4.0 Achieving the Vision: Organizational Issues

One of the three parts of our charge is to *recommend an implementation plan to enact any changes anticipated in the recommendations for new areas of emphasis*. Our response has been to recommend a major Advanced Cyberinfrastructure Program (ACP) to create, provision, and apply advanced cyberinfrastructure to advance, and ultimately revolutionize, the conduct of scientific and engineering research and allied education. Success for this far-reaching ACP will require synergy among constituencies with varied expertise as well as incentives for participation. The goal of this section is to help NSF leaders create an organizational and leadership structure (some of which, because of its foundation-wide nature, are unusual to NSF) that effectively realizes the goals of the ACP. The Panel has given extensive attention to this part of our charge. We recommend a number of basic principles, processes, and incentives while avoiding being overly prescriptive as to the details so as to allow flexibility for NSF in its implementation of the ACP.

Two complementary activities are to be organized. The first is programs within NSF, which prescribe how resources are allocated to the various activities, evaluate proposals and make awards, and assess outcomes. These programs also represent and advocate for the ACP within the governmental and NSF budget process. The second involves the science and engineering community itself – the researchers, developers, and operational organizations that carry out the missions defined in the ACP. NSF can have significant influence on the organization of the community through setting priorities, defining programs, establishing evaluation criteria for proposals, and then evaluating proposals.

4.1 Elements of the Program

The key elements of the ACP are shown in Figure 4.1. The proximate outcome is new ways of conducting research through the application of information technology. The conduct of science and engineering research is built (in part) on these applications, which are tailored to the specific needs of people, groups, organizations, and communities conducting that research. Thus, the ACP directly funds activities resulting in the conceptualization, implementation, and use of such applications—it is not focused on cyberinfrastructure alone. Some applications are generic (such as distributed collaboration), and many others are discipline specific (like distributed community access to a specific scientific instrument).

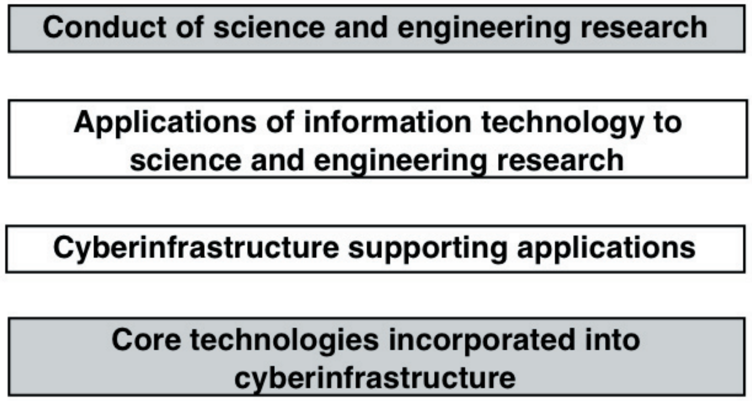

Figure 4.1. A layered architectural view of the ACP. The shaded boxes fall outside the scope of this report.

Applications are enabled and supported by the cyberinfrastructure, which incorporates a set of equipment, facilities, tools, software, and services. The ACP supports the creation and operation of advanced infrastructure tailored to specific domains, but it obviously does not include the core funding for the research (the top shaded box). Likewise, the ACP includes support for research on systems issues relevant to bringing together a heterogeneous mix of technologies (hardware, software, communications, storage, processing) to support advanced applications. Core technologies in the lower shaded box encompass the bulk of the current CISE research budget and should be preserved rather than reallocated to the ACP.

4.2 Technology Research and Technology Transfer

While the ACP is about revolutionizing the conduct of research, an equally important opportunity is to transform information technology itself. To illustrate this important aspect of the ACP, a second technology-transfer dimension is added in Figure 4.2. The three major phases of technology transfer (further elaborated and subdivided in Appendix C) are *applied research* (conceptualizing and bringing new application and infrastructure ideas to fruition), *development* (creating new technology artifacts ready for deployment), and *operations* (installing these software artifacts and enabling facilities and equipment, integration, keeping them running, and supporting end users). These phases are all relevant to both applications and cyberinfrastructure.

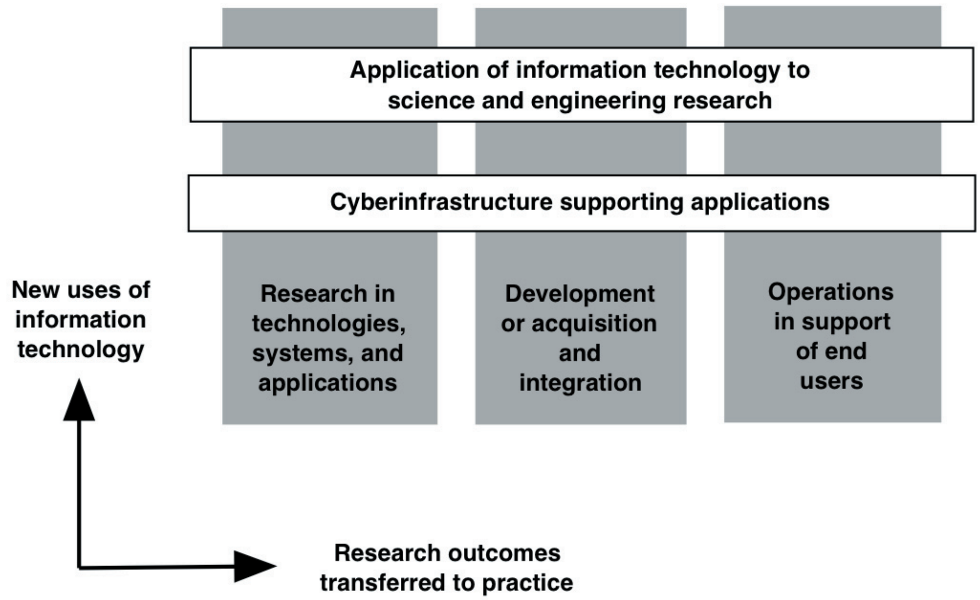

Figure 4.2. Technology transfer adds another dimension, where operations are supported by development, which is based on research outcomes.

4.3 Some Challenges

This ACP is ambitious, and as a starting point for considering its organization, we must recognize the most serious challenges to its success.

Only domain scientists and engineers can revolutionize their own fields. At its core the ACP involves rethinking the processes and methodologies underlying individual scientific and engineering fields. Domain scientific and engineering researchers must step up and enthusiastically create and pursue a vision.

Computer scientists (and allied technological fields, such as information science, and electrical engineering) must be involved. The substantial and ongoing involvement of information technology specialists is required to ensure that innovative new uses of technologies are identified, existing technologies are molded in new ways, and research into new technologies and new applications of technology is informed by opportunities and experiences in science and engineering research.

Taken together, these two issues present a serious challenge to any organizational structure. If the organization is weighted too heavily toward the domain scientists, the focus overemphasizes procurement of existing technologies, and computer scientists become viewed as “merely” consultants and implementers. If the weight shifts too heavily toward computer science, the needs of end users may not be sufficiently addressed, or effort shifts too heavily toward creating new technologies with insufficient attention to stability and user support.

Commonalities across science and engineering disciplines must be captured. Absent appropriate levels of coordination and sharing of facilities and expertise, there would be considerable duplication of effort, inefficiency, and excess costs.

Collaboration across science and engineering disciplines must be empowered and enabled, not impeded. Too often information technology becomes a source of Balkanization and an obstacle to collaboration or innovative change. The goal of the ACP is to make the cyberinfrastructure and applications an enabler (not an obstacle) to opportunistic and unanticipated forms of collaboration across disciplines, as well as encourage the natural formation of new disciplines. As in achieving commonalities, realizing this goal requires a largely collective effort.

Social scientists must work constructively with scientists and technologists. The social scientists can assist in understanding social and cultural issues underlying the direction of the ACP and, like technologists, can aid research in their own disciplines based on the experience gained.

4.4 Organization within NSF

The ACP will be retrofitted to an NSF organization whose primary mission, the conduct of science and engineering research and education, remains unchanged. It will be important and challenging to pursue major changes in the organization and processes underlying NSF’s primary missions to promote innovative application of information technologies, while avoiding significant organizational disruptions. Thus, we suggest that the organization of the ACP be overlaid in a matrix fashion on the existing organizational structures with the addition of a new coordinating ACP Office (ACPO).

As a starting point, the structure of Figure 4.1 is modified to align better with the research disciplines represented at NSF and becomes Figure 4.3.

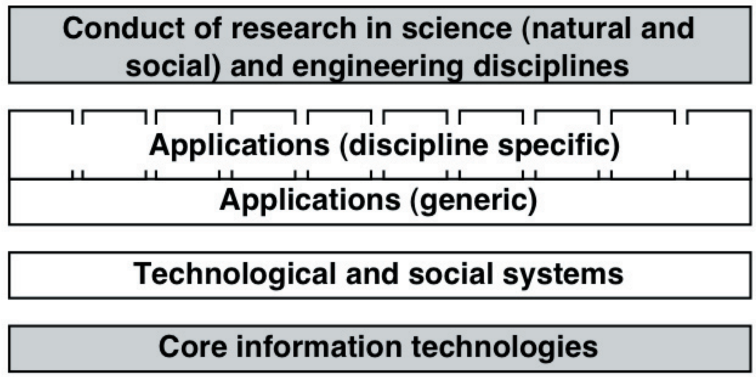

Figure 4.3. Relationship of the layers of Figure 4.1 to underlying disciplines. Applications are a hybrid case with shared responsibility between technological and disciplinary programs.

Cyberinfrastructure brings together many technologies (hardware, software, processing, storage, communication, etc.) to provide a coherent end-to-end functionality in support of applications; that is, at its heart cyberinfrastructure is a *technological system*. Many core technologies have themselves a system flavor, but we distinguish technological systems at the top level of hierarchy—where technology meets applications and uses—and observe that systems in this sense have special significance to both cyberinfrastructure and to applications. Figure 4.3 also emphasizes that, in the context of the fundamentally social enterprise of science and engineering research, technological systems as defined here and social systems (groups, organizations, and communities) are fundamentally intertwined.

Insofar as possible, applications should be generic, seeking to serve a variety of disciplines, but with sufficient flexibility, configurability, and extendibility to accommodate local variations and extensions. This contributes to both commonality (enabling future cross-discipline collaboration) and efficiency (through sharing of resources and expertise). On the other hand, there are clearly discipline-specific needs as well, with many organizational and process changes not readily transferred to other disciplines. A common cyberinfrastructure encourages commonalities and opens the door to future cross-disciplinary collaboration.

The organization within NSF should mirror the types of players (deliverers of research, development, and operations) illustrated in Figure 4.4.

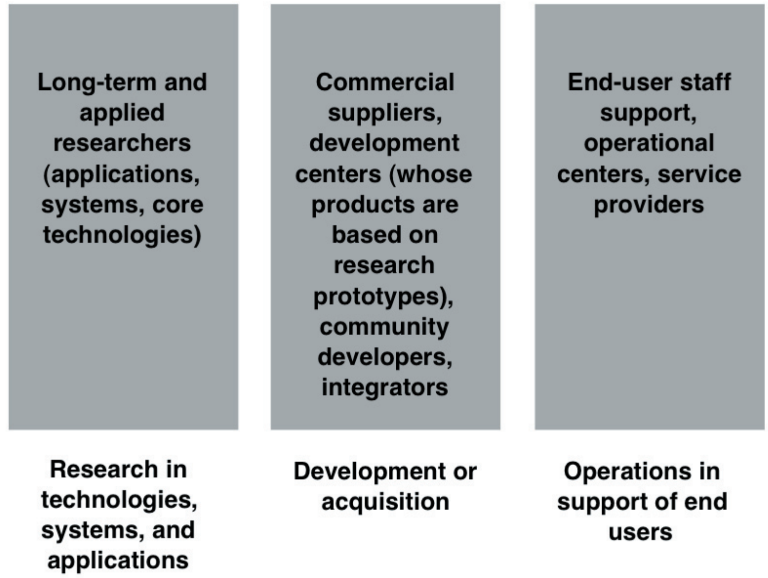

Figure 4.4. Summary of specific players delivering parts of the ACP

In terms of internal organization, our proposed division of responsibility is illustrated in Figure 4.5 (for applications) and Figure 4.6 (for cyberinfrastructure). We envision the initiative being served by a matrix management structure with the direct involvement of all of the NSF directorates. Some overriding principles (referencing Figures 4.5 and 4.6) can be stated.

Domain science and engineering directorates must take the lead in revolutionizing their respective fields through new research organization and processes, supported by new applications of information technology. We envision a program in each interested directorate (and we hope they will all be interested) that takes primary responsibility for formulating and implementing a vision, fostering buy-in and participation of its respective scientific or engineering research community, and creating a coherent program. Such efforts need to be open and oriented toward mutual coordination among directorates and should emphasize common standards and employ a common cyberinfrastructure.

CISE must be deeply involved both in serving as a technology leader for the overall initiative and in using scientific applications and experience of application users to inform its own technology research. CISE should be primarily responsible for both cyberinfrastructure and generic applications (much as it has managed the PACI program in the past) while also improving specific areas as outlined in Section 5. A primary goal of cyberinfrastructure is to capture the major technology requirements and provide tools to aid in application development, thus minimizing the need for technology-specific activities in other directorates. CISE will take responsibility for identifying commonalities among the needs of different disciplines. It should also lead the effort to define common infrastructure and standards that ensure that commonalties are captured and that future interdisciplinary collaboration is encouraged. CISE should be responsible for ensuring that the ACP is founded on a vibrant research agenda in technological systems and applications and that the research feeds the development of prototypes, production services, and commercially valuable end products. Finally, CISE should include and cooperate with SBE in conducting underlying research in the social aspects of both systems and applications.

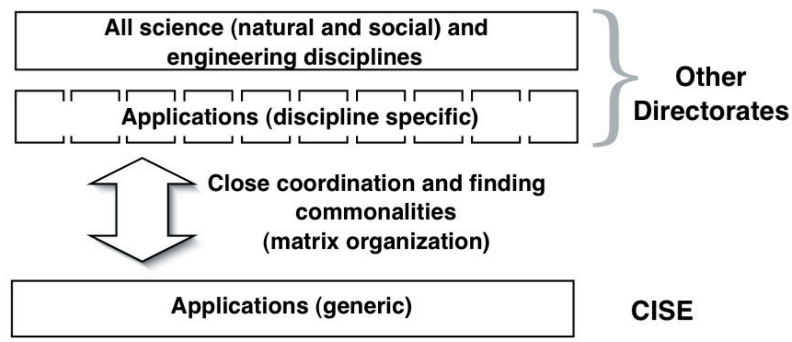

Figure 4.5. Assignment of responsibility for the vision and governance of applications to the NSF directorates.

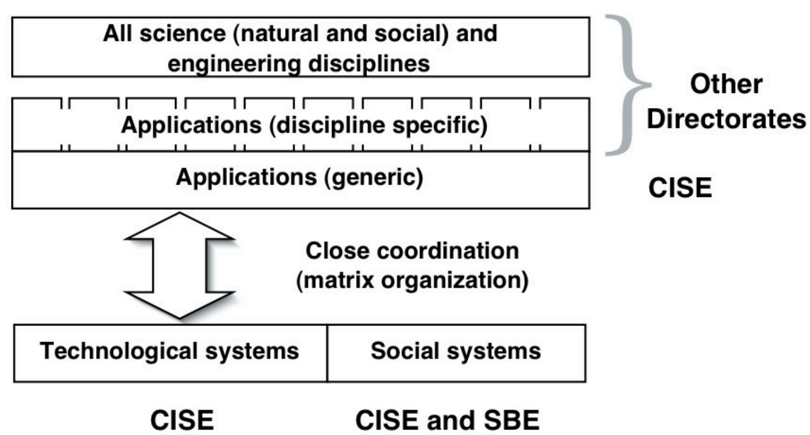

Figure 4.6. Assignment of responsibility for the vision and governance of cyberinfrastructure to NSF directorates.

To meet the challenges of achieving commonalities and collaboration, it is critical that the constituent programs within each directorate be viewed as parts of a foundation-wide initiative, while seeking to ensure that each respective community is served well.

Maintenance of sufficient coordination within the proposed matrix management structure will be formidable. **We therefore recommend that a single coordinating ACP Office (ACPO) be established to provide overall vision and guidance and exercise budgetary planning and responsibility.** (This office may or may not be an "Office" in the usual NSF meaning of the word. It could be administratively hosted in CISE or elsewhere, but it needs significant autonomy as described in this section.) The ACPO defines the overall vision of the ACP and represents and advocates this vision internally and externally to NSF. It develops budgets for the ACP, including the overall budget and sub-budgets for the various activities and the directorates. It serves as a central point of coordination among the complementary activities, including the identification and pursuit of commonalities and achieving uniformity and consistency where appropriate.

The directorates are the primary source of vision for their respective disciplines, and they formulate proposals to the ACPO for new programs and solicitations, insofar as appropriate in collaboration with (and as appropriate jointly with) other directorates. The ACPO evaluates the merit of those proposals consonant with the coordinated direction of the ACP, including assessing past efforts and seeking advice from the community. The ACPO then determines (or at least recommends to the Director of the Foundation) budgetary allocations to the various directorates based solely on the merit of their proposals and an evaluation of how these pieces fit together constructively in the overall coordinated activity of the ACP. The ACPO also represents the ACP in coordination with various other agencies and international bodies.

It is important that the ACPO view itself as the leader of a revolution in the conduct of research, and not primarily as an "information systems" or "information technology procurement" organization (a common organizational construct in government and industry).

The ACPO will not directly evaluate or fund projects in the community, this being the responsibility of the individual directorates. The reporting relationship of the ACPO should maintain budgetary independence from other programs in the directorates and place the ACPO in a position to strongly represent the budgetary needs of the ACPO within NSF and the government.

The leader of the ACPO is an especially important responsibility. **Its leader must have fundamental responsibility for achieving these goals, with sufficient credibility, power, and authority to succeed**.

This highly qualified person should be visible and highly placed, able to lead a large and complex matrixed operation with a substantial budget. Whether this leader is a discipline or computer scientist or engineer is secondary (a broad background and interests is ideal); most important is that he or she be deeply committed to successfully achieving the vision of a revolution in science and engineering research and be willing to explore and learn in the process.

The leader of the ACPO, although perhaps attached to an existing directorate, should be a functional peer with the assistant directors of the NSF directorates. A position at this high level is necessary to attract the right combination of visionary and manager, and to represent NSF as the leading U.S. agency in cyberinfrastructure when dealing with other federal agencies and international partners. An example of a structure that can be considered is as follows:

- The leader of the ACPO would report to an ACP Steering Committee consisting of the assistant directors of all involved directorates and chaired by the CISE AD (in recognition of the special role of CISE in the ACP).

- The Steering Committee would meet regularly with the leader of the ACPO and assume collective responsibility for the success of the initiative. The leader, working with the Steering Committee, would be delegated primary responsibility over a budget allocated to the ACPO.

- The ACPO leader would work with the Steering Committee in program generation, allocation of budget to directorates, awards, and oversight. The appropriate directorates working with their respective communities would carry out the details of this work.

- The leader of the ACPO would also be responsible for NSF liaison to other relevant programs in federal agencies and international bodies.

- The ACPO is intended to be the coordinator of an effective matrix organization, not a large organization duplicating or replacing the normal directorate activities. The ACPO would have a modest staff to help in budget and program development and performance reviews.

Appendix C includes more discussion of roles and organizational options.

4.5 Organization of the Community

Much of the work of the ACP will be carried out by individual research groups in the science and engineering research community, who will provide vision and experimentation and who will ultimately conduct research in new ways. The development and integration portion of the ACP, as well as operations of a common infrastructure and centralized user support, will be carried out by new or existing centers in the community funded by NSF. These centers are divided into several categories, including development centers, generic centers serving the science and engineering communities broadly, and disciplinary centers.

The ACP requires an organization for internal NSF coordination, as well as a central point of coordination in its external implementation. One or more development centers should be devoted to activities at the core of the initiative. These core activities include the planning, acquisition, integration, and support of the major software platforms and components at the foundation of the cyberinfrastructure. This includes choice of commercial software for underlying computing platforms and, where available and appropriate, for middleware and application components. These core activities will call for close coordination with industry, including the possible use of industrial products or prototypes as a basis of the ACP, and assistance with the transfer of successful technologies developed within the initiative into commercialization. Other core activities include the productizing of research prototypes, the development of new capabilities, and the integration of all these elements into a uniform software release with subsequent maintenance, support, and upgrade. In some cases, software may be maintained and upgraded by the community (e.g., open source), in which case the core activity includes governance of the process, such as choosing patches or upgrades to include in the releases. Development centers may be contracted industrial firms, existing laboratories, or new centers set up for this expressed purpose. The ongoing NSF Middleware Initiative provides (on a smaller scale) valuable experience and guidance in the organization of this portion of the ACP.

Generic centers focus on operations and user support for applications and infrastructure serving the broad research community, and discipline centers focus on applications and infrastructure more specialized and dedicated to particular disciplines, and include strong expertise in a discipline and its particular needs and challenges. Generic centers are needed to pursue broad commonalities, while disciplinary centers can accumulate disciplinary skills and thereby better meet specific disciplines needs.

There is no intention that these activities be strongly separated; development, generic, and disciplinary activities may be co-located or even grouped within common centers. One appealing organizational model, for example, is a development or generic center that maintains and integrates a collection of disciplinary groups.

Processes - As emphasized in Figure 4.2, several distinct activities each make essential contributions to the ACP. One such contribution is research—a traditional emphasis of the NSF—but there are others, broadly defined as development, operations, and use. These are decidedly not independent activities. Technology transfer seeks to benefit science and engineering research by employing the best ideas arising from research. But research agendas also should be influenced by the vision for the future conduct of science and engineering research. Similarly, there is a vertical flow of ideas and influence. Applications are influenced by emerging or anticipated capabilities in cyberinfrastructure, which are influenced in turn by advances in core technologies. And core technology research should be informed by anticipated cyberinfrastructure requirements, which in turn is influenced by capturing commonalities among application opportunities.

The research supporting applications in Figure 4.2 will increase the collaboration among computer scientists (and related disciplines, such as information science and electrical engineering) and domain scientific and engineering researchers (including the social sciences) to the benefit of all sides. Similarly the research supporting technological and social systems will increase the visibility of research into information technology systems in the broad sense, incorporating processing, storage, and communication into holistic social-technical systems solutions.

It is informative to examine the internal organization of the CISE directorate in light of these changing and magnified responsibilities. The vertical organizational structure of Figure 4.3 would focus attention most squarely on the greatest challenges mentioned earlier and highlight research into systems and applications. However, care should also be exercised that research efforts devoted to advancing core technologies receive continued high priority, as these efforts remain a critical underpinning of both the ACP and the nation's industry and economy.

Following the successful Internet experience and the more recent NSF middleware initiative, we expect that the development process leading to structure shown in Figure 4.2 will focus on the productizing and integration of a combination of commercially available software and research prototypes. The ACP must maintain a balance between deploying and gaining experience with emerging technologies, while providing users with a stable environment that is well documented and supported. The goal of development is thus to create and evolve a unified software distribution that is well maintained and supported. Of course, the development and operations are undertaken by experienced organizations funded by NSF, normally under cooperative agreements. The longer-term goal should be the commercialization of successful cyberinfrastructure and applications, with NSF continuing to fund development at the frontiers of noncommercially available solutions.

The operations stage will mix two models, as appropriate: a software distribution that can be installed, operated, and supported within the end-user organizational context, and software that is centrally operated to provide services over the network. NSF will fund organizations prepared to develop, maintain, and upgrade software distributions made available to end-user organizations and also organizations that operate cyberinfrastructure and/or applications provided as services invoked over the network. A proper and evolving balance should be maintained between professional staff supporting centralized operations and end-user operations, taking into account tradeoffs between the greater accountability and familiarity of local staff versus the efficiency and sharing of resources and expertise arising from centralization.

Incentives - The three primary activities identified in Figure 4.2 have very different metrics for evaluating proposals and outcomes.

Research is a competition of ideas. Allocation of resources starts with the program announcement and evaluation of the resulting proposals. This is bottom-up, stating the evaluation criteria with detailed initiatives arising from the research community. Overlap or duplication is acceptable where different researchers pursue competing visions for accomplishing similar ends. Post-evaluation is based on the intellectual quality and impact of the research outcomes.

Development is a competition of plans. An overriding goal of development is to limit duplication of effort, and concentrate resources on a set of integrated and maintained software distributions collectively covering the scope of the ACP. Thus, development is partitioned and assigned to organizations based on the responsiveness to needs and credibility of their plan for pre-defined concrete outcomes. Post-evaluation is based on how effectively the plan has been implemented and also on how extensively the outcomes are adopted and used and on user satisfaction.

Operations is a competition for users. Operations serve end-users, domain scientists, and engineering researchers, responsively providing service and support. There should be two or more competitive operational options available to users. A primary point of post-evaluation should be the satisfaction of the users who are served, and to a lesser extent the number of users who are served, based on input from the user community.

These distinct evaluation criteria should not suggest that these activities must be strongly separated organizationally; to the contrary, there may be advantages to grouping applied research, development, and operations (or some subset of these activities) within a common organization and geographic location.

Continuity - Human resources are critical to getting cyberinfrastructure and applications working, keeping them working, and providing user support. In the interest of funding more grants, NSF has arguably undersupported the recurring costs of permanent staff, preferring to focus resources on acquiring “hard” or “tangible” assets or the support direct research costs. In the ACP, human resources are the *primary* requirement in both development and operations, and success is clearly dependent on adequate funding both in centers and in end-user research groups.

Where possible, off-the-shelf commercial technologies and services should be acquired, but advanced and experimental capabilities will require NSF support of applied research, development, and operations. Success depends on specialized skills not readily available in the job market; rather, the most valuable staff will arrive with generalized programming and system administration skills and then learn valuable specialized skills through years on the job. A starting assumption in the funding of development and operations organizations should be continuity and long-term commitment. Absent significant problems and negative evaluations, funding initiatives in these areas should work from a base assumption of at least a ten-year lifetime for each participating organization. This is not to minimize the importance of ongoing evaluation and feedback, nor is it intended to preclude the redirection of funding from poorly performing organizations.