Cyberinfrastructure and Digital Government

Yigal Arens, USC/ISI, arens@isi.edu
Jamie Callan, CMU, callan@cs.cmu.edu
Sharon Dawes, CTG, Albany, sdawes@ctg.albany.edu
Jane Fountain, KSG, Harvard, jane_fountain@harvard.edu
Eduard Hovy, USC/ISI, hovy@isi.edu
Gary Marchionini, UNC, march@ils.unc.edu

Introduction

The NSF's proposed Cyberinfrastructure (CI) Program addresses the creation of electronic infrastructure to enable more ubiquitous, comprehensive knowledge environments that provide complete functionality for the science and engineering research communities in terms of people, data, information, tools, and instruments, while providing unprecedented capability and capacity for computation, storage, and communication. Such an infrastructure will enable research, education, and other knowledge-intensive communities to share and collaborate over time, distance, organizations, and disciplines.

It is important however to recognize that the CI being created, enhanced and expanded today will impact all expressions of social behavior and most specifically the workings of government. It makes no more sense to plan for the support future CI will provide to science and engineering in isolation from its social, cultural and governmental value than it does to consider only the academic purpose of the Internet and the Web as they exist today. All decisions made in the course of building up our nation's CI will impact social and governmental functions. The interests of the latter must therefore be considered in the process.

The above is not just an abstract declaration of principle. Government is in a unique position as a supporter of research. It collects and disseminates data and provides services that are unavailable from any other source. Digital Government (DG) in and of itself is a topic of ongoing research which has increased in importance, scope and visibility as the roles of government in information societies around the world have become clearer. Its scope is inherently inter-disciplinary, spanning many areas of social science and technology. Naturally, computer science is an essential subfield of research. The NSF's DG research program has supported research in databases, natural language processing, user interfaces, data privacy, and more. But given DG's ultimate reliance on utilization of technology, sociology, business and law are inextricably involved. The countless applications to different government services require the participation of researchers from fields as diverse as communications, statistics, transportation, social work and architecture—to name only some.

As a key sector of society, government is a large, influential, and potentially high-profile tester and user of CI. Government can provide many challenges in the form of specific research opportunities for planning, building and evaluating a 21st century cyberinfrastructure. Most CI elements are relevant in digitally-enabled government functions and democratic processes, including networks, distributed computation, distributed large heterogeneous datasets and long-term storage and archiving, middleware, necessary security capabilities, sensors, shared common tools (e.g. visualization, spatial data manipulation, knowledge discovery), sustainability, and the understanding of societal impact of information technologies.

Government is not only a user of CI, however. Government has several characteristics that make it unique, and thereby pose challenges for CI that no other sector of society is likely to. Frequently, these characteristics are extremes (government must serve *everyone*, it must be and be recognized as *completely trustworthy*, it must be and be recognized as *completely fair*). CI that successfully serves government needs will therefore likely also satisfy the needs of less-demanding research, education, and even commercial sectors.

Finally, government is ultimately directly responsible for much of CI, just as governments today are responsible for many physical infrastructure components either directly (e.g., Interstate system, water systems) or through regulation (e.g., power grid, broadcasting) and for aspects of social infrastructure (such as food safety and health care).

In this White Paper we strongly recommend that the NSF consider seriously the unique role of government, its unique potential as a challenge provider and as a user of CI, and the likely high-benefit rewards of creating cyberinfrastructure that enables government to serve society—all of us—better.

What makes Government Unique?

The special nature of government brings to light problems and research opportunities for CI that are not fully addressed anywhere else. These opportunities derive from government's multi-faceted structure and unique powers to shape and influence our social and physical worlds. Because government is not a monolithic entity but a collection of thousands of jurisdictions and agencies sharing responsibility for the public good, infrastructure must be extended to situations at both the largest and smallest scales of operation for both government and civil society. Every community from the largest state to the smallest municipality needs to be connected in useful and affordable ways to a robust and flexible cyberinfrastructure. These problems of scaling up and scaling down, in both technical and economic terms, deserve close attention.

Government's regulatory and coercive powers are intimately tied to information. From the recognition of issues that call for government intervention, to the development of statutes and regulations that address them, to the monitoring and enforcement process, government relies on elements of cyberinfrastructure to gather, analyze and use information in many formats derived from a wide variety of sources. Cyberinfrastructure research should be concerned with the sources, type, quality, and completeness of data used in each of these activities, as well as with policies and means of access and preservation.

Homeland security has dramatically raised the visibility of information as a crucial element of governmental action. Proposals to harvest and analyze information from widely disparate information sources, and to make decisions affecting national security as well as civil liberties based on the results are now being considered and implemented. This makes it more important than ever that research focus equally on technical tools and the policies and methods of governance and citizen engagement in democracy that will ensure their effective and controlled use.

In particular, government's specific requirements demand new collaborative research across disciplines and place special demands on the developing cyberinfrastructure. These requirements include:

- *Ubiquitous service and universal access*: The government may not offer service in only profitable market segments nor choose to ignore certain sectors of society. FedEx may elect not to serve remote or sparsely populated areas. The US Postal Service must deliver to everyone, everywhere. As a result, IT solutions for rare situations must be considered in government contexts. The Americans with Disabilities Act mandates government ensure equal access to its information and services regardless of citizens' individual abilities and capabilities. This requirement introduces fundamentally unique challenges and complexities to the equation of ubiquity and access.
- *Confidentiality*: The government has access to information that simply may not ever be made public. Credit card companies, for example, can balance the cost of protecting your credit card number against the cost of compensating you for fraudulent use. The SSA cannot do the same with Social Security Numbers.
- *Unique and objective assessments*: Voting, census taking and many other information collection and evaluation endeavors by the government put it in the position of a trusted repository and source. This requires special measures of confidence, fairness, reliability, and objectivity that commercial organizations are not subject to. Should another narrow Florida-like election occur, all IT involved must be verifiable, and its role transparent.
- *Information sharing*: Information integration solutions for government are more complex and need to be better controlled. Commercial enterprises that obtain data about individuals are essentially free to exchange and cross correlate it. This is not the case for government, not even for the Census Bureau and the IRS, and not even within individual agencies. In certain circumstances, a socially sanctioned limited amount of sharing may however greatly improve services. Even the CIA and FBI, whose charters *include* the collection of data about individuals, are not free to exchange information to that degree. This issue is of particular importance as homeland security gains centrality in the national agenda.
- *Security watch*: Government has a responsibility toward citizens to provide safety from crime, terrorism, and even natural disasters. Its special powers of information gathering, correlating, and usage in action must be open to

supervision and accountability. These powers as well as access needs by individuals change in drastic ways in times of crisis, and then revert to normal. The unique problems posed by such events must be understood and handled effectively.

Some Broad Research Directions

Because government is unique in some fundamental respects, it is unrealistic to expect that direct application of scientific or commercial hardware and software products will adequately satisfy its IT needs. In this section we outline four general directions for cyberinfrastructure research enabled by and specifically attuned to government.

1. Government as a Provider of Data

During the last decade our view of the cyberinfrastructure has changed from a focus on computer hardware and networks alone to include large information repositories and large and diverse user communities. The Internet, for example, had little public impact until the arrival of Web search engines. It is no longer possible to do serious research in some fields without access to datasets that contain millions of records or user communities that contain thousands of active users. The need for such resources is in many cases a barrier to entry for young scientists and scientists from small universities or research labs.

The federal government is one of the largest publishers of information in the country, and it is a publisher with unique characteristics. The diversity of the information it creates, collects, and publishes is unmatched by the private sector, and much of what it publishes is unencumbered by copyright restrictions. Specialized user communities have grown around specific government Web sites; recent E-Government initiatives at Federal, State, Local and Tribal levels will eventually make government one of the largest providers of services that use the cyberinfrastructure.

In spite of recent efforts, government use of information technology remains far inferior to that of the private sector. Information is published with little thought about how to make it accessible beyond the core community it was written for. E-Government services are struggling to reach levels achieved by E-Commerce sites in the middle-to-late 1990s. Public policy is increasingly influenced by data-driven research, but many government agencies do not have sufficient ability to analyze large volumes of data quickly and accurately. Integration and analysis of information collected by different agencies, where legal and appropriate, remains one of the weak links in homeland security.

Partnerships between government agencies and academic researchers solve problems for both partners. Government agencies have expert knowledge of policy domains and their challenges and can supply the large and varied datasets and the large user communities that are crucial to modern research in many disciplines. Such partnerships significantly leverage research funding by eliminating the need for researchers to recreate resources the government already has. They also focus research attention on problems that are important to government agencies and society at large. Academic researchers become a source of new thinking, technology, and expertise that government agencies don't have.

Partnerships between government agencies and academic researchers should be a national priority. The sudden availability of information during the last decade has caused a dramatic shift towards data-driven research and large-scale data analysis across a wide range of academic disciplines, and yet a scarcity of large research datasets often forces scientists to search under a small number of lampposts. The government collects large amounts of information in support of varied policy-making and regulatory efforts and for national security reasons. Effective government use of the cyberinfrastructure offers the possibility of dramatically improving the quality of data-driven research on a wide range of technical and social problems to improve the effectiveness of government at all levels, and more generally to improve the quality of life in the United States.

2. Data Management and Digital Preservation

The federal, state, and local governments have unique data management problems because their activities are distributed across a large number of agencies and departments. Navigating this maze is difficult for ordinary citizens and small businesses, resulting in niche systems—often manually-maintained and non-scalable—such as GovBenefits.gov, that provide cross-agency "one stop shopping" services. The integration and sharing of information among government IT systems, when allowed by law, is weak and *ad hoc*. One research challenge is to create federation layers that provide integrated services across a large number of diverse information systems that are governed by different legislative, privacy, and other requirements.

One of the foundations of democratic government is that the government must be responsive to its citizens. New communications technologies such as email, online discussion forums, and wireless devices have been adopted widely and now make communication with others a much more frequent event than at any other time in history. The government has been slow to respond. Information retrieval, text analysis, and natural language processing are required to enable government employees to respond intelligently and efficiently to large volumes of email, public commentary, and other citizen communications, and to organize and summarize it for use by policy- and decision-makers.

Democratic governments are also required to keep accurate records, make them available to the public, and preserve the essential records of government for the life of the republic. The government's records are now largely digital, so digital preservation must become a core government competency. Digital preservation is important across the cyberinfrastructure, but the government's needs are unique and foundational and thus must play an important role in shaping the technology. Businesses may have the luxury of discarding records or using derived forms when the costs exceed the business value; democratic governments don't.

3. Privacy and Trust

Government's special role in life will be mirrored in CI. People cannot opt out of government—one must adapt one's behavior to laws and regulations. The quid pro quo for our requirement to participate is the notion that in a democracy, we are all part of the government—we can speak out, vote, and be elected to office. The implications for CI are myriad, ranging from online voting mechanisms to the electronic versions of the thousands of forms federal, state, and local agencies require. Research is needed on secure and easy to use mechanisms for interacting with government at all levels.

Beyond the systems that support citizen interaction is the need for government to maintain trust at multiple levels. At an obvious level, the distributional, regulatory, and coercive powers of government demand that personal information and transactions be held in confidence—that is governments must sustain trust that individuals' privacy will be protected while their needs are being served. Additionally, the many different views that government has for each individual are more intimate and diverse than any set of corporate entities will have and these views must be easily auditable by individuals on demand. Commercial interests may take care of basic privacy and auditing needs in CI, but the special relationships among people and governments demand more research and special solutions.

At a more problematic level, governmental powers must be balanced by CI mechanisms that help people trust that they truly are a part of our governments—that our voices matter. The CI systems that support decision making (e.g., public debate, voting, rulemaking, etc.) must be believable, transparent, and archival. Transparency and confidentiality are two sides of the trust coin and represent a delicate balance that CI must address since the high hurdles for access and anonymity that physical record-keeping often provides is moot in CI. The research community and government must work together to insure that the technical and socio-political challenges these multiple levels of trust present are met in CI. This requires research beyond technical issues such as scale or efficiency to include governmental interfaces that help people see how their inputs fit into the eventual behavior of government. CI should support the right of people to trace the flow of their inputs into the aggregate pools of evidence that shape policy and law.

4. Social Issues: E-Democracy and E-Rulemaking

It is no accident that the three key computer science research directions outlined immediately above—government as a provider of data, data management and preservation, and privacy and trust—are fundamental not only to CI but also to democratic governance in a knowledge-based society. Access to information is critical for citizens in a democracy. Data management and preservation are rapidly becoming essential elements of policymaking at every level of government not only in the United States but globally. Privacy and trust are fundamental to liberty and democratic freedom.

Beyond these fundamentals of information and computer sciences, a wide range of pressing social science challenges have emerged with the growth of digital government. To note a few examples, science and technology researchers and government officials know little about the development, management and governance of web-enabled cross-agency and intergovernmental arrangements. For such organizational and political arrangements to become feasible, essential questions of governance regarding accountability, jurisdiction and control must be worked out. These conceptual and empirical challenges indicate the close interrelationship between CI and the social, organizational, group-level, and individual behaviors anticipated in information-based environments. Similarly, little is known about how individuals and groups communicate and build alliances and networks in the type of environment associated with CI. It is not only science and technology researchers who will function in a new environment; all types of deliberative communities of people, data, information, tools and institutions will increasingly rely on and be influenced by CI. Citizens and policymakers increasingly use geographical information systems (GIS) in ways that deeply influence decision-making processes. Electronic voting and electronic rulemaking are similarly reshaping political and policymaking functions as deliberative and decision making processes that depend critically upon CI. These and other basic processes of democratic systems depend upon the security, reliability, privacy, and transparency of CI.

The Role of Digital Government

As noted earlier, the nation's CI impacts all expressions of social behavior. CI for science and engineering cannot, therefore, be considered in isolation from its social, cultural and governmental value.

The NSF recognized this dynamic in the creation of its Digital Government (DG) program in 1999. Digital Government is a topic of ongoing research. Its scope is inherently interdisciplinary, spanning many areas of technology and social science. Naturally, computer science is an essential subfield of its research. The NSF's DG program has supported research in database access and integration, natural language processing, user interfaces, data validation, privacy, and security, long-term archiving of information, and more.

However, government is inherently a social enterprise. The problems of digital government are not entirely technical and often not common to typical businesses and service providers, and little commercial effort can be found addressing the non-technical and uniquely governmental aspects of government needs. In particular, as noted above, decisions on target constituencies and functionality in government services (such as issues of privacy and trustworthiness) are legal and political, not subject to simple economic or business tradeoffs. Such intertwining of social, legal and technical issues is something that can best be addressed by research efforts targeted specifically at digital government. Therefore, in addition to DG's ultimate reliance on utilization of technology, disciplines such as sociology, political science, public administration, business, and law are inextricably involved. The countless applications of information technology to different government services require the participation of researchers from fields as diverse as communications, statistics, transportation, social work, criminal justice, and architecture, to name only a few. Researchers in the NSF's DG program have collaborated at the federal, state, or local levels with government employees in a wide variety of areas, including:

- National Statistical Information Infrastructure (Census, Bureau of Labor Statistics, Department of Energy);
- Public safety and law enforcement (data and process integration, data quality, interorganizational and intergovernmental relations);
- Crisis Management and emergency response (infrastructure needs, data quality and availability, public information processes, monitoring , sensing);
- Democratic practices – citizen engagement, voting, information on candidates, legislative and judicial processes and E-Rulemaking;
- Environmental data integration, analysis, monitoring and management;
- Public benefit programs (social security, veterans affairs, housing, welfare);
- Shared physical infrastructure (highways, rail, shipping);
- Public decision making (modeling, simulation, citizen–government dialog, auditing trails, public datasets, transparency); and
- Security (international, homeland).

This research is already establishing aspects of cyberinfrastructure for government, in a nascent form. Devoting some portion of CI research specifically to government can provide a massive boost of technology and overarching infrastructure to these efforts, and simultaneously reap the benefit of providing contacts, demonstrations, and in some cases solutions to real government problems.

Conclusion

The report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure emphasized the dangers of not acting quickly and with sufficient decisiveness. Among other issues, it noted the threat of proliferating data formats, loss of observational data, increased technological “balkanization”, wasteful redundancy, and more. Building up the nation's cyberinfrastructure while ignoring the very central and significant digital government constraints and contributions outlined in this document runs precisely the same risks. A misstep at this time will result in incalculable harm to governmental research and societal functions that will use the national CI as their basic medium. Furthermore, the investment in CI that the NSF is currently contemplating is very large. If we do not consider digital government issues from the start, the cyberinfrastructure that results will not serve the unique needs of one of its largest and most important information and service providers. Such an oversight now would require the investment of a comparable sum again later to correct incompatibilities and overcome failures. This would constitute a colossal waste of national resources.