

TOWARDS THE DIGITAL GOVERNMENT OF THE 21ST CENTURY



A Report from the Workshop on Research and Development Opportunities in Federal Information Services

Herbert Schorr
USC/ISI
and
Salvatore J. Stolfo
Columbia University

June 24, 1997

Available on the World-Wide Web at
<http://www.isi.edu/nsf/final.html>



PREFACE

This report was developed by the Organizing Committee of the Workshop on R&D Opportunities in Federal Information Services Workshop held in May 1997. The Workshop goals were to:

- identify immediate targets of opportunity in Federal information services that could benefit from a collaborative relationship with academic and industrial computer and information sciences R&D;
- draft an R&D agenda that addresses fundamental issues of applying and extending advanced information technology and developing innovative Federal information services by using Advanced Technology Pilots and testbeds to accelerate the process; and
- recommend mechanisms for involving Federal agencies with information intensive services in appropriate R&D activities.

The Workshop was supported by:

- the Applications Council of the National Science & Technology Council Committee on Computing, Information, and Communications,
- the National Science Foundation's Directorate for Computer and Information Science and Engineering,
- the President's Government Information Technology Services Board, and
- the National Institute of Health's National Center for Research Resources

This report is available in several formats on the World-Wide Web at:

<http://www.isi.edu/nsf/final.html>



ORGANIZING AND FEDERAL COORDINATING COMMITTEE

Organizing Committee

Herbert Schorr (Chair)

Executive Director
Information Sciences Institute
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292-6695
Phone: (310) 822-1511
Fax: (310) 823-6714
email: schorr@isi.edu

Eliot Christian

Chief, Data and Information Management Staff
US Geological Survey
802 National Center
Reston, VA 20192
Phone: (703) 648-7245
Fax: (703) 648-7069
email: echristi@usgs.gov

Salvatore J. Stolfo (Co-chair)

Professor, Department of Computer Science
Columbia University
606 CEPSR
New York, NY 10027
Phone: (212) 939-7080
Fax: (212) 666-0140
email: sal@cs.columbia.edu

Colin Crook

Senior Technology Officer
Citibank, N.A.
153 E. 53rd Street 23/1
New York, NY 10043
Phone: (212) 559-2393 (voice mail - 2379)
Fax: (212) 559-5274 (backup is 559-5138)
email: Colin.Crook@citicorp.com

Alan Blatecky

Vice President
MCNC
P.O. BOX 12889
RTP, NC 27709-2889
Phone: (919) 248-1149
Fax: (919) 248-1455
email: alanb@mcnc.org

Patricia Edfors

Champion for Security of the
Government Information Technology Services
Board
Department of the Treasury
1425 New York Avenue, NW, Room 2150
Washington DC 20220
Phone: (202) 622-1552
Fax: (202) 622-1595
email: patricia.edfors@cio.treas.gov

John Cavallini

Director, Technology and Planning
Dyncorp
12750 Fair Lakes Circle
Fairfax, VA 22033
Phone: (703) 222-1403
Fax: (703) 222-1699
email: jc@dyniet.com

*Towards the Digital Government
of the 21st Century*

Dr. Edward A. Fox

Professor of Computer Science and
Associate Director for Research of Computing
Center
Virginia Polytechnic Institute and State
University
660 McBryde Hall
Blacksburg, VA 24061-0106
Phone: (540) 231-5113
Fax: (540) 231-6075
email: fox@vt.edu

Joseph Hardin

Associate Director
Software Development Group
National Center for Supercomputing
Applications
University of Illinois
Champaign, IL 61820
Phone: (217) 244-7802
Fax: (217) 333-5973
email: hardin@ncsa.uiuc.edu

Thomas A. Kalil

Senior Director to The National Economic
Council
for Science & Technology
Executive Office of the President
Old Executive Office building, Room 233
Washington D.C. 20500
Phone: (202) 456-2802
Fax: (202) 456-2223
email: KALIL_T@a1.eop.gov

Constance McLindon

Director of Systems Deployment
Corporation for National Research Initiatives
1895 Preston White Dr., Suite 100
Reston, VA 20191-5434
Phone: (703) 620-8990
Fax: (703) 620-0913
email: cmclindo@cnri.reston.va.us

Robert C. Musser

Project Director
Private Sector Council
1101 16th Street, N.W., Suite 300
Washington D.C. 20036-4803
Phone: (202) 822-3910
Fax: (202) 822-0638
email: psci@capaccess.org

Daniel Schutzer

Vice President
Citicorp Technology Office
909 Third Ave.
New York, NY 10022
Phone: (212) 559-1876
Fax:
email: Dan.Schutzer@citicorp.com

Stephen L. Squires

Special Assistant for Information Technology
Defense Advanced Research Projects Agency
3701 North Fairfax Dr.
Arlington, VA 22304-1714
Phone: (703) 696-2263
Fax: (703) 696-2202
email: squires@darpa.mil

Federal Coordinating Committee

Lawrence E. Brandt

Program Manager for Advanced Information
Systems
Division of Advanced Scientific Computing
National Science Foundation
4201 Wilson Blvd., Room 1122
Arlington VA 22230
Phone: (703) 306-1963
Fax: (703) 306-0632
email: lbrandt@nsf.gov

Melvyn Ciment

Deputy Assistant Director
Computer and Information Science and
Engineering
National Science Foundation
4201 Wilson Blvd.
Arlington VA 22230
Phone: (703) 306-1900
Fax: (703) 306-0577
email: mciment@nsf.gov

Mickey Duniho

Senior Technical Advisor for Intelligent Agent
Technology
National Security Agency
Q03/NBP1
9800 Savage Road
Fort Meade, MD 20755-6000
Phone: (301) 688-9893
Fax: (301) 688-9169
email: madunih@afterlife.ncsc.mil

Valerie Gregg

Internet Collaborative Advisor for the
Associate Director
U.S. Bureau of the Census
Washington D.C. 20233
Phone: (301) 457-4102
Fax: (301) 457-1229
email: vgregg@census.gov

Luis G. Kun

Senior Advisor, Center for Information
Technology
Agency for Health Care Policy and Research
2101 East Jefferson Street, Suite 602
Rockville, MC 20852
Phone: (301) 594-1483
Fax: (301) 594-2333
email: lkun@po5.ahcpr.gov

Carlynn Thompson

Director
Directorate of Research, Development, and
Acquisition Information Support
Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Fort Belvoir, VA 22060-6218
Phone: (703) 767-9175
Fax: (703) 767-9174
email: cthompson@dtic.mil

Bill T. Turnbull

Director of High-Performance Computing and
Communications
National Oceanic and Atmospheric
Administration (NOAA)
1315 East West Highway, Room 15300
Silver Spring, MD 30910
Phone: (301) 713-3573
Fax: (301) 713-4040
email: wturnbull@hpc.noaa.gov

Judith L. Vaitukaitis, M.D.

Director
National Center for Research Resources
National Institute of Health
Building 12A, Room 4007
12 South Drive, MSC 5662
Bethesda, MD 20892-5662
Phone: (301) 496-5793
Fax: (301) 402-0006
email: vaitukaitis@nik.gov



ABSTRACT

A partnership between Government agencies and the information technologies research community has succeeded in the past for the benefit of the Nation. The most notable example is the emergence of the Internet as the basis for broad scientific, cultural, civic, and commercial discourse, evolving from what was originally a Government-supported networking research project. The collaborative development of a new applied research domain is critical to help meet the Nation's growing information service demands. Applied research that considers real world operating constraints can provide valuable new problems and insights for the academic research domain, leading to new demonstrable and deployable systems. This applied research domain is a **National Challenge** to provide a transition strategy for migrating Federal Information Services from legacy systems, through the interoperable systems of the Internet, and toward more advanced integrated global systems. A unique opportunity exists for a new paradigm for interaction between Government and citizen; an opportunity to invent the

Digital Government for the citizens of the 21st century.

This report recommends that the Federal Government invest in a new applied research program to realize this vision.



TABLE OF CONTENTS

<i>EXECUTIVE SUMMARY</i>	1
Recommendations	4
Conclusion	5
<i>1. BACKGROUND</i>	6
1.1 The Federal Information Services Context	6
1.2 Relevant Federal R&D Efforts	7
<i>2. GOVERNMENT NEEDS</i>	9
2.1 The Federal Information Services Laboratory	10
2.2 A Conceptual Domain for R&D in Federal Information Services	12
<i>3. APPLIED RESEARCH OPPORTUNITIES</i>	18
3.1 Storing and Archiving Information	18
3.2 Find and Access Information	19
3.3 Integrate Information from Multiple Sources	19
3.4 Mine and Discover New Knowledge	20
3.5 Validate and Visualize Information	20
3.6 Universal Access to Information	20
3.7 Security, Trust, Electronic Commerce	21
<i>4. RECOMMENDATIONS</i>	22
Recommendation 1. Coordinate Multi-agency Efforts to Focus on Priority Problems	22
Recommendation 2. Inform the Research Community	22
Recommendation 3. Bridge the Culture Gap Between Researchers and Federal Information Service Providers	22

*Towards the Digital Government
of the 21st Century*

Recommendation 4. Initiate Advanced Technology Pilot Projects ----- 23

Recommendation 5. Fund a substantial number of Advanced Technology Pilot projects and basic exploratory research projects through Broad Agency Announcements. ----- 24

Recommendation 6. Provide Stable Funding for a Focused Program of Applied Research, Technology Transfer, and Training/Education Activities in the Federal Information Services Domain ----- 24

APPENDIX A - Details on the Workshop on R&D Opportunities in Federal Information Services ----- 26

Workshop Thematic Areas ----- 27

ACKNOWLEDGMENTS----- 28

INVITED PAPERS----- 29

EXECUTIVE SUMMARY

At the end of this extraordinary century of scientific achievement and technological innovation, American society is poised for a radical shift in the manner in which individuals learn and work, conduct their business, and are entertained and informed, arising from the innovation of new information technologies. While this shift affords unprecedented opportunities, it also challenges Government agencies to dramatically improve the information services they provide to the American public. Federal, State and local governments are all striving to meet citizens' rising expectations for easy, fast, secure and accurate interactions, even in an environment of shrinking staff and budgets. There are expectations that information technologies will allow the Government to improve operations and services delivery, but these hopes are tempered by the failure of several recent widely-publicized modernization efforts costing the American taxpayer billions of dollars with few discernible results.

A related challenge is cultural: most industrial and Government information service managers have management and goals that do not usually value investment in R&D. Moreover, many Government agencies procure expensive and complex information systems without the benefit of sufficient interaction with each other or with the R&D community. At the same time many U.S. industries have reduced their investment in basic and applied research. Hence, suppliers of information technology may produce fewer innovations from their laboratories. Federally funded basic and applied research in academia and national laboratories will need to assume a larger role in the development of Federal Information Services to address this issue.

The disconnect between providers of private and public information services and the information and computer technologies research community is of great concern. Because information plays many roles in Government agencies (as a currency of exchange, as an important by-product of their missions, as the carrier of work flows and services), it is essential that as agencies restructure and re-engineer to exploit these new efficiencies, the R&D community and other sectors form new relationships.

The new network-centric world will present other societal challenges beyond the pragmatic concerns of efficient and affordable Government information services. For example, the public's right to know is a fundamental tenet of any democratic society, yet there is a growing concern about the potential dangers to individual rights and privacy arising from electronic access to a variety of personal information. This concern recently caused the Social Security Administration to shut down its on-line pension benefit information system. A second important example is the new opportunity for civic discourse made possible by the wired-citizen directly interfacing with Government. A **Digital Government** will allow public access to Government information and services, and group participation in discussions at any time and from anywhere on the globe with the required security and trust.

This vision is a distinct possibility but will require Government, in collaboration with the R&D community and information service providers from all sectors, to define and implement a bold new research agenda, supported by new investments in information technology. These issues are clearly resonating within the broad Federal policy arena, as exemplified by the establishment of the Federal Information Technology Innovation Fund, Vice-President Gore's "Access America" report, and the strategic plan of the Committee on Computing, Information and Communication of President Clinton's National Science and Technology Council.

Realizing the **Digital Government** presents great technical and intellectual research challenges, but also promises great value and can provide valuable new insights and interesting new applied research problems, leading to deployable new systems. Thinking of Federal Information Services as a laboratory means that a

*Towards the Digital Government
of the 21st Century*

huge variety of sources and systems are available to invent new and very advanced applications. Typically, academic research laboratories simply don't have sufficient tools or sources of information and data to explore real world problems. Instead, abstractions are studied in closed, controlled environments. Research that considers real world operating constraints can lead to new insights that reduce the complexity of real world problems. Students immersed within a real world context can learn how to apply the knowledge acquired from their academic studies to large problems of great potential impact, thus helping to ensure the next generation of knowledge worker can impact the Federal enterprise. In summary, applied research in Federal Information Services offers some unique challenges unavailable in basic theoretical research. Building, nurturing and leveraging relationships with Federal Information Service providers and the research and education community supported by sufficient long term funding for applied research, training and advanced education, is key to developing a critical mass of researchers and students to work on truly large scale demonstrable applications of benefit to the public.

Research topics with a special focus on the unique characteristics of the Federal information services domain might include the following:

- Storing and Archiving Information
- Finding and Accessing Information
- Integrating Information from Multiple Sources
- Mining and Knowledge Discovery
- Universal Access to Information
- Validating and Visualizing Information
- Security, Privacy and Electronic Commerce

These and other research topics will enable a whole new range of advanced applications. Several examples are:

- Crisis management for natural disasters can be driven by information services that link, summarize and present critical information on demographics, utility maps and geospatial data to guide and assist in rapid evacuations and force deployment, reducing loss of life and property. Concomitantly, collaboration technologies may bring together the special expertise of remote individuals to participate in group decision making to jointly solve and manage a live crisis.
- Economic and statistical data integrated across multiple Federal agencies to provide more accurate near-term view of economic activity and to inform the public for better understanding of its Government's activities in health, education, and welfare.
- A new Federal Information Infrastructure for the Digital Government will lead to new ways in which people and Government employees can interact with each other, make decisions, share ideas and collaborate on common problems. It is conceivable that "Virtual Agencies" might develop that bring together spheres of common interests in multiple agencies with interested groups of citizens to solve a common problem. Traditional lines of responsibility among agencies can thus be broken down and accommodate new interests and special problems more effectively. Policy-makers may be able to easily tap in to the talents and expertise of people from the private sector, and across agencies in the public sector.
- Intelligent transportation systems are a distinct possibility. Real-time integration of information from databases, sensors and "toll point-of-sale terminals" might make it possible to optimize at huge cost savings travel and distribution of goods.

Research and development in computer and information science and technology, as well as research in statistics and the social sciences, is funded by many Federal agencies, universities and businesses. What is necessary now is to define a new range of cross-disciplinary applied R&D objectives with the needs of the

*Towards the Digital Government
of the 21st Century*

Federal information services community in mind, applying that R&D to the interesting problems and data within that arena.

Although many Federal agencies understand the potential of an integrated and inter-operable Internet and intra-net approach to support their missions, most Federal information still resides in legacy vertically-integrated systems designed for specific applications, data sets, and mission areas. These systems have been maintained in stovepipe fashion and have grown individual cultures in the service community without attracting the interest of the R&D community. This disconnect between information services and R&D is by no means limited to the Federal sector, but is mirrored in the private sector and causes concerns there also. Information resource managers, both Federal and private, are painfully aware that improvements and fundamental changes within these stovepipes have been difficult, e.g. cross-stovepipe integration. A horizontally-coordinated alternative to stovepipe systems is emerging, with the potential to provide a transition strategy from legacy systems, through the interoperable layers of the Internet protocols, and toward a more advanced multi-dimensionally integrated Federal enterprise. This system structure separates the applications that support their users from the underlying objects (databases) and reconnects them through an intermediate layer of services and middleware functions such as authentication, synopsisizing, and translation. To effect this move from vertically integrated legacy systems toward multi-dimensional integrated information sets is a significant research problem with very high potential payoff; it represents a **National Challenge for Federal Information Systems**.

Many applications will share common middleware component technologies in this new structure. What these middleware component technologies may ultimately be serves as a conceptual domain for applied R&D research. For example, protected interfaces and wrappers to legacy systems will be needed to support new trusted services and enable the graceful transition to Internet-based applications when they are upgraded. Existing databases can be encapsulated with object interfaces so that they are available with appropriate security to the Intranet of the Federal enterprise. Applications can thus be developed effectively through middleware and services that access underlying object bases which are mediated by authentication, access control, auditing and accountability mechanisms, indexing, synopsisizing, translation, and searching services. These horizontal systems provide a scalable, distributed and portable capability heretofore not possible nor previously contemplated by Federal information service providers. The middleware can move easily to new platforms and systems preventing its own potential legacy problem to keep pace with growth in function and scale of new systems.

Recommendations

To move toward the creation of the **Digital Government**, a number of steps must be taken:

Recommendation 1: Coordinate Multi-agency Efforts to Focus on Priority Problems

Establish mechanisms to coordinate multi-agency R&D efforts to maximize transfer and realize the full benefits of any new information technology innovations broadly. This coordination should be collaborative with Federal mission agencies and their information service requirements.

Recommendation 2. Inform the Research Community

Initiate workshops and information exchanges to educate the information technology R&D community about current capabilities, future requirements, and unique challenges of Government information services; foster innovation in the increasingly important Government information services.

Recommendation 3. Bridge the Culture Gap Between Researchers and Federal Information Service Providers

Develop and maintain a community of researchers from several disciplines (Computer Science, Statistics and the Social Sciences) devoted to Federal information services and related applied research through technical organizations, symposia, and formal journals with academic relevance and excellence, and with a strong education component.

Recommendation 4. Initiate Advanced Technology Pilot Projects

Support applied research programs through Advanced Technology Pilot projects driven by Government data and missions, with continual updating and enhancement of systems and services.

Recommendation 5. Fund a substantial number of Advanced Technology Pilot projects and basic exploratory research projects through Broad Agency Announcements.

Two types of projects are envisioned, Advanced Technology Pilot projects of substantial size and cost, and more focussed applied and exploratory research. A substantial number of projects will be needed to address the scope of issues and problems.

Recommendation 6. Provide Stable Funding for a Focused Program of Long Term Applied Research, Technology Transfer, and Training/Education Activities in the Federal Information Services Domain

Leverage existing investments in new computer and information science research and technologies by Federal, academic and industrial laboratories, through multi-sector partnerships to improve the timeliness and the impact of new technologies on i) Federal information services delivery, ii) individual agency information technology operations, and iii) strategic information services planning.

Conclusion

A partnership between Government agencies and the information technologies research community has succeeded in the past for the benefit of the Nation. The most notable example is the emergence of the Internet as the basis for broad scientific, cultural, civic, and commercial discourse, evolving from what was originally a Government-supported networking research project. The collaborative development of a new applied research domain is critical to help meet the Nation's growing information service demands. Applied research that considers real world operating constraints can provide valuable new problems and insights for the academic research domain, leading to new demonstrable and deployable systems. This applied research domain is a national challenge to provide a transition strategy for migrating Federal Information Services from legacy systems, through the interoperable systems of the Internet, and toward more advanced integrated global systems. A unique opportunity exists for a new paradigm for interaction between Government and citizen; an opportunity to invent the **Digital Government for the citizens of the 21st century**.

1. BACKGROUND

At the end of this extraordinary century of scientific achievement and technological innovation, the American society is poised for a radical shift in the manner in which individuals learn and work, conduct their business, and are entertained and informed. This new era will also herald in the age of the **Digital Government**, whereby the American public can access information, participate in group discussions and decision making, and transact its business with Government anywhere and any time over secure networks, on a foundation of trust while using a broad range of information appliances.

In 1993, the Vice President defined National Performance Review goals to develop interoperable, Internet-based Government services. The report "Reengineering Through Information Technology" has served as the impetus for many Federal agencies to begin moving towards new information technologies. A Government Services Information Infrastructure was proposed as the means of connecting Government employees to work together more efficiently to serve the public's interests. The goals laid out include the establishment of a coordinating office bridging various Federal agencies, the creation of an interoperable information infrastructure and a Government wide "blue book" directory service. Interconnection of services are also defined and coordinated by the Government Information Technology Services Working Group (recently established as the GITS Board (GITSB)). This is a bold vision of a new way that Government operates and delivers services.

The global marketplace driven by new information services is a technological call to arms to maintain the US's competitiveness in information technologies. The Nation needs to continue advancing the frontiers of technology, to stimulate innovation and reduce the barriers to technology transition from laboratory to industry, so that its benefits are available to all sectors of society and all individuals who may wish to use it. In the age of the multi-national corporation, the focus on competitiveness needs to move to our Nation's greatest resource, the "knowledge worker". It is essential that the Nation train the next generation of U.S. workers to compete in an emerging Web-based service economy.

The US economy has seen dramatic changes over the last few decades. The services sector has become a dominant share of the GNP, and information services is a component of crucial importance across the entire economy. The US's major trading partners are rapidly growing consumers of the US information technology base creating a highly competitive global marketplace. The US policies of funding basic research in the physical sciences and mathematics generated enormous innovation in the manufacturing sector. However, today and in the foreseeable future, computer and information science is playing a more dominant role while perhaps being neglected in its share of Federal investment. Although the US business community has invested in industrial R&D of computer and information science, most of these investments have been focused on short term gains. Targeted Government investment in basic and applied long term R&D in computer and information science can contribute significantly to insuring the future dominance of the US technological base and the workers trained in these technologies.

1.1 The Federal Information Services Context

Concomitantly, Government agencies are under enormous pressure to improve their services and operate more efficiently. Recent legislation, executive orders and declining budgets have caused Federal agencies to reformulate their processes, to transact business more efficiently, and to provide new and improved services to the citizenry. The confluence of these pressures, the globalization of the economy and the demand for efficient

Towards the Digital Government of the 21st Century

Government, provide new urgencies for investment in basic and applied research in computer and information science and engineering. Some Federal agencies have been agile in taking early advantage of Internet and Web technologies. This experience has raised expectations and interest that the Government further utilize technological advances to improve operations and services delivery. Similar improvements are expected of commercial companies by their customers. On the Federal side, however, this enthusiasm is tempered by the failure of several recent widely-publicized modernization efforts costing the American taxpayer billions of dollars with no discernible results. It is a great challenge to more effectively introduce new technologies into Federal and other large organizations to increase the chance of success of fielding new and efficient services.

A series of recent high-level panels, councils and boards (several are cited below) have recommended that Government continue its role, not only as a supporter of long term R&D, but also that Government become a more active convener of the sectors to promote innovation in information technology. These issues have been documented in recent reports. (Web references can be found for these reports at the URL: <http://nii.nist.gov/>.)

One of the more interesting outcomes of these discussions and recommendations is the "Access America" report (see <http://www.gits.fed.gov/htm/gsii.html>). In the report, Vice President Gore notes that the recommendations "...*paint a picture of the kind of Government we should have as we begin the next century.*" He envisions "... *a Government where all Americans have the opportunity to get services electronically and where, aided by technology, the productivity of Government operations will be soaring.*"

The GITS Working Group Accomplishment Report on The Information Infrastructure Task Force (IITF) provides further directives to realize this vision. The GITSB, formed by the White House, articulated policies and initiatives to accelerate deployment of the National Information Infrastructure (NII). Within the IITF, and with the support of the Vice President, the GITSB was created to coordinate efforts to improve the application of information technology by Government agencies. The GITS mission is to promote the improvement of agency performance through the use of information technology, accelerate the deployment of advanced networking technologies, and, in conjunction with the Office of Management and Budget (OMB) and General Services Administration (GSA), establish procurement and implementation directives to improve productivity and reduce costs.

The GITS Accomplishment reports delineate the current status of these information technology initiatives, define the actions required to carry out implementation, and update the progress made with respect to each of the initiatives. (GITS reports are available on the Web at: <http://nii.nist.gov> and through the draft Government Services Information Infrastructure (GSII) Plan at: <http://www.er.doe.gov/production/octr/mics/gsiiplan.html>).

The Congress and the Executive Branch have also recently recognized the importance of information services in accomplishing Government missions through the Information Technology Management Reform Act of 1996. This law promotes multi-agency and multi-program cooperation and acquisitions to capitalize on emerging private sector technology advances. This law contemplates and authorizes \$750 million for Advanced Technology Pilot projects over a five year period.

1.2 Relevant Federal R&D Efforts

Substantial investments are already being made by the Federal Government into research initiatives focusing on underlying technologies. Unfortunately, few are devoted to or even take into consideration the particular needs of Federal information services, especially with respect to legacy migration problems and the scale of problems that will be encountered when all Federal data are network-accessible. Clearly, there is the opportunity for joint activities and initiatives of various agencies to be leveraged and exploited. For example:

- The Committee on Computing, Information, and Communications (CCIC) and the Applications Council within the President's Office of Science and Technology Policy, has evolved from the multi-

*Towards the Digital Government
of the 21st Century*

agency High-Performance Computing and Communication initiative (see <http://www.hpcc.gov/blue96/section.2.4.html>).

- CCIC has recently reorganized itself along five R&D program component areas and associated working groups which will provide a framework for multi-agency coordination. These program component areas of potential benefit to Federal information services are High-End Computing and Computation, Large-Scale Networking, High Confidence Systems, Human-Centered Systems, and Human Resources, Education and Training.
- The Applications Council, one of the sponsors of the R&D Opportunities in Federal Information Services Workshop, promotes early application of advanced computing, information and communications technologies and R&D capabilities to critical Federal Government missions. The Council has appointed several working groups (Federal Information Services, Universal Access, Crisis Management, Next Generation Internet) to promote inter-agency collaboration and Advanced Technology Pilot projects that demonstrate the potential of using advanced information technologies to achieve common goals. The Council is responsible for addressing "Access America" report goal A15: Integrate The Government Services Information Infrastructure. Item 5 of that goal reads as follows:

Improve the coordination of information technology research and development activities with the work of the GITS Board.

The GITS Board works jointly with the National Science and Technology Council's (NSTC) Applications Group to bring representatives of various information technology R&D communities together. This joint effort investigates ways to coordinate R&D efforts with GITS Board programs to build an integrated infrastructure for the Government. Including promising new developments, such as Next Generation Internet technologies, should be part of these discussions. Recently, an individual was named to serve on both the GITS Board and the National Science and Technology Council Applications Group. The representative works with the GITS Board and Applications Group members to identify and define the specific research and development projects that will help turn the recommendations in this "Access America" report into reality.

- The Defense Advanced Research Projects Agency Information Technology Office (ITO) is responsible for research into breakthrough information technologies for use in advanced defense applications. The office's mission is to provide the networking and computing hardware, software, systems and management technologies vital to ensuring DOD military superiority. These technologies often have non-military applicability. (Further information can be found at <http://www.ito.darpa.mil/>.)
- The National Science Foundation CISE Directorate supports research to enable the rapidly emerging information society. The principal goals are to improve basic understanding and design of information and knowledge processing mechanisms; to facilitate information and knowledge exchange between humans and machines; and to understand better the impact of advanced information technologies in distributed work environments. Ongoing programs support research in such areas as the design of database and knowledge-based systems, information systems in distributed and networked environments, virtual reality and multimedia environments, human language technology, and sensor-based control in intelligent robots. (Further detail can be found at <http://www.cise.nsf.gov/iris/>.)
- Of particular note is Digital Libraries. Two Federal agencies, DARPA and NASA, have teamed with NSF to support six research projects developing new technologies for Digital Libraries, storehouses of information available through the Internet. The projects' focus is to dramatically advance the means to collect, store, and organize information in digital forms, and make it available for searching, retrieving, and processing via communication networks, all in user-friendly ways. (Details can be browsed at <http://www.cise.nsf.gov/iris/DLHome.html>.)



2. GOVERNMENT NEEDS

The unique demands of Federal information services provide a rich suite of applied research problems of great intellectual challenge. For example, the special relationship between citizen and Government presents interesting issues that await a concerted national scale research effort.

The citizen as customer is perhaps a useful metaphor but, unlike business, Government cannot target a particular preferred customer base to maximize its profit and reduce its losses to the exclusion of others. Whether they are able-bodied, English-speaking, sighted, or not, information literacy is an enabling requirement for all, demanding innovative approaches to training and education. Multiple interface modalities which enfranchise all may require considerable training and sophisticated interface technologies that seamlessly transcend modality boundaries.

Along with the business world, the move from legacy information systems is slow, expensive and politically constrained. Implementation of integrated information systems requires fundamental shifts in business models, such as in management structures and organizational rewards to foster innovation along new business technical lines.

Government agencies need to cross-leverage each other's resources, integrate their systems and data collection, and benefit from economies of scale that collaboration may afford. Information technology solutions required for integrating Government sources include tools to discover, search and access data and information from disparate Federal sources, identify and ameliorate hot spots, and produce summary responses to reduce network loads. Machine readable, and human digestible taxonomic descriptions and ontologies of available Federal Information Sources, their meaning and relationship with each other would enable automated searching, load balancing and proper dissemination of appropriate information to the public.

Scaling is a fundamental challenge for Government agencies. Today, users of the Internet and World Wide Web (WWW) are a minority of the American public, but this is changing with remarkable speed. When the majority of the American public becomes active users, existing Federal information services may collapse under the expanded demand for information and services. No one can presently predict what new services will be demanded by the public and the concomitant load on resources that would be generated. Furthermore, continuous collection of data, and the integration of heterogeneous sources of very large size will certainly stress the limits of information seeking applications. An individual query initiated at one PC by a single user, for example, may involve scores if not hundreds of search tasks spread over a large number of information sources increasing the chances for bottlenecks that slow response time and render services useless.

A broad information technology infrastructure is needed to inform policy formation and decision making to ensure accountability for the data and knowledge it collects and warehouses. Meta-data standards which cross domains and information collections are needed to accurately describe the sources of information, how they were collected and were validated. Scalable collaboration technologies are needed in meeting the demands of large groups of people who may wish to participate in group discussions with Government.

Privacy, security and trust are critical elements in the relationship between Government and the public. Large credit card companies have fielded easy to use payment systems with a modest level of security. A certain amount of undisclosed loss due to fraud is tolerated since business can relatively easily spread the margin between profit from fees and interest and loss due to stolen cards and fraudulent transactions. No simple "margin spread" is available to a democratic Government. All citizens must be fully protected. Maximal trust

*Towards the Digital Government
of the 21st Century*

in accessing and transacting with Federal services is of paramount importance to achieve wide adoption by the public (and relevance).

Efficient and trusted electronic benefits transfer with fraud detection stands to produce considerable efficiencies in Government processing environments to deliver benefits faster and more accurately to the public with maximal trust. The traditional means of securing sensitive information or assets, based upon encryption technologies, secure passwords and PINS, must be augmented with new security features that raise the bar of protection. Technologies that aim to detect fraud, as well as those that attempt to prevent fraud, can work together to improve the trust between citizen and Government. Migration of legacy systems from batch orientation to modern online transaction processing architectures would achieve several concurrent goals of efficiency, application development and scalability. Government systems should explore a new generation of information systems including "agent-based" infrastructures for rapid deployment of new services and security and protection mechanisms to a wide range of sites when they become available.

Government needs to streamline its complex procurement processes by taking full advantage of electronic marketplaces and automated process engineering systems. By analyzing, recording and representing existing Government transaction processes, bottlenecks and "pinch points" may be identified and ameliorated by automated means of restructuring process and information flows.

2.1 The Federal Information Services Laboratory

Federal information services offer great technical and intellectual challenges. Purely theoretical research may reduce the complexity of real world problems to discover general principles. Applied research that considers real world operating constraints can provide valuable new problems and insights for the academic research domain, leading to new demonstrable and deployable systems. Students immersed within a real world context can learn how to apply the knowledge acquired from their academic studies to large problems of great potential impact. Building, nurturing and leveraging relationships with Federal information service providers and the research and education community from several disciplines, supported by sufficient long term funding for applied research, training and advanced education, is key to developing a critical mass of researchers and students to work on truly large scale demonstrable applications within the Government context.

Government information services provide unique opportunities that differ from big business. For instance, matters of security, privacy and integrity are absolutely critical design criteria for systems since public trust is at stake. The size of Federal databases, their distribution, the vast number of agencies, and the aforementioned security/privacy concerns all contribute to added complexities of Federal information services. Government provides numerous services which in-turn generate many unique systems (crisis management, law enforcement, environmental protection, patents and trademarks, air traffic control, and regulatory affairs). Government must assure universal access to its information thereby introducing an array of issues. Once technical solutions are identified or developed, the Government must go through a lengthy, complex, competitive procurement process as opposed to vendor-specific acquisitions. These considerations and others offer more than enough applied research challenges of a cross-disciplinary nature involving Computer Science, Statistics and the Social Sciences.

By way of summary, the following chart details the unique aspects of Federal Information Services.

A Summary Of The Unique Aspects Of Federal Information Services

Security, privacy and integrity as prime architectural and design criteria

- Authentication
- Audit
- Journaling
- Observation: Government, unlike companies, cannot trade dollar losses for security, privacy exposure

Scale

- Size of databases and repositories
- Inter-agency data access
- Distributed systems network
- Observation: Government, unlike companies, does not have a core business; its in every business.

Citizen as a customer

- Equality of access, shared kiosks, computational utilities
- Service level approaching commercial practice
- Virtual agencies
- 24X7 operation
- Quality of data, ontologies
- Multiple language support
- Freedom of Information
- Observation: Government must always be open to everyone.

Government as an Information Technology customer

- Systems are procured from many vendors
- Competitive procurements
- Observation: Government must foster cooperation, innovation and competitiveness.

Unique systems and applications

- Law enforcement
- EPA
- Patent and Trademark Office
- Emergency Management
- Regulatory Agencies
- Observation: Government is the only provider of unique services.

2.2 A Conceptual Domain for R&D in Federal Information Services

The Government, like most other large organizations, is still providing information services using technology that is in large measure several generations behind the current Internet and Web-style of information technology. Most Federal information services are built upon vertically-configured, legacy-embedded architectures (stovepipes) designed for mission-specific applications with datasets that are typically incongruent with other related Federally-held data. While they can be efficient in sharply-defined domains, stovepipe systems are very hard to maintain and upgrade with new applications and services, or integrate with other stovepipes. These monolithic and hierarchical systems have grown technical and management cultures in the service community that have little or no connection with information technology researchers and developers. This disconnect between the information services community and the R&D community is by no means unique to Federal systems. Large private corporations and organizations have the same cultural complexities and difficulty in effectively upgrading their services and systems to meet new demands from their customers for network-based access.

An alternative system structure, just beginning to be embraced by agile organizations and businesses, is horizontally coordinated across stovepipe systems (see Figure 1). These modern engineered systems separate the applications that support users and mission critical tasks from the underlying objects (databases in legacy systems) and reconnects them through a flexible intermediate layer of services and middleware. This layer enables access to multiple sources of information, and can mediate queries among heterogeneous data types and information sources on the lower layer while delivering data streams in multiple modalities to applications and presentation tools at the uppermost layer.

Many applications will share common middleware component technologies in this new structure. What these middleware component technologies may ultimately be serves as a conceptual domain for applied R&D research. For example, protected interfaces and wrappers to legacy systems will be needed to support new trusted services and enable the graceful transition to Internet-based applications when they are upgraded. Existing databases can be encapsulated with object interfaces so that they are available with appropriate security to the Intranet of the Federal enterprise. Applications can thus be developed effectively through middleware and services that access underlying object bases which are mediated by authentication, access control, auditing and accountability mechanisms, indexing, synopsising, translation, and searching services. These horizontal systems provide a scalable, distributed and portable capability heretofore not possible nor previously contemplated by Federal information service providers. The middleware can move easily to new platforms and systems preventing its own potential legacy problem to keep pace with growth in function and scale of new systems.

Figure 1: A Conceptual Domain for R & D in Federal Information Services.

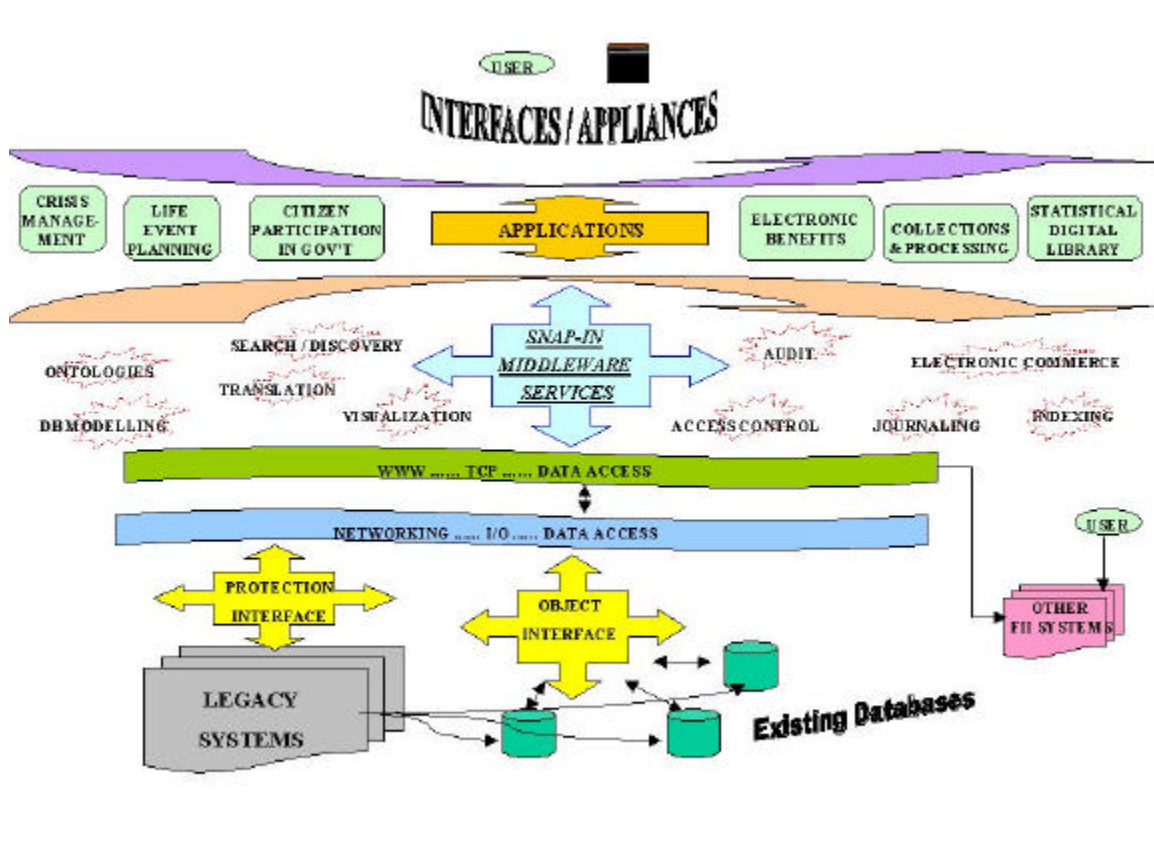
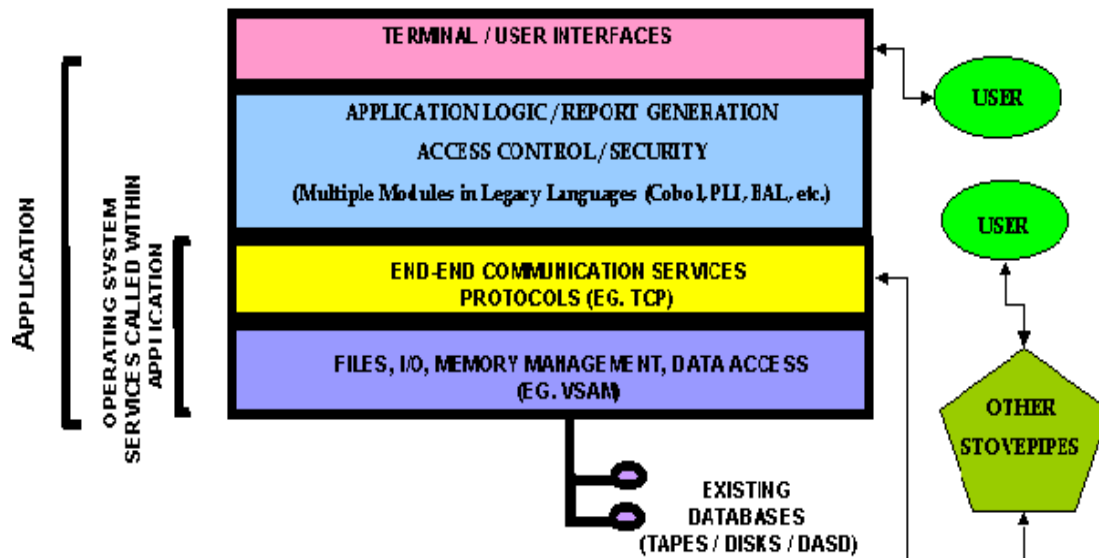


Figure 2: A Typical Stovepipe (Legacy) Application.



*Towards the Digital Government
of the 21st Century*

Present day systems are written in a stovepipe manner and interact with their environment by directly involving operating system services (see Figure 2). Such services include the bit-level network services, as well as I/O routines and memory allocation. The next layer up is called the transport level in the communication view of the system and includes end-to-end services needed by applications. This layer provides delivery and includes services such as TCP, data access, methods and databases. Most existing applications are written using this level of software. As a result, much code is written as part of the application that, while tailored to the particular application, needs to be repeated in each and every major application subsystem. The goal here is to find middleware components that define a new set of shared services and resources above the existing transport layer/upper level operating system services, that can be easily and conveniently used by application code. This "middleware" layer raises the level of common services and introduces new types of higher level services. By using this set of services several objectives are being pursued:

- Reduce the amount of code needed to be uniquely provided by a distributed system application.
- Make the application more independent of any specific operating system and hardware platform.
- Make it easier to write other applications that use the same data and services, thus helping to eliminate the inability that stovepipe applications have had in communicating and interacting with one another easily.

In Figure 1 above, the sample middleware services identified are:

- Authentication - insures that the user of a system is indeed who the person represents he or she is. This is usually done with passwords, public keys and/or signature verification systems.
- Journaling - provides transaction recovery in case of system failure. This is usually done in a database system by causing a duplicate copy of a transaction to be made at a synchronous point defined in the code as a "commit" point.
- Audit - in case the system security is compromised, this service attempts to maintain enough information so that the intruder can be traced and eventually identified.
- Database Modeling Services - knowledge-based systems which allow model semantics of legacy databases to be described.
- Ontological Services - description of terms to be used in models as well as tools to customize an ontology to describe a given domain.
- Indexing - a service which builds a representation of information in a digital library. Today this is usually key-word index based.
- Visualization - a service which is a framework for visualization and includes tools for presentation of information in multiple modalities.
- Translation Services - translation into different natural languages.
- Search/Discovery - find resources on information as in digital libraries based on text and/or descriptors provided by users, or compute useful patterns in data that reveals hidden structure in raw data.
- Access Control - determine if a user or system after they have been authenticated can have access to the data or processors on the system.
- Electronic Commerce Services - sources to enable electronic commerce such as brokerage, catalogs and payment mechanisms.

This set of services is by no means definitive or exhaustive and is an active area of research. Moreover, the intent is to define a set of interfaces, rather than a fixed system. Hence, the implementation of the middleware

*Towards the Digital Government
of the 21st Century*

and lower layers can evolve and include new services/interfaces while allowing the applications written on top of them to move to new platforms over time and to enhance the applications.

In addition, the conceptual framework shown in Figure 1 integrates existing legacy systems, wherever possible, into the new Federal Information Infrastructure (FII). The legacy systems are treated as data repositories front-ended by the new middleware based systems (MBS). These middleware based systems will be the prime interface for both Government employees and the citizenry. Thus, a query which requires data to answer it from multiple legacy systems will be converted into sub-queries against each of the legacy systems, and the individual results are then fused to produce a comprehensive answer. Protection interfaces in front of each middleware based system (there should be multiples of these distributed geographically and across agencies) are needed to check authorization for both individual users and other middleware based systems. A scheme like the above should allow for the addition of new services (modernization) while allowing the gradual migration of legacy systems eventually to middleware based systems (provided compatible hardware versions of the existing legacy systems can be obtained); the legacy system scheme roughly follows part of this industrial practice.

The object interface in front of existing legacy databases is displayed in Figure 1 so that the middleware based system can have direct access to existing databases without using legacy applications. Thus, new applications will not have to invoke legacy systems and is another migration path for legacy systems; this structure allows a two step process by first replacing legacy applications and then replacing the database system (if the new applications work properly).

There will probably be no single solution to any of these middleware components, but rather a variety of choices that will need to interoperate seamlessly; one application may use a specific authentication mechanism for example, while another cooperating application may use another. Furthermore, many applications will need all of these services, others perhaps only a few of these services. No single middleware service will fit all applications. Instead there will likely be many choices each providing a unique flavor of service. This is different than in the commercial world where one company would attempt to have its product the "industry standard" that all other components will have to interface to. A "logical crossbar switch" is therefore needed to provide seamless interoperation between different services so, for example, the user who is accessing multiple applications would only need to be authenticated once.

Many of these components are presently active areas of research and development with many companies offering products tailored to specific platforms or environments; how these may seamlessly interact and operate correctly is unknown. But this is an important desideratum to foster competition and innovation to ensure continual improvement of middleware services.

Many large corporations within their own enterprises are undertaking similar activities to those proposed here. Others are developing "industry wide" products to meet the demands of other, larger businesses. Experiences from industry and current vendor solutions can provide considerable guidance to Federal Information Services providers but cannot be depended upon to solve the entire problem.

Generic middleware services cannot operate effectively without having considerable knowledge about the underlying objects and upper layer applications they access or interact with. For this reason, considerable intelligence is needed in these components. For example, "indexing" cannot be supported only by the current generation of "keyword based" search engines; rather ontological representations that aim to model the "meaning" of the information, thesauri and other semantic-based representations are needed to accurately index into a huge collection of information. The meaning of information is a public enterprise, not a private dictum, and will necessarily need to be open to continual refinement and enhancement. Furthermore, the underlying objects contain the "public's repository" of information, which is huge and highly dynamic and must remain fully open. Services that access this repository are thus expected to be fluid and will need to be

*Towards the Digital Government
of the 21st Century*

constantly updated and maintained. Hence, these indexing schemes that are the primary means of "finding" information must themselves be open and continuously updated as their underlying sources change. Multiple proprietary indexing schemes simply cannot easily coexist within this environment. It is unlikely any commercial entity can manage the legal and financial risks associated with maintaining an accurate index to the public's repository, and interoperability and consistency between competing companies is unlikely. These research and development activities are thus best conducted in an open research environment with the widest participation possible.

The unique characteristics of Federal Information Services require new approaches and a concerted research and development effort leveraging the existing state-of-the-art:

- The scale of the public's repository is far larger than any individual corporate enterprise knowledge base;
- The scope of the public's knowledge repository is far more diverse than any individual commercial organization; the potential range of uses of federally-held information is enormous. There is no single "core business" function that is solely supported by the federal enterprise, it is in every conceivable business;
- The public's knowledge repository is open to any and all interested parties "24X7" and globally. Thus, secured corporate environments and proprietary networks which limit access and manage "loads" (by way of customary business "9 to 5" schedules), are not good models to provide access to citizens anywhere and anytime;
- Outsourcing, or privatization, of the entirety of Federal Information Services to commercial organizations exposes the public's knowledge repository to an unwarranted level of risk. Concomitantly, it is not clear any commercial organization can assume the legal liability inherent in managing the public's information. Corporations contracted to serve the public's interests may not be immune from law suits and prosecutions from unintended, or purposeful breaches, misuses or mistakes;
- The level of security and trust demanded by the citizen is far more than what business can guarantee and deliver.

By way of summary, in considering the unique characteristics of Federal Information Services, its current state of practice in using information technologies to develop public services, and its relationship with the citizenry, the following generic technical concerns and issues are relevant.

- The widest variety of information on nearly every conceivable topic and service needs to be integrated and available to all interested individuals. Thus, the research topic of "information integration" is a complex problem for Federal Information Services that goes considerably beyond the current state of the art in research. Cross-disciplinary research is needed involving the fields of Computer Science, Statistics and the Social Sciences to ensure adequate treatment of the scope of problems involved.
- The scale and inherent distributed nature of the Federal enterprise raises unique challenges in managing very large and rapidly growing data and information sources, and how to improve the acquisition of this information to aid in its management.
- Having a large and diverse, and ultimately integrated, repository openly available at all times will generate considerable interest and new applications will be sought by many users to summarize, analyze and digest this information for their own specific needs and purposes. Hence, efficient and effective analysis techniques applied to such a large scale of available information is a unique concern that potentially dwarfs the scale of business concerns.
- The citizen user will want access to the Federal enterprise in ways they demand and can afford.

*Towards the Digital Government
of the 21st Century*

Thus, their own individual styles of interaction will demand that services, information and collaboration tools must be provided to support the widest range of information appliances in all modalities to match their own capabilities. To those citizens who cannot afford, or chose not to own, their own appliances that access the Internet medium, publicly available kiosks or "cyberbooths" are required to provide fair access to all.

- One of the most important services provided by Government involves informational and financial transactions with the citizenry. Business activity reporting, tax payments and reimbursements, patent searching and filing and entitlement program benefits payments, to name a few, consume the lion's share of activities and interactions between the citizen and Government. These commerce-based activities of the Internet are presently the main impetus for the explosive growth in network computing by businesses. It is undoubtedly the case that the citizen user will demand the same transaction environment be available from Government as they use in their daily lives with their bank, retail store and restaurant. Hence, solutions for effective, secure and efficient electronic commerce functions between citizen and Government are required.



3. APPLIED RESEARCH OPPORTUNITIES

Thinking of Federal Information Services as a laboratory means that a huge variety of sources and systems are available to invent new and very advanced applications. Typically academic research laboratories simply don't have sufficient tools or sources of information and data to explore real world problems. Instead, abstractions are studied in closed, controlled environments. Thus, applied research in Federal Information Services offers some unique challenges unavailable in basic theoretical research, such as:

- The scale and scope of Federal information sources and services is far larger than the problems typically challenging the research community. For example, most software researchers deal with perhaps up to 10,000's of lines of code, while in the Government arena, millions of lines of (legacy) code are found.
- The complexity of the operating environment is more important than a theoretical understanding and solution of "cleaned up" problems.
- Even polynomial time algorithms are infeasible if one is dealing with a billion element database;
- Systems are not "stand alone", but rather are parts of large complexes of sub-systems where their connections are not immediately apparent;
- The metrics for success are usually cost-based and cannot be determined theoretically but rather are dictated by the operating environment; these costs could substantially change the way one solves a problem.

The following are areas where Federal information services raise leading edge issues that need applied research.

3.1 Storing and Archiving Information

Modern organizations have come to realize the enormous benefit to putting all of their data into large knowledge repositories from which requests for increasingly complex queries can be made. Multimedia capabilities provide the potential for supporting new educational, medical, informational, and entertainment applications. Building knowledge repositories is especially problematic for Federal information service providers now operating on new executive and congressional orders to provide high quality services within a short time frame.

The new networked multimedia information environment presents substantial technical problems before it can deliver its full benefits to society. Designers of information services and systems need to understand how to design storage hierarchies that can deliver multimedia information in a cost-effective manner with the requisite quality of service. Designing multimedia storage systems has recently been an active research area, but most of the efforts have concentrated on video-on-demand servers. In designing and building large high performance multimedia storage systems, one must consider a whole spectrum of applications, from relatively low bandwidth, high throughput, and "just-in-time" delivery of video-on-demand servers to very high bandwidth, relatively low volume, and ASAP delivery of supercomputing/scientific applications. Such systems must be able to accommodate the various storage, performance, and reliability requirements of the different

types of media and applications. A primary goal of research and development activities envisioned here ought to study cost-effective designs of multilevel multimedia storage hierarchies to support the needs of a variety of applications centrally focused on search and query processing tasks.

3.2 Find and Access Information

Once the information is captured and archived, it is essential to be able to find, use and interpret it. Research is needed on the development of algorithms and experimental systems to facilitate the effective representation, storage, and access of various forms of multimedia data. How might a user or system developer be empowered with tools and systems that effectively and efficiently find and present relevant information from a large sea of expanding multimedia information sources?

For individuals, better query support is required. They may wish to find an expert (or agent), a database (source), a (special) collection, discover resources, documents, passages, or find answers to questions they pose. Better algorithms are needed, that depend upon representation and indexing, as well as analysis and categorization. Better interfaces are required to facilitate translation between information needs and queries, concepts and terms, documents and passages, as well as links and other relationships.

Research and development is needed on ontological representations of data, information sources and available expertise, and the means of integrating multiple query results into a comprehensible whole. Sharable ontologies providing domain-specific assistance are needed for searching sources, systems, practices and problem-solving capabilities of various agencies. Machine readable, and human digestible taxonomic descriptions of available information sources, their meaning and relationship would benefit automated searching, and dissemination of appropriate information. Data objects which "know" about themselves, how they were collected, filtered, when they were collected, what security/access control was used, help ensure quality of service.

3.3 Integrate Information from Multiple Sources

Information integration is a serious research problem. There is close connection to areas like information retrieval, hypertext and computational linguistics. This research must operate in the broad context of the Internet, as well as in narrower contexts like WWW and digital libraries. To serve user needs, intelligent integration from various sources is required. This is difficult enough when trying to merge from multiple distributed subsets of an information collection, with reasonable ranking and combination of evidence. Further research on that problem is promising. Additional integration research is needed to support a variety of views, multiple interfaces, and focusing on specific classes of tasks. Researchers with different backgrounds must break the barriers between focusing on data, information, and knowledge; they must bridge the boundaries between text, hypertext and multimedia. Other researchers from the fields of Statistics and the Social Sciences must bring perspectives on cognitive/affective issues, and on social, group, collaboration and cultural concerns to understand users as well as to facilitate management and organizational cooperation. Studies are needed to consider bias, quality assessment and multiple interpretations.

Information-seeking tasks would be better served by understanding the cognitive process employed by people so that, for example, uncertainty in statistics and their concomitant limitations can be effectively communicated, and understood.

Information integration also must be studied in the context of long term learning, as well as short term real-time best effort decision making. Research is needed to span time frames from simple independent transactions, to long sessions, to ongoing problem solving activities feeding upon local collections of semi-persistent information structures.

3.4 Mine and Discover New Knowledge

Certain searches initiated by users may involve ad hoc queries that require fast query processing over massive, inherently distributed warehouses of data. Systems that provide these capabilities need to be architected by programming environments that manage the integration of large complexes of cooperating information systems. The research themes identified here include information systems and infrastructures that can easily deploy (parallel) systems that analyze remote, inherently distributed data and information. New approaches to developing effective search engines and analysis systems perhaps involve building agent-based architectures that launch machine learning programs to remote databases to learn models or classifiers of that data. Integrated collections of these data models might then be used as complex patterns for searching over remote data sources more intelligently than simple keyword or text retrieval searches.

There are a series of research opportunities with respect to the efficiency of Government operations. In addition to the challenges of long term data archiving, the sheer size of many data sets needed in real time demands improvement in data transport which includes the development of scalable I/O architectures and interfaces. Research in very high algorithm complexity for advanced analytics includes developing analysis algorithms for unstructured data, scaling algorithms to work with larger data sets, and data reduction through sampling.

3.5 Validate and Visualize Information

A major research area concerns how to evaluate quality of data issues, such as the error in estimating spatial relationships or in disaggregating spatially referenced data to smaller grids. The general problem is that it is not known how best to characterize and communicate quality, especially given a wide range of user communities. An interesting possibility may be to analyze data usage records to provide clues to data inadequacies that could be exploited by system managers to enhance the appropriate application of the data and information products.

Better information visualization is needed to support searching, browsing, navigating, and discovery. Developing guidelines and fundamental insights on which visual techniques are most effective in particular contexts is an interesting problem. It would also be useful to explore mechanisms for the visualization of fuzzy geographic data.

3.6 Universal Access to Information

There is an emerging opportunity to attack a broad range of access issues that affect both disabled and abled citizens through the integration of a number of technologies that are already well developed. One rubric for these efforts is InterMedia, which focuses on the need to be able to cross media and modality boundaries at will when dealing with networked information and services. The need to be able to move easily from, say, image to text is variously driven by processing power restrictions, network constraints, varying display capabilities, and variable user sensory abilities. The combination of methods, protocols and translation tools that would allow this capability to be overlaid on the current network or WWW environment. Users may request such choices, or transformations between modalities, in the process of customizing their environment for their momentary capabilities and needs.

This research area may build upon much of the work that has been done to support access to variously disabled users by moving easily between images, text, and sound (and perhaps new modalities such as olfactory). The

next generation of this work points to a valuable generalization of the cross modal methods and standards already developed. The appeal of such a direction for a Federal information services program should be apparent: it focuses on the traditional Government issue of encouraging universal access, and support for a disabled portion of the population, through the support of research and development of a number of basic technologies. These include networking protocols, multiple media session transcription, capture, search and replay methods, conversational synopsis and gisting technologies, and various instantiations of agent and brokering services which would support negotiation between user and system service capabilities. Only Government has commitments to such groups where profitability and cost-effectiveness are not the driving concerns. History has shown that many of these technologies are likely to find a broad basis for use among the fully enabled in circumstances where abilities are impaired by natural circumstances.

3.7 Security, Trust, Electronic Commerce

Web-based electronic commerce is emerging with several consortia of big businesses each striving to develop a common standard for electronic payments. There has been a proliferation of proposed (proprietary) payment mechanisms and media including a variety of card-based mechanisms (as exemplified by the SmartCard Forum), electronic tokens of one variety or another, and even electronic funds transfers over multiple proprietary networks (as for example, the DARPA-supported E-Check project involving the Financial Services Technology Consortium). Commercenet seeks to involve a wide range of telecommunications, computer and financial companies to develop a common electronic marketplace with interoperable payment functions fielded by members. The Electronic Benefits Transfer program of the Federal Government with services contracted to Citibank signals a new means to transact with the Government. Although there is much activity in these R&D and standardization efforts, no clear winners and losers have yet emerged. However, a new range of technical problems have surfaced; issues of security and scale are paramount. For example, the U.S. consumer generates 62 Billion check transactions a year. Current electronic commerce accounts for less than .1% of this amount. The scaling issues here may demand that the Federal information services providers fully embrace legacy migration strategies from batch to modern On-line Transaction Processing (OLTP) environments. One of the key difficulties here is a lack of understanding how to more effectively introduce new technologies and new work-flow processes in organizations. Research in work flow technologies to identify bottlenecks and inefficient processes to automate the migration to streamlined and efficient processes is highly desirable.

The traditional means of securing transactions with card and PIN-based mechanisms over proprietary networks cannot be trusted over public (and global) networks. New technological solutions to enhance security and authentication, raising the bar higher to ward off fraudulent transactions, is of paramount importance. New protection mechanisms that not only secure and authenticate transactors but also detect fraudulent transactions in real time, ought to be investigated to provide higher levels of protection for the citizen. Advanced security/privacy/fraud detection systems trusted by the public are crucial to realize a new generation of electronic payments systems. Furthermore, new mechanisms to rapidly upgrade large numbers of network computing sites would greatly improve real-time protection mechanisms for the citizen. Hence, research is needed in new secured agent-based infrastructures for rapid deployment of newly available technologies and systems.

4. RECOMMENDATIONS

There is a clear indication of growing Federal awareness of the potential impact of these technologies within the Government sector, as exemplified by the Information Technology Innovation Fund, Access America report and the CCIC strategic plan. These opportunities and sources of support should be immediately exploited. However, even though R&D opportunities in Federal Information Services are clear, a major hurdle is for the R&D community to gain access to the actual problems and source material which comprise Federal Information Services. We believe that the following balanced goals will lead to a better informed environment for collaboration:

Recommendation 1. Coordinate Multi-agency Efforts to Focus on Priority Problems

To maximize the chance for broad applicability and uptake of innovative technologies across multiple agencies, a coordination body should serve as the primary conduit for R&D activities in Federal information services. Coordination with other Government efforts involving the National Information Infrastructure should be stressed.

Recommendation 2. Inform the Research Community

Initiate a number of workshops and information exchanges to familiarize the information technology R&D community from all sectors about the capabilities of and unique problems faced by Government agency information services. This process should result in a broad understanding of the applied research agenda within Federal Information Services. Invite researchers, developers and students from universities, industry and national laboratories, hosted by collaborating Federal agencies to identify targets of opportunity.

Recommendation 3. Bridge the Culture Gap Between Researchers and Federal Information Service Providers

The research community will not become connected with Federal information service providers unless the means for professional advancement and career development are clearly available to sustain their professional interests. Therefore, a community of researchers from several disciplines (Computer Science, Statistics and the Social Sciences) devoted to Federal information services applied research and student education should be encouraged, nurtured and supported to establish new courses, technical organizations, conferences and symposia, and formal journals with academic relevance and excellence.

Internship programs in providing summer research opportunities to students should be established to broaden their academic studies. We believe that the issues confronted by the operational and development centers of the Federal Information Services community could form a broader model for cooperation between industry, Government, and universities to help speed the development of the knowledge worker who will need to compete in the 21st century.

Recommendation 4. Initiate Advanced Technology Pilot Projects

Advanced Technology Pilot projects, infrastructure testbeds and other prototype activities are central to bridging the range of activities from R&D to information service areas which have little R&D background. The recommended Advanced Technology Pilot projects suggest how the Government should initiate a process of community linkage. These suggestions should not be viewed as a call for a formal procurement program for specific Advanced Technology Pilot projects of limited function and duration. Rather, these pilots should serve as a first step in a long term process, nurtured by a funding program that seeks continual renewal of fielded systems, incremental improvements and new services, all driven by persistent public input. Several immediate opportunities exist, as follows:

- All Advanced Technology Pilot projects will have important security and privacy aspects which must be addressed; technologies must be implemented that build and maintain the public's trust across all fielded systems.
- The statistical program of the Federal Government is a likely target of opportunity. Efforts have already started to integrate the various statistical data sources culminating in the recently announced Web site <http://www.fedstats.gov>. Various Advanced Technology Pilots could be initiated to demonstrate substantial benefits to the public such as improved statistical information dissemination and increased timeliness and quality of service.
- Crisis management assistance. Emergency management professionals in the field (who may be less computer literate than desirable) would benefit from automation that helps guide the process of finding, acquiring, integrating, summarizing and visualizing appropriate information to manage a live crisis. For example, online utility maps integrated and overlaid with demographic data may provide a potent aid in developing optimal strategies for water management in wide spread fire emergencies. For related information see the National Research Council report "Computing and Communications in the Extreme: Research for Crisis Management and Other Applications (1996) at <http://www.nap.edu/readingroom/enter2.cgi?CO.html>.
- Demonstrate innovative use of information technology to re-engineer work processes or to link across programs and agencies for proposal and research procurements, e.g., the NSF Fastlane project, a recent project at the Office of Naval Research, and comparable projects of the National Institutes of Health and the Department of Energy. A formal research program emanating from this activity may serve as a bootstrap Advanced Technology Pilot project for a proposal brokerage function.
- Demonstrate benefit to one or more agencies or programs through lower operating costs by selecting a legacy, stovepipe system for migration to a modern network-based, middleware driven operating environment. Examples of current projects that may be leveraged include International Trade Data System, National Environmental Data Index, NTIS Open Source Information System, Public Key Infrastructure Pilot for Information Infrastructure Protection Services, National Computer Security Resource Clearinghouse, Federal Internet Security Support Program, and the Government-wide Electronic Messaging Program.

Recommendation 5. Fund a substantial number of Advanced Technology Pilot projects and basic exploratory research projects through Broad Agency Announcements.

We believe that the "DARPA model" of funding or procuring research is the best model to consider for Federal Information Services R&D. Here, a program need is identified, and through multi-agency funding and cooperation, a Broad Agency Announcement becomes the most effective way to focus attention to solve the problems. Two types of projects are envisioned, Advanced Technology Pilot projects of substantial size and cost, and more focussed applied and exploratory research.

Advanced Technology Pilot projects are expensive, cannot necessarily be done solely by graduate students and must have enough demonstrable results to persuade agencies to procure a final deployable system. Furthermore, the research and development activities cannot be done "in place". Federal agencies cannot be expected to provide development capability within their operational centers, but rather additional facilities will be needed to build systems along side or possibly remotely over high speed communication lines.

We envision Advanced Technology Pilot projects requiring several senior researchers or faculty principal investigators, several graduate students, full time programmers and system designers, and substantial equipment, and storage and communication costs. Such projects will cost on the order of \$1-2 million each per year. Moreover, projects involving multiple agencies will need larger investments.

More basic exploratory research projects involving a senior researcher, several graduate students or post docs, and modest equipment expenses are typically funded at at least the \$500,000 level.

It is difficult to imagine smaller grants could effectively address the complex issues addressed in this document. Furthermore, it is likely that a substantial number of projects will be needed to address the range of issues and problems. It is also important that these projects be conducted concurrently and collaboratively since the field of computer and information science is moving too rapid to force a sequential effort.

We emphasize that this effort needs to be funded incrementally, to keep pace with new developments and larger mandates that may emanate from successful projects.

Recommendation 6. Provide Stable Funding for a Focused Program of Applied Research, Technology Transfer, and Training/Education Activities in the Federal Information Services Domain

Leverage and direct the Information Technology Innovation Fund administered by GSA to broaden its range of activities from near-term production improvements to shared infrastructure, technology transfer, and Advanced Technology Pilot projects. This will assist multiple agencies in re-engineering work processes to deploy the technology solutions and artifacts developed. Seek other sources of R&D funding; solicit agency resources and identify cost sharing opportunities.

*Towards the Digital Government
of the 21st Century*

The best investment strategy to realize the full potential of information technology is one that is based on a constant level of funding over long periods of time with occasional larger investments when breakthroughs are discovered. At a constant level of investment, the nation will be able to benefit from the continual improvements in performance and cost of fielded systems. It is also quite natural for people to continually develop, innovate and grow. The systems fielded by the Federal Government ought to keep pace with these new developments and expectations.



APPENDIX A

Details on the Workshop on R&D Opportunities in Federal Information Services

Building upon earlier studies, a Workshop on R&D Opportunities in Federal Information Services was held May 12-15, 1997, in Washington, D.C. in close cooperation and with the support of:

- the Applications Council of the National Science & Technology Council Committee on Computing, Information, and Communications,
- the National Science Foundation's Directorate for Computer and Information Science and Engineering,
- the President's Government Information Technology Services Board,
- the National Institute of Health's National Center for Research Resources, and
- other interested Federal agencies.

The Workshop was hosted by the Information Sciences Institute of the University of Southern California, and was formulated and organized to discuss and recommend ways to involve the information technology R&D community with Federal information services providers. The seventy-eight invited participants were drawn from the information technology research and management communities from academia, industry and Government. (Background materials on the workshop can be found on the WWW at the URL <http://www.isi.edu/nsf>.)

A number of papers were sought and presented from Federal agencies, including the Social Security Administration, the Federal Emergency Management Agency, the National Institutes of Health, the National Security Agency, the Department of the Interior, The Office of Management and Budget, the Department of Agriculture, the General Services Administration, the National Oceanic and Atmospheric Administration, a number of statistical agencies (represented by the OMB), and the Patent and Trademark Office. These papers served to acquaint the workshop participants with the range of issues and problems faced by Federal Information Services providers in their current efforts to modernize their systems and provide new services to the public.

Several of the papers described on-going efforts by Federal agencies who are modernizing their systems to provide the same services now offered to the citizenry, but through direct interaction over the Internet. Others considered new applications that would be possible with a new Federal Information Infrastructure. Indeed, a range of advanced applications were envisaged. For example, crisis management for natural disasters can be guided by automation driven by information services that link, summarize and present critical information on demographics, utility maps and geospatial data to guide and assist in rapid evacuations and force deployment reducing loss of life and property. Economic and statistical data integrated across multiple federal agencies may provide an accurate near-term view of economic activity. Rather than basing policy on stale data on

*Towards the Digital Government
of the 21st Century*

"leading economic indicators", that measure an economy that existed months ago, rapid integration, aggregation and summarization of economic data while it is collected may allow agile policy evaluation and change. A new Federal Information Infrastructure for the Digital Government will lead to new ways in which people and Government employees can interact with each other, make decisions, share ideas and collaborate on common problems. It is conceivable that "Virtual Agencies" might develop that bring together spheres of common interests in multiple agencies with interested groups of citizens to solve a common problem. Traditional lines of responsibility among agencies can thus be broken down and accommodate new interests and special problems more effectively. Policy-makers may be able to easily tap in to the talents and expertise of people from the private sector, and across agencies in the public sector. Intelligent transportation systems are a distinct possibility. Real-time integration of information from databases, sensors and "toll point-of-sale terminals" might make it possible to optimize at huge cost savings travel and distribution of goods. These and many other advanced applications are made possible by a new information infrastructure for **The Digital Government**.

Workshop Thematic Areas

At its first meeting in December 1996, the Workshop Organizing Committee identified five thematic areas to structure and focus the Workshop. Papers were solicited and presented in each category. These five areas were:

Information integration: techniques to define, design and maintain shared taxonomies and ontologies, or the means of mediating queries among multiple data and information sources; collaboration tools for network based information systems to allow widely distributed groups of citizens or Government personnel to interact remotely to achieve common goals; adaptive planning systems for crisis management environments.

Very large scale data and information acquisition and management for geospatial and multidimensional data: technologies to cost-effectively acquire, integrate and view geographic, biological, environmental, social, and economic data and meta-data of all types;

Advanced Analytics for large datasets/information collections: infrastructures to allow the broadest range of analysis techniques to be applied to user selected views of any data and information sources;

Electronic transaction and electronic commerce technologies: common transaction media between Government and the public; demonstrably successful migration strategies from batch-oriented transaction systems to scalable and efficient online systems; security and authentication mechanisms to maintain the highest levels of privacy;

Information services for the public; human-computer interactions: attracting and keeping the citizen connected (e.g., human-centered factors), visualization and presentation technologies to accommodate the widest range of interaction environments and modalities, multimedia objects, tools sets and user's needs and requirements.

Further detail can be browsed on the Web, including:

- The List of Organizing Committee Members
- The List of Workshop Participants, their Organizations and Presented White Paper Abstracts

The Workshop Agenda



ACKNOWLEDGMENTS

Many individuals from Federal agencies and Universities helped to make the Workshop a success. In particular, we wish to thank Melvyn Ciment of the National Science Foundation, serving in his role as Co-Chair, Applications Council and member of the GITS Board. His vision and leadership provided the impetus for the Workshop. Thanks are also due to Larry Brandt of NSF and Stephen Squires of DARPA for their guidance and thoughtful reading of earlier drafts of this report. We also wish to thank the members of the Workshop Organizing Committee who provided many of the technical ideas and much of the written material that this report is based upon, especially Edward Fox of the Virginia Polytechnic Institute, Eliot Christian of the U.S. Geological Survey, Joseph Hardin of the National Center for Supercomputing Applications, Daniel Schutzer of Citicorp Technology Office, and Valerie Gregg of the U.S. Bureau of the Census. Please see <http://www.isi.edu/nsf> for a complete list of names, titles and institutions.



INVITED PAPERS

Nabil R. Adam, Vijay Atluri, and Kathleen R. McKeown
Rutgers University
Clinical Information Systems: Making Use of Research in Digital Libraries
http://cimic.rutgers.edu/~atluri/prop_nsf/prop_nsf.html

Dr. George Araya
Desert Sands Unified School District
Microwave Metropolitan Area Network
<http://www.dsusd.k12.ca.us/smithsonian.htm>

Yigal Arens
University of Southern California
The Universal Information Base
<http://www.isi.edu/sims/arens/UIB.html>

John Aronis
University of Pittsburgh
Knowledge and Data Integration for Discovery
<http://www.cs.pitt.edu/~aronis/>

William Bainbridge
National Science Foundation
Information Systems for Crisis Management in Areas of Controversy
<http://cti.itc.virginia.edu/~jkh8x/soc805/wsb.html>

Kimberly Barnes
Oak Ridge National Laboratory
Information Technology for the Collaborative Management Environment
<http://www.epm.ornl.gov/~kimberly/fis.html>

Susan Brummel and William E. Smith, Ph.D.
The Challenge Toward Science-Enabled Governance: Realizing Embedded Levels of Engagement in a Smart Process
<http://www.isi.edu/nsf/papers/nsfsb.html>

Cavan Capps and William Rankin
Bureau of the Census
A Distributed Data Object Architecture for One Stop Shopping for Government Statistics on the Internet
<http://www.bls.census.gov/dataobject.htm>

Claude Christensen
Department of the Interior
"The Department of Everything Else" - A Study of Information Services, Practices, Directions, Barriers, and Opportunities in a Federal Agency
<http://www.isi.edu/nsf/papers/christensen.htm>

*Towards the Digital Government
of the 21st Century*

James J. Cimino
Columbia University
Access to Health Care Information Resources through Intelligent Controlled Terminologies
<http://www.isi.edu/nsf/papers/cimino.htm>

Cathryn Dippo
Bureau of Labor Statistics
National Statistical Information Infrastructure
<http://www.isi.edu/nsf/papers/dippo.htm>

William Eddy
Carnegie Mellon University
Advanced Methods for Visualization and Analysis of Very Large-scale and Multiple-media Data
<http://www.stat.cmu.edu/~bill/fedstat/fedstat.html>

Stephen Eick and Alan Karr
Bell Laboratories
Visualization: Toward Understanding Information
<http://www.bell-labs.com/user/eick/bibliography/1997/toward/toward.html>
Visualization: Toward Understanding Information (postscript version)
<http://www.bell-labs.com/user/eick/bibliography/1997/toward.ps.gz>

Shelley Ford
Defense Technical Information Center
Survivable, Timely Information Access, Analysis and Dissemination in Emergency Situations
<http://www.dtic.mil/staff/cthoms/CFP1.html>

Michael Goodchild
UC Santa Barbara
The Alexandria Digital Library for Geospatial Data
<http://www.ncgia.ucsb.edu/~good/WhitePaper.html>

Robert Groves
University of Michigan and Joint Program in Survey Methodology
Using Technology to Adapt Government Statistical Data Acquisition to an Increasingly Diverse Economy and Society
<http://www.isi.edu/nsf/papers/groves.htm>

Dr. Amar Gupta
Massachusetts Institute of Technology
Transportation Information Access Dual-Use Initiative: TRIAD
<http://www.isi.edu/nsf/papers/gupta.htm>

Dana Hallman-Bama
Cooperative State Research, Education and Extension Service, USDA
Designing a Process Centered Intranet
<http://www.isi.edu/nsf/papers/bama.html>

Eduard Hovy
University of Southern California
A Standard for Large Ontologies
<http://www.isi.edu/nsf/papers/hovy2.htm>

*Towards the Digital Government
of the 21st Century*

Dean H. Judson, PhD.
University of Nevada
Assessing the Quality of Data in Data Warehouses and Administrative Records
<http://www.isi.edu/nsf/papers/judson.htm>

Alan Karr
National Institute of Statistical Sciences
Turning Data into Knowledge: Pilot Projects
<http://www.niss.rti.org/papers/fedinfo.html>
Turning Data into Knowledge: Pilot Projects (postscript version)
<ftp://server.niss.rti.org/pub/papers/fedinfo.ps.gz>

Gary Marchionini
University of Maryland at College Park
Public Access and Use of Government Statistical Information
http://www.glue.umd.edu/~march/govinfo_paper.html

Sallie Keller-McNulty
Kansas State University
Data Access and Data Confidentiality in a Decentralized Federal Statistical System
<http://www.isi.edu/nsf/papers/sallie.htm>

Hendrik Meij
SEDAC/CIESIN
MABLE+
<http://oseda.missouri.edu/plue/geocorr/doc/mablewp.html>

Reagan Moore
San Diego Supercomputer Center
Information Based Computing
<http://www.sdsc.edu/MDAS/Proposals/IBC.html>

Ron Musick
Lawrence Livermore National Lab
Large-Scale Data Mining; Pilot Project in Human Genome
<http://www.isi.edu/nsf/papers/musick.htm>

Walt Scacchi, John Noll, Cedric Knight, Capt. Felton "Jay" Miller
University of Southern California
(Re)Engineering Research Grants Management: From Acquisition Reform to Knowledge Brokering
<http://www.usc.edu/dept/ATRIUM/NSF-FIS-Workshop.html>

Gary Secrest
National Security Agency
Establishing the Zone of Cooperation Between Government, Industry, and Academia
<http://www.isi.edu/nsf/papers/secrest.htm>

*Towards the Digital Government
of the 21st Century*

Salvatore J. Stolfo

Columbia University, Department of Computer Science

Data Mining over Distributed Databases and its Application to Fraud and Intrusion Detection in Financial Information Systems

<http://www.cs.columbia.edu/~sal/JAM/PROJECT/nsfwhitepaper.html>

Data Mining over Distributed Databases and its Application to Fraud and Intrusion Detection in Financial Information Systems (postscript version)

<http://www.cs.columbia.edu/~sal/JAM/PROJECT/nsfwhitepaper.ps>

William T. Turnbull

Deputy Director, HPCC, NOAA

Next Generation Internet and Federal Applications

<http://www.hpcc.noaa.gov/pubs/NGIandFIS.html>

Rodney Van Meter

USC/Information Sciences Institute

Latency Management and Quality of Service in Storage Systems

<http://www.isi.edu/~rdv/sleds/posn.html>

Latency Management and Quality of Service in Storage Systems (postscript version)

<http://www.isi.edu/~rdv/sleds/posn.ps>