 0910847 to the University of California San Diego and awards 0451237 and 0504075 to Indiana University). The TeraGrid home page as of the date of this report is http://www.teragrid.org/.

> *TeraGrid XD* refers to the suite of services to be offered with NSF support to the US research community (and when appropriate beyond) as defined in NSF 08-571 "TeraGrid Phase III: eXtreme Digital Resources for Science and Engineering (XD)" [1].
>
> *XROADS* refers to a particular set of philosophies, services, and approaches to implementation of TeraGrid XD as put forth in proposal prepared under a planning grant (number 0928542) funded by the National Science Foundation and formally proposed as the organizing entity for TeraGrid XD in proposal number 1053614 submitted to the National Science Foundation with Richard Moore, San Diego Supercomputing Center, University of California San Diego as PI. *XROADS* is a collaborative effort with partners including University of California San Diego, University of California Berkeley, University of Chicago/Argonne National Laboratory, Clemson University, University of Delaware, Elizabeth City State University, University of Florida, Hispanic Association of Colleges & Universities, Indiana University, Internet2, Lawrence Berkeley National Laboratory, National Center for Atmospheric Research, Purdue University, Raytheon, and Redelfs LLC.

The Campus Cyberinfrastructure and Campus Bridging REM took place after the announcement of the formation of an NSF Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging [9] – a topic clearly related to the general theme of campus cyberinfrastructure. The question "What does campus bridging mean?" was discussed at some length. While the requirements elicitation meeting did not reach a firm definition of campus bridging, the discussion was consistent with the definition subsequently adopted by the NSF Advisory Committee for Cyberinfrastructure Campus Bridging Task Force [2]:

> *Campus bridging* is the integrated use of cyberinfrastructure operated by a scientist or engineer, other cyberinfrastructure on the scientist's campus, at other campuses, the regional, national, and international levels in a seamless manner as if they were proximate to the scientist. A key aspect to this definition is the crossing of boundaries, both internal to and among organizations, as seamlessly as possible and when working within the context of a Virtual Organization (VO) making the 'virtual' aspect of the organization irrelevant (or helpful) to the work of the VO.

The Campus Cyberinfrastructure and Campus Bridging REM took place in the particular context of a competition between two teams awarded planning grants to submit full proposals for the overall organization and operation of TeraGrid XD. As stated in the solicitation itself,

> "As a recent external assessment of the TeraGrid noted, 'The TeraGrid is among the best grids in the world with respect to providing resources to a broad and inclusive audience. The TeraGrid is: enabling leading edge science and engineering, providing access to world class heterogeneous resources for a large and diverse community of researchers and educators, [and] leading in innovative approaches to exploiting cyber-infrastructure.'" [1]

This statement is affirmed in a variety of ways by a variety of sources, including highly positive satisfaction surveys from users of the TeraGrid from 2005 through 2010.

The XROADS collaboration includes several organizations that have been involved with the TeraGrid [10] – in some cases from its inception. The XROADS collaboration believed, based on this past experience, the things the TeraGrid did well would continue to be done well in the future if the XROADS collaboration was ultimately selected to implement TeraGrid XD. The requirements elicitation meetings were thus very much focused on areas of perceived need for improvement or expansion of services to be delivered to the US science and engineering community.

In general, the requirements and recommendations identified by the REM participants in this report identify needs general to any implementation of a TeraGrid XD as defined above. This REM took place after the NSF released the Dear Colleague Letter titled Cyberinfrastructure Framework for 21st Century Science and Engineering (now referred to as CIF21) [3], which was considered as part of the background for and input into the discussion at this REM. In this report we describe the needs identified that apply in general to any implementation of TeraGrid XD overall, CIF21, or the general US science and engineering research and education community.

The remainder of this document provides additional information about the meeting and its outcomes. Section 3 provides details about the requirements elicitation meeting. Section 4 presents a formal system engineering statements of needs derived from this REM.

This report draws from the experience and expertise of a small focus group of US experts in cyberinfrastructure. It should be of value to the US science and engineering community and to the NSF as regards TeraGrid XD and its more general efforts related to implementation of its vision for a Cyberinfrastructure Framework for 21st Century Science and Engineering.

3 Preparation for meeting and in-person meeting activities

The XROADS collaboration identified a small group of experts and community representatives who could reflect the needs of the greater US science and engineering community regarding the topics of campus cyberinfrastructure and campus bridging. Because this REM was conducted in the midst of a competition for NSF funding, the XROADS collaboration chose not to invite faculty or staff from institutions known to be participating in the competing XSEDE proposal [11]. Nine nationally known experts were invited to participate in this REM; seven were able to participate. In this report 'REM participants' refers to these seven experts (not identified by name in this report). Eight members of the XROADS collaboration also attended as observers, facilitators, or presenters; these individuals are referred to in this report as 'XROADS attendees.' The names of the XROADS attendees and their affiliations are given in Appendix 1. This study was approved by the Indiana University Office of Research Administration Human Subjects office as an exempt research study as protocol #1102004821.

The objectives for this meeting were identified in advance to all REM participants and XROADS attendees as follows:

- Better understand current realities on campuses with respect to cyberinfrastructure / high performance computing / TeraGrid.
- Anticipate needs for next 3-5 years – how can TeraGrid XD be of greater value from the campus viewpoint? How can TeraGrid XD and campuses bridge the technical gaps for a more seamless user experience?
- Identify challenges and opportunities, particularly networking issues/options.
- Investigate resource sharing – extending the pool of resources available to TeraGrid XD in a win-win way and in ways that match US research needs more broadly.
- Investigate resources geared specifically toward training, education, and outreach services (TEOS). Undergrads, grad students, and high school students don't really need to run on Kraken. Meet those needs in a meaningful way.

3.1 Information collection prior to requirements elicitation meeting

REM participants completed a questionnaire (via the Web) in advance of the meeting. The results of that questionnaire were used to create summaries used as one of several forms of input for discussion at the in-person meeting in Indianapolis. The questions in the questionnaire are presented in Appendix 2.

3.2 In person meeting at the IUPUI campus

The Resource Elicitation Meeting was held on the Indiana University–Purdue University Indianapolis (IUPUI) campus. The REM participants generally arrived in Indianapolis on the evening of 14 October, and stayed during the course of the meeting at the University Place Conference Center and Hotel on the IUPUI campus. Travel expenses were reimbursed, and REM participants and XROADS attendees were provided all meals on site save breakfast. REM participants and XROADS attendees were provided a per diem for expenses based on US Government rates implemented according to Indiana University travel policy. REM participants were not otherwise compensated or paid for their time and expertise.

The actual course of events, as they took place, was as follows:

Thursday 15 October 2009

9:00am – Welcome and introductions
9:20am – XROADS team introductions
9:34am – Meeting agenda and logistics
9:37am – Question for invitees: What brought you here?
10:08am – XROADS overview
10:45am – Break
11:00am – Meeting objectives
12:10pm – Lunch
1:42pm – Lunch discussion debrief
2:00pm – Questionnaire responses summary
3:30pm – Break
3:49pm – Breakout sessions
4:36pm – Breakout group #2, report out
4:49pm – Breakout group #1, report out
5:06pm – Seeding the dinner discussion
6:00 pm – Dinner and discussion

Friday 16 October 2009

8:09am – Dinner discussion recap
9:00am – Bridging the last mile to campus
9:38am – Break
10:12am – Last mile for data users
10:30am – Topics that haven't been covered; topics that need to be revisited
11:22am – Big project support?
11:40am – Wrap-up: What are you taking away?
11:45am – Wrap-up: What should TeraGrid XD take away?
12:00 Box lunch to go; participants depart IUPUI campus

The meeting began with personal introductions. Participants were asked to describe their current position, research, use of high-performance computing, professional background, intellectual passion, and, where appropriate, role in XROADS. General introductions concluded with invitees individually answering the question, "What brought you here?" The Campus Cyberinfrastructure and Campus Bridging REM was different than many other REMs since it was based on organizational structures and functional roles within organizations, rather than a particular type of user or scientific need. This was evident in the

answers to this question, which included, “to have a chance to influence the TeraGrid so that it meets needs of researchers on my campus more effectively,” “have a chance to understand what TeraGrid offers to campuses now,” and “better understand how to connect campuses to the TeraGrid in terms of networking.”

Richard Moore, Acting Division Director, UC San Diego/SDSC, San Diego, CA, offered an overview of the TeraGrid XD initiative and the XROADS project.

During the discussion at this meeting, it was noted that first author Craig Stewart was involved at that time in the general issue of campus bridging in two contexts. One context was as a participant in the XROADS collaboration and the other was as chair of the NSF Advisory Committee for Cyberinfrastructure Campus Bridging Task Force. In response to suggestions made at the requirements elicitation meeting, Stewart committed that after the TeraGrid XD proposals have been submitted the key outcomes from the Campus Cyberinfrastructure and Campus Bridging REM would be made more generally available to the community in the form of a report. The REM participants thought it valuable that the results of their time and effort at the campus cyberinfrastructure and campus bridging REM be available to the US science and engineering research and education community generally. Participants thought it particularly likely that this report would be of value to the ACCI Campus Bridging Task Force in the development of its reports and recommendations [2]. This report fulfills first author Stewart’s commitment to the REM participants.

Because the organization of this REM was more structural than topical, discussion moved from topic to topic throughout the discussion. Early discussion focused on networking technology, networking needs, and the gap in resources between what is realistically available at the campus, state, or regional level and what must be provided at the national level.

In discussions there were, initially in particular, very divergent opinions even among the small number of participants in this REM. There were wide ranging discussions during the day and a half meeting in Indianapolis. This is to be expected in any discussion of facilities as large and encompassing as the existing TeraGrid or the planned TeraGrid XD facility. By the end of a day and a half of discussion, the REM participants had arrived at a set of consensus requirements for TeraGrid XD. This does not mean there was unanimous support for each of the consensus requirements. These requirements represented consensus – in the sense of general sense of the group – without unanimity. Most requirements agreed upon were relevant to TeraGrid XD, although there was one finding of a more general nature that received strong support in discussion at this REM.

4 Development of formal systems engineering statements of needs

After the requirements elicitation meeting concluded, XROADS staff converted the consensus statements and requirements into a format expressed properly within systems engineering practice. Needs statements were given a criticality ranking according to the number of participants expressing interest in and desire for the need, and the intensity of their expression. Criticality ratings of the needs expressed at this requirement elicitation meeting were categorized according to the following criteria:

- *Critical.* Multiple REM participants appeared to indicate that this is essential to their ability to use the system effectively.

- *Highly Desirable.* Multiple REM participants appeared to indicate that this would enhance their use of the system significantly.

- *Desirable.* Multiple REM participants appeared to express interest or desire, or one participant indicated that this was essential or highly desirable.

- *Not rated.* This rating typically represents a need that emerged from analysis of meeting notes but which was not discussed directly and fully at the meeting with all participants present.

These needs statements are listed below grouped by topic and ordered by criticality, with most critical first. In addition, for each need there is a rationale written by a member of the XROADS collaboration that provides explanation or context for a particular need.

4.1 TeraGrid XD overall

N1. TeraGrid XD must create a suite of resources that fill the gap between resources that are feasibly provided at the campus level, and the very large Track I and II systems scheduled to be the vast majority of the TeraGrid by the middle of 2010.

Criticality: Critical

Explanation: The upper bound of what can be expected reasonably at the level of campus, state, or regional facilities is far from the lower bound of resources scheduled for availability in the TeraGrid in the 2010-2011 time range (e.g. Ranger [12], Kraken [13], and Blue Waters [14]). Many of the existing, smaller supercomputers within the TeraGrid are slated to leave service as part of the TeraGrid in 2010 or 2011. Future solicitations from the NSF for major computing resources are more likely to be at a scale larger than the existing Ranger and Kraken. REM participants viewed filling the gap that exists right now between the upper end of what a campus might have as a local resource and the low end of what is planned as resources within the TeraGrid as primarily a national need, not something that can be done at the campus or regional level. Need N3 below suggests one way in which TeraGrid XD could fulfill this need.

N2. TeraGrid XD must help NSF occupy a leadership position in cyberinfrastructure within the US and world.

Criticality: Highly Desirable

Explanation: NSF has for decades had a leadership role in cyberinfrastructure. In recent years, other organizations – particularly the Department Of Energy (DOE) – have taken on more of a leadership role in high performance computing. This has been largely a result of DOE investments in very large-scale parallel supercomputers focused on simulation. NSF-funded science and engineering research is becoming more data-centric. This is true of research uses of cyberinfrastructure in general. It is not that large-scale simulation is becoming unimportant. Rather, data-centric computing is expanding in importance with the advance of the data deluge at high rates. TeraGrid XD, if it is successful in fulfilling the goals set out in the NSF solicitation, must and will have the effect of putting NSF in more of a leadership position in cyberinfrastructure in the US and world generally.

N3. TeraGrid XD should create a reasonable way for campuses to donate compute resources and in so doing facilitate interoperability between local and national resources.

Criticality: Critical

Explanation: There was considerable discussion at this REM of the success of the Open Science Grid [15] and the use of contributed computer cycles. The REM participants felt that many campuses would be willing to donate compute resources to TeraGrid XD if there were a reasonable way to do so. The REM participants felt in particular that there would be high value to providing a way to deploy and use a consistent environment within donated facilities. This would be a way to match needs in the gap between current Tier 3 vs. Tier 1 and 2 systems. Such clusters might fall into units of something between 64 and

256 cores generally. Participants agreed that something akin to the old "cluster in a box" software [16] would be valuable particularly if it provided a way to deploy a set of VMs that could be used for TeraGrid XD users. A combination of Rocks [17] and Grid Appliance [18] approaches were suggested as possible technical solutions. It was also agreed among REM participants that this would be good for TeraGrid XD, good for the national science community, and would have the side effect of improving consistency and quality of cluster management at colleges and universities generally.

N4. In order to provide a better, more easily usable system, TeraGrid XD should provide a consistent software environment, and in particular a high-quality scripting environment to support workflows and job management.

Criticality: Critical

Explanation: A great deal of time was spent discussing the obstacles in moving from campus to national level cyberinfrastructure such as the existing TeraGrid. For a variety of reasons – some having to do with the variety and quality of some local environments, some having to do with the current TeraGrid – this transition is simply more too difficult. Creating a consistent software environment, in ways that allow the software environment to be mimicked at the local level, the barriers to adoption of national CI and integration of national and local CI will be greatly reduced. The REM participants recognized that some of the largest supercomputers in the TeraGrid are going to be unique in certain regards. The interconnect and processors, for example, are likely to form a unique combination for any of the largest supercomputers in the TeraGrid. For that reason, REM participants felt it was unrealistic to set a goal of a completely consistent environment across all TeraGrid XD systems. The idea of a high quality scripting environment was appealing to REM participants as a way to improve researchers ability to run workflows across multiple systems (and types of systems) within TeraGrid XD. A high quality scripting environment was also seen as a rational response to the essential fact of system heterogeneity within TeraGrid XD.

N5. In order to provide a better, more easily usable system, TeraGrid XD should provide good mechanisms for interoperability of load between TeraGrid XD and commercial cloud providers.

Criticality: Highly Desirable

Explanation: The REM participants made note of two empirical facts: cloud computing providers are implementing facilities that get significantly larger each year; and more and more researchers are using cloud providers as one option among several to meet their computing needs. Particularly from the viewpoint of campus CI and campus bridging, the REM participants supported the importance of having good ability to interoperate between TeraGrid XD and commercial cloud providers. What this implies is that if there were good ability to interoperate between campus CI and TeraGrid XD, and good interoperability between TeraGrid XD and commercial cloud providers, then from the standpoint of an individual researcher this would create good interoperability among all of these resources – campus, TeraGrid XD, and commercial clouds.

N6. TeraGrid XD should have a clear focus that is reasonably balanced between research, discovery, and learning.

Criticality: Desirable

Explanation: REM participants discussed different types of needs faced by faculty and students at different campuses. The Carnegie Classifications [19] are intended to and do something about the type of research done at a particular institution. And while it seems almost tautological, the majority of very large research projects in higher education are led by faculty at very large research universities. But if TeraGrid

is to be broadly useful to campuses throughout the US is must take a more balanced approach to research (in CI through its implementations), discovery (in science disciplines supported by TeraGrid XD), and teaching and learning. The participants generally agreed that TeraGrid – by virtue of its mission – has been focused very heavily in supporting research, and that more support for learning would clearly be good. It was also acknowledged that this would be a deviation from past history and this would be difficult in a financially constrained environment.

4.2 Resources for Training, Education, and Outreach Services (TEOS)

N7. TeraGrid XD should create a newsletter that focuses specifically on medium and small schools, to convey information of practical use to them and to demonstrate real interest in broader engagement.

Criticality: Critical

Explanation: This need was the result of discussion about the challenges of learning about TeraGrid if you are currently at a small college or university, or the challenge of getting someone at a small college or university to become interested in the TeraGrid. The REM participants felt that the current TeraGrid Web presence is too hard to understand if one is not already a user of the TeraGrid. Both the fact that one has to go to the TeraGrid Web page to learn about the TeraGrid (a "pull" for information) and the Web site itself were viewed by some workshop participants as creating obstacles to adoption. A high-quality newsletter targeted at small to medium size institutions of higher education could increase adoption of TeraGrid XD. First, such a newsletter – if done well – would demonstrate real commitment to outreach beyond the current TeraGrid user community. A newsletter would also be a 'push' form of communication, and one targeted to researchers and educators at small to medium colleges and universities would also allow a focus on material of specific interest to this user community.

N8. An Open Science Grid-style of contributed virtual clusters could provide a highly valuable resource for TeraGrid XD Training, Education, and Outreach Services (TEOS).

Criticality: Highly Desirable

Explanation: This is a specific way to implement need N2 for TEOS purposes.

4.3 Networking

N9. TeraGrid XD must work more actively with the applied networking community, particularly as represented at EDUCAUSE, and look for economically feasible ways to implement dedicated networking where needed by heavy-duty TeraGrid users. TeraGrid XD should have a networking structure fundamentally different than available today – perhaps making some use of dynamic lambda provisioning. The focus of the TeraGrid XD network design has to be on end-to-end solutions valuable to researchers starting at the desktop and going to and through the next generation of the TeraGrid.

Criticality: Critical

Explanation: There were several components to the discussion. The discussion of the current state of affairs had the following three main components:

- The original TeraGrid network – with a 4 x 10 Gbps 'backbone' – was long ago recognized as not sufficient in many regards to meet the peak needs of heavy-duty TeraGrid users. On the one hand, the average utilization of some of the main parts of the original TeraGrid network as in the past

been low enough that the bandwidth in some portions has decreased over time. The TeraGrid network has become more star-like in topology over time as new resource partners have been added, with a main hub in Chicago and often independent links to the various current Resource Partners. The bandwidth of the existing TeraGrid network connections from RPs to each other is insufficient to support movement of petascale data sets around within the TeraGrid.

- The TeraGrid network is private and essentially isolated from other research networks. This makes movement of data from outside the TeraGrid to inside the TeraGrid more difficult than might otherwise be the case.

- For many campuses the critical bottleneck in bandwidth is the last mile – or perhaps more accurately the last two miles. One part of the 'last mile' problem is the connectivity from a high-speed network to the core router on a campus. The second component is the connectivity from the main campus connecting point to the desks and labs of individual researchers on campus.

There are multiple elements to this need statement. REM participants suggested that working out the technological design of a TeraGrid XD network was not the best use of their time in a requirements elicitation meeting. Discussion focused instead on the following observations and suggestions:

- Without taking time to make recommendations on specifically how to restructure the existing TeraGrid network, the REM participants felt that restructuring was needed, and that a key aspect of such a restructuring should be the abandonment of the use of a private, isolated network.

- One observation made and generally agreed upon by REM participants is that there is not nearly as much interaction as there should be between staff of the TeraGrid and two other groups of people: the applied networking community in general, and the community of Information Technology (IT) and CI staff and faculty who participate regularly in EDUCAUSE [20] in particular. As a result, the TeraGrid staff do not have as good an understanding of some of the real, practical networking problems faced 'on the ground' in smaller and mid-sized campuses. May of these are state-specific, in some cases tied to state regulations on telecommunications and IT infrastructure. REM participants felt that more interaction with EDUCAUSE and the community of EDUCAUSE members would be a particularly good approach to correcting this situation.

- Again, without trying to design a new network for TeraGrid XD, the REM participants stressed the importance of accounting for and attending to 'last mile' performance issues. They suggested that use of dynamic bandwidth services as one useful component of a new TeraGrid XD network (either available or announced from two major national backbone providers at the time of the REM).

4.4 Identity management

N10. TeraGrid XD should depend upon InCommon (or perhaps, more generally, SAML) and Shibboleth (or its successors) for authentication.

Criticality: Critical

Explanation: The discussion of current authentication within the TeraGrid included two general themes:

- Each TeraGrid user generally has multiple usernames on multiple accounts. In fact, the number of different usernames held by a particular user of the TeraGrid may often equal the number of different resource partner sites at which that user has authorization to use systems.

- The recent TeraGrid Single-Sign-On capability [21] is a tremendous step forward in ease of use of the TeraGrid.

From the perspective of campus cyberinfrastructure and campus bridging, REM participants noted that the current arrangement of usernames and authentication methods is a significant obstacle to implementation of complex workflows that strive to use a mix of TeraGrid and non-TeraGrid (e.g. campus or commercial cloud) resources.

The REM participants were all aware of the recommendation made in a recent joint workshop report published by EDUCAUSE and CASC (Coalition for Academic Scientific Computation) [7]. This need is essentially an endorsement of a very similar recommendation made in the EDUCAUSE/CASC workshop report. While Shibboleth [22] was discussed explicitly by the REM participants, the sense of the discussion clearly extended to newer successors such as the newer CILogon service being developed at the National Center for Supercomputing Applications [23].

4.5 Globally accessible file systems

N11. TeraGrid XD must include a globally accessible file system and better tools for moving data within TeraGrid XD and between campus systems and TeraGrid XD while protecting as appropriate the security of data.

Criticality: Critical

Explanation: Many of the REM participants were familiar with past globally accessible file system efforts based on IBM's General Parallel File System (GPFS) software or the open source Lustre file system [24-27]. There was strong support for a globally accessible file system within TeraGrid XD based on the following factors:

- A globally accessible file system ameliorates the impact of differences in network connectivity to and within the TeraGrid. A researcher with relatively poor connectivity can move data to a globally accessible file system as best possible, and once the data are moved to that point, they may easily be moved around within TeraGrid XD.

- This need represents an outcome of another part of the general discussion of complex workflows spanning TeraGrid resources and non-TeraGrid resources. REM participants supported a globally accessible file system as very helpful in facilitating such workflows.

While recognizing and endorsing the value of globally accessible file systems, they noted that current tools for movement of data do not meet researcher needs for functionality or expectations regarding ease of use.

4.6 Advanced support

N12. TeraGrid XD should consider embedding consultants (paid for via AUSS) within a very few, very large Virtual Organizations (VOs) – with co-investment of personnel provided by such VOs.

Criticality: Highly Desirable

Explanation: The REM participants were highly supportive of Virtual Organizations as an important way to organize scientific research. The focus of the National Science Foundation on VOs as important to science, and cyberinfrastructure as a way to support the activities of VOs, were both well supported by comments made in discussion. As in other segments of the discussion, the Open Science Grid was viewed

as a good example – in this discussion, in terms of its use of VOs as an organizing tool. Embedding *expert* consultants in VOs is implicit in this statement of need by virtue of the comment that the consultants should be paid from the Advanced User Support Services portion of a planned TeraGrid XD budget. The embedding of expert consultants was explicit in the discussion. The REM participants viewed this sort of embedded consultant as likely to be highly effective and cost-effective in terms of facilitating use of TeraGrid XD. The idea of co-investment of personnel provided by a VO to aid support of TeraGrid XD usage by VO members was regarded as useful in several ways. The most important two ways discussed were leveraging investment by TeraGrid XD with co-investment (match) by the VO, and in calling for a strong level of commitment on the part of the VO. This commitment was regarded as highly important to the effectiveness of a relationship between TeraGrid XD and a VO. In general REM participants agreed that this sort of embedded consultant would be highly effective as a way to transmit information and training expertise into large VOs, thus leveraging the structure of the VO.

N13. TeraGrid XD should offer on-site training [at a major facility of TeraGrid XD]– perhaps with a variety of timing options – one week on site, once per month for 4 months perhaps.

Criticality: Desirable

Explanation: This need was ranked as "Desirable" in the sense of the definition given above: "Multiple participants appeared to express interest or desire, or one participant indicated that this was essential or highly desirable." There was general support, but the strongest support came from one REM participant who had participated in training at a major TeraGrid site. The two main benefits of a training experience on site at a major XD facility were the following:

- Getting a person out of their home environment and the distractions and interruptions associated with that, so that they could focus on training about TeraGrid XD, was viewed as very important.
- Embedding people in an environment filled with experts on TeraGrid XD was viewed as a very effective way to train. Such training might be for local support personnel as well as researchers who have to do their own support.

The above enthusiasm for training away from home, at a TeraGrid XD site, was tempered by the realization that it probably takes four weeks of training for a TeraGrid XD novice to become an effective supporter of or self-sufficient research user. One month away from home in a single block was viewed as simply not feasible. The idea of one week a month for four months was accepted by the REM participants as a good happy medium – a month away, but broken into segments to make it feasible for the trainee and the trainee's colleagues and / or employer.

N14. TeraGrid XD should establish a certification program, with different levels, including: User, Supporter, Educator, and Domain-Specific certification.

Criticality: Highly Desirable

Explanation: This discussion focused on two themes. One was the value to the science and engineering research community of a certification process as a way to assure quality of support. The other was the natural motivating power of anything that has elements of a 'merit badge' program. Certifications provide a level of assurance to end users and to campus administration that trainers and support personnel are qualified to do what they assert they can do. A certification process creates an incentive for support personnel to take tests and verify their level of knowledge. At the same time, the REM participants were uniform in believing that this sort of approach would provide a positive feedback loop that motivates everyone to create a better support environment.

N15. TeraGrid XD should provide good online training tools, to ease the challenges to small schools in particular.

Criticality: Critical

Explanation: Like the discussion of InCommon, discussion of online training largely echoed needs already widely expressed in other venues. The problem is simple and well understood, and has the following general components:

- High-quality online training tools are important first to make effective use of the resources invested in the TeraGrid and cyberinfrastructure generally.
- There is considerable duplicated effort in creating training materials, and the training materials that exist are of variable quality.
- Smaller schools in particular are not able to create online tools or new curricula from scratch because it is difficult to provide faculty members the release time required to prepare good materials.

The REM participants were uniform in supporting the idea that putting high quality training tools online helps everyone – from end users to educators.

N16. TeraGrid XD should establish a "proposal shepherd" process to aid people through the allocations request process [and improve the allocations request process to boot].

Criticality: Desirable

Explanation: This need was the result of a discussion that in some ways paralleled the discussion of the TeraGrid network. In years past the TeraGrid has recognized shortcomings in the process of applying for and receiving allocations of resources from the TeraGrid. The TeraGrid has considerably improved that process in the recent years. Yet the current allocation process is still widely agreed to be too difficult and arcane. The term 'proposal shepherd' was invented by an REM participant to designate a person with a role of navigating an applicant through the TeraGrid allocation application process. The REM participants noted that if the process were made radically easier under TeraGrid XD, a 'proposal shepherd' might become less important, but someone to guide requestors through the process would be greatly helpful at present.

4.7 Finding not directly pertinent to TeraGrid XD – Computational Science curriculum

In addition to the above consensus statements of requirements regarding TeraGrid XD, there was a general and interesting discussion at this REM on the topic of computational science as a discipline. This resulted in the following statement that captures the very strong consensus of the REM participants:

> Computational Science exists, as a distinct discipline, in addition to and distinct from computer science. The US science and engineering community needs a curriculum for computational sciences that includes problem decomposition, workflows, modeling, rapid prototyping, scalability, and validation and verification.

An important point made in discussion is that this is not just a matter of a group of people asserting that "what we do constitutes a discipline and we want to be recognized." Rather, the most important thing about identifying this area as a discipline is that so doing signals to young students that they can plan an academic career around training in and pursuing a given discipline. REM participants strongly affirmed the many existing reports that talk about more need for highly trained IT professionals skilled in

computational science. They felt that identifying this area as a discipline would help attract more talented people to pursue this area of study, and that establishing an agreed-upon curriculum would help training and education be highly effective. This finding is consistent with the recommendation made by the NSF Advisory Committee on Cyberinfrastructure (ACCI) that: "The National Science Foundation should create a program in Computational and Data-Intensive Science and Engineering (CDS&E), based in and coordinated by the NSF Office of Cyberinfrastructure." A copy of the ACCI recommendation, endorsed by then-Director Dr. Arden Bement, is available [28].

5 References

[1] National Science Foundation. TeraGrid Phase III: eXtreme Digital Resources for Science and Engineering (XD). 2008. Available from: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503282 [cited 31 Jan 2011]

[2] Indiana University Pervasive Technology Institute. Campus Bridging. Available from: http://pti.iu.edu/CampusBridging [cited 31 Jan 2011]

[3] National Science Foundation. Dear Colleague Letter: Cyberinfrastructure Framework for 21st Century Science and Engineering (CF21). 2010. Available from: http://www.nsf.gov/pubs/2010/nsf10015/nsf10015.jsp?org=OCI [cited 31 Jan 2011]

[4] Indiana University. IUScholarWorks Repository: XROADS Community. Available from: https://scholarworks.iu.edu/dspace/handle/2022/12994 [cited 1 Mar 2011]

[5] Masters, M. An Overview of Requirements Elicitation. 2010. Available from: http://www.modernanalyst.com/Resources/Articles/tabid/115/articleType/ArticleView/articleId/1427/An-Overview-of-Requirements-Elicitation.aspx [cited 31 Jan 2011]

[6] Sommerville, I. and P. Sawyer Requirements Engineering - A Good Practice Guide. John Wiley & Sons. Available from: http://www.knovel.com/web/portal/browse/display?_EXT_KNOVEL_DISPLAY_bookid=1537 [cited 31 Jan 2011]

[7] EDUCAUSE Campus Cyberinfrastructure Working Group and Coalition for Academic Scientific Computation. *Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies*. February 2009. Available from: http://www.educause.edu/Resources/DevelopingaCoherentCyberinfras/169441 [cited 31 Jan 2011]

[8] Stewart, C., S. Simms, B. Plale, M. Link, D. Hancock and G. Fox. What is Cyberinfrastructure? In: *Proceedings of SIGUCCS 2010* (Norfolk, VA, 24-27 Oct, 2010). Available from: http://portal.acm.org/citation.cfm?doid=1878335.1878347 [cited 31 Jan 2011]

[9] National Science Foundation. Event: NSF Advisory Committee for Cyberinfrastructure. 2010. Available from: http://www.nsf.gov/events/event_summ.jsp?cntn_id=114728&org=null [cited 31 Jan 2011]

[10] TeraGrid. Home page. Available from: http://www.teragrid.org/ [cited 31 Jan 2011]

[11] Towns, J., R. Roskies, J. Boisseau and P. Andrews. *Planning for XSEDE: the eXtreme Science and Engineering Discovery Environment* (Award number 0941686). National Science Foundation, 2009. Available from: http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0941686 [cited 31 Jan 2011]

[12] Texas Advanced Computing Center. HPC Systems. Available from: http://www.tacc.utexas.edu/resources/hpc/ [cited 31 Jan 2011]

[13] National Institute of Computational Sciences. Kraken. Available from: http://www.nics.tennessee.edu/computing-resources/kraken [cited 31 Jan 2011]

[14] NCSA. About the Blue Waters project. Available from: http://www.ncsa.illinois.edu/BlueWaters/ [cited 31 Jan 2011]

[15] Open Science Grid. Home page. Available from: http://www.opensciencegrid.org/ [cited 31 Jan 2011]

[16] Green, K. In-A-Box is Out of the Bag. 2001. Available from: http://access.ncsa.illinois.edu/Stories/IAB/ [cited 31 Jan 2011]
[17] Rocks Clusters. Home page. Available from: http://www.rocksclusters.org/ [cited 31 Jan 2011]
[18] Joomla! Grid Appliance User Interface. Available from: http://www.grid-appliance.org/http://www.grid-appliance.org/ [cited 31 Jan 2011]
[19] Carnegie Foundation for the Advancement of Teaching. The Carnegie Classification of Institutions of Higher Education. 2010. Available from: http://classifications.carnegiefoundation.org/ [cited 31 Jan 2011]
[20] EDUCAUSE. What is EDUCAUSE? Available from: http://www.educause.edu/ [cited 31 Jan 2011]
[21] TeraGrid. TeraGrid Login Quick Start: Single Sign-On (SSO). Available from: https://http://www.teragrid.org/web/user-support/login_quickstart [cited 31 Jan 2011]
[22] Shibboleth. Home page. Available from: http://shibboleth.internet2.edu/ [cited 31 Jan 2011]
[23] CILogon Service. Home page. Available from: https://cilogon.org [cited 31 Jan 2011]
[24] Simms, S.C., G.G. Pike and D. Balog. Wide Area Filesystem Performance Using Lustre on theTeraGrid. In: *Proceedings of Proceedings of the TeraGrid 2007 Conference* (Madison, WI, 2007). Available from: http://pti.iu.edu/pubs/wide-area-filesystem-performance-using-lustre-teragrid [cited 15 Dec 2009]
[25] Simms, S.C., C.A. Stewart and S.D. McCaulay. Cyberinfrastructure resources for U.S. Scholarship: the TeraGrid. In: *Proceedings of Proceedings of the 36th annual ACM SIGUCCS conference on User services* (Portland, OR, USA, 2008). ACM. Available from: http://portal.acm.org/citation.cfm?id=1450057 [cited 10 Feb 2010]
[26] Simms, S., S. Tiege, G. Pike, B. Hammond, Y. Ma, C. Westneat, L.L. Simms and D. Balog. Empowering Distributed Workflow with the Data Capacitor: Maximizing Lustre Performance Across the Wide Area Network. In: *Proceedings of Proceedings of the Workshop on Service-Oriented Computing Performance: Aspects, Issues, and Approaches* (Monterey, CA, 2007). Available from: http://portal.acm.org/citation.cfm?id=1272465 [cited 31 Jan 2011]
[27] Walgenbach, J., S.C. Simms, J.P. Miller and K. Westneat. Enabling Lustre WAN for Production Use on the TeraGrid: A Lightweight UID Mapping Scheme. In: *Proceedings of TeraGrid'10* (Pittsburgh, PA, 2010). Available from: http://dx.doi.org/10.1145/1838574.1838593 [cited 28 Feb 2011]
[28] NSF Advisory Committee for Cyberinfrastructure Task Force on Grand Challenges. *Final Report*. March 2011. [PDF] Available from: http://www.nsf.gov/od/oci/taskforces/TaskForceReport_GrandChallenges.pdf [cited 31 Mar 2011]
[29] Orcutt, J., G. Fox, I. Foster and R. Moore. *XROADS: Planning the Future eXtreme Digital Cyberinfrastructure* (Award number 0928542). National Science Foundation, 2009. Available from: http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0928542 [cited 31 Jan 2011]
[30] Stewart, C.A., A. Shankar, T. Hacker, E. Wernert, M. Bruhn, D.S. McCaulay and B. Voss. *SCI: TeraGrid Resource Partners: Indiana University* (Award number 0504075). National Science Foundation, 2005. Available from: http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0504075 [cited 31 Jan 2011]
[31] Stewart, C.A., A. Shankar, M. McRobbie, S. Simms and B. Voss. *SCI: ETF Early Operations - Indiana University* (Award number 0451237). National Science Foundation, 2005. Available from: http://nsf.gov/awardsearch/showAward.do?AwardNumber=0451237 [cited 31 Jan 2011]
[32] Berman, F. and P. Messina. *The TeraGrid: Cyberinfrastructure for 21st Century Science and Engineering* (Award number 0122272). National Science Foundation, 2001. Available from: http://nsf.gov/awardsearch/showAward.do?AwardNumber=0122272 [cited 31 Jan 2011]

[33] Sheddon, M. *SCI: Extended Terascale Facility: Early Operations* (Award number 0451566). National Science Foundation, 2005. Available from: http://nsf.gov/awardsearch/showAward.do?AwardNumber=0451566 [cited 31 Jan 2011]

[34] Berman, F., A. Kamrath, R. Moore and V. Samalam. *SCI: Delivering Cyberinfrastructure: From Vision to Reality* (Award number 0438741). National Science Foundation, 2005. Available from: http://nsf.gov/awardsearch/showAward.do?AwardNumber=0438741 [cited 31 Jan 2011]

[35] Moore, R. and M. Sheddon. *SCI: TeraGrid Resource Partners* (Award number 0503944). National Science Foundation, 2005. Available from: http://nsf.gov/awardsearch/showAward.do?AwardNumber=0503944 [cited 31 Jan 2011]

[36] Norman, M. and A. Snavely. *Flash Gordon: A Data Intensive Computer* (Award number 0910847). National Science Foundation, 2009. Available from: http://nsf.gov/awardsearch/showAward.do?AwardNumber=0910847 [cited 31 Jan 2011]

[37] Stewart, C.A., R. Bramley, B. Plale, C. Pilachowski, S. Simms and T. Hacker. *MRI: Acquisition of a High-Speed, High Capacity Storage System to Support Scientific Computing: The Data Capacitor* (Award number 0521433). National Science Foundation, 2005. Available from: http://nsf.gov/awardsearch/showAward.do?AwardNumber=0521433 [cited 31 Jan 2011]

[38] Barnett, W. and C.A. Stewart. *An InCommon Roadmap for NSF Cyberinfrastructure* (Award number 1040777). National Science Foundation, 2010. Available from: http://www.incommon.org/nsfroadmap.html [cited 31 Jan 2011]

[39] Stewart, C.A., J. Bottum and D. Jent. *Campus Bridging Taskforce Meeting* (Award number 1002526). National Science Foundation, 2009. Available from: http://nsf.gov/awardsearch/showAward.do?AwardNumber=1002526 [cited 31 Jan 2011]

[40] Dreher, P., C.A. Stewart, J. Pepin, G.T. Almes and M. Mundrane. *Campus Bridging Taskforce Sponsored Workshop: Campus Leadership Engagement in Building a Coherent Campus Cyberinfrastructure* (Award number 1059812). National Science Foundation, 2010. Available from: http://nsf.gov/awardsearch/showAward.do?AwardNumber=1059812 [cited 31 Jan 2011]

Appendix 1. XROADS collaboration staff participating in workshop

Nine nationally-recognized experts were invited to participate in the campus cyberinfrastructure and campus bridging REM; seven of the nine were able to participate. The participants at this meeting represented a campus-centric view of TeraGrid XD users: campus CIOs, deputy CIOs, experts in campus cyberinfrastructure, and experts in networking.

In addition to the invited participants in the REM, representatives of the XROADS collaboration present at this meeting were:

1. **Brian Davenport**, System Engineer, Raytheon, observer and note-taker.
2. **Jack E. Kleinert**, Senior Principal Systems Engineer, Raytheon Intelligence and Information Systems, Garland, TX, observer and note-taker.
3. **David Hart**, Assistant Director, UC San Diego/SDSC, San Diego, CA, observer.
4. **Susanne Jul**, President, Amaryllis Consulting, San Francisco, CA, meeting facilitator.
5. **Dale Lantrip**, Project Manager, Indiana University – Purdue University Indianapolis, Indianapolis, IN, meeting support.
6. **Richard Moore**, Acting Division Director, UC San Diego/SDSC, San Diego, CA, presenter and observer.
7. **D. Scott McCaulay**, Director, Research Technologies (Applications) Indiana University, Bloomington, IN, observer.
8. **Craig Stewart**, Executive Director, Indiana University Pervasive Technology Institute, and Associate Dean for Research Technologies, Indiana University, Bloomington, IN, meeting host.

Appendix 2: Participant Questionnaire Text

REM participants completed a questionnaire (via the Web) in advance of the meeting. The results of that questionnaire were used to create summaries used as one of several forms of input for discussion at the in-person meeting in Indianapolis. The questions were as follows:

Section 1: Introduction

Your Name

The institution(s) and/or organization(s) that you represent at this workshop

Please rate the following:

How familiar are you with the TeraGrid for the type of work that you do?

Not at all *Somewhat* *Fairly* *Very* *Don't Know*

How important is scientific computing to research on your campus?

Not at all *Somewhat* *Fairly* *Very* *Don't Know*

How important is scientific computing to education on your campus?

Not at all *Somewhat* *Fairly* *Very* *Don't Know*

How important is computational science to research on your campus?

Not at all *Somewhat* *Fairly* *Very* *Don't Know*

How important is computational science to education on your campus?

Not at all *Somewhat* *Fairly* *Very* *Don't Know*

Approximately how many of the researchers on your campus use the TeraGrid now?

Approximately how many of the researchers on your campus do you think have the technical training and skills to be able to use the TeraGrid?

Approximately how many of the researchers on your campus do you think could benefit from use of the TeraGrid as it is currently implemented (assume they are able to use it)?

Section 2: Current Engagement with CI/HPC/TeraGrid on Your Campus

Do you or your campus have existing HPC resources? (Please describe briefly.)

Are you or your campus currently supporting HPC/CI users? (Please describe briefly.)

Is the TeraGrid viewed as delivering services broadly useful to your campus, or only to a few particular researchers and are these researchers on your campus? What do you think are the reasons for those perceptions?

Are you supported by a professional support staff? How many staff members are there? How many CI/HPC users do they support?

How do you connect to the national CI? How does your campus CI connect to it?

Does your campus have a comprehensive plan for engaging with and using the TeraGrid, other grid projects (for example, OSG or SURAGrid), or other CI projects and resources? (If so, please describe briefly.) Is the intent to support research only, or research and education?

In what disciplines on your campuses (if at all) have you seen successful examples of TeraGrid/HPC/CI supporting your campus's education, training, or outreach (capacity-building) goals or tasks? Why do you consider them successful? What attributes or elements were most critical to their success?

Thinking of your campus as a whole: What have you found to be the most challenging aspects of using TeraGrid/HPC/CI for researchers and/or educators on your campus? How might TeraGrid XD lessen these challenges? (Depending upon the challenges, some possible suggestions might be: domain-specific education/training portals; a single educational access portal; comprehensive links to grid-based resources of all sizes and types (not just HPC systems); single sign-ons for a course; individual student short-term allocations; global file systems, data transfer tools, workflow tools, workshop support platforms, course management systems, scheduled network bandwidth, or any other solutions you envision.)

What do you think is the main problem in integrating campus CI to the national CI?

Section 3: Data and Interface Needs

What kinds of national-level data management requirements do your campus research and education activities require or anticipate? (Please indicate if these are current requirements or anticipated ones.)

Do you need or anticipate needing access to shared data? Do you move data, and from and to where? (Please indicate the approximate estimated average size of the data and if the transfers are between national, regional/system or local data storage resources.) How much data do you keep long term, and how do you decide what to keep?

For those data that you think your campus should maintain on national storage systems: Are there conditions placed on their maintenance in shared data archives, and if so, what kinds? (Considerations here could include: Are they subject to IRB conditions for collection, access, storage, and curation? How do you indicate the provenance, access, or curation requirements and restrictions of shared data? Do you collect any data associated with student and/or subject identities that must be anonymized before they can be used in a shared data environment? Do you collect photographic/video/image data that require permissions to post or share? Is the data costly in terms of time or other resources to reproduce?)

Do you have or anticipate a data challenge/crisis to your research or educational needs in the next 3-5 years? (Please describe briefly.)

Do you, your group, or your constituents find web portal interfaces such as the TeraGrid User Portal, Science Gateways, or the HPC University useful? If so, what features do you use most?

Are there features that should be improved or added to make them more effective for training, education, or outreach, or that would be helpful to involving more researchers or educators? (E.g., Google gadgets for job monitoring, course management systems, curriculum libraries, texting when a job is complete or all students complete an activity, user forums, collaboration tools.)

Section 4: Future Needs

What is most important to you and your campus with regards to connections between local computational environments (campus, organization, or research lab) and future TeraGrid XD cyberinfrastructure?

Where is the "last mile" (reaching to local resources) for researchers on your campus? Is there more than one "last mile"? What "last mile problems do researchers on your campus face in using the current TeraGrid?

What are the most challenging issues for people on your campus who are considering becoming TeraGrid users (e.g., time to learn the TeraGrid system, complexity of the TeraGrid system, lack of technical or professional computational scientific background, lack of information, lack of training or educational opportunities, not seeing the potential payoff for the science or engineering involved, the TeraGrid user interface, finding out where and how to get help when needed, etc.)? What could TeraGrid XD do to help with these issues?

What kinds of training do you think would benefit your campus or your group the most? How should training be delivered; that is, what are effective ways of delivering training and which would you consider to be most effective and why? What topics are most important in regard to training?

What sort of in-depth support for advanced users (and advanced uses) do you think would be valuable in aiding TeraGrid XD improve the efficiency of scientific discovery and in enabling discoveries that would not be possible without use of TeraGrid XD?

Do you anticipate new needs emerging as necessary for you to meet your campus's educational, training, or outreach goals in the next 3-7 years? For example, what would you need in order to incorporate emerging technologies such as cloud computing or instrument data, into your teaching?

What problems do you anticipate for the future of TeraGrid XD?

Section 5: Overall

What are the top three things missing from TeraGrid services that would benefit your campus's researchers as a whole?

What are the top three things missing from TeraGrid services that would benefit your campus's educational activities as a whole?

Do you have additional comments or suggestions you would like to make either about TeraGrid XD or the upcoming meeting?

Appendix 3. Statement of multiple interests by XROADS participants

The advanced cyberinfrastructure community in the US is relatively small, and it is difficult to assemble a team of experts without some of those experts having multiple interests in a given topic. None of the following statements of multiple interests rises to the level of an institutional conflict of interest. However, for clarity and transparency the following multiple interests are stated for the benefit of the reader:

- Everyone involved in the XROADS collaboration has an institutional interest in the outcome of NSF Solicitation 08-571, and all authors received funding directly for their own salaries or salaries of staff via NSF grant 0928542 to UCSD [29].

- Stewart and McCaulay work together at IU, and their groups have received funding via the TeraGrid and the Open Science Grid via NSF awards including [30, 31].

- Moore and his colleagues at SDSC have received funding via the TeraGrid via awards including [32-36].

- Stewart and Moore have both been involved in (and received funding for) the operation of wide area file systems used by their own institutions, the TeraGrid, and/or others, via the aforementioned awards to IU and UCSD, and also via award NSF 0521433 to Indiana University [37].

- Stewart is a member of the InCommon Board of Directors, although that status postdated his advocacy of InCommon as a standard for authentication of local and national cyberinfrastructure. (Indeed, the invitation to join the Board was based largely on Stewart's prior advocacy). Stewart has subsequently been co-PI on a grant award from the NSF to develop documentation related to InCommon [38].

- Stewart chairs the NSF ACCI Task Force on Campus Bridging, and in this context has received or been Co-PI on the following awards from the NSF: [39, 40].