Appendix A: More about "What is Cyberinfrastructure?"

Cyberinfrastructure makes applications dramatically easier to develop and deploy, thus expanding the feasible scope of applications possible within budget and organizational constraints, and shifting the scientist's and engineer's effort away from information technology development and concentrating it on scientific and engineering research. Cyberinfrastructure also increases efficiency, quality, and reliability by capturing commonalities among application needs, and facilitates the efficient sharing of equipment and services.

Historically, infrastructure was viewed largely as raw resources like compute cycles or communication bandwidth. As illustrated by many activities in the current PACI centers and by the recent NSF middleware program, the scope of infrastructure is expanding dramatically beyond this narrow definition. For purposes of the ACP, infrastructure will comprise of a diverse set of technologies, facilities, and services and intangibles like design processes and best practices and shared knowledge. A major technological component is software that participates directly in applications and software tools that aid in the development and management of applications. A critical non-technological element is people and organizations that develop and maintain software, operate equipment and software as it is used, and directly assist end-users in the development and use of applications. The ACP seeks to bring about dramatic and beneficial change in the conduct of science and engineering research. Applications will greatly expand their role and become increasingly integral to the conduct of science and engineering research.

Cyberinfrastructure, as it captures commonalities of need across applications, incorporates more and more capabilities integral to the methodologies and processes of science and engineering research. Cyberinfrastructure will become as fundamental and important as an enabler for the enterprise as laboratories and instrumentation, as fundamental as classroom instruction, and as fundamental as the system of conferences and journals for dissemination of research outcomes. Through cyberinfrastructure we strongly influence the conduct of science and engineering research (and ultimately engineering development) in the coming decades.

Technologists are naturally the first to embed leading-edge technologies integrally with their research. The Internet—an inspirational example of this—was a new infrastructure defined initially with the narrow purpose of enabling new research in distributed systems, but which has now deeply impacted all research disciplines. The ACP seeks to replicate this type of dramatic change across a wide spectrum of disciplines and a wide spectrum of applications.

The ACP emphasizes infrastructure and applications. The overriding goals of infrastructure are several-fold:

More applications. Dramatically reduce the effort and the required expertise required to develop, deploy, and operate new distributed applications, encouraging more extensive development and use of such applications.

More capabilities. Provide facilities and supporting services that allow the community to do things not feasible otherwise.

More efficiency. Expand what can be accomplished for a fixed budget through sharing, reuse, and reduced duplication of both effort and facilities.

Reuse and multiple-use of designs. Infrastructure tries to capture commonalities across a range of applications. Functionalities and capabilities thus captured (manifested typically in a well-maintained software distribution) can be subsumed into many applications, reducing development time and effort and increasing quality.

Spread of best practices. Infrastructure can be a way to promulgate the best ideas, leveraging them in multiple places.

Achieving interoperability. Infrastructure offers a reference point for mediating the interaction among applications, defining common interfaces and information representations. The alternative of asking applications to interact directly with one another results in a combinatorial explosion of mutual dependencies, creating a house of cards that eventually falls of its own weight.

Tools. Infrastructure provides a set of software tools that make it easier to develop applications.

Services. Infrastructure provides, as an alternative to software that can be 'designed into' an application, services that can be invoked over the network by applications. When this approach is adopted, responsibility for the installation and administration of software and supporting equipment (and more generally provisioning and operations, as described below) is shifted to a service provider, where an aggregation of expertise and experience increases efficiency and effectiveness.

Shared facilities. Infrastructure allows the sharing of common facilities and equipment and instrumentation. This can be more efficient, due to statistical multiplexing[1] and as a way to reduce expensive duplication. Examples include sharing an expensive right-of-way for fiber optic cables or the sharing of a high-performance supercomputer with massive memory and input-output performance.

Assistance and expertise. Infrastructure provides direct assistance to end-users in making use of the available software, tools, and services, and does this efficiently and effectively through an experienced and shared pool of expertise.

Thus, it is critical to think of infrastructure as having several foundations:

Technological artifacts. These human-constructed artifacts include facilities (computers, mass storage, networks, etc.) and software. These artifacts sometimes provide services, and sometimes they are simply available to be 'designed into' applications.

Technological services. Various capabilities are provided as services available over the network rather than as software artifacts to be deployed and operated locally to the end-user.

Services from people and organizations. These include everybody who is providing a shared pool of expertise leveraged by the entire scientific and engineering research community to develop and operate the technological artifacts and provide advice and assistance to end users in making use of them.

[1] Statistical multiplexing is an efficiency advantage arising from sharing a set of work units in a single higher-performance facility rather than splitting them over multiple resources. Examples of resources are a communication link and a processor, where examples of work units are packet transmission or processing tasks. It is important to separate two performance factors: *throughput* (work per unit time) and *delay* (time elapsing from work request until that work is completed). For a fixed resource, average delay increases with average throughput (this is called *congestion*). The efficiency arises because for the same average delay, a larger shared resource can operate at higher average throughput. Alternatively, for the same average throughput, a shared resource can reduce the average delay. Thus, information processing and communication resources display increasing returns to scale even beyond any direct unit cost advantages from higher performance.

Appendix B: Analysis of Web Survey Results

A quantitative analysis of the Web survey is valuable to supplement the more qualitative conclusions presented in Sections 2 and 3.

User and Application Profiles

Comparisons with the Hayes Report, when available, are shown on the right. Where appropriate, each survey result is followed by a brief analysis and conclusions.

5. Principal categorization as a researcher:	Number of Responses	Response Ratio
University (regular or research) faculty	321	45%
University staff	82	12%
Post-Doctoral	79	11%
Graduate Student	99	14%
Undergraduate Student	9	1%
Federal laboratory	66	9%
Private sector	34	5%
Other	21	3%
Total	711	100%

6. Field of Research matched to NSF Divisions or Directorates: *(check all that apply)*	Number of Responses	Response Ratio	Hayes Report (1995)
Physics	190	26%	14%
Chemistry	119	16%	11%
Astronomy	65	9%	7%
Materials Research	64	9%	5%
Mathematical Sciences	67	9%	3%
Earth and Atmospheric Sciences	194	27%	Not included
CISE - Computer Science	136	19%	20%
ENG - Engineering	118	16%	19%
BIO - Biological Sciences	117	16%	6%
GEO - Geosciences	75	10%	4%
SBE - Social, Behavioral, and Economic Sciences	8	1%	1%
Other *(please specify)*:	51	7%	9%

7.	Do you presently use, or anticipate using in the future, digital libraries or federated data repositories?	Number of Responses	Response Ratio
I use them now		175	34%
No, but plan to do so in the future		117	23%
No, and have no plans to do so in the future		101	19%
Unsure or no opinion		127	24%
	Total	520	100%

11.	If your code(s) is/are platform-specific, which of the following platforms do you use?	Number of Responses	Response Ratio
Compaq Alpha/Linux		58	17%
Compaq Alpha/Tru64		75	22%
Cray/UNICOS		118	34%
IBM/AIX		111	32%
Intel/Linux		175	51%
Intel/MS Windows		61	18%
Intel/Solaris		19	6%
Hewlett-Packard/HP-UX		36	10%
PowerPC/Mac OS		29	8%
SGI/IRIX		159	46%
SGI/Linux		38	11%
Sun/Solaris		133	39%
Other *(please specify)*:		37	11%
	Total	718	100%

10.	Is/are your code(s) portable or platform-specific (e.g., must run on an Intel Linux system -- see next question for a sample list of machines)? If some are and some are not (e.g., principal application is portable but the visualization code is not), please so state in the comment line.	Number of Responses	Response Ratio
Portable		511	72%
Platform-specific		113	16%
Mixed *(please comment)*:		145	20%

12. Is/are your code(s) parallelized? If some are and some are not (e.g., principal application is parallelized but the visualization code is not), please so state in the comment line.	Number of Responses	Response Ratio
Yes	382	54%
No	160	23%
Mixed *(please comment)*:	196	28%

13. If your application is parallelized, which of the following tools are used? *(select all that apply)*	Number of Responses	Response Ratio
MPI	443	81%
OpenMP	136	25%
Shared memory parallelism via platform-specific directives	113	21%
Parallelism via language constructs (e.g., High-Performance Fortran)	66	12%
Automatic (i.e., compiler-based) parallelization	89	16%
Other *please specify*:	71	13%

14. Which of the following is/are the primary factor(s) limiting the execution speed of your application(s)? *(Select all that apply.)*	Number of Responses	Response Ratio
CPU speed	506	71%
Cache size	202	28%
Memory bandwidth	290	41%
Number of processors	317	45%
Internal communication network bandwidth (e.g., message passing)	257	36%
External network bandwidth (e.g., Internet2)	51	7%
I/O bandwidth	149	21%
Data archive bandwidth	71	10%
Other	31	4%
Don't know	65	9%

15. Which of the following is/are the primary factor(s) limiting the size or duration of your runs? *(Select all that apply.)*	Number of Responses	Response Ratio
Cache size	80	12%
Memory size per processor	227	33%
Total memory size	219	32%
Number of processors	300	43%
Disk space availability	143	21%
Queue configuration (e.g., short jobs that are delayed substantially because of large memory requirements)	166	24%
Other	70	10%
Don't know	76	11%

16. Do you generally perform the analysis of output generated by your application on a machine different from that on which the application runs?	Number of Responses	Response Ratio
Yes	531	74%
No	184	26%
Total	715	100%

17. Do you generally perform, or have a need to perform, the analysis of output generated by your application as the application is running?	Number of Responses	Response Ratio
Yes	312	44%
No	405	56%
Total	717	100%

18. Please express the resources used for analysis/mining/visualization of output generated by your application as a percentage of the resources required for the application itself.	Number of Responses	Response Ratio
0-25%	471	66%
25-50%	116	16%
50-75%	38	5%
75-100%	17	2%
100-150%	8	1%
150-200%	3	0%
Over 200%	9	1%
Don't know	47	7%
Total	709	100%

A total of 677 respondents specified the size of their data or output files that are analyzed/mined/visualized. Multi-gigabyte data sets was quite common and 100+ gigabyte sizes were rare. However, two respondents noted that their data sets range in size from between 1-10 terabytes.

2. Have you (or your collaborators or students) used any PACI Center during the past year?	Number of Responses	Response Ratio	Hayes Report (1995)
Yes	389	71%	75%
No	159	29%	25%
Total	548	100%	

3. Which of the following facilities have you used during the past year (check all that apply)?	Number of Responses	Response Ratio	Hayes Report (1995)
NCSA - National Center for Supercomputing Applications	175	34%	
PSC - Pittsburgh Supercomputing Center	117	23%	
SDSC - San Diego Supercomputer Center	187	36%	
NCAR - National Center for Atmospheric Research	115	22%	
Other Federal (NASA, DoE...) Supercomputer Center(s)	118	23%	31%
State or regional Supercomputing Center	36	7%	10%
My University's Supercomputing System or Facility	139	27%	26%
My own, or my department's, supercomputing facility	191	37%	22%
VIEW Other (please specify):	65	13%	

4. To what degree has the use of national supercomputing centers (NSF or otherwise) influenced your research? In addition to hardware, storage, and software, be sure to consider issues such as consulting and scientific collaboration.	Number of Responses	Response Ratio	Hayes Report (1995)
No Influence	70	13%	8%
Moderate Influence	151	27%	15%
Significant Influence	126	23%	29%
Essential Influence	203	37%	48%
Total	550	100%	

The above statistics indicate an increasingly bi-modal structure in the use of high-performance computing resources: large supercomputing centers and departmental or research-group facilities. Furthermore, the response to question 4 indicates a substantial increase in the impact or importance of national supercomputing centers on research. This and other information suggests that users desire increased investments in high-end computing as well as local facilities to facilitate their usage. Increasing emphasis is being placed on the use of the latter, especially in light of affordable technology (e.g., 160 gigabytes of disk space available for $300, which means that affordable terabyte storage capability already is available on the desktop).

6.	How would you rate the PACI Centers in providing the resources you need (this includes cycles, tools, consulting, etc)?	Number of Responses	Response Ratio
Poor 1.		24	5%
Fair 2.		47	10%
Good 3.		142	31%
Very Good 4.		163	36%
Excellent 5.		75	17%
	Total	451	100%

Comments regarding the above noted that PACI staff were helpful, with problems centering around an inadequate number of cycles, long queues, poor turn around, overcrowding, the need for more memory, and the increasing difficulty of actually using parallel machines (compared to the autoparallelizing compilers on, for example, Crays) and obtaining performance that represents a reasonable fraction of theoretical machine peak.

8.	What specific limitations, if any, have impeded your progress at the PACI Centers?	Number of Responses	Response Ratio
Resources oversubscribed in general		186	52%
Poorly-designed job queueing		76	21%
Inadequate consulting support		44	12%
Poor management and administration		23	6%
Cumbersome policies		57	16%
VIEW Other *(please elaborate)*:		142	40%

9.	Are the strategies for resource allocation at the PACI Centers appropriate, fair, and effective?	Number of Responses	Response Ratio
Yes		236	49%
No		30	6%
Unsure or no opinion		216	45%
	Total	482	100%

With regard to allocations, PACI needs more cycles, users noted difficulty in obtaining sufficient time as well as large number of processors (e.g., 64-128 processors for several days). A preponderance of respondents also noted the lack of multi-year grants of time as a major limitation to research grants that cover multiple years.

11.	Does an appropriate balance exist at the PACI Centers with regard to hardware, software, tools, and personnel?	Number of Responses	Response Ratio
Yes		194	40%
No		44	9%
Unsure or no opinion		243	51%
	Total	481	100%

Most of those providing written responses to Question 11 noted that too much emphasis is placed on raw hardware performance compared to tools and especially personnel. Specifically, users commented that many tools never proceed beyond the experimental stage to full deployment, and that although existing personnel are excellent, they are spread far too thin compared to the sophistication of the hardware and software environments they are tasked with supporting. Overwhelmingly, users support significant increases in support personnel. Further, users note that investments in new directions (e.g., Grid technologies) appear to be slow in yielding tangible benefits to the broader community.

13.	If you are new or relatively new to high performance computing, are you satisfied with the mechanisms available at the PACI Centers to bring new users on board?	Number of Responses	Response Ratio
Yes		172	44%
No		31	8%
Unsure or no opinion		188	48%
	Total	391	100%

Only 84 written comments were received out of the 391 individuals who responded to Question 13. Most already had been working with high performance computing, and several noted that the switch from vector-based machines to other paradigms had been difficult. For those who clearly were new to high performance computing, the experience in using the PACI centers was judged to be positive.

15.	Do you feel that practices/infrastructures/modalities used in other countries could be effective if adopted by the US?	Number of Responses	Response Ratio
Yes		49	10%
No		45	10%
Unsure or no opinion		378	80%
	Total	472	100%

For those users who expressed familiarity with infrastructures in other countries, the overwhelming sentiment expressed in written responses to Question 15 is that the United States is leading the way in the provision of hardware, network connectivity, supporting software, and collaboration tools. Further, the PACI infrastructure clearly is unique (other countries are adopting it) and thus has a significant impact on science and engineering progress. Several respondents expressed regret that the US does not have access to Japanese supercomputers

Disciplinary Impacts

There exists universal sentiment in the community that significant discovery has been enabled by the PACI centers, and that many, even more significant discoveries will be possible in the future. A good portion of these are anticipated to occur at the intersection of disciplines as well as in the context of societal implications, and made possible by Grid and related capabilities. Multidisciplinary teams will continue to proliferate, and efforts must be made to support them. Likewise, respondents noted that the proliferation of powerful, affordable desktop and departmental or research-group computers have had a dramatic impact on the ability to perform exploratory research, analyze data, and extend research in new directions. A sample of individual responses to a question regarding disciplinary impacts to date is given below. In many cases, respondents noted that the impacts were too numerous to mention.

- Simulation of newly discovered planets.
- Discovery and explanation of critical phenomena in gravitational collapse.
- First molecular dynamics simulations to show the sequence of actions of a polymerase and its role in the DNA repair process.
- Complete simulations of solid rocket booster firings.
- Demonstration of the practical predictability of individual thunderstorms and their related weather.
- Simulation of the large-scale structure of the universe.
- Fully coupled climate model simulations showing the agreement between observed and predicted increases in the global mean temperature.
- Largest simulations to date of Einstein's equations of general relativity.
- Simulation of fully three-dimensional coupled flow and heat transfer in a turbine engine.

A sample of individual responses to a question regarding anticipated discoveries for the future is given below.

- Mining of massive data sets across disciplines (e.g., weather and population).
- Upscaling fine resolution ecosystem models to broader scales.

• Determining the nature of dark energy.
• The generation of gravitational waveforms from numerical simulations of colliding black holes and neutron stars.
• Simulation of the complete life cycle of a tornado.
• A complete representation of the coupled magnetosophere-ionosphere-thermosphere system.
• Prediction of code performance on high-end computers.
• Use of the Grid to solve massive large practical problems.
• Understanding of protein and DNA folding and unfolding.
• First principle predictions of structures for protein domains.
• Large modeling of interactions among regions of the human brain.
• Tools for data analysis and knowledge extraction.
• Fully 3-D imaging of small-scale structures deep within Earth interior.
• Major advances in the treatment of subgrid scale turbulent processes in large eddy simulations.
• Extremely long-term and more realistic simulations of the entire coupled climate system.
• New insights into chaotic systems and properties of fluid turbulence.
• Complete simulations of the Earth-Sun system.
• Greatly improved tropical cyclone predictions.

A long-standing question regarding the provision of resources for the community is whether centers should serve primarily one or multiple disciplines. As shown below in Question 19, approximately 1/4 of all respondents believe that a PACI center or program organized around a specific discipline would be of greater value to them as a user compared to the present multidisciplinary organization of the centers. However, those who responded in the affirmative also noted that, despite possible advantages, the necessary division of resources to create such centers would lead to an overall reduction of quality and capability. Further, such centers would tend to maintain historical boundaries between traditional disciplines, which is incongruent with the future of science and engineering research and education. In the context of the current PACI framework, however, respondents expressed a clear desire for greater depth of consulting expertise within specific disciplines. They further noted that PACI centers should be able to dedicate significant resources to large disciplinary projects. defined periods of time.

19. Would a PACI Centers Program organized around disciplines (i.e.; a center for Physics, another for Chemistry, etc.) be of greater value to you as a user?	Number of Responses	Response Ratio
Yes	118	24%
No	206	42%
Unsure or no opinion	171	35%
Total	495	100%

Considerable emphasis has been given by the PACI program to facilitating interactions among disciplines. Question 21 shows that users have widely differing views regarding the effectiveness and even the appropriateness of such a role. Based upon written responses, most users view the establishment of interdisciplinary collaborations as the responsibility of individual scientists, and many don't identify the PACI centers as the first point of reference for linking with other disciplines. The greatest value of the centers as a "melting pot" of disciplines appears to be the linking of domain scientists with computer scientists.

21. Have the PACI Centers been effective in facilitating your interactions with researchers from other disciplines?	Number of Responses	Response Ratio
Yes	123	26%
No	155	33%
Unsure or no opinion	195	41%
Total	473	100%

Anticipated Use of Emerging Capabilities

This portion of the survey sought information regarding future use of emerging technologies such as the Grid, federated data depositories, and digital libraries. It also sought input about special needs, such as real time and on-demand availability of resources.

2. Do you plan to run your application(s) across long-haul networks in a distributed fashion using the Computational Grid?	Number of Responses	Response Ratio
Yes	116	22%
No	258	49%
Unsure or no opinion	153	29%
Total	527	100%

The response to Question 2 above indicates, the written responses confirm, that the community as a whole is not aware of the Grid or concepts related to it and distributed Web services. Many noted that they lack expertise to modify their codes for execution across distributed resources, and that poor performance on existing parallel platforms might suggest equally poor performance across the grid. Further, numerous respondents appeared skeptical of the practicability of distributed methodologies, at least within the present management and facilities environments.

4. Do you presently use collaboratory or knowledge networking (e.g., the Access Grid)?	Number of Responses	Response Ratio
Yes	61	12%
No	461	88%
Total	522	100%

Likewise for Question 4, most respondents expressed lack of understanding about collaboratories and knowledge networks, though many noted that they presently are using, or soon plan to be using, the Access Grid to facilitate remote collaboration. Several commented that the Access Grid needs to become more reliable and cost effective to be practicable for community-wide use. Interestingly, when asked of their requirements for networked collaborations, most respondents said they had none.

7. Do you presently use, or anticipate using in the future, digital libraries or federated data repositories?	Number of Responses	Response Ratio
I use them now	175	34%
No, but plan to do so in the future	117	23%
No, and have no plans to do so in the future	101	19%
Unsure or no opinion	127	24%
Total	520	100%

9. If you use digital libraries or federated data repositories, what are the primary impediments, if any?	Number of Responses	Response Ratio
Slow network speeds	118	55%
Ineffective data cataloging and/or difficulty locating required data	107	50%
Disparate formats among data types	112	52%
Inadequate documentation	80	37%
Inadequate or missing information on quality control	56	26%
Inadequate analysis tools	48	22%
Inadequate local storage for data analysis	34	16%
VIEW Other *(please elaborate)*:	40	19%

Questions 7 and 9, and their associated written responses, indicate a potentially significant increase in the need for digital libraries and federated data repositories. Several noted the lack of easy accessibility to historical data holdings, and the difficulty of dealing with multiple formats and data characteristics. However, most who provided written responses feel that data repositories are among the most important and challenging aspects of high performance computing and should receive considerable attention in the future.

10. Do you conduct research that requires running codes and analyzing their output in real time (e.g., weather prediction software)?	Number of Responses	Response Ratio
Yes	124	24%
No	403	76%
Total	527	100%

12. Do you conduct research that requires management or control of remote devices or instruments?	Number of Responses	Response Ratio
Yes	52	10%
No	470	90%
Total	522	100%

14. Do you conduct research that requires real time data acquisition and/or cataloging?	Number of Responses	Response Ratio
Yes	113	22%
No	408	78%
Total	521	100%

16. Is network quality of service an important issue in your research?	Number of Responses	Response Ratio
Yes	327	63%
No	97	19%
Unsure or no opinion	93	18%
Total	517	100%

Questions 10, 12, 14 and 16 above dealt with timeliness and related quality of service issues. A remarkable 24% of respondents noted that they conduct research that requires real time analysis of results, i.e., analysis that must be conducted as soon as the results are available, with the topic areas ranging from weather prediction (dominant response) to visualization and nano-materials research. A similar response was found for real time data acquisition and cataloging, while a smaller percentage of respondents noted the need for remotely controlling instruments. Changes in these percentages are difficult to anticipate, though the written comments suggest that several who do not require such capabilities now most likely will within the next 5 years.

A clear majority of respondents noted that network quality of service is important in their research, mostly in the context of speed, reliability, and security, generally in that order. We were surprised that nearly one fifth of those responding had no opinion or were unsure.

18. Would your research benefit from advanced visualization technology (e.g., immersive or virtual reality systems)?	Number of Responses	Response Ratio
Yes	269	52%
No	121	23%
Unsure or no opinion	128	25%
Total	518	100%

Slightly more than half of those responding to Question 18 indicated the need for advanced visualization technology. Many noted that existing systems are not suited to their needs, are too slow, and are too expensive and not practical (e.g., cave, power wall). Several indicated the need to visualize in dimensions greater than 4, and that visualization tools lag significantly both hardware and scientific application codes. In that context, it was noted that advanced visualization technologies are slow to move from the prototype phase (e.g., demonstrations) to practical implementation for use by the broader community. Finally, most of those providing written responses noted that visualization is a key component of their research methodology.

Other Future Needs

20. Do you expect that your area of research will require access to supercomputing resources in the future?	Number of Responses	Response Ratio
Yes	479	91%
No	17	3%
Unsure or no opinion	28	5%
Total	524	100%

The need for continued access to high end resources is underscored by the remarkable response to Question 20 above (93% responded in the affirmative in the Hayes report). Numerous written responses contained the phrases "stretching the limits," "supercomputing is essential to my research," and "I see no end to the need." Many also noted that their most significant discoveries have been facilitated by use of the most power computing resources available (e.g., weather prediction, turbulence research, materials research, chemistry, bioinfomatics).

22.	During the next several years, what fraction of your computing needs do you think could met by:					
Percentage indicates total respondent ratio and parenthesis indicate actual number.	1 0-20%	2 21-40%	3 41-60%	4 61-80%	5 81-100%	6 Do not know
1. Desk top workstations	40% (210)	24% (125)	14% (73)	9% (48)	9% (45)	2% (10)
2. Mid Range Systems (Department/University owned with a market value from $100K to $2M)	25% (133)	28% (147)	20% (105)	10% (54)	6% (33)	3% (14)
3. Highest Performance Systems (National)	25% (133)	18% (94)	16% (85)	15% (80)	18% (94)	4% (22)

23.	For the "Highest Performance Systems", how do you rank your projected needs in the following areas:				
Percentage indicates total respondent ratio and parenthesis indicate actual number.	1 No Opinion	2 Unimportant	3 Important	4 Very Important	5 Critical
1. Processing speed	3% (15)	1% (7)	17% (86)	32% (167)	46% (238)
2. Large Memory size	3% (13)	7% (34)	21% (110)	33% (172)	35% (180)
3. Large Mass Storage availability	3% (18)	12% (64)	26% (135)	30% (155)	26% (132)
4. Networking Bandwidth for remote access	5% (24)	15% (76)	37% (191)	28% (142)	13% (67)
5. Large I/O Bandwidth (to local peripherals)	9% (45)	20% (102)	34% (173)	20% (104)	14% (73)
6. Network Bandwidth for inter-node communication	9% (45)	11% (56)	18% (94)	25% (128)	34% (174)
7. System availability and reliability	2% (8)	2% (12)	18% (91)	34% (177)	43% (221)
8. System security	5% (27)	23% (121)	35% (180)	19% (96)	15% (78)
9. Education and training	8% (42)	23% (118)	42% (215)	17% (88)	9% (44)
10. User support tools	6% (31)	16% (80)	41% (212)	24% (126)	11% (58)
11. Optimization	7% (34)	10% (54)	34% (173)	31% (162)	16% (82)
12. Consulting	6% (29)	17% (88)	42% (216)	22% (112)	12% (62)
13. Dedicated resources	13% (65)	25% (130)	33% (168)	17% (88)	10% (52)
14. Visualization	6% (33)	22% (111)	36% (185)	22% (114)	12% (63)
15. Data services	14% (73)	27% (138)	35% (182)	14% (70)	7% (38)
16. Code migration	13% (65)	24% (123)	34% (176)	18% (95)	8% (43)

The responses to Questions 22 and 23 above largely mimic those presented in the Hayes report, though with a general shift at the present time toward dependence upon personal workstations and departmental systems. Anticipated use of high performance national systems was found to be nearly identical to that in the Hayes report, and the percentages in Question 23 tended to shift slightly overall toward greater importance, with new categories (e.g., dedicated resources) clearly viewed as important.

24. Rate below your view of the importance of national centers in providing the listed services (located at the centers) in relation to your research or educational projects:					
Percentage indicates total respondent ratio and parenthesis indicate actual number.	1 No Opinion	2 Unimportant	3 Important	4 Very Important	5 Critical
1. Diversity of computer architectures	11% (58)	26% (133)	31% (159)	21% (109)	10% (51)
2. Visualization facilities and software (including Virtual Environments)	10% (50)	28% (146)	34% (176)	20% (104)	6% (33)
3. Information processing	16% (80)	29% (148)	33% (170)	15% (79)	5% (24)
4. Consulting services	8% (39)	17% (87)	43% (221)	21% (107)	11% (58)
5. Third party or commercial software	11% (55)	31% (160)	34% (177)	15% (77)	7% (37)
6. Expertise in your research topic/Research Team Formation	11% (55)	37% (189)	28% (145)	17% (85)	6% (32)
7. Assistance in Code Porting and Optimization	10% (51)	18% (95)	37% (189)	21% (109)	13% (65)
8. Training	11% (55)	19% (98)	39% (203)	21% (108)	8% (41)
9. Communication software	15% (78)	22% (112)	38% (197)	16% (83)	5% (27)
10. Data Services (massive databases)	12% (64)	25% (130)	26% (132)	19% (100)	15% (75)
11. Repository for Data and Data Dictionaries	12% (64)	24% (123)	27% (138)	18% (94)	16% (80)
12. Internet software development	15% (79)	36% (183)	29% (150)	13% (66)	5% (24)

The responses to Question 24 also largely mimic those presented in the Hayes report, though again with a general shift at the present time toward greater importance of all items.

25. What specific modalities do you feel are most appropriate for the future (check all that apply)?	Number of Responses	Response Ratio
Large essentially autonomous national centers	156	31%
Large coordinated national centers	290	58%
Interconnected collaborative regional centers	203	41%
Local centers	192	39%
Discipline-specific facilities	120	24%
Data-only facilities	47	9%
VIEW Other *please specify*:	29	6%

Finally, the responses to Question 25 indicate an overall sentiment toward non-discipline specific, interconnected collaborative centers and alliances.

Open-Ended Comments

Suggestions for meeting the needs of high-performance computing:

- Smoother transition of scales from local to regional to national
- Better coordination between NSF, DOE and NIH
- Need more powerful local/regional machines
- NSF grants should be linked to CPU allocations
- Integrate climate modeling with observations
- "PACI should have the middle ground between pure production and computer science research centers enabling collaborative research in high performance computing"
- Enhance network speed/bandwidth for data archives
- More high bandwidth links to facilities beyond national centers
- Concern/worry about the lag of academic computing versus DOE resources
- "PACI needs ambitions plans to ensure that the next generation of students are trained on state-of the art machines"

General Comments on the PACI program

- "PACI is very important"
- "Need to maintain the diversity of PACI centers"
- "Flexible powerful national centers are very useful"
- "PACI provides high end machines for the high end user"
- Current machines and strategy are not supporting high end science and engineering applications
- "Need to streamline allocations"
- "Terrible turnaround"
- "Turnaround too slow"
- "PACI needs capability, not just capacity"
- "Improve Queues"
- "Queues too long, machines overloaded"
- "Yearly grant applications a burden"
- PACI needs to maintain data archives, historical data and make them available"
- PACI needs to integrate "digital libraries data collections and persistent archives so as not to lose knowledge"
- "need data access accounts"
- "PACI needs to support a diversity of high-end users"
- "PACI needs many processors and multi-terabyte memory"
- "Too few users of 1,000+ processors"
- "Need large numbers of nodes with reasonable latency"
- "PACI should encourage usage of large processor sets"

Appendix C: More on Organizational Issues

This appendix provides some additional detail surrounding the organization of the ACP within NSF.

Organizational Alternatives

Alternative organizational structures can be made to work given the appropriate level of cooperation across organizational boundaries. Given sometimes competing objectives, no single organizational approach can capture seamlessly every aspect of what is to be accomplished. Having said this, the organizational approach can be a useful tool for emphasizing and promulgating the most fundamental, cherished, and/or difficult-to-achieve goals, and communicating to everyone involved (NSF and the research community it serves) the goals and their priorities. The greatest challenges were discussed in the body of the report, and our organizational recommendations emphasize successfully addressing these challenges with minimum disruption of the existing NSF organization.

There are many ways the ACP could be organized, and it is useful to list some alternatives considered and their perceived shortcomings:

Overlay the ACP on the current organization. The Panel believes that the INITIATIVE, to be fully successful, must be an agency mission with the highest priority and the highest visibility, and the organization of the ACP within NSF should reflect this. In our opinion, the ambitious goals of the INITIATIVE cannot be achieved by business as usual, but neither does it demand radical changes.

Making fine-tuning changes. For example, a fine tuning might consist of simply combining ACIR and ANIR into a single division within CISE with responsibility for all infrastructure. Again, the Panel believes this doesn't place adequate importance and visibility on ACP to be successful.

Centralize the ACP in a single organization. The Panel considered the option of creating a separate organization within the Office of the Director to centrally manage an infrastructure program on behalf of all directorates, similar to the Office of Polar Programs. This has the disadvantages that it does not involve either computer scientists or domain scientists integrally in the ACP. We believe that such an organization, in spite of best intentions, could evolve toward a typical 'information systems' organization that focused on procurement and operations, and might give inadequate attention to new technological opportunities and would not be sufficiently responsive to the needs of the end-user communities.

Distribute the ACP among all Directorates. In this approach, each directorate would take complete and exclusive responsibility for the infrastructure and applications supporting its respective community. The problems with this approach are very clear from the preceding discussion. Like the Office of the Director solution, it might focus too much on procurement and operation of current technologies, may result in excessive duplication of effort, and over time could create a serious balkanization of infrastructure becoming a serious obstacle to interdisciplinary collaboration and programs.

Technology Transfer

It is useful to specify in slightly more detail the horizontal technology-transfer dimension in Figure 4.2. The research actually includes two distinctly different flavors of research:

Fundamental, longer-term research in information technology and its applications. This type of research pursues revolutionary new ideas and fundamental understanding without being constrained by the current environment. This type of research, while extremely important, generally falls outside the scope of the ACP, with the exception of new long-horizon research on systems (social and technological) specifically supporting cyberinfrastructure and applications.

Applied, nearer-term research in information technology and its applications. This type of research seeks nearer-term outcomes that take strong account of and explicitly try to change and enhance the current environment. Its outcomes often include working prototypes fitting within an existing environment that can later be leveraged as a starting point for development, after they prove their mettle and after refinement through end-user experience.

Research outcomes that are deemed promising for the science and engineering research community are moved into the development and operational phases. A more detailed description of these phases includes:

Development of applications and infrastructure includes a set of activities resulting in a set of working, interoperable, and maintained implementations of working infrastructure and applications. The outcome is a set of software distributions that are interoperable and work within a prescribed environment of equipment and other software, including commercially available software.

Operations of applications and infrastructure include a set of activities resulting in a production environment where executing infrastructure and applications support end-users while they conduct science and engineering research. This includes provisioning, wherein the required facilities and equipment and various software modules are acquired and integrated and tested, and support of users in effectively using the capabilities.

A role of infrastructure is to reduce the time and effort and cost required for the development, provisioning, and operations of applications. Experience indicates that these phases do not follow sequentially, but rather it is most effective to repeat them in a process of successive refinement (in the context of software development, this is sometimes called the *spiral* model).

The development phase can be further subdivided into some constituent functions:

Conceptualization and analysis. Identification of an opportunity, assessment and analysis of needs, and development of a detailed set of requirements. It is particularly important in this ACP that this incorporate the outcomes and prototypes from NSF-sponsored information technology research.

Design. Choose an architecture (how to divide and conquer the implementation) and develop a plan (including identification of what designs can be reused, what can be purchased off the shelf, and what needs to be developed).

Implementation and testing. Programming new software or adapting existing software (e.g. from a research prototype), testing and refinement in the intended operational environment. In many cases a starting point may be a prototype arising from research. The outcome is a single software distribution to be used everywhere.

Maintenance and upgrade. Repair defects identified during provisioning and operations, and add new capabilities and features based on user needs identified during operational experience.

The Role of CISE

To be sure that the ACP leverages the most advanced technologies, utilizes state-of-the art software development processes and methodologies, remains fixated on advancing the information technologies themselves utilizing the opportunity to conceptualize and experiment with new applications in research, captures commonalities of need, and enables rather than hinders cross-disciplinary collaboration and programs, we recommend a continuing strong involvement of CISE in the management and implementation of ACP. In addition, like the other directorates, CISE's own research should also be a target of new applications and an opportunity to utilize the expanded infrastructure.

Activities within CISE comprise the three layers shown in Figure 4.3 of Section 4: core technologies, social and technological systems, and applications. The core technology layer includes a diverse set of individual technologies, available commercial products, processes, and best practices. The two higher layers comprise a set of integrated and coordinated activities, each activity dealing with three related activities: research, development, and provisioning and operations. The systems layer focuses predominantly on the infrastructure supporting applications, and the applications layer on new ways of conducting science and engineering research that are built upon this infrastructure. This architecture thus preserves both the vertical (core technology through infrastructure through applications) and horizontal (technology transfer from research through development and use) structure described earlier.

This layered architecture suggests that the horizontal grouping take precedence over the vertical, primarily because it will be more effective at capturing commonalities and coordinating activities across end-user communities. These are the most difficult goals to achieve, particularly so within NSF because of the separation of scientific and engineering disciplines in NSF and in the research community served by NSF. This organizational structure places these goals of commonality and coordination as the most explicit and visible within the organizational structure.

There are some additional points of evaluation that emphasize technology transfer:

Applied research seeks to influence the direction of development. Thus, a point of proposal evaluation should be the appropriateness of the research in light of end-users needs and a roadmap for serving those needs over the coming years, and the post-evaluation should be based in part on success in moving the ideas and prototypes into development.

A point of evaluation for development proposals should be ongoing processes for the technology transfer from research outcomes. Since specific research outcomes cannot be anticipated, there is an element of uncertainty, suggesting annual adjustments in direction. (The cooperative agreement mechanism for funding serves this end well.) Post-evaluation should focus on whether the development activity resulted in a stable and supported software distribution, whether and how proactively it has been provisioned, and (allowing a reasonable time for diffusion) and user satisfaction as well as (secondarily) how many users have been attracted.

A further point of evaluation for provisioning and operations is how effectively and proactively the organization has worked with developers to make the distributions available to users, and how effectively user experience and problems have been fed back to developers for maintenance, upgrade, and new capabilities.

There also needs to be an application program within CISE, collaborating closely with the other Directorates. This has several important purposes:

Like the other Directorates, the CISE research community can itself be served by the innovative application of information technology.

CISE involvement ensures the proximate and ongoing engagement of technology experts in identifying, formulating, and implementing new applications.

CISE takes primary responsibility to identify commonalities of need among different scientific and engineering disciplines that can be served by shared applications and infrastructure, to avoid unnecessary duplication of effort, and to empower future scientific collaboration across disciplines.

CISE takes responsibility for identifying and promulgating generic applications of wide interest across the NSF community. Examples include collaboration, data storage and archiving, digital libraries, numerical tools, and similar capabilities, all realized keeping in mind customization and extension to meet discipline-specific needs.

The Panel envisions an applications program in SBE as well. As in CISE, this has more than one purpose. One goal is to identify new applications to serve social scientists. A second goal is to involve social scientists in studying the application of information technology to groups and organizations, both how this can be done effectively and its impact.

The systems layer of the proposed architecture of ACP will fall predominantly within CISE (as well as the social sciences, see below). It groups a set of activities relating to heterogeneous compositions of diverse technologies with social constructs (groups, organizations, and communities). The idea is to serve several interests:

Infrastructure and applications emphasize systems composed of many technologies, including processing, storage, communications, and software. While there is considerable relevant activity in systems in many (and perhaps even all) of the current programs in CISE, the Panel believes it is time (and the challenges of the ACP focus attention on this) to view systems as a first-class target for research. By providing a focused effort on systems, more research attention can be focused here[1].

The infrastructure portion of ACP prominently involves systems issues. Thus, the systems layer is where activities surrounding infrastructure development, provisioning, and operations could reside.

Social systems figure prominently in the context of both the applications and infrastructure portions of the ACP. There is a deep integration of social systems with technological in both applications and infrastructure (the latter less obvious, but relating to the human organizations involved in development, provisioning, and operations). Thus, a systems activity should include social system issues, and thus collaborate with SBE.

Both applications and infrastructure suffer from a serious disconnect in fundamental objectives between technological researchers and end-user communities. The typical attitude of users is “we need it right away”, while technologists appropriately assert that “we don’t know the right way to do this until we do the research, and developing this right now will set in stone premature and suboptimum assumptions”. The Internet is an inspirational example of how these legitimate competing interests can reach a compromise by coupling both deployed infrastructure and its applications to *both* research outcomes and to end-user experience. The NSF middleware ACP is a recent example of how a program can be designed to base development of deployable infrastructure on a coupled and coordinated program of applied research, prototyping, and productizing of research outcomes.

The Role of Other Directorates

The post-evaluation of programs within the non-CISE directorates should focus on how substantially and beneficially the actual conduct of research has been changed, and how widely and effectively the new applications are actually used. Over time, resources should flow to directorates most successful in effectively using information technology to beneficially change the conduct of research.

This aspect of the ACP is a substantial change from the current Partnerships portion of the PACI program. In particular, as described in Section 5, the direction and funding of some portion of 'enabling technologies' and 'applications' would shift to the Directorates, where they would be subject to ordinary peer review. Of course, we expect the present PACI's to participate in these competitions, often involving collaborations with domain scientists and engineers. Rather than placing the burden on Centers to find partners to participate in applications, this would shift the direct control to Directorates and base awards on a competitive peer review process. If proposals are initiated by the domain experts, and if successful proposals demonstrate directly the enthusiasm and commitment to revolutionizing the conduct of research in the discipline and not simply serving their own narrow requirements, much more will be accomplished. The direct involvement of domain-expert program managers within the directorates will stimulate interest and involvement among more domain scientists and engineers, and they will also serve as coordinators to make sure that the aggregate activity funded out of the Directorate forms a coherent and complete ACP serving an entire domain of science or engineering.

The future direction of the PACI program is discussed in Section 5.

These programs within the directorates are also expected to work closely with each other and with CISE. All proposals to a Directorate should be evaluated in part on the credibility of its plan to execute its vision by working with NSF-funded centers or others, and also in its coherence to the overall ACP. For example, does the proposed activity make appropriate and maximum use of centrally developed infrastructure, does it anticipate opportunities to serve a larger community, and does it avoid duplication of effort with related activities in other Directorates? Does it have a credible plan to develop and support production technology and applications for the benefit of the entire discipline? For this reason, all such programs should be considered joint programs with the relevant portions of CISE, and proposal evaluation should be a joint responsibility involving both domain and technology experts.

The Role of the PACI's

The impetus for this 'matrix' form of organization surrounding applications shifts responsibility for initiating application research and development away from the PACI's and toward the scientific and engineering directorates. This is motivated by some perceived shortcomings in the current organization surrounding application initiatives in CISE. While domain scientists and engineers are encouraged to participate through 'partnerships', these are largely ad hoc collaborations driven by individual initiative rather than any common vision or direction. If the conduct of science and engineering research is to be revolutionized, this will be based on leadership for creating and executing a vision emanating from the non-CISE directorates within programs focused on this objective led by committed and visionary managers who are themselves domain experts. If the non-CISE directorates and the communities they serve have a major stake in the outcome, they can provide the necessary leadership to rally researchers around creating and executing a vision. We envision ACP program directors within each of the directorates to be domain scientists or engineers who possess a deep and abiding interest in revolutionizing the conduct of research in their respective fields through information technology. We expect them to motivate and lead their respective communities, as well as define coherent programs that systematically approach this challenge.

All this applies to the funding side of the equation, but when it comes to delivery the role of the PACI's and other centers serving this ACP may not be greatly changed. This separation of funding should not preclude the grouping of activities within centers where this makes sense, such as software development and operations. Investigators within the scientific and engineering research communities will likely seek the involvement of centers in prototyping and productizing their application ideas, to bring in needed software engineering expertise and to lend credibility to technology transfer.

Industry Involvement

Not only should commercial technologies be acquired in preference to development of similar technologies, but the goal of the ACP should be commercialization of both cyberinfrastructure and applications that prove to be widely used and beneficial. This will not be practical for some more esoteric and specialized applications. However, most of the cyberinfrastructure technologies and many applications should attract industrial interest, and longer term government support for ongoing development and operations should prove unnecessary.

To the extent suitable off-the shelf technologies are available, they should be acquired and used; researchers and NSF program managers need to be well connected to current and emerging commercial activities and seek alliances with them as appropriate. One centralized activity should systematically choose and license commercial solutions to avoid multiple (and incompatible) choices and to obtain favorable licensing terms. For example, prototypes and experimental results may originate from self-supported activities in industry as well as from NSF-supported researchers.

The information technology researchers participating directly or indirectly with this ACP have limited ability to integrate, maintain, and support their own research prototypes. This is the primary role of development organizations that start with these prototypes and end with an integrated and supported software distribution. It is generally healthy to consider alternative approaches and encourage competition among prototype solutions before choosing one to develop and deploy.

The intermediate- to long-term goal should be to commercialize all infrastructure and many application solutions developed in this ACP that are successful and gain a significant following, and withdraw from those that don't. It should not be necessary for NSF to support the development activities indefinitely in any particular area; rather, the goal should be to migrate those development dollars to new areas. Thus, a growing portion of the supported infrastructure and applications are expected to be off-the-shelf commercial technologies licensed with financial support from NSF, with NSF funding for prototyping and development continually redirected to the moving frontier of new (non-commercially supported) capabilities.

Infrastructure suffers from a 'chicken and egg' conundrum in the commercial world: Which comes first, the infrastructure or the applications? It is difficult to invest in new infrastructure with no applications available to provide value to users, and application investment usually follows existing infrastructure. Following the inspirational example of the Internet, this ACP seeks to use NSF investment coordinated across both infrastructure and complementary applications to 'jump start' new commercial markets, and later move those applications and supporting infrastructure together into commercial practice. The ultimate beneficiary will be not only the science and engineering research communities, but the U.S. economy and industry as a whole.

Participation by Other Agencies and Governments

The nature of the ACP is that it should provide value to all science and engineering research, regardless of whether it is funded by NSF and regardless of whether it is conducted in the U.S. or abroad. In fact, the more universally these applications and supporting infrastructure are adopted, the more value they offer to each participant. Thus, this should be viewed as a government-wide initiative and include strong international cooperation. NSF will be the leader, but it should seek broad participation by others.

Appendix D: Names and Affiliation of Those Giving Testimony to the Panel

NAME	TITLE	AFFILIATION	
Anthes, Rick	President	University Corporation for Atmospheric Research	2
Avery, Paul	Professor of Physics	University of Florida	2
Baker, Keith	Professor of Physics	Hampton University	3
Bell, Gordon	Senior Researcher	Microsoft Corporation	3
Berman, Fran	Professor of Computer Science and Director	San Diego Supercomputer Center & NPACI	1
Bernholc, Jerzy	Professor of Physics	North Carolina State University	1
Billy, Carrie	Director of Information Technology Initiatives	American Indian Higher Education Consortium	2
Bradburn, Norman	Assistant Director, Directorate for Social Behavioral & Economic Sciences	National Science Foundation	1
Bush, Aubrey	Director, Division of Advanced Networking Infrastructure and Research	National Science Foundation	2
Cassatt, James	Director, Division of Cell Biology & Biophysics	National Institutes of Health	3
Cavanaugh, Margaret	Staff Associate, Office of the Director and Chair of the Working Group on Environmental Research and Education	National Science Foundation	1
Cherniavsky, John	Senior Advisor for Research, Division of Research, Evaluation & Communication	National Science Foundation	1
Colvin, Mike	Team Leader, Computational Biology Group & Biotechnology Research Program	Lawrence Livermore National Laboratory	1
Connolly, John	Director, Center for Computational Sciences	University of Kentucky	3
DeRosa, Marc	Scientist	Lockheed Martin Solar National Physics Laboratory	2
Djorgovski, George	Professor of Astronomy	California Institute of Technology	1
Dunning, Thom	Director, William R. Wiley Environmental Molecular Sciences Laboratory	Pacific Northwest National Laboratory	1
Eisenstein, Robert	Assistant Director, Directorate for Mathematical & Physical Sciences	National Science Foundation	1
Ellisman, Mark	Professor of Neurosciences	University of California San Diego	1
Erb, Karl	Director, Office of Polar Programs	National Science Foundation	2
Feiereisen, William	Chief, NASA Advanced Supercomputing	NASA Ames Research Center	2
Finholt, Tom	Assistant Professor of Psychology, Director of Collaboratory for Research on Electronic Work	University of Michigan	1
Foster, Ian	Professor, Department of Computer Science	University of Chicago	1

Fulker, David	Director, National Science Digital Library & Unidata	University Corporation for Atmospheric Research	1
Giles, Roscoe	Professor of Electrical and Computer Engineering	Boston University	2
Graves, Sara	Professor of Computer Science	University of Alabama Huntsville	2
Grossman, Robert	Research Associate, Astrophysical, Planetary and Atmospheric Sciences	University of Colorado	1
Gulari, Esin	Acting Assistant Director, Directorate for Engineering	National Science Foundation	1
Harvey, Newman	Professor of Physics	California Institute of Technology	1
Hey, Tony	Professor of Computation and Director of United Kingdom e-science Core Programme	University of Southampton and The Engineering and Physical Sciences Research Council	1
Jacobs, Clifford	Section Head, Division of Atmospheric Sciences	National Science Foundation	1
Johnson, Chris	Professor of Computing	University of Utah	1
Kaplow, Wes	Chief Technology Officer	Qwest Government Services Division	2
Karin, Sid	Professor of Computer Science and Engineering	University of California, San Diego	1
Karniadakis, George	Professor of Applied Mathematics	Brown University	1
Karshmer, Arthur	Director, Program for Persons with Disabilities	National Science Foundation	3
Kennedy, Ken	Professor of Computer Science	Rice University	2
Killeen, Tim	Director	National Center for Atmospheric Research	1
Kubiatowicz, John	Professor of Computer Science	University of California, Berkeley	2
Landau, Rubin	Professor of Physics	Oregon State University	3
Levine, Michael	Professor of Physics	California Institute of Technology	1
Lynch, Clifford	Executive Director	Coalition for Networked Information	2
Martin, William	Professor of Nuclear Engineering	University of Michigan	3
McRobbie, Michael	Vice President for Information Technology, Professor of Computer Science	Indiana University	1
Moore, Reagan	Associate Director, Data-Intensive Computing	San Diego Supercomputer Center	3
Myers, Gene	Vice President for Research Informatics	Celera Genomics	2
Oliver, Ed	Director, Advanced Scientific Computing Research	Department of Energy	3
Prudhomme, Tom	Deputy Project Manager, NEESgrid	National Center for Supercomputing Applications	1
Ramirez, Alex	Executive Director for Information Technology Initiatives	Hispanic Association of Colleges and Universities	2

Reed, Dan	Professor and Director	NCSA, University of Illinois at Urbana-Champaign	1
Roskies, Ralph	Professor of Physics & Scientific Director, Pittsburgh Supercomputing Center	University of Pittsburgh	1
Roskoski, Joann	Executive Officer, Directorate for Biological Sciences	National Science Foundation	1
Shulenburger, David	Provost	University of Kansas	3
Smarr, Larry	Professor of Computer Science and Engineering	University of California San Diego	1
Stevens, Rick	Professor of Computer Science; Division Director, Mathematics and Computer Science	University of Chicago; Argonne National Laboratory	1
Strawn, George	Chief Information Officer, Office of Information & Resource Management	National Science Foundation	1
Sugar, Bob	Professor of Physics	University of California, Santa Barbara	1
Taylor, John M.	Director General of Research Council	United Kingdom Office of Science and Technology	1
VanHouweling, Doug	President	University Corporation for Advanced Internet Development, Internet 2	1
Wallach, Steve	Vice President, Office of Technology	Chiaro	1
Wladawsky-Berger, Irving	Vice President Technology & Strategy	IBM	1
Wolff, Steve	Advanced Research and Technology Initiatives	Cisco Systems	2
Woodward, Paul	Professor of Astronomy	University of Minnesota	2

Appendix E: Charge to the Blue Ribbon Advisory Panel on CyberInfrastructure

Recent advances in technologies such as the Internet, digital libraries, data mining, visualization, tele-instrumentation, tele-presence, and grid computing have opened new opportunities to enhance our computational research capacity. The pace of research and innovation in these areas persuades us that NSF must draw much more aggressively on these new technology opportunities. Additionally, we must establish a setting for Terascale Research and Infrastructure that is better able to utilize new technology opportunities. With expected enhancements to computing systems, and the ability to store and transmit data, our goal now is no less than Teraflops, and Terabytes ultimately via Terabit networks.

The PACI Program, now in its fourth year, was formulated based on the recommendations of the Hayes Report issued in September 1995. The Blue Ribbon Panel is being convened to:

Goal-A) evaluate the performance of the PACI Program in meeting the needs of the scientific research and engineering community;

Goal-B) recommend new areas of emphasis for the NSF Directorate for Computer and Information Science and Engineering that will respond to the future needs of this community; and

Goal-C) recommend an implementation plan to enact those changes.

Goal-A: In assessing the impact of the PACI program the Blue Ribbon Panel should consider the following questions:

A-1 How well does the current PACI Program meet the needs of the science and engineering communities served by the NSF for providing access to high-end computational resources, information intensive resources, visualization resources and network access to those resources?

How successful has the PACI program been in the engaging new communities, e.g. social scientists, in the use of high performance computing as a tool for the conduct of their research and education activities?

A-2 In the context of the current and projected state of the technology (computer, storage, networking, etc.) are the current PACI partnerships meeting the recommendations of the Hayes Report.

A-3 How well does the current PACI Program fit within the NSF's objectives for providing the cross-Foundational infrastructure needed to enable further advances and lead to breakthroughs across all areas of science and engineering?

To what extent does the current program complement or overlap similar resources made available to academic scientists through NSF and other agencies such as NASA, DOE, DOD, etc.?

A-4 What impact have the Enabling Technologies (ET), Applications Technologies (AT), and Education, Outreach and Training (EOT) components of the PACI Program had on the scientific community both within the current PACI Partnerships, and in the larger research community?

Has support for these activities been adequate in the current program?

A-5 How successful have the current PACI Partnerships been at leveraging PACI support through interactions with other programs within the NSF, within other federal and state agencies, through partnerships with technology vendors, within partnering universities, and through industrial partners?

A-6 What are the international aspects of PACI?

Goal-B: How can NSF best take advantage of the significant changes in technology that have occurred since the issuance of the Hayes Report?

B-1 How have changes in computing, storage, and networking technology affected the methodologies, practices, and needs of the science and engineering community?

How have the continuing needs of the science drivers affected, and been realized in, areas of IT?

What impact will the availability of high performance cyber-infrastructure have on enabling cross-disciplinary research?

What societal applications (customized manufacturing, edutech, eldertech, etc.) have been impacted by these changes?

What developments are expected in the near future that will further change these needs?

What implications do these changes have for the high performance cyber-infrastructure that NSF will support?

B-2 Is the current NSF investment in computational, networking, storage, and visualization infrastructure sufficient to address the current and future high end demands of the science and engineering research community provided by NSF?

B-3 How can NSF better support computer scientists who develop tools that accelerate the efficient and effective use of high-end computing and communications infrastructure for simulation, data acquisition, storage and display, etc.?

B-4 What are the barriers that confront potential HPC users that wish to take advantage of state-of-the-art computational, storage, networking and visualization resources in their research? What can be done to remove these barriers?

B-5 What can be done to improve the education and outreach activities to broaden access to high-end computing? How can the number of scientists and others who have the knowledge and skills necessary to be able to use high-end computing be increased?

Goal C: Finally, the Blue Ribbon Panel should offer suggestions on how to implement any recommended changes:

C-1 What are the highest priority cyber-infrastructure investments NSF needs to make for high performance users?

C-2 What research "infrastructure" should be coupled with cyber-infrastructure?

C-3 What new science opportunities are likely to produce further developments in IT?

C-4 How should any new investments in infrastructure/technologies be administered? What proportion of infrastructure operating expenses needs to be planned for to assure the best utility of such infrastructure?

C-5 Should the Cooperative Agreements for the current PACI Partnerships be renewed for an additional 5 year period? With recommended changes to their current missions?

C-6 Should the current NSF infrastructure programs in high performance computing and networking be, extended or modified so that NSF is better poised to deal with the future needs of the science and engineering community?