Designing a Digital Future

NOTE: Sidebars, Appendices, and Rest of Report

Report
About the President’s Council of Advisors on Science and Technology
The President’s Council of Advisors on Science and Technology
EXECUTIVE OFFICE OF THE PRESIDENT
Executive Report
Introduction
Recommended Initiatives and Investments in NIT R&D to Achieve America’s Priorities and Advance Key NIT Research Frontiers
The Importance of Government Leadership
Improved Effectiveness of NITRD Coordination
Crosscutting Themes
PCAST NITRD Program Review Working Group
Table of Contents
1. Introduction
   1.1 The Organization of this Report
   1.2 A Preview of the NITRD Portfolio and the NITRD Coordination Process and Structure
2. The Impact of Networking and Information Technology
3. Recent Technological and Societal Trends
4. The Role of Advances in NIT in Achieving America’s Priorities
   4.1 NIT for Health
   4.2 NIT for Energy and Transportation
   4.3 NIT for National and Homeland Security
   4.4 NIT for Discovery in Science & Engineering
   4.5 NIT for Education
   4.6 NIT for Digital Democracy
      A Picture is Worth a Thousand Numbers
5. Recommendations: Initiatives in NIT R&D to Achieve America’s Priorities
6. NIT Research Frontiers
   6.1 NIT and People
6.2 NIT and the Physical World
6.3 Large-Scale Data Management and Analysis
   Extracting Worldly Knowledge from the World Wide Web
6.4 Trustworthy Systems and Cybersecurity
6.5 Scalable Systems and Networking
6.6 Software Creation and Evolution
6.7 High Performance Computing

7. Recommendations: Investments in the NIT Research Frontiers
   Introduction
   Privacy and Confidentiality
   The Ubiquitous Role of Privacy
   NIT and People
   NIT and the Physical World
   Large-Scale Data Management and Analysis
   Trustworthy Systems and Cybersecurity
   Scalable Systems and Networking
   Software Creation and Evolution
   High Performance Computing

8. Technological and Human Resource Requirements
   8.1 Hardware, Software, and Data Infrastructure
   8.2 Education and Human Resources

9. Recommendations: Technological and Human Resources
   Hardware, Software, and Data Infrastructure
   Education and Human Resources

10. Strengths and Limitations of the NITRD Coordination Process and Structure
11. Recommendations: NITRD Coordination Process and Structure
12. The Role of Federal Investment in NIT R&D
   12.1 The Critical Role of Federal Investment
   12.2 The Incremental Investment Implied by this Report

Appendices:
   A: Expert Input into the PCAST NITRD Review
   B: Acknowledgments
   C: Abbreviations used in this Report

Footnotes
   1
   2
   3
   4
# The President’s Council of Advisors on Science and Technology

## Co-Chairs

<table>
<thead>
<tr>
<th>Name</th>
<th>Title and Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>John P. Holdren</td>
<td>Assistant to the President for Science and Technology</td>
</tr>
<tr>
<td></td>
<td>Director, Office of Science and Technology Policy</td>
</tr>
<tr>
<td>Eric Lander</td>
<td>President, Broad Institute of Harvard and MIT</td>
</tr>
<tr>
<td>Harold Varmus*</td>
<td>President, Memorial Sloan-Kettering Cancer Center</td>
</tr>
</tbody>
</table>

* Dr. Varmus resigned from PCAST on July 9, 2010 and subsequently became Director of the National Cancer Institute (NCI).

## Members

<table>
<thead>
<tr>
<th>Name</th>
<th>Title and Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosina Bierbaum</td>
<td>Dean, School of Natural Resources and Environment</td>
</tr>
<tr>
<td></td>
<td>University of Michigan</td>
</tr>
<tr>
<td>Christine Cassel</td>
<td>President and CEO, American Board of Internal Medicine</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Chad Mirkin</td>
<td>Rathmann Professor, Chemistry, Materials</td>
</tr>
<tr>
<td></td>
<td>Science and Engineering, Chemical and Biological Engineering and Medicine</td>
</tr>
<tr>
<td></td>
<td>Director, International Institute of Nanotechnology</td>
</tr>
<tr>
<td>Mario Molina</td>
<td>Professor, Chemistry and Biochemistry</td>
</tr>
<tr>
<td></td>
<td>University of California, San Diego</td>
</tr>
<tr>
<td></td>
<td>Professor, Center for Atmospheric Sciences</td>
</tr>
<tr>
<td></td>
<td>Scripps Institution of Oceanography</td>
</tr>
<tr>
<td></td>
<td>Director, Mario Molina Center for Energy and Environment, Mexico City</td>
</tr>
<tr>
<td>Christopher Chyba</td>
<td>Professor, Astrophysical Sciences and International Affairs</td>
</tr>
<tr>
<td></td>
<td>Director, Program on Science and Global Security</td>
</tr>
<tr>
<td></td>
<td>Princeton University</td>
</tr>
<tr>
<td>S. James Gates, Jr.</td>
<td>John S. Toll Professor of Physics</td>
</tr>
<tr>
<td></td>
<td>Director, Center for String and Particle Theory</td>
</tr>
<tr>
<td></td>
<td>University of Maryland</td>
</tr>
<tr>
<td>Shirley Ann Jackson</td>
<td>President, Rensselaer Polytechnic Institute</td>
</tr>
<tr>
<td>Richard C. Levin</td>
<td>President, Yale University</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Ernest J. Moniz</td>
<td>Cecil and Ida Green Professor of Physics and Engineering Systems</td>
</tr>
<tr>
<td></td>
<td>Director, MIT’s Energy Initiative</td>
</tr>
<tr>
<td>Craig Mundie</td>
<td>Chief Research and Strategy Officer</td>
</tr>
<tr>
<td></td>
<td>Microsoft Corporation</td>
</tr>
<tr>
<td>Daniel Schrag</td>
<td>Sturgis Hooper Professor of Geology</td>
</tr>
<tr>
<td></td>
<td>Professor, Environmental Science and Engineering</td>
</tr>
<tr>
<td></td>
<td>Director, Harvard University-wide Center for Environment</td>
</tr>
<tr>
<td></td>
<td>Harvard University</td>
</tr>
<tr>
<td>David E. Shaw</td>
<td>Chief Scientist, D.E. Shaw Research</td>
</tr>
<tr>
<td></td>
<td>Senior Research Fellow, Center for Computational Biology and Bioinformatics</td>
</tr>
<tr>
<td></td>
<td>Columbia University</td>
</tr>
</tbody>
</table>

http://semanticommunity.info/A_NITRD_Dashboard/Designing_a_Digital_Future

Updated: Sat, 19 Sep 2015 01:01:15 GMT

Powered by mindtouch™
Dear Mr. President,

We are pleased to send you this report, Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology, prepared by your President’s Council of Advisors on Science and Technology (PCAST) acting in its role as the President’s Innovation and Technology Advisory Council (PITAC). This report fulfills PCAST’s responsibilities under Executive Order 13539 and the High-Performance Computing Act of 1991 (Public Law 102-194) as amended by the Next Generation Internet Research Act of 1998 (Public Law 105-305) and by the America COMPETES Act of 2007 (Public Law 110-69).

The Networking and Information Technology Research and Development (NITRD) Program is the primary mechanism by which the Federal government coordinates its unclassified networking and information technology (NIT) research and development (R&D) investments. Fourteen Federal agencies, including all of the large science and technology agencies,
are formal members of the NITRD Program, with many other Federal entities participating in NITRD activities. The program helps ensure that the Nation effectively leverages its strengths, avoids duplication, and increases interoperability in such critical areas as supercomputing, high-speed networking, cybersecurity, software engineering, and information management.

To provide a solid scientific basis for its assessment of NITRD, PCAST appointed an expert 14-member Working Group, which consulted with more than 50 individuals, including government officials, industry representatives, and experts from academia.

PCAST finds that NITRD is well coordinated and that the U.S. computing research community, coupled with a vibrant NIT industry, has made seminal discoveries and advanced new technologies that are helping to meet many societal challenges. Importantly, however, PCAST also finds that a substantial fraction of the NITRD multi-agency spending summary represents spending that supports R&D in other fields, rather than spending on R&D in the field of NIT itself. As a result, the Nation is actually investing far less in NIT R&D than the $4 billion-plus indicated in the Federal budget. To achieve America’s priorities and advance key research frontiers to support economic competitiveness in NIT, this report calls for a more accurate accounting of this national investment and recommends additional investments in NIT R&D, including research in networking and information technology for health, energy and transportation, and cyber-infrastructure, among others.

NIT has yielded enormous benefits for the Nation’s economic competitiveness, national security, and quality of life. To maintain America’s leadership in NIT in an ever more competitive global environment, the Federal Government must be bold in its investments, including funding of high risk/high reward research with the potential to move this essential field in unanticipated directions. PCAST believes that execution of the recommendations in this report will enable us to address critical priorities and challenges in the years ahead.

Sincerely,

John P. Holdren  Eric Lander  Shirley Ann Jackson  Eric Schmidt
PCAST Co-chair  PCAST Co-chair  PITAC Co-chair  PITAC Co-chair

Executive Report

Introduction

From smartphones to eBook readers to game consoles to personal computers; from corporate datacenters to cloud services to scientific supercomputers; from digital photography and photo editing, to MP3 music players, to streaming media, to GPS navigation; from robot vacuum cleaners in the home, to adaptive cruise control in cars and the real-time control systems in hybrid vehicles, to robot vehicles on and above the battlefield; from the Internet and the World Wide Web to email, search engines, eCommerce, and social networks; from medical imaging, to computer-assisted surgery, to the large-scale data analysis that is enabling evidence-based healthcare and the new biology; from spreadsheets and word processing to revolutions in inventory control, supply chain, and logistics; from the automatic bar-coding of hand-addressed first class mail, to remarkably effective natural language translation, to rapidly improving speech recognition – our world today relies to an astonishing degree on systems, tools, and services that belong to a vast and still growing domain known as Networking and Information Technology (NIT). NIT underpins our national prosperity, health, and security. In recent decades, NIT has boosted U.S. labor productivity more than any other set of forces.
The United States has a proud history of achievement and leadership in NIT. The Federal Government has played an essential role in fostering the advances in NIT that have transformed our world. Steady Federal investment in NIT research over the past 60 years has led to many of the breakthroughs noted above, often a decade or more after the research took place. The Federal investment in NIT research and development is without question one of the best investments our Nation has ever made. In order to sustain and improve our quality of life, it is crucial that the United States continue to innovate more rapidly and more creatively than other countries in important areas of NIT. Only by continuing to invest in core NIT science and technology will we continue to reap such enormous societal benefits in the decades to come.

Recent technological and societal trends place the further advancement and application of NIT squarely at the center of our Nation’s ability to achieve essentially all of our priorities and to address essentially all of our challenges:

- **Advances in NIT are a key driver of economic competitiveness.** They create new markets and increase productivity. For example, an investment in the National Science Foundation’s Digital Library Initiative in the 1990’s led to Google, a company with a market capitalization of nearly $200 billion that has transformed how we access information.

- **Advances in NIT are crucial to achieving our major national and global priorities in energy and transportation, education and life-long learning, healthcare, and national and homeland security.** NIT will be an indispensable element in buildings that manage their own energy usage; attention-gripping, personalized methods that reinforce classroom lessons; continuous unobtrusive assistance for people with physical and mental disabilities; and strong resilience to cyber warfare.

- **Advances in NIT accelerate the pace of discovery in nearly all other fields.** The latest NIT tools are helping scientists and engineers to illuminate the progression of Alzheimer’s disease, elucidate the nature of combustion, and predict the size of the ozone hole, to cite just a few examples.

- **Advances in NIT are essential to achieving the goals of open government.** Those advances will allow better access to government records, better and more accessible government services, and the ability both to learn from and communicate with the American public more effectively.

Both the science and the practice of NIT have seen dramatic changes during the sixty-year history of the field. The ability of the computing research community, coupled with a vibrant NIT industry, to deliver those changes – to discover and advance new areas of NIT research and development (R&D) that stimulate technological progress and meet societal challenges – has been essential to the Nation’s success. There are enormous opportunities for future transformations. To meet the challenge of change, America must continue to make R&D investments in new areas of NIT.

Of course, the Government is not alone in investing in NIT R&D. Industry has made, and continues to make, major contributions. It is important, however, not to equate the very large industry R&D investment in NIT with fundamental research of the kind that is carried out in universities and a small number of industrial research labs. The vast majority of industry R&D in NIT is focused on development – on the engineering of future products and product versions. Few major NIT companies have formal research organizations, and even those that do invest relatively little in research compared to their investment in development activities. Fundamental research with the potential for future transformational application represents a small fraction of overall industry R&D in NIT – a situation that is both appropriate and unlikely to change. For that reason, among others, Federal investment in NIT R&D is and will remain essential.

As a field of inquiry, NIT has a rich intellectual agenda – as rich as that of any other field of science or engineering. In addition, NIT is arguably unique among all fields of science and engineering in the breadth of its impact. Computer science research, carried out to a great extent in America’s research universities with funding from Federal agencies such as the National Science Foundation (NSF) and the Defense Advanced Research Projects Agency (DARPA), lies at the heart of our Nation’s leadership. It is this research – which ranges from the design of computers and networks to robotics, software, and algorithms – that has repeatedly led to the introduction of entirely new product categories that became multi-billion-dollar industry sectors. The “extraordinarily productive interplay of federally funded university
research, federally and privately funded industrial research, and entrepreneurial companies founded and staffed by people who moved back and forth between universities and industry” has been well documented.

Essentially all unclassified federally funded R&D activities in NIT and related fields fall within the scope of the Networking and Information Technology Research and Development (NITRD) Program. The term “NITRD Program” refers both to the mechanism by which the Federal Government coordinates its unclassified R&D investments in NIT, and to the unclassified Federal NIT R&D portfolio itself. The NITRD member agencies report aggregate NIT R&D investments in excess of $4 billion annually. The largest investments are reported by the National Institutes of Health (NIH) and NSF (roughly $1 billion each), followed by the Office of the Secretary of Defense and the Department of Defense Service research organizations (OSD/DoD), the Department of Energy (DoE), and DARPA (roughly $500 million each). However, analysis indicates that a substantial fraction of the NITRD crosscut budget (the multi-agency spending summary) represents spending on NIT that supports R&D in other fields, rather than spending on R&D in the field of NIT itself. For example, an expert review of the top 100 awards (by award size) in NIH’s NITRD portfolio – totaling nearly $600 million, roughly half of NIH’s NITRD crosscut total – concluded that only between 2% and 11% (by dollar value) should be considered NIT R&D. The remainder is spent on various forms of NIT infrastructure that provide essential support for biomedical research, but not on NIT R&D. We have used NIH as an example only because the laudable transparency of its records and reporting allowed such an analysis to be performed. Although other agencies do not report NIT R&D spending in sufficient detail to make the same analysis possible, it seems likely that in many cases a similar confusion in classification of NITRD investment occurs. An important finding of this report is that the Nation is actually investing far less in NIT R&D than is shown in the Federal budget.

In summary, the transformative NIT research that fuels innovation and achievement and strengthens our Nation needs to come from Government investment, yet it is currently difficult to ascertain the magnitude of that investment. Furthermore, going forward, the participating agencies in the NITRD Program must more aggressively embrace the expanding role that advances in NIT play in America’s future. A broad spectrum of Federal agencies – those currently participating in NITRD and some which are not yet doing so – must recognize that their abilities to accomplish their missions are inextricably linked to advances in NIT, and must invest in NIT R&D to catalyze the advances that are critical to their missions. Strategic leadership must come from the top – from those within the Federal Government with the authority to implement new strategies.

The PCAST NITRD Program Review Working Group was asked to assess not only the coordination function of NITRD but also the investment portfolio itself. In the remainder of this Executive Report, and in greater detail and breadth within the body of the report, we describe some of the compelling and important scientific and technical problems that must be addressed in order to maintain and strengthen the transformative effect of NIT on the Nation and the world, and we describe some of the essential research that will be needed to solve those problems. A bottom-up analysis of some of the key initiatives that we recommend in this report suggests that an investment of at least $1 billion annually will be needed for new, potentially transformative NIT research. Uncertainty regarding the precise nature of current expenditures makes it difficult to determine how much of this investment can be obtained through repurposing and reprioritization and how much will require new funding. We believe, however, that a lower level of investment in this critically important area could seriously jeopardize America’s national security and economic competitiveness.

---

**Recommended Initiatives and Investments in NIT R&D to Achieve America’s Priorities and Advance Key NIT Research Frontiers**

The Federal Government’s investment in NIT R&D dates from the birth of the field more than sixty years ago. NITRD as a coordination effort, though, had its genesis in the High-Performance Computing Act of 1991 – “An Act to provide for a coordinated Federal program to ensure continued United States leadership in high-performance computing.” Its scope was broadened by the Next Generation Internet Research Act of 1998, and again by the America COMPETES Act of 2007.

In its early years, NITRD’s role was seen as coordinating research in the fundamentals of computing, while the use and advancement of the resulting technology to address our national priorities was left to individual agencies. In recent
years, the value and importance of multi-agency coordination in the development and application of NIT to achieve the Nation’s priorities has become apparent, and has led to the creation of NITRD Senior Steering Groups in the vital areas of Cyber Security and Information Assurance, and Health IT. NITRD is well-positioned to facilitate similar coordination in NIT for other important national priorities, among them energy and transportation, and education and life-long learning.

The role of NIT in addressing our national priorities, and the NIT research frontiers that contribute to making progress in strengthening our NIT capabilities, raise many important research questions that must be tackled. It is essential that short term needs not crowd out the longer term research that anticipates future needs. It is also essential that some NIT research explore bold, unconventional ideas that would have enormous impact if they could be realized. A recent report from the American Academy of Arts & Sciences describes both the benefits of such transformative research and the mechanisms that can be used to foster it.

The Federal Government must invest in new multi-agency NIT R&D initiatives in areas of particular importance to our national priorities. Such investments should include funding for high risk/high reward research with the potential to move these areas in unanticipated directions. Some of this research will require large project teams and sufficiently long time horizons to allow ambitious goals to be achieved. We see three areas in which such initiatives are particularly timely and important.

Recommendation [Section 5]: The Federal Government, under the leadership of NSF and Health and Human Services (HHS), with participation from the Office of the National Coordinator for Health Information Technology (ONC), the Centers for Medicare and Medicaid Services (CMS), the Agency for Healthcare Research and Quality (AHRQ), the National Institute of Standards and Technology (NIST), the Veterans Health Administration (VHA), DoD, and other interested agencies, should invest in a national, long-term, multi-agency research initiative on NIT for health that goes well beyond the current national program to adopt electronic health records. The initiative should include sponsorship of multi-disciplinary research on three themes:

• to make possible comprehensive lifelong multi-source health records for individuals;

• to enable both professionals and the public to obtain and act on health knowledge from diverse and varied sources as part of an interoperable health IT ecosystem; and

• to provide appropriate information, tools, and assistive technologies that empower individuals to take charge of their own health and healthcare and to reduce its cost.

This program should build on national activities promoting the adoption and meaningful use of electronic health records that are usable by all appropriate organizations; it should complement the shorter-term ONC programs; and it should augment the research investments that the various agencies are currently able to make. In addition to increased attention on using NIT for wellness and for addressing chronic conditions, the departments and agencies mentioned above should continue to investigate novel uses of NIT, such as NIT-assisted surgery, to deliver care for acute conditions. They should continue to pursue advances in the innovative use of NIT, such as sensing and monitoring, to understand the basic biological and psychological mechanisms that underlie disease. And they should continue to address NIT research opportunities that support current and continuing work by HHS and NSF on transformational innovation in healthcare delivery and basic research in health and wellness.
Recommendation [Section 5]: The Federal Government should invest in a national, long-term, multi-agency, multi-faceted research initiative on NIT for energy and transportation. As part of that initiative:

- DoE and NSF should be major sponsors of research for achieving dynamic power management in applications ranging from single devices to buildings to the power grid.

- NIST should organize the multi-stakeholder formulation of interoperable standards for real-time control. Interoperability facilitates repeated cycles of innovation by multiple vendors, promoting the development of versatile and robust NIT.

- DoD should continue to be a major sponsor of research on using NIT to achieve low-power systems and devices.

- The Department of Transportation (DoT) should sponsor ambitious NIT research relevant to surface and air transportation.

Current research in the computer simulation of physical systems should be expanded to include the simulation and modeling of proposed energy-saving technologies, as well as advances in the basic techniques of simulation and modeling.

Recommendation [Sections 5 and 7]: The Federal Government should invest in a national, long-term, multi-agency research initiative on NIT that assures both the security and the robustness of cyber-infrastructure. NSF and DoD, in collaboration with the Department of Homeland Security (DHS), should aggressively accelerate funding and coordination of fundamental research

- to discover more effective ways to build trustworthy computing and communications systems,
- to continue to develop new NIT defense mechanisms for today’s infrastructure, and most importantly,
- to develop fundamentally new approaches for the design of the underlying architecture of our cyber-infrastructure so that it can be made truly resilient to cyber-attack, natural disaster, and inadvertent failure.

Infrastructure to be protected includes the Internet and the national telecommunication system as well as computing systems controlling such national resources as the electric power grid and the financial system. Where fundamental NIT advances are needed to support these initiatives, mission agencies should invest in fundamental research in NIT, either alone or in collaboration with NSF, and should not limit their programs to application-specific research.

Effective use of NIT in increasing our economic competitiveness and achieving our other national priorities depends not only on incorporating innovative NIT into a wide variety of domains, but also on ensuring that the basic science and engineering of NIT remain vibrant and strong. At the time of the High-Performance Computing Act of 1991, the importance of high performance computing and communication (HPCC) to scientific discovery and national security was a major factor underlying the special attention given by Congress to NIT. Although HPCC continues to contribute in important ways to scientific discovery and national security, many other aspects of NIT have now risen to comparable levels of importance. Among these NIT areas are the interactions of people with computing systems and devices, both individually and collectively; the interactions between NIT and the physical world, such as in sensors, imaging, robotic and vision systems, and wearable and mobile devices; large-scale data capture, management and analysis; systems that protect personal privacy and sensitive confidential information, are robust in the face of malfunction, and stand up to cyber-attack; scalable systems and networking (i.e., systems and networks that can be either increased or decreased in
complexity, size, generality, and cost); and software creation and evolution. HPCC is but one of many important areas of NIT, and America’s prowess in HPCC is but one of many measures of our international competitiveness in NIT.

To achieve our national priorities, and to stimulate the next generation of transformative advances in NIT, we must ensure that the modern and emerging research frontiers are well supported. Investment in those areas must include funding for high risk/high reward research with the potential to move these areas in unanticipated directions.

Recommendation [Section 7]: The Federal Government must increase investment in those fundamental NIT research frontiers that will accelerate progress across a broad range of priorities. Among such investments:

- NSF and DARPA, with the participation of other relevant agencies, should invest in a broad, multi-agency research program on the fundamentals of privacy protection and protected disclosure of confidential data. Privacy and confidentiality concerns arise in virtually all uses of NIT.

- NSF, DARPA, and HHS should create a collaborative research program that augments the study of individual human-computer interaction with a comprehensive investigation to understand and advance human-machine and social collaboration and problem-solving in a networked, on-line environment where large numbers of people participate in common activities. Understanding such collective human-NIT interactions is increasingly important for defense, for health, and for the activities of daily life.

- NSF should expand its support for fundamental research in data collection, storage, management, and automated large-scale analysis based on modeling and machine learning. Our ever-increasing use of computers, sensors, and other digital devices is generating huge amounts of digital data, making it a pervasive NIT-enabled asset. In collaboration with NIT researchers, every agency should support research, to apply the best known methods and to develop new approaches and new techniques, to address data-rich problems that arise in its mission domain. Agencies should ensure access to and retention of critical community research data collections.

- NSF and DARPA, in collaboration with those agencies tackling problems whose solution entails instrumenting the physical world – including the Environmental Protection Agency (EPA), DoE, DoT, parts of DoD other than DARPA, NIH, the Department of Agriculture (USDA), and the National Oceanic and Atmospheric Administration (NOAA) – should increase research in advanced domain-specific sensors, integration of NIT into physical systems, and innovative robotics in order to enhance NIT-enabled inter-action with the physical world.

At the same time, new investments must not supplant continued investment in important core areas such as high performance computing, scalable systems and networking, software creation and evolution, and algorithms, in which government-funded research is making important progress. Topics of importance within these more established core areas continue to change in response to advances in technologies and applications. High performance computing (HPC) is a case in point. Although HPC plays a critical role in ensuring our national security, our economic competitiveness, and our scientific and technological leadership, the United States must anticipate and adapt to the broadening of its high-end computational needs and changes in the underlying technologies available to address them. Highly influential comparative rankings of the world’s fastest supercomputers are for the most part based on metrics relevant to only some of our national priorities, and must not be regarded as the sole measure of our continued leadership in this essential area. Although it is important that we not fall behind in the development and deployment of HPC systems that address pressing current needs, it is equally important that we not allow either the funding allocated to the procurement of large-scale HPC systems, or undue attention to a simplistic measure of competitiveness, to “crowd out” the
fundamental research in computer science and engineering that will be required to develop truly transformational next-generation HPC systems. To lay the groundwork for such systems, we will need to undertake a substantial and sustained program of fundamental research on hardware, architectures, algorithms and software with the potential for enabling game-changing advances in high-performance computing.

The Importance of Government Leadership

Many of our recommendations address multiple agencies – sometimes in collaborative roles, sometimes in coordinated roles, and sometimes in addressing different issues within an overall area of need. A successful coordinated attack on the Nation’s most challenging and important problems requires focused attention on multi-disciplinary, problem-driven research in NIT. That focus must come from Federal leadership. NITRD is chartered and staffed to coordinate multi-agency programs, not to create them. Strategic leadership, when necessary, must come from those with the authority to implement new strategies, namely the Office of Science and Technology Policy (OSTP) and the National Science and Technology Council (NSTC), to which NITRD reports. That leadership must have continuity, breadth and depth, and a focus on NIT.

Both the need for leadership and the need for broad multi-disciplinary research require action on the part of the Federal Government.

Recommendation [Section 11]: The Federal Government must lead in ensuring that strong multi-agency R&D investments are made in NIT to address important national priorities:

- OSTP should establish a broad, high-level standing committee of academic scientists, engineers, and industry leaders dedicated to providing sustained strategic advice in NIT.
- The NSTC should lead in defining and promoting the major NIT research initiatives that are required to achieve the most important existing and emerging national priorities.

In addition to ensuring that NIT research in support of the Nation’s priorities is conducted and that the results are translated into practice, it is essential that appropriately motivated and educated individuals are available as both researchers and practitioners. All indicators – all historical data and all projections – argue that NIT is the dominant factor in America’s science and technology employment, and that the gap between the demand for NIT talent and the supply of that talent is and will remain large. Increasing the number of graduates in NIT fields at all degree levels must be a national priority. Fundamental changes in K-12 education are needed to address this shortage. Here too the Federal Government must take the lead.

Recommendation [Section 9]: The NSTC’s Committee on STEM Education proposed in a recent PCAST report must exercise strong leadership to bring about fundamental changes in K-12 STEM education in the United States, among them the incorporation of computer science as an essential component.

Improved Effectiveness of NITRD Coordination

Thus far, we have focused primarily on the Federal NIT R&D portfolio and the need for multi-disciplinary collaboration in many areas. We now turn to the government coordination process for those investments.
The NITRD inter-agency coordination mechanism is widely – and we think correctly – viewed as successful and valuable. The collection of NITRD working groups has, over the years, enabled government research managers to become familiar with the activities of their colleagues in other agencies, and to formulate joint programs in areas of mutual interest. Nonetheless, steps can and should be taken to improve the effectiveness of the coordination process.

**Recommendation [Section 11]: The effectiveness of government coordination of NIT R&D should be enhanced:**

- The number of NITRD member agencies should be increased. The duration, management levels, and topic areas of the NITRD coordinating groups should be flexible. Budget reporting categories should be decoupled from the coordinating structure.

- The National Coordination Office (NCO) for NITRD should create a publicly available database of government-funded NIT research, and should provide regular detailed reporting to the Director of OSTP.

- The Office of Management and Budget (OMB) and OSTP should reflect NITRD priorities in their annual Budget Priority Memorandum. In addition, it is important to recognize the inherent limitations of any such process. In particular, each agency’s representatives are charged with advancing that agency’s mission, and not with devising a broader national strategy. As recommended previously, the NSTC must provide strategic leadership where necessary.

Continued attention must also be given to stable, evolvable, state-of-the-art shared NIT infrastructure for research, as well as new forms of infrastructure to support new research areas and paradigms. Shared NIT infrastructure – whether computational resources, communication networks, community databases (e.g., PubMed and the Protein Data Bank), or collaboration tools – has become essential to research in virtually all fields. NIT is one such field; NIT infrastructure that supports NIT research is a crucial component of NIT R&D, essential to achieving advancements in networking and information technology, which (among many other benefits) will yield the next generation of NIT infrastructure for all fields.

The Federal investment currently included in the NITRD crosscut budget includes NIT R&D, NIT infrastructure that supports NIT R&D, and NIT infrastructure that supports R&D in other fields. PubMed and the Protein Data Bank are examples of NIT investments that provide essential shared infrastructure for biomedical R&D; they do not represent NIT R&D. Similarly, high-end computing facilities, while essential for many types of research, are for the most part shared NIT infrastructure for physical, biological, and engineering fields other than NIT.

It is appropriate that investments in shared NIT infrastructure for R&D be included within the NITRD Program. However, it is important that investments in NIT that support R&D in other fields be clearly differentiated from investments in NIT R&D. A large portion of the “High End Computing Infrastructure and Applications” budget category, which accounts for roughly $1.5 billion of the $4.3 billion NITRD crosscut total, is attributable to computational infrastructure used to conduct R&D in other fields, and not to NIT R&D or infrastructure for NIT R&D. In addition, as illustrated earlier by the analysis of the NIH NITRD portfolio, various agencies include in their reports for other NITRD budget categories investments in NIT that support R&D in non-NIT fields. Thus the aggregate NITRD crosscut budget significantly overstates the actual Federal investment in NIT R&D. By leading policymakers to believe that we are spending much more on such activities than is actually the case, this discrepancy contributes to a substantial, systematic underinvestment in an area that is critical to our national and economic security.
Recommendation [Section 11]: The NCO and OMB should redefine the budget reporting categories to separate NIT infrastructure for R&D in other fields from NIT R&D, and should ensure more accurate reporting of both NIT infrastructure investment and NIT R&D investment.

In summary: The United States has a proud history of achievement and leadership in NIT that has yielded enormous benefits for our economic competitiveness, our national security, and our quality of life. Execution of recommendations in this report will play an essential role in ensuring the vitality of our Nation's NIT endeavors and enabling us to address our priorities and meet our challenges.

Crosscutting Themes

The five broad themes listed below recur throughout this report, and are of great importance to the future of all Federal agencies:

• Data volumes are growing exponentially. There are many reasons for this growth, including the creation of nearly all data today in digital form, a proliferation of sensors, and new data sources such as high-resolution imagery and video. The collection, management, and analysis of data is a fast-growing concern of NIT research. Automated analysis techniques such as data mining and machine learning facilitate the transformation of data into knowledge, and of knowledge into action. Every Federal agency needs to have a “big data” strategy.

• Engineering large software systems to ensure that they are secure (behaving as expected in the presence of an adversary) and trustworthy (behaving as expected in the absence of an adversary) remains a daunting challenge. The growing complexity of the systems we are building and our increasing societal reliance upon them outpace our ability to reason about them, and to engineer them to be secure and trustworthy.

• As NIT increasingly pervades daily life, systems are storing and processing a greater volume and diversity of private information about individuals. Privacy is a critical issue in all societal applications of NIT – most obviously in areas such as healthcare and electronic commerce, but also in areas such as energy, transportation, and education. Privacy challenges do not and must not require us to forgo the benefits of NIT in addressing national priorities. Rather, we need a practical science of privacy protection, based on fundamental advances in NIT, to provide us with tools we can use to reconcile privacy with progress.

• Interoperable interfaces – the means by which components of the smart grid can talk to each other, for example, or by which electronic health records can be shared and added to by many parties – are an important stimulus to technology innovation and adoption. Optimally, these interfaces would be open: anyone may create products that use the interfaces without paying fees; and a public, transparent process is used to establish and revise the standards that define the interfaces.

• The NIT supply chain is vulnerable. The hardware and software components used to build systems are sourced worldwide. We must anticipate and be prepared for various forms of threats to supply, quality, and security.

PCAST NITRD Program Review Working Group

Co-Chairs

David E. Shaw* Edward D. Lazowska
<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chief Scientist, D. E. Shaw Research</td>
<td>Bill &amp; Melinda Gates Chair in Computer Science &amp; Engineering</td>
<td>University of Washington</td>
</tr>
<tr>
<td>Senior Research Fellow, Center for Computational Biology and Bioinformatics</td>
<td>Director, eScience Institute</td>
<td></td>
</tr>
<tr>
<td>Columbia University</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* PCAST member

**Members**

<table>
<thead>
<tr>
<th>Francine Berman</th>
<th>Edward W. Felten</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vice President for Research</td>
<td>Professor of Computer Science and Public Affairs</td>
</tr>
<tr>
<td>Professor of Computer Science</td>
<td>Director, Center for Information Technology</td>
</tr>
<tr>
<td>Rensselaer Polytechnic Institute</td>
<td>Policy</td>
</tr>
<tr>
<td></td>
<td>Princeton University</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stephen Brobst</th>
<th>Susan L. Graham</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chief Technology Officer</td>
<td>Pehong Chen Distinguished Professor of Electrical Engineering and Computer</td>
</tr>
<tr>
<td>Teradata Corporation</td>
<td>Science Emerita and Professor of the Graduate School</td>
</tr>
<tr>
<td></td>
<td>University of California, Berkeley</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Randal E. Bryant</th>
<th>William Gropp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dean of the School of Computer Science</td>
<td>Paul and Cynthia Saylor Professor of Computer Science</td>
</tr>
<tr>
<td>Carnegie Mellon University</td>
<td>Deputy Director for Research, Institute for Advanced Computing Applications and Technologies</td>
</tr>
<tr>
<td></td>
<td>University of Illinois Urbana-Champaign</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mark Dean</th>
<th>Anita K. Jones</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM Fellow and Vice President</td>
<td>University Professor Emerita</td>
</tr>
<tr>
<td>IBM Research</td>
<td>University of Virginia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deborah Estrin</th>
<th>Paul Kurtz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jon Postel Professor of Computer Science</td>
<td>Managing Partner</td>
</tr>
<tr>
<td>Director, Center for Embedded</td>
<td>Good Harbor Consulting, LLC</td>
</tr>
<tr>
<td>Networked Sensing</td>
<td></td>
</tr>
<tr>
<td>University of California, Los Angeles</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Michael Kearns</th>
<th>Robert F. Sproull</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor of Computer and Information Science</td>
<td>Vice President and Director of Sun Labs</td>
</tr>
<tr>
<td>Founding Director, Market and Social Systems Engineering Program</td>
<td>Oracle</td>
</tr>
<tr>
<td>University of Pennsylvania</td>
<td></td>
</tr>
</tbody>
</table>

**Staff**

<table>
<thead>
<tr>
<th>Mary Maxon</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Deputy Executive Director</td>
<td></td>
</tr>
<tr>
<td>President’s Council of Advisors on Science and Technology</td>
<td></td>
</tr>
</tbody>
</table>
# Table of Contents

Executive Report  
PCAST NITRD Program Review Working Group  
1. Introduction  
1.1 The Organization of this Report  
1.2 A Preview of the NITRD Portfolio and the NITRD Coordination Process and Structure  
2. The Impact of Networking and Information Technology  
3. Recent Technological and Societal Trends  
4. The Role of Advances in NIT in Achieving America’s Priorities  
4.1 NIT for Health  
4.2 NIT for Energy and Transportation  
4.3 NIT for National and Homeland Security  
4.4 NIT for Discovery in Science & Engineering  
4.5 NIT for Education  
4.6 NIT for Digital Democracy  
5. Recommendations: Initiatives in NIT R&D to Achieve America’s Priorities  
6. NIT Research Frontiers  
6.1 NIT and People  
6.2 NIT and the Physical World  
6.3 Large-Scale Data Management and Analysis  
6.4 Trustworthy Systems and Cybersecurity  
6.5 Scalable Systems and Networking  
6.6 Software Creation and Evolution  
6.7 High Performance Computing  
7. Recommendations: Investments in the NIT Research Frontiers  
8. Technological and Human Resource Requirements  
8.1 Hardware, Software, and Data Infrastructure  
8.2 Education and Human Resources  
9. Recommendations: Technological and Human Resources  
10. Strengths and Limitations of the NITRD Coordination Process and Structure  
11. Recommendations: NITRD Coordination Process and Structure  
12. The Role of Federal Investment in NIT R&D  
12.1 The Critical Role of Federal Investment  
12.2 The Incremental Investment Implied by this Report  

Sidebars:  
Crosscutting Themes  
The Pervasiveness of NIT  
NIT and the Retail Revolution  
Interoperable Interfaces and Demonstration Testbeds Drive Innovation and Economic Growth  
Terrorists and Crooks: Internet-Enabled  
A Picture is Worth a Thousand Numbers  
Extracting Worldly Knowledge from the World Wide Web  
Improving Software Quality: "No Silver Bullet"  
Breaking the Speed Limit  
Progress in Algorithms Beats Moore’s Law  
The Ubiquitous Role of Privacy  
The NITRD Crosscut Budget Significantly Overstates the Actual Federal Investment in NIT R&D  
The Research Component of Industry R&D in NIT  
Why We’re Able to Google  

http://semanticommunity.info/A_NITRD_Dashboard/Designing_a_Digital_Future  
Updated: Sat, 19 Sep 2015 01:01:15 GMT  
Powered by mindtouch™
1. Introduction

1.1 The Organization of this Report

1.2 A Preview of the NITRD Portfolio and the NITRD Coordination Process and Structure

2. The Impact of Networking and Information Technology

3. Recent Technological and Societal Trends

4. The Role of Advances in NIT in Achieving America’s Priorities

4.1 NIT for Health

4.2 NIT for Energy and Transportation

4.3 NIT for National and Homeland Security

4.4 NIT for Discovery in Science & Engineering

4.5 NIT for Education

4.6 NIT for Digital Democracy

Information technology is transforming government operations and opening new communication channels between
government and citizens. A broad vision going by the name of digital democracy envisions the use of information
technologies to improve public discourse, increase dialogue between citizens and government, make government more
open and transparent, improve the operation of government, and bring the benefits of technology to everyone. Existing technologies have much to offer, and governments – from national to local levels – have been alert to many of these opportunities. However, unlocking the full benefits of digital democracy will require NIT research to address fundamental challenges.

**What digital democracy can deliver for society and government.** Digital democracy can transform our society over the next decade or two. In a world where digital democracy is fully realized:

- Government will be more responsive and accountable because citizens can see and measure how it responds to requests and problems and can more fully participate in planning processes. Personal technologies and social computing will enable real-time citizen reporting on civic needs and government service quality. Government will have unprecedented awareness and access to public opinion and public expertise, thanks to new methods of soliciting and gathering input from citizens. These tools will be available to all people, whenever and wherever needed, via Internet and mobile devices.

- The Internet will become a forum for substantive public collaboration and deliberation, thanks to new social technologies that foster positive interactions and are not polluted by spam and flame wars. All citizens will be able to make effective use of government data, thanks to breakthrough tools supporting access, analysis, and visualization for non-experts. Archivists, historians, journalists, and the public will have better and more convenient access to government records, including information previously available only in paper form.

- Digital tools will foster institutional innovation within government, making government services and processes faster, better, and cheaper. Government employees will have vastly better access to information and expertise, inside and outside government. Information and knowledge will be able to flow where they are needed, even as agencies maintain their distinct missions and cultures. Government regulation will be more effective and overregulation will be reduced because regulators have better information about where their actions are effective, and about the costs of overregulation. The availability of excellent, flexible tools and well-documented best practices will bring these benefits of digital democracy to all levels of government, including state and local.

- Publication of data will become a valuable public policy tool. By opening data sets to the public, government will foster transparency and competition, improving outcomes in the market and in public life.

- Election processes will be secure, convenient, inclusive, and strongly error-resistant, thanks to judicious use of NIT in voting and tabulation.

- Access to government and access to justice will be improved for all citizens.

**NIT research to support digital democracy.** Experience teaches that digital democracy requires much more than simply deploying technology. Success depends on judicious use of technology, on management and consensus building, and on building a culture of continual and iterative institutional innovation inside and outside government. At present, much effort is spent wrestling with the technical details of data formats, system architectures, and technology management.

What may be less obvious is that the practical difficulties of deploying transformative technology are often manifestations of deep technical challenges that NIT research can address. An NIT R&D program aimed at these challenges can open new opportunities for digital democracy that are not possible today. Some examples of these research opportunities are:

*Privacy and security:* Methods for putting datasets online without compromising citizen privacy. Methods for verifying, without compromising privacy, that an online commenter is a real person, or is not the same person using a different identity. Effective methods for users to verify the authenticity of information from government datasets.

*Social computing:* Understanding how to make online interaction engaging for citizens. Mechanism design for public deliberation: understanding how to structure and (semi-)automatically manage discussion forums to keep the discussion on-point and productive. Helping citizens find and help each other. Measuring public opinion and...
sentiment in ways that resist gaming and strategic behavior. Specializing social media and crowdsourcing technologies for use in government.

**E-voting**: Methods for voting from remote locations, e.g., for overseas military personnel, that fully address inherent security risks. Making voting more convenient without compromising security. Protecting the secrecy of the ballot in an environment of ubiquitous, high-resolution cameras and sensors.

**Computer-aided text analysis and processing**: Methods for government employees to sift through large sets of comments and discussion to find relevant information. Methods for automatically summarizing discussions and extracting the most important arguments and facts.

**Data analysis tools**: Tools to aid export of legacy datasets, making them available to the public with whatever documentation exists, as well as porting of data, which connotes greater effort to make the data more easily understandable, standardize formats, create thorough explanatory documentation, and so on. Tools for analyzing the format and semantics of unstructured or poorly documented data sets. Tools for finding likely errors in diverse datasets. Processes for managing errors, including error handling and metrics for responsiveness.

**Data tools for non-experts**: Tools to let non-experts effectively access, analyze, and visualize government data, and to publish the results of their analyses. (See the sidebar “A Picture is Worth a Thousand Numbers on the previous page”).

---

**A Picture is Worth a Thousand Numbers**

Data visualizations can illuminate trends in large and complex data sets. Visualizations have typically required significant time and expertise to create. But new tools allow anyone to create useful visualizations, broadening participation in public debate by letting ordinary citizens explore government datasets and create persuasive arguments from them.

Two examples of recent highly usable data visualization tools are Tableau and Many Eyes (a publicly available experiment by IBM Research). Many Eyes, as of this writing, features citizen-created visualizations of carbon emissions by G-20 vs. non-G-20 countries, sources of Pakistan flood aid, and teacher starting salaries by state, among many others. These visualizations are built using data-exploration tools on the site, from data sets provided by sources such as the Census Bureau, the Bureau of Labor Statistics, and the Federal Reserve. Visualizations can be embedded in external web sites, so they can enter the public conversation.

As open government initiatives put more data online, the value of further advances in easy-to-use visualization tools will only increase. What was once available only to experts will now be open to any citizen.

---

**Shorter-term development for digital democracy**. There are many things the Federal Government can do, and is doing, to advance digital democracy. The Office of Science and Technology Policy (OSTP) Open Government Initiative, which includes the opening of government datasets at Data.gov and the publication of government spending data, is a valuable first step. Progress toward opening data and documents at OSTP, the National Archives and Records Administration (NARA), and elsewhere, deserve continued effort and attention.

The Community Health Data Initiative (CHDI) is just one example where OSTP and the Department of Health and Human Services (HHS) are taking a leadership role in targeting NIT innovation toward today’s information needs in community and public health. The effort engages private and public players and promotes engagement by a broad range of organizations, from county health departments to patient advocacy groups to social media startups. This strong beginning could be further enhanced by leveraging already available technical capabilities in new ways in a form of technical and social innovation. For example, moving beyond the integration of existing institutional data sets, the CHDI
could create a channel for citizen participation and dissemination of citizen-generated data using mobile, social, and traditional web media. Similar efforts throughout government deserve continued support.

Success in digital democracy requires participation from both government and the public (including companies and non-profits). Some things can only be done by government, e.g., publication of government-held data as in the Data.gov initiative and the USAspending.gov dataset. Other things are best done by non-government actors, e.g., organizing political discussions related to government data and activities. In some areas, such as defining data-driven metrics for quality of healthcare, both government and private parties may have things to offer. It is important to enable innovation both inside and outside of government.

Much can be learned from the impact of NIT on business. Initially, organizations transfer their legacy paper-based processes into the electronic realm. This offers relatively modest benefits. Over time, processes are re-engineered to truly take advantage of what the digital world can offer. This is a slow process – often taking a decade or more – but it unlocks the big benefits of going digital. This will surely be the pattern in government, with the greatest benefits of digital democracy only becoming evident over time.

The difficulty of changing government processes presents additional challenges. But getting this transformation right offers huge benefits: by making government more efficient, more responsive, and more transparent, we can make it more effective at addressing the challenges of the 21st century.

5. Recommendations: Initiatives in NIT R&D to Achieve America’s Priorities

6. NIT Research Frontiers

6.1 NIT and People

6.2 NIT and the Physical World

6.3 Large-Scale Data Management and Analysis

Collecting, storing, preserving, managing, analyzing, and sharing exponentially increasing quantities of data present a variety of significant NIT challenges that research must address. A vast range of data sources, from webcams and weblog posts to telescopes and supercomputer simulations, is flooding our world with enormous amounts of data in many different forms. These data are stored in many different formats and many different environments, from computer hard drives to large-scale data warehouses. Numerous unresolved issues arise in attempts to ensure that all of this data maintains its integrity and availability, both now and long into the future. Making use of the data presents another set of challenges. Advanced machine-learning algorithms enable sophisticated analyses of data sets, leading to breakthroughs in science, medicine, commerce, and national security (see sidebar “Extracting Worldly Knowledge from the World Wide Web”, on page 50 for one example). Graphical visualization and other methods allow people to gain valuable insights from large collections of data. To gain maximum advantage from data sets that continue to grow larger and more complex, new techniques are needed in both these areas. There is also a critical need for better techniques for sharing both data and the results of data analyses while respecting the privacy rights of individuals and the needs for
government and corporate confidentiality. Effective use of data will be critical to meeting every one of this report’s technical priorities. Below we describe the fundamental elements of working with data and some of the key R&D challenges and coordination needs they engender.

### Extracting Worldly Knowledge from the World Wide Web

Although the field of machine learning has made enormous progress in recent years, the scope for further research remains almost unlimited.

People learn throughout their lifetimes, and at an accelerating pace—accumulated knowledge facilitates further learning. For example, to understand the difference between the sentences “Clarissa went to the store in her car” and “Clarissa went to the store in her neighborhood,” we need to know that a car is a vehicle while a neighborhood is a place. Various computer science research projects over the years have attempted to manually create “knowledge bases” containing essential information of this sort, but the task has proved impractical.

Instead, the Never Ending Language Learning (NELL) project at Carnegie Mellon University is extracting these facts from the World Wide Web. Starting with a small set of categories and examples, NELL seeks statistical reinforcements for the facts it has already learned, and uses what it knows to extract new facts. For example, when it sees the statement “Michael McGinn is the Mayor of Seattle,” it can infer that Michael McGinn is a person and that Seattle is a city. When it later sees “Dallas is bigger than Seattle,” it can infer that Dallas is also a city. NELL uses a collection of 1 billion pages extracted from the World Wide Web by means of a system that Google provided to NSF-sponsored researchers. Each iteration of its analysis requires around four hours on a massive computer system made available to university researchers by Yahoo!. Funding for the project has come from DARPA and Google. As of October 2010, NELL had extracted nearly 500,000 facts with an estimated accuracy of at least 87%. Other researchers are starting to use NELL’s knowledge base to improve natural language understanding programs and to create better World Wide Web search engines.

We use two additional practical examples to illustrate both the promises and the challenges of large-scale data collection and analysis. First, consider a consortium of cancer research institutes creating and managing a data repository based on data collected from millions of patients. These data include copies of medical histories in textual and spoken form, x-ray and MRI images, and test results from biopsies and genetic microarrays. Assembling and managing all of these data is a monumental task, but it could lead to much a deeper understanding of disease processes and the way different medications affect different populations. As a second example, consider the case of the FBI collecting data from many different sources in many different forms: video surveillance data, intercepted email and telephone calls, law enforcement records, and even online information, such as web pages, videos, and weblog posts. Among the masses of information are small traces indicating the activities of crime rings and terrorist organizations. This data must be collected in ways that preserve individual rights against unreasonable searches and seizures, but by analyzing this information, the FBI may be able to uncover and disrupt major threats to the safety and security of U.S. citizens. In both these examples, management and preservation of the data to ensure its future accessibility is an important element of our ability to gain new insights and understanding.

**Research challenges in data collection, storage, and management.** It is estimated that around 1.2 zettabytes (1.2 billion terabytes) of digital data are generated worldwide each year by numerous devices in numerous forms: remote sensors, online retail transactions, text documents, email messages, web posts, camera and video images, computers...
running large-scale simulations, and scientific instruments such as particle accelerators and telescopes. The core technology for data storage, especially magnetic disks, has progressed rapidly, enabling government, research, and corporate organizations to create massive data warehouses that can store much of the data as fast as it is created. But storing raw data is just a small part of the larger issues of creating and maintaining data repositories. We must view data repositories as archives requiring long-term stewardship based on sustainable economic models. Important issues include:

**Representations**: How to adopt and evolve standards for important categories of information. These representations must allow different companies and organizations to create software tools that generate, manipulate, and analyze societally important data. Left on its own, the software industry is likely to create a number of incompatible, proprietary standards that become obsolete. (Consider, for example, the case of word-processing formats and the fact that the Federal Government still mandates using WordPerfect format for official documents, long after most organizations have transitioned to other software.)

**Detecting and correcting errors or inaccuracies in the data**: Although various forms of outlier detection have been developed and applied, these methods need to be more sophisticated and comprehensive when applied to data sets of societal importance.

**Support for data management policies**: Systems to support data privacy and access limitations, retention requirements, requirements for mechanisms to reduce the risk of data loss or damage, and other aspects of increasingly stringent data policy and regulatory requirements.

**Data provenance**: Tracking how, where, and when data are created and modified. This is an important and often overlooked aspect of data stewardship.

**Data integrity**: Ensuring that data are not corrupted either accidentally or maliciously.

**Data storage engineering**: Ensuring reliability, reducing power consumption, incorporating new technology. Management of data across multiple storage technologies and multiple hierarchies, and with replication across multiple geographic locations. Continued research is required to adapt to changing technology (e.g., nonvolatile RAM), performance requirements, and the need to provide consistent views of data worldwide.

**Development of sustainable economic models**: Necessary for supporting data access and preservation over the long term, especially beyond the durations of typical research grants.

Many of these requirements appear in our cancer research institute example. Interoperability for electronic health records, considering both current and future needs and technology, are of critical national importance both to realize the promise of better healthcare and to enable the use of patient data for medical research. As for recording and tracking data provenance, suppose it was determined that an automated blood analysis instrument had been giving faulty readings over a one-month period. It should be possible to identify all patients who might need to be retested and all scientific analyses that may have been tainted by faulty data. It is crucial to maintain original data for medical and scientific research to enable validation of results and to support longitudinal studies. In addition, regulations such as the Health Insurance Portability and Accountability Act (HIPAA) create many requirements for data access and retention. Similarly, the FBI must collect and manage data in ways that satisfy rules of evidence. It must keep track of data provenance so that it can later carry out deeper investigations of critical information and use the information when...
prosecuting legal cases. Errors either in the initial data or due to subsequent corruptions could have devastating effects on innocent people, as well as on the ability of the FBI to carry out its mission.

**Research challenges in data analysis.** Increasingly sophisticated methods of data mining and machine learning allow us to extract more and more useful insights from many data sources. Prominent recent examples include search engines, automated language translation, customer recommendation systems, and credit card fraud detection. This is an area of very active research with ever-increasing capabilities, but also with increasingly high expectations. Important issues include:

**Systems:** Engineering computer systems that can perform complex processing of data on very large scales. Internet-based industries are building computer systems of unprecedented size to house and analyze their data and to serve millions of customers worldwide. Comparable systems could also provide powerful capabilities for scientific research, for making government data available to citizens, and for national security.

**Algorithms:** Developing more sophisticated machine learning techniques, especially ones that apply to very large data sets. Machine learning is still in its infancy, and we can reliably predict that great strides will be made in creating algorithms with new capabilities that can scale to handle the very large data sets being generated now and in the future.

**Programming:** Computational models and languages suited for expressing data analysis algorithms that map onto large-scale, parallel systems. Recently developed proprietary and open-source programming tools have demonstrated greatly enhanced scalability and programmer productivity. These tools and models must be extended and refined to handle wider classes of applications and to make them easier for non-specialists to use.

**Cross-media information extraction:** Understanding speech, images, video, and unstructured data; translating speech and text to other languages. Although these topics have been the subject of decades of research, new data-driven approaches promise to be much more effective.

**Information fusion:** Analysis that combines multiple data sources in multiple different forms. Many important insights can be gained by analyzing different representations of a single event or phenomenon.

Using our cancer research institute as an example, we can imagine important medical breakthroughs being made by systematically analyzing imaging data (x-ray, MRI) along with patient histories, including automatic transcription of dictations from patients and caregivers. These can lead to new diagnostic regimens that are far more efficient and effective than today’s methods, which rely heavily on the experience and judgment of medical specialists. Dealing with the data from millions of patients will require computing capabilities far beyond those currently being used in medical research, and will require collaborations between medical researchers and a wide range of computer scientists and engineers. Interoperability will permit the analysis of much larger patient populations.

For the case of the FBI, we can see that the activities of a criminal or terrorist organization will show up in many different forms. The patterns of communication between different individuals, via phones and email, can be analyzed as a social network, revealing its command structure and ways that it can most effectively be disrupted. It would be possible to track the movements of individuals through the locations of their phone calls, their use of public transportation and credit cards, and from video surveillance. Creating a comprehensive picture of these activities from the many different forms of data requires much higher levels of automation and sophistication than exists today.
**Research challenges in controlled and effective data sharing.** Sharing different forms and different aspects of data offers important societal and organizational benefits. However, numerous instances in which researchers have been able to breach the confidentiality of supposedly anonymized data sets demonstrate the difficulty of sharing data in the face of increasingly sophisticated analytic techniques. Some key issues include:

*Models and algorithms for controlled data sharing:* Current statistically-based anonymization methods provide no real guarantees for data privacy. For example, such methods often assume that all data come from a single dataset, whereas many breaches of privacy result from correlating multiple data sources. The recently devised differential privacy model, on the other hand, considers the existence of additional data sources. Developing and applying algorithms based on such models is crucial to taking full advantage of the wealth of data sources available to society.

*Data presentation and visualization:* Making complex data understandable to people, specialist and non-specialists alike. This involves determining what information to feature, and how. Visualizations of cancer tumors, Internet traffic, weather patterns, sociological data, etc., are important to facilitate new insights critical to understanding.

Again, we can see how these issues arise for our cancer research institute example. Although medical record privacy is considered very important, current policies and regulations are a patchwork of poorly specified and ineffective rules. With healthcare data, as with other data sets concerning individuals and organizations, we must greatly improve our ability to derive and share useful insights from data while preserving privacy and confidentiality. Otherwise, there is a major risk that either private data will be disclosed, or that we will have to impose such strict controls that useful information cannot be shared.

For the case of the FBI, there are many instances where data must be shared with other organizations – local law enforcement agencies, the security services of other countries, and even the public – but this must be done in ways that protect the rights of individuals and do not accidentally leak information about covert sources and methods. Current methods of classifying information, unfortunately, are not sufficiently reliable in the face of sophisticated data analysis methods. They are also far too labor intensive.

6.4 Trustworthy Systems and Cybersecurity

6.5 Scalable Systems and Networking

6.6 Software Creation and Evolution

6.7 High Performance Computing

7. Recommendations: Investments in the NIT Research Frontiers
Introduction

Advances in NIT rest on a broad and deep foundation of more than 60 years of fundamental research. That foundation, which is divided for convenience into a collection of core areas, continues to evolve as changes in technologies and new uses of NIT stimulate new breakthroughs and deeper understanding. In order to make progress in the uses of NIT, continuing research in core areas is essential.

The following summary recommendation, which appears in the Executive Report, highlights the most important elements of the more detailed recommendations that appear later in this section. We note again the importance of high risk/high reward research with the potential to move these areas in unanticipated directions.

Recommendation: The Federal Government must increase investment in those fundamental NIT research frontiers that will accelerate progress across a broad range of priorities. Among such investments:

- NSF and DARPA, with the participation of other relevant agencies, should invest in a broad, multiagency research program on the fundamentals of privacy protection and protected disclosure of confidential data. Privacy and confidentiality concerns arise in virtually all uses of NIT.
- NSF, DARPA, and HHS should create a collaborative research program that augments the study of individual human-computer interaction with a comprehensive investigation to understand and advance human-machine and social collaboration and problem-solving in a networked, on-line environment where large numbers of people participate in common activities. Understanding such collective human-NIT interactions is increasingly important for defense, for health, and for the activities of daily life.
- NSF should expand its support for fundamental research in data collection, storage, management, and automated large-scale analysis based on modeling and machine learning. Our ever-increasing use of computers, sensors, and other digital devices is generating huge amounts of digital data, making it a pervasive NIT-enabled asset. In collaboration with NIT researchers, every agency should support research, to apply the best known methods and to develop new approaches and new techniques to address data-rich problems that arise in its mission domain. Agencies should ensure access to and retention of critical community research data collections.
- NSF and DARPA, in collaboration with those agencies tackling problems whose solution entails instrumenting the physical world – including the Environmental Protection Agency (EPA), DoE, DoT, other parts of DoD, NIH, the Department of Agriculture (USDA), and the National Oceanic and Atmospheric Administration (NOAA) – should increase research in advanced domain-specific sensors, integration of NIT into physical systems, and innovative robotics in order to enhance NIT-enabled interaction with the physical world.

The recommendations that follow describe some areas in which additional investment is needed in order to realize the advances that our national priorities require, and other areas in which attention must be directed to particular challenges. These investments should supplement ongoing research in more established areas such as algorithms and computer graphics that are not called out explicitly in this report.

Recommendation: New investments must not supplant continued investment in important core areas in which government-funded research is advancing. Continued attention must also be given to sustained high-quality shared research infrastructure, including new forms of infrastructure to support new research areas and paradigms.

Privacy and Confidentiality

Preserving personal privacy is a critical need that pervades NIT. Our democratic society puts a high value of protection of personal privacy, while the protection of corporate information is a key element of competitiveness. Controls on data disclosure are essential to protecting the safety of individuals and of our Nation. All NITRD agencies should evaluate how considerations of privacy and confidentiality will affect the deployment and use of the technologies emerging from their NIT R&D efforts. In some cases the issues are not well understood. The Federal Government needs to support R&D to better understand, address, and ameliorate the privacy and confidentiality issues that are identified.

http://semanticommunity.info/A_NITRD_Dashboard/Designing_a_Digital_Future
Updated: Sat, 19 Sep 2015 01:01:15 GMT
Recommendation: NSF and DARPA, with the participation of other relevant agencies, should invest in a broad multi-agency research program on the fundamentals of privacy protection and protected disclosure of confidential data. The program should address at least the following important issues:

- developing methods that allow agents – that is, individuals or software acting in well-defined and suitably constrained roles – to perform analytics on large datasets while preserving privacy and confidentiality;
- creating and investigating formal models of privacy that combine concepts from statistics and computer science; these models should be characterized in terms of what guarantees they provide, what adversaries they can withstand, and what forms of sharing they permit;
- understanding the consequences for privacy and confidentiality of technology trends such as large-scale data gathering, analytics, correlations of multiple sources, machine learning, and ubiquitous sensors, as well as of protective regimes such as cybersecurity;
- devising methods that give individuals knowledge of what data about them is held and appropriate control over the use of that data;
- exploring the privacy-preserving design of human-centered systems that create and use information about people, in financial, medical, demographic, and residential domains;
- creating ways to educate users about, and protect users against, actions they might take that inadvertently compromise privacy;
- using the fruits of research into privacy and confidentiality protection to enable privacy- and confidentiality-related policies to be stated in application-relevant terms, and enforced in application-specific ways; one example is coordination with agencies including NIH, HHS, the Department of Commerce, and the Department of Justice (DoJ) to ensure that policies and regulations concerning medical records, census data, and other socially important datasets are based on sound scientific principles.

The Ubiquitous Role of Privacy

Online privacy is already a significant issue for many Americans – witness the debate over privacy on social networks and the rise of identity theft. Technology trends can only raise the stakes.

Privacy challenges clearly arise for electronic medical records, but in fact they come up in all of the national priority areas. The smart grid will save energy by instrumenting, analyzing, and optimizing power usage within a home – but these actions convey information about activities within the home. Smart transport will reduce congestion and save energy by optimizing the movement of individual vehicles and tracking people’s travel needs – but these details convey information about personal activities. Personalized education will use data about a student’s education history and progress to offer the best instruction – but this, too, is sensitive information.

We cannot afford to forgo the benefits of NIT in addressing national priorities. Rather, we need fundamental advances in NIT – a practical science of privacy protection – to give us the tools to reconcile privacy with progress.

Privacy challenges arise whenever we want to allow access to information for some purposes but not for others. A public health researcher may want to search for subtle trends hidden in a large collection of patients’ health records, but we shouldn’t allow the researcher to learn the details of any specific patient’s record. Ideally, we could scrub or anonymize data before giving it to the researcher, but doing this safely, without scrubbing away the very trends the researcher is seeking, is difficult in theory and risky in practice.

Privacy challenges are complicated by problems of inference and side information. Revealing a fact implicitly reveals everything that can be inferred from it. Revealing a patient’s prescriptions, for example, could implicitly reveal the patient’s medical conditions. Inferred facts lead to further inferences; one seemingly innocuous fact can
trigger a cascade of inferences. It is difficult to catalog and control all methods of drawing inferences. Worse yet, an analyst can combine revealed facts with all of the side information available from other sources, so that our concerns about the possible extent of inferences must take into account all the side information that might be available. These are difficult problems, but there is hope that fundamental NIT research can address them.

Understanding how to reconcile privacy with the application of NIT to national goals, both in general and with respect to specific goals, will improve Americans’ privacy while enabling ever more beneficial application of NIT.

NIT and People

The modes and the ease with which people interact with computers have improved as richer forms of interaction and better understanding of human capabilities have informed the design of interactive systems. The advent of widely available networking and the introduction of digital consumer products have further empowered people. We are now experiencing another spurt of growth – into the realms of social computing and media, NIT-enabled social science, and collective interaction.

Recommendation: NSF, DARPA, and NIH should create a research program that augments the study of individual human-computer interaction with a comprehensive investigation to understand and advance human-machine collaboration and problem-solving in a networked, online environment. The program should:

- create a science of social computing that, for example, gives insight into how to organize human contributions, how to incentivize participants, and how to design generic social-computing frameworks that could be used by different organizations for diverse purposes;
- foster research that pushes the field beyond the current examples of crowd-sourcing;
- encourage theoretical, algorithmic and engineering foundations that guide the design of peer-production systems (in which large groups of individuals, sometimes tens or even hundreds of thousands, collaborate online) for a wide variety of tasks;
- design novel mission-specific uses of collaborative computing;
- create shared privacy-preserving research platforms to enable researchers in computational social science to share and exchange experimental designs, behavioral experimental data, and human subject panels and subjects. For example, a promising application area for such experimental research is the study of human decision-making regarding security and privacy issues, so as to inform technology and design considerations in those areas.

NIT and the Physical World

The 2007 PCAST assessment 63 of the NITRD Program spurred a new emphasis on Cyber-Physical Systems. New and valuable research activities were launched, and important programs and collaborations were fostered involving NSF, DARPA, and industry. We recommend expanding and deepening these efforts in three particular areas: sensor development, robotics, and open architectures.

Recommendation: NSF, in collaboration with those agencies tackling problems whose solutions entail instrumenting the physical world – including EPA, DoE, DoT, DoD, NIH, USDA, and NOAA – should conduct research to design, fabricate, and test sensors that are problem-domain specific and that are cheaper, smaller, better packaged, lower powered, and more autonomous than those available today. These advances would lead to new applications and new markets that would be sustained in the long term through traditional commercial activities.
**Recommendation:** DARPA, NSF, NIH, and DoE should continue to sponsor research on large-scale modular robotics and computer vision and should collaborate to increase the rate of innovation, usability, and scalability of autonomous actuation in environmental, medical, manufacturing and defense contexts.

**Recommendation:** Interoperability is essential for enabling NIT to be broadly embedded in the physical world. Early adoption of both will create an open and fair playing field for Federal agencies to see competition among their suppliers, and for commercial products built by different companies to operate compatibly. If the United States creates the open architectures and standards, U.S. industry will gain an early advantage. NIST, in consultation with NSF, should lead an interagency effort to build consensus and fund reference implementations.

**Large-Scale Data Management and Analysis**

Virtually every Federal agency has opportunities to capitalize on the analysis of large volumes of data to further its mission. The data might be human-generated in electronic form, it might be obtained from sensors or other observational tools, it might be derived from computer simulations, or it might be the by-product of other kinds of NIT applications. As data becomes increasingly abundant, fundamental research to advance our expertise in collecting, analyzing, understanding and using that data becomes increasingly urgent. In addition to the recommendations below, controlled data sharing and data privacy are critical issues; these topics are addressed in the "Privacy and Confidentiality" recommendations above.

**Recommendation:** NSF should expand its support for fundamental research in data collection, storage, management, and analysis. Its programs should address topics such as:

- contextual metadata for data obtained from sensors in the physical world;
- information derived from cross-correlation;
- information fusion algorithms that combine data from diverse sources, differing scales, differing types and metadata;
- long-term preservation;
- data provenance and integrity;
- inference from incomplete and uncertain data;
- deep analysis of the information contained in the data;
- abstraction, summarization, and visualization of complex data and information.

**Recommendation:** Every agency should engage in R&D to apply the best existing methods and to develop new approaches and new techniques to address data-rich problems in its mission domain. Collaboration between NIT researchers and domain experts is essential to this work.

**Recommendation:** Under the leadership of NIST, processes and policies should be established for the NITRD agencies to publish real-world primary data sources such as detailed logs of events or sensor readings, in order to facilitate research into techniques for mining and abstracting those sorts of data.

The private sector should be encouraged to participate as well. Examples include a “click stream” of users trying to find data on a government web site, detailed readings of water flows and levels, or video surveillance of busy highways. Care must be taken to release data in a privacy-preserving way based on sound scientific principles.

**Trustworthy Systems and Cybersecurity**

The trustworthiness and security of NIT systems are characteristics that transcend the uses to which these systems are put. All systems must be secure against unintended behaviors, against unauthorized access, and against threats to their availability and integrity.
Recommendation: NSF and DARPA should aggressively accelerate their initiatives to fund and coordinate fundamental research to find more effective ways to build trustworthy systems and to assure cybersecurity. These initiatives should include programs to:

- Advance the art and the practice of designing and implementing trustworthy systems, which act only as users expect them to act even in the face of failures. Develop methods to analyze the trustworthiness of designs and implementations. The research should focus explicitly on systems critical to society;
- Develop fundamentally new “clean slate” designs and outside-the-box approaches that will provide a new basis for assuring the security of systems and data. These new approaches should provide a basis for relating classes of attacks, specific (possibly new) defense designs, and security policies, so that more careful reasoning and analysis can be performed;
- Develop methods suitable for implementing a small, rigorously isolated set of very basic capabilities that can be relied upon with a high degree of confidence to provide truly essential NIT-based services in the event of, for example, a catastrophically damaging cyber-attack.

Scalable Systems and Networking

Research on the design and implementation of scalable systems has a long history in computer science. Both NSF and DARPA have supported relevant research for many years. This is a critical core activity for computer science research, and continued investment is required to keep up with changing application needs and technology capabilities. Continued progress will draw on expertise ranging across many different system layers – an effort that only collaborative, multidisciplinary teams can provide. The overall system environment today consists of many layers under the control of different companies, governments, and standards organizations. Support for the fundamental advances the Nation needs demands higher levels of coordination between these entities than now exists.

Recommendation: NSF, DARPA, and other organizations should continue their leadership and funding of core research into scalable systems in order to ensure that networked systems will adapt to the ever-changing needs of applications, to the capabilities engendered by new technology, and to evolving needs for security and privacy.

Recommendation: To foster an innovative ecosystem in NIT, government agencies concerned with networked systems operations, including DoD, NSF, and the Federal Communications Commission (FCC), must continue to encourage and invest in open systems development. They must coordinate with standards organizations (IETF, W3C) and the NIT industry to ensure that standards pertaining to the different interfaces within and among the layers of the networked systems environment are defined and kept up to date.

Recommendation: In the area of wireless systems, NSF, FCC, and the National Telecommunications and Information Administration (NTIA) should partner to create, sustain, and promote the use of a nationwide infrastructure for spectrum monitoring that cuts across commercial, public safety and DoD applications. NSF, DHS, and NTIA should partner to create programs that promote innovative use of public safety frequencies. NSF, DHS, and DARPA should jointly articulate the synergies among their individual needs and programs in wireless spectrum management.

Software Creation and Evolution

Over the past several decades, software research has made major advances in our ability to create increasingly large, complex, and critical software systems. The continuing emergence of new challenges requires a steady stream of new advances. Investment in software research must be sustained.

Recommendation: NSF, DARPA, and other organizations that need software tailored to their mission requirements should continue their leadership and funding of core research in methods to improve the design, development, modification, and maintenance of all varieties of software. That research should address language design, tools, analysis methods, methods for collaborative design and development, and techniques that provide security and robustness. Attention must be given to system design and programming for scalability, paradigms for parallelism at
multiple levels of granularity, software for heterogeneous systems involving interaction with the physical world, and software for systems that incorporate human interaction. Long term evaluative research is required to determine which tools and techniques yield sustainable improvement in software creation.

**High Performance Computing**

HPC is increasingly important for research in many areas of science and engineering. It is essential to national security, and is a major tool in addressing other important national priorities. In order to maintain its historical leadership in the design and effective utilization of HPC, the United States must anticipate and adapt to the broadening of its high-end computational needs and to changes in the underlying technologies available to address them. The primary focus must be on advances that will address important national needs, and not on the relative ranking of each country’s fastest supercomputer on the Top500 list.

**Recommendation:** NSF, DARPA, and DoE should invest in a coordinated program of basic research on architectures, algorithms and software for next-generation HPC systems. Such research should not be limited to the acceleration of traditional applications, but should include work on systems capable of (a) efficiently analyzing vast quantities of both numerical and non-numerical data, (b) handling problems requiring real-time response, and (c) accelerating new applications. Specific areas of investigation should include:

- Novel system architectures for massively parallel computing
- High-bandwidth, low-latency processor interconnection networks
- Reliability and fault-tolerance in massively parallel computer systems
- Hardware and software design techniques for the dramatic reduction of power consumption
- Data-intensive computing, including non-numerical applications
- Programming models and languages for massively parallel machines
- Systems software for massively parallel systems
- Improved approaches for system management

In addition to designing next-generation systems, significant effort must be devoted to R&D focused on extracting the greatest possible scientific benefit from current leading-edge systems.

**8. Technological and Human Resource Requirements**

**8.1 Hardware, Software, and Data Infrastructure**

**8.2 Education and Human Resources**

**9. Recommendations: Technological and Human Resources**

**Hardware, Software, and Data Infrastructure**

NIT infrastructure – be it computational resources, communication networks, community databases, or collaboration tools – has become essential to research in virtually all fields. Although some infrastructure is acquired and managed exclusively by individual research projects, many research fields benefit from access to large-scale shared infrastructure – shared because of the considerable expense of acquiring and maintaining it, because of the long-term need, or because of the desire that multiple researchers use a common base. High-end computing systems, such as those made
available the NSF Supercomputer Centers, and collections of curated data, such as PubMed and the Protein Data Bank, are examples of large-scale shared NIT infrastructure. High-end computing infrastructure provided by the NSF Centers and similar facilities has a long history, but shared infrastructure that supports data access, management, use, and preservation is in its early stages. The health of the Nation’s research enterprise depends on sustained and reliable infrastructure of both kinds.

An important observation is that virtual or physical computing centers that provide infrastructure services for general R&D can often boost their value by hosting some NIT research activities as well. Some NIT research that serves to advance the technology underlying the infrastructure can be conducted using that infrastructure without disruption to other users. In addition, some science and engineering research using computer centers may confer extra value by stress-testing the infrastructure and providing a cadre of skilled consultants to help other users.

**Recommendation:** With NSF taking the lead, the NITRD agencies should develop an improved framework for the development and support of shared large-scale research infrastructure with the following properties:

- Proposed NIT infrastructure projects should be evaluated not only on whether they satisfy a demonstrated need, but also on their adaptability, reliability, adoptability, stability, size of user base, capability, and other appropriate metrics of infrastructure success, as well as on their plans for sustaining the infrastructure over time.

- Shared large-scale infrastructure is best managed with robust rather than minimal levels of support. To that end, the organizations that develop and manage large scale computation and data infrastructure should be constituted so that researchers working on problems for different agencies can perform their work at these common centers. (Occasionally, mission agency constraints will preclude such sharing, for example when data is classified.)

- In budgetary summaries, large-scale infrastructure costs for NIT resources devoted to R&D in areas other than NIT (e.g., in physics or medicine) should be clearly designated as infrastructure for those disciplines rather than mislabeled as NIT R&D. NIT R&D should be explicitly called out in budget summaries, as is R&D for other user communities.

- Plans and practices should be defined and implemented to manage the curation and preservation of long-lived and large data sets, so that data infrastructures survive beyond the lifetime and the boundaries of the projects that generated them.

**Recommendation:** NITRD should initiate a proactive approach for supporting data-driven research and the preservation of research data. We propose that NITRD work with each agency to designate critical data collections important to their communities and “best of breed” repositories to foster sustainable data infrastructure. This data infrastructure should follow best practices in curation and community standards, offer broad access for the research community, and ensure the sustainability of community data needed for new discovery. Programs should exploit the capabilities of the private sector, university libraries, government repositories, and other facilities that can support best practices and sustainable business models for broad access to long-lived digital information.

### Education and Human Resources

The ever-expanding role of NIT in our society creates an ever-increasing demand not only for NIT professionals, but also for individuals who can utilize NIT flexibly and creatively and who can apply NIT “modes of thought” in a wide variety of endeavors. The Nation must take concrete steps to ensure that the American people have the education and skills to meet that demand.

**Recommendation:** The NSTC’s Committee on STEM Education proposed in a recent PCAST report must exercise strong leadership to bring about fundamental changes in K-12 STEM education in the United States, among them the incorporation of computer science as an essential component.
Research is needed to inform the necessary changes to STEM education. That research must address both curriculum content and understanding of the motivations and incentives that will encourage students to seek and persevere in STEM education.

**Recommendation:** NSF and ED should fund research to determine an age-appropriate progression of concepts for STEM education in computer science that generates strong skills in fluency, computational thinking, and the science and engineering aspects of computer science. That research should include the creation and assessment of the best ways to enable students to learn those concepts. The agencies should work with the academic community to determine and continuously update the appropriate concepts. 83

**Recommendation:** NSF and ED should fund research to analyze why people do or do not choose to become computing professionals and why students of all ages, from childhood to post-graduate, do or do not choose to study computer science. That research should identify factors that inhibit greater participation in NIT and should propose and evaluate remedies.

10. Strengths and Limitations of the NITRD Coordination Process and Structure

11. Recommendations: NITRD Coordination Process and Structure

12. The Role of Federal Investment in NIT R&D

12.1 The Critical Role of Federal Investment

12.2 The Incremental Investment Implied by this Report

Appendices:

A: Expert Input into the PCAST NITRD Review

B: Acknowledgments

C: Abbreviations used in this Report

Footnotes


4. See the Section 12 sidebar “Why We’re Able to Google” (page 107).

5. See Section 12.


President’s Council of Advisors on Science and Technology. (September 2010). Prepare and Inspire: K-12 Education in Science, Technology, Engineering, and Math (STEM) for America’s Future. http://www.whitehouse.gov/sites/defa...med-final.pdf

http://www.tableausoftware.com/

http://www-958.ibm.com/software/data/cognos/manyeyes/

http://www.hhs.gov/open/plan/opengov...nitiative.html

http://rtw.ml.cmu.edu/

http://www.nytimes.com/2010/10/05/sc...05compute.html

http://gigaom.com/2010/05/04/we-cant...gh-tiny-pipes/


http://www.top500.org/
President’s Council of Advisors on Science and Technology. (September 2010). Prepare and Inspire: K-12 Education in Science, Technology, Engineering, and Math (STEM) for America’s Future. http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCOVID.pdf