5.0 Partnerships for Advanced Computational Infrastructure: Past and Future Roles

5.1 The Past and Present

We have described prior NSF sponsored investments that have collectively created a platform for major science-driven expeditions to develop and apply advanced cyperinfrastructure. The longest running and largest investment has been a series of initiatives to advance U.S. science and engineering by providing computational resources, including the Partnerships for Advanced Computational Infrastructure (PACI) program. Our charge specifically asks us to assess the effectiveness of the PACI program and to make recommendations about its future in the context of any new directions we propose.

Advanced computing programs began in the early 1980s, when the most powerful machines at that time —"supercomputers"— were not available to the entire U.S. scientific community. Hence the predominant need was for access to computing cycles at the highest end, and as a result five NSF Supercomputer Centers were founded in 1986 and 1987. The PACI program, established in 1997, was the next step. The goals of the two PACI partnerships (hereafter called "the PACIs") — the National Partnership for Advanced Computational Infrastructure (NPACI) and the National Computational Science Alliance (hereafter called "the Alliance") — were much broader than furnishing access to high-end compute power and the associated services. Their missions included provision of data storage and networking, education and outreach, and fostering of interdisciplinary research. At the center of each PACI partnership is a leading-edge site — the National Center for Supercomputing Applications (NCSA) for the Alliance, and the San Diego Supercomputer Center (SDSC) for NPACI. The PACI program is explicitly not allowed to support basic research.

Following the guidelines of the original PACI solicitation, the activities of the PACI partnerships have addressed multiple needs and served multiple purposes, some of which we highlight:

- During the five years of the current program, the two PACI partnerships have fulfilled their mission of providing high-end computing cycles. This conclusion is based on systematic, regularly conducted user surveys that are reported to NSF, and on the survey conducted as part of this panel's information-gathering process (Appendix B).

- The PACIs have supported, engendered, and supplied software tools to help users take advantage of architecturally diverse, increasingly complex, and distributed hardware. In addition to

joining and enhancing pre-existing software activities such as Globus[40] and Condor[41], the PACIs have initiated diverse projects involving all aspects of high-end computing. Two examples are the Access Grid[36], used at more than 100 sites worldwide, and the Cactus[42] programming framework, an open-source environment that enables parallel computation on different architectures along with collaborative code development.

- Through a joint Education, Outreach, and Training[32] activity, the PACIs have broadened access to computational science and engineering by encouraging the participation of women and under-represented groups at all educational levels.

- Many successes in domain science and engineering have been enabled as well as supported in part by PACI funding. In particular, some PACI-enabled collaborations among domain scientists and computer scientists have been exemplars of interdisciplinary interactions in which information technology becomes a creative, close partner with science. To name one among many, the recently funded National Virtual Observatory[10] which includes participants from the Alliance and NPACI, was described as a top priority in the 2001 U.S. National Academy of Sciences decadal survey of astronomy and astrophysics[43]. To a degree beyond anything anticipated even five years earlier, the National Virtual Observatory links astronomy with cyberinfrastructure in the forms of grid computing and federated access to massive data collections. The National Virtual Observatory concept grew from collaborations associated with the PACI program, and illustrates how advances in computer science and information technology can inspire new methodologies and directions in science, not just traditional science that is bigger and faster.

- International collaboration is an inherent part of computational science and engineering, and the PACIs are regularly involved with leading international consortia such as the Global Grid Forum.[22] Individual scientists supported in part by PACI are leaders in visible international projects such as GridLab[44], which involves Grid computing and numerical relativity.

The PACI partnerships have been reviewed annually by a program review panel convened by NSF. These reviews have been consistently positive with respect to the overall achievements of the Alliance and NPACI as defined by the criteria of the PACI program. However, not surprisingly for such a large and complex program, different aspects of the program have had different degrees of success. This is not meant as a criticism; it would be unrealistic to expect perfection in every element of the PACI program, which created new organizations with notable differences from the original supercomputer centers mentioned earlier.

Turning now to issues of concern, the PACI program has exhibited, from its beginning, a tension between two needs that cannot easily be reconciled: providing production systems for the current generation of high-end users, and moving to the next highest level of computing capability. Since the program's core funding has never been adequate to support more than one generation of computer system, tradeoffs have been inevitable.

In addition, the annual program review panels have expressed repeated concerns about the overall effectiveness and responsiveness of PACI activities in discipline-specific codes and infrastructure ("application technologies") and, to a lesser extent, generic software and infrastructure for high-end computing ("enabling technologies"). We discuss these concerns further below.

The PACIs have unquestionably had significant success and impact. Nonetheless, we believe that certain changes, described in the next section, should take place so that the PACIs, or their successors, become an integral part of the ACP proposed here.

5.2 Rationale for the Future

Part of the charge to the present panel was to evaluate the performance of the PACI program in meeting the needs of the scientific and engineering research communities. Given our broad definition of cyberinfrastructure we have interpreted this charge as an opportunity to consider potential roles for the PACI partnerships in a greatly expanded context. Since the Pittsburgh Supercomputing Center (PSC)[5] was selected by NSF in 2000 as the site for the Terascale Computing System[45], we include PSC as well as the PACIs in our discussion of the future.

The panel believes that today's science and engineering research continues to require computing resources at ever-higher levels and in ever-wider dimensions.

- The need remains, exactly as described in the 1995 Hayes Report[36], for the U.S. science and engineering research community to have access to machines that are substantially more powerful than those available at typical research universities, and for support services to enable those machines to be used most effectively.

- We anticipate increasing demand for advanced networking capabilities (including speed, bandwidth, quality of service, and security) for the indefinite future.

- The importance of data in science and engineering continues on a path of exponential growth; some even assert that the leading science driver of high-end computing will soon be data rather than

processing cycles. Thus it is crucial to provide major new resources for handling and understanding data; the National Virtual Observatory (briefly described in Section 5.1) emerged from recognition that the data avalanche in astronomy requires digital archives, metadata management tools, data discovery tools, and adaptable programming interfaces.

- Finally, sustained work is needed on software tools and infrastructure that enable general use of computing at the highest end, as well as on discipline-specific codes and infrastructure. It is universally agreed that producing and maintaining widely usable, reliable software is at least one, possibly several, orders of magnitude more difficult than generating an initial high-quality prototype.

As described in Section 2, the Panel is recommending a broad Advanced Cyberinfrastructure Program whose goal is to transform the conduct of science and engineering research, and which includes significant, sustained new funding for both discipline-specific and generic enabling infrastructure. Since the ultimate drivers of cyberinfrastructure are the needs of the scientific and engineering research communities, we believe strongly that those needs will be addressed most effectively by ensuring that enabling and application infrastructure projects associated with the ACP receive rigorous peer review. This is a fundamental change from the all-in-one structure of a PACI partnership, whose activities have been funded and reviewed as a unit. Our view is based on both philosophical and practical reasons.

Organizationally, this would be accomplished by creating new applications-focused programs within each interested NSF Directorate, as discussed in Section 4. These programs would also create any discipline-specific cyberinfrastructure required to support these applications, often based on extensions to the more generic cyberinfrastructure. Each of these programs would seek to create and execute a broad vision for revolutionizing research within their respective disciplines through the support of peer-reviewed projects. In many cases, we expect participation in these projects by the PACIs and other ACP-supported centers in partnership with disciplinary experts. The justification for this is the belief that disciplinary experts, in close partnership with computer scientists, are best able to judge the merits, impact, and importance of applications and specialized cyberinfrastructure focused on their field, and that these projects should be peer reviewed rather than initiated by the centers. In addition, reviewers who have substantial experience with software development, who take a broad view of high-end computing, and who will pay attention to opportunities for complementary activities and unnecessary duplication, should assess the quality of cyberinfrastructure projects.

The practical motivation for recommending separate peer review of application and enabling technology activities rests on the following observations, frequently made during the panel's information-gathering phase:

- The PACIs are not standalone, but partnerships involving many partners. Commitments have been made, explicitly or implicitly, to a number of partners, and these partners are represented in the PACI management structure. Thus it is difficult to phase out activities of existing partners or add new partners.

- There has been only limited review of enabling and application technology activities, particularly in assessing their impact on the relevant users and communities.

The peer review process that we envisage must always include consideration of the quality of each proposal's computer science and information technology aspects. To be specific, infrastructure projects in application areas need to be peer-reviewed by both domain and computer scientists, as are the current Information Technology Research (ITR)[46] proposals, to assess their quality based on criteria defined by the needs of cyberinfrastructure for the particular scientific community. In this regard, it is important that there should be no artificial distinction, as there was in the original PACI program, between research and development; the best enabling and application infrastructure projects, almost without exception, include both. Enabling and application infrastructure projects can be proposed by researchers and teams from any eligible institution or group of institutions, including, of course, the current PACI leading-edge sites and/or their partners. It is essential, however, that non-PACI teams to be given an opportunity to compete for this funding.

It is entirely consistent to believe, as the panel does, that the PACI program has had many successes, and at the same time to recommend a new structure for the future. We repeat our awareness of the outstanding results that have been achieved in both application and enabling technologies by PACI-supported efforts. In no sense are we advocating that such efforts be curtailed; in fact, our expectation is the opposite. Given the expertise developed at leading-edge PACI sites, proposals involving these groups should have a high success rate in peer-reviewed settings. Peer review of application and enabling infrastructure projects is therefore unlikely to be harmful to the best teams currently supported by PACI funding, while opening funding opportunities to a wider field.

5.3 The Future of the PACI program

To preserve the many accomplishments and talented personnel associated with the PACI program while the ACP is being defined, the panel recommends a two-year extension of the current PACI cooperative agreements. After those two years, until the end of the original ten-year lifetime of the PACI program, the panel believes that the two existing leading-edge sites (NCSA and SDSC) and PSC

should continue to be assured of stable, protected funding to provide the highest-end computing resources. In addition, the two PACI partnerships should continue their activities in education, training, and outreach. At the end of this period, there should be another competition for the roles of "leading-edge sites", possibly renamed, with (if appropriate) revised missions and structures.

Based on the assumption that sufficient new funding is in place, the new, separately peer-reviewed enabling and application infrastructure part of the ACP would begin in 2004 or 2005, after the two-year extension of the current cooperative agreements. New funding is absolutely essential to retain experienced PACI staff and to maintain already-established successful collaborations in enabling and application technologies. As observed in Section 4, *trained and knowledgeable people are the single most important component of cyberinfrastructure.*

With this timeline – a two-year extension of the current agreements and a major infusion of new funding in 2004 of 2005 for separately funded, peer-reviewed infrastructure projects – coupled with a partial disaggregation of functions through 2007, the panel believes that stability will be ensured for parts of the PACI program where it is most needed. Our further hope is that this schedule will reduce the energy and anxiety associated with submission of the annual program plan.