Reflections on the NSF Supercomputer Center Program
The Good, the Bad, and the Ugly
Larry Smarr, Founding Director of NCSA
December 2, 2009

In response to a request from Thomas Zacharia, in lieu of my attending the NSF HPC Workshop on December 4, I have written down some informal thoughts looking back over the 25 years of the NSF Supercomputer Center program. My proposal in 1983 was the first to NSF to propose what became the centers program, shortly followed by Sid Karin's. I was founding director of NCSA and held that position for 15 years. For the last decade I have built a new cross-disciplinary institute, the California Institute for Telecommunications and Information Technology, and have had an opportunity to work closely with SDSC on the UCSD campus.

While most of the technical work at the workshop will focus on the next generation of supercomputers and their applications, I believe there are some important lessons to be drawn on the institutional and cultural successes and failures of the last 25 years. I offer these reflections for your consideration as you think about how best to organize the NSF HPC program going forward. I have divided my thoughts into three sections: The Good, the Bad, and the Ugly. The Good are the accomplishments of the NSF SC centers, many of which were unexpected in 1985. The Bad are the cultural and institutional shortcomings of that program. The Ugly are the missed opportunities, largely caused by the Bad.

The Good

Increased the number of academic supercomputer users. It was estimated that before the 1985 launch of the NSF SC centers there were ~100 academic supercomputer users. After the first five years of the centers program a two orders of magnitude increase, as measured by those that logged onto one or another of the centers machines, was induced in the national academic HPC human resource pool. This vastly increased the scale of academic research using HPC and provided a pool for industry and the labs to hire from.

Stimulated use of HPC simulation in industry. Each of the centers recruited industrial partners and trained them on the use of HPC. NCSA developed an industrial partner program which attracted leading companies from over a dozen categories of the Fortune 500 classification. One notable example is Eli Lilly, which trained over 200 of their staff by total immersion sessions at NCSA, then became the first pharmaceutical company to purchase their own supercomputer (Cray-2), and within a year most major pharmas had followed and acquired HPC resources.

Brought an HPC Garden of Architectures to the community. In a short period of time the NSF centers working jointly with DARPA and NSF acquired almost all major HPC parallel architectures and made them available to the academic HPC community. This drove a rapid evolution of exploring new algorithms for key applications which were most efficient on the new hardware architectures.

Incubated the global Internet and Web. Although the Internet protocols were over a decade old when the NSF centers program began, the decision of the networking section of the Office of Advanced Scientific Computing to only support TCP/IP, led to the NSFnet backbone, buildout of the regionals, and extension to early adopter campuses. The NSF networking division, formed after CISE was created, continue to aggressively upgrade the NSFnet. The vBNS program brought high speed shared Internet to many campuses. These activities led directly to today's global Internet. NCSA Mosaic, developed only three years after Tim Berners-Lee created the WWW protocols, exponentially grew the nacent web community. Indeed in 1994 NCSA was the most hit web site on the planet and as a result we were forced to invent the first parallel web server. The NCSA Mosaic programmers left UIUC to form Netscape, Microsoft licensed Mosaic to form the basis of Internet Explorer, and Apache moved the Mosaic server code through open source to form the Apache server. Together this led to one of the largest NSF-induced transformations of the global economy in the history of NSF grants.

Drove Scientific Visualization. The need for visualization of the massive datasets generated by the NSF centers drove the development of computer graphics teams at a number of centers. The concept of data-driven scientific visualization quickly swept the academic community, but also had a major impact, largely through SIGGRAPH, on Hollywood and later the gaming community. For instance, Stefen Fangmeier, who was NCSA scientific visualization project manager in 1987, went on to spend over 15 years as a visual effects supervisor at Industrial Light and Magic, working on such films as Terminator 2, Jurassic Park, Dreamcatcher, Perfect Storm, and Master and Commander,

Pioneered Collaboration Technologies. Because HPC applications often involve teams with members spread across multiple institutions, the NSF SC centers were natural locations for the development of collaborative technologies. There was also a need for center consulting staff to collaboratively analyze complex data output and code with remote users. As a result, the NCSA Software Development Group developed one of the first cross-platform (Windows, Mac, Unix) synchronous desktop collaboration software systems, NCSA Collage, focused on collaborative data analysis in 1990. Five years later this was replaced by NCSA Habanero, one of the largest Java applications yet written at the time, which was automatically cross platform. Under the PACI NCSAlliance, ANL led development of the Access Grid, which enabled many remote sites to share real-time video conferencing over the Internet, becoming widely used around the world. In addition, high end experiments in novel collaboration technologies also were explored, such as linking CAVEs or PowerWalls so that avatars represented the location of remote collaborators in a shared data space. This foreshadowed the use of the OptIPuter to link scalable OptIPortals with HD video streams, which is becoming commonplace today.

The Bad

Lack of institutionalization of the centers. In spite of constant requests from the centers, NSF never institutionalized the centers program as it had NCAR, NRAO, NOAO, etc. Those centers are, respectively, where the nation computes atmospheric

sciences, observes with radio waves, and observes at optical wavelengths. The SC centers should be the sites where the academic community computes and where the staff support for things computational are housed. That is, select a few sites and give them the same multi-decadal guarantee of existence, with periodic reviews to maintain quality and user responsiveness. This would reduce a great deal of the endless rounds of existential worry and report writing which characterized the centers, at least during my 15 years as a director.

NSF induced a competitive culture between centers. A corollary of the above point is that the centers, by NSF design, were forced into a secretive and competitive posture relative to one another. Because one never knew when the next competition would come down from NSF, one hoarded any possible advantage to use in that next round. If the centers had been institutionalized they could relax and afford to be open and sharing. As one example of the disincentive to collaborate, it took me several years to convince the other centers to come together to form a joint national peer review board, because it undercut the ability of centers to recruit application "stars" and claim exclusivity with them. I believe the country would have seen the emergence of a national cyberinfrastructure during the PACI era if the centers had been institutionalized and incentives had been put in place for sharing and joint projects.

Narrowing Rather than Broadening Mission. One of the reasons that so many of The Good things happened was the flexibility that was inherently part of the original SC centers mandate. Yes, first and foremost the mission was centered on acquiring, installing, operating, and user consulting for HPC resources, but in addition there was funding opportunity to hire application domain experts, software tools developers, computer graphics and digital arts wizards, etc. In the PACI era this was broadened even more by the partnering with many other universities, national labs, and industrial partners. However, it seems to me that in the last decade the NSF has drastically narrowed the scope of the SC centers until finally the centers seem to be being dealt with as if they were contractors for installing and operating machines only. This had naturally led to a systematic "brain drain" away from the centers and a major lowering of their innovation opportunity space. I think it highly unlikely today that many of the successes of the first decade could occur in the centers as they are funded and reviewed currently.

The Ugly

Lack of balanced user-to-HPC architecture. From the beginning of the NSF centers program, a basic architectural concept was building a balanced end-to-end system connecting the end user with the HPC resource. Essentially, this was what drove the NSFnet build-out and the strong adoption of NCSA Telnet, allowing end users with Macs or PCs the ability to open up multiple windows on their PCs, including the

supercomputer and mass storage systems. Similarly, during the first five years of the PACI, both NPACI and the Alliance spent a lot of their software development and infrastructure developments on connecting the end-user to the HPC resources. But it seems that during the TeraGrid era, the end-users only have access to the TG resources over the shared Internet, with no local facilities for compute, storage, and visualization that scale up in proportion with the capability of the TG resources. This sets up an exponentially growing data isolation of the end users as the HPC resources get exponentially faster (thus exponentially increasing the size of data sets the end-user needs access to), while the shared Internet throughput grows slowly if at all.

NSF drops support for national networking. After 15 years of leadership in increasing Internet backbone speed and connectivity to campuses, NSF has essentially removed itself from supporting the needed growth in capability of the Internet for the increasing data-intensive requirements of the end-users of the TG resources, with the notable exception of the IRNC. This is in spite of the creation and growth of the National LambdaRail and more recently the Internet2 Dynamic Circuits, both of which provide clear channel IP fiber optic connections at 10,000 Mbps. Although the NSF did support several 10G connections BETWEEN the TG sites, the NSF has essentially withdrawn from the national backbone, regional, and local support for dedicated or on-demand large data pipes to the end-users of the TG. Imagine that NSF had only supported the Internet links between the five centers in the late 1980s and hadn't supported the build-out of the regionals and the access to the early adopting campuses!

No systemic cyberinfrastructure plan with centers having key role. In spite of 15 years of development of components of CI, there is still no NSF-wide layered CI defined and being used broadly. MREFCs are individually defining and building their own CI (NEES, OOI, NEON), as well as Division-level grants (eg: iPLANT). I have always believed that the NSF SC centers, as the original data-intensive generators would be in the ideal position to come together with the CS and applications communities (the intersection they have always worked at) to define a national CI system and support it for the major NSF opportunities. However, to have done this would have required The Bad not to have existed. Namely, defining and supporting an NSF-wide national CI would be natural if the SC centers had institutional stability and longevity, a collaborative rather than competitive culture, and a broadening rather than narrowing mandate. With the formation of an Office of CI, there is a chance to try and change all this, but without a robust and flexible set of NSF SC centers, there are no obvious sites to house the software engineers and consultants to support a national NSF CI program.

My hope is that these remarks can help inform the discussions of the NSF HPC Workshop. I am happy to engage with the process in the future if it would be helpful.