IDENTITY IN DIGITAL GOVERNMENT

A RESEARCH REPORT
OF THE

DIGITAL GOVERNMENT CIVIC SCENARIO WORKSHOP

CONVENED BY
L JEAN CAMP

SPONSORED BY
THE NATIONAL SCIENCE FOUNDATION
&
THE KENNEDY SCHOOL OF GOVERNMENT

Identity in Digital Government

A report of the 2003 Civic Scenario Workshop

Jean Camp
Organizer

An Event of the
Kennedy School of Government
Harvard University
Cambridge, MA 02138

Sponsored by the
National Science Foundation
Arlington, VA

For additional copies of this report, please contact:

Jean Camp
Kennedy School of Government
79 John F. Kennedy St.
Harvard University
Cambridge, MA 02138

Acknowledgements

My research team deserves first and foremost acknowledgement. The doctoral students Warigia Bowman, Allan Friedman, Sabine Schaeffer, Sara Wilford, and Rachel Greenstat all provided important assistance with the event. In addition, Warigia Bowman and Allan Friedman were critical in the development and review of this document.

For the workshop itself, it would have been nothing without the attendees, all of whom provided exceptional and unique expertise. The leaders of the scenario groups and the authors of the technology descriptions deserve particular thanks. These include Ari Schwartz, Paul Syverson, Georgia K. Marsh, Stuart Schechter, and Allan Friedman with Sara Wilford. The leaders of the technology description groups were Bennet Yee, Elaine Newton, Carl Ellison and Roger Dingledine. Anna Lysyanskaya provided an excellent description of the capacities of cryptography for the event. Elaine Newton offered specific contributions to the report, including biometrics best practices, methods for creating privacy-enhancing biometrics, and the relationship between false positives and negatives in a biometric system.

Lori Uhland of Rare Events made the entire event run smoothly from soup to nuts. Thank you, Lori.

Thanks to Jane Fountain and the Center for Digital Government at the Kennedy School of Government. Also, I would like to thank catering and security services at the Kennedy School. Thanks to Catherine Gorodentsev and Michael Johnson for organizational navigation.

I am especially grateful to the National Science Foundation. Communicating complex organizational requirements to technologists and technical subtleties to decision makers are perennial problems. These difficult problems are compounded when combined. The National Science Foundation had the vision to fund a scenario-based workshop not only to examine identity in digital government, but also to evaluate the method for crossing disciplinary, institutional and organizational barriers.

My most lasting appreciation goes to each of those people who paused in their demanding positions to work on the critical problem of identification and authentication for digital government. It was a rare honor to work with such excellent minds for two days. Thank you.

Introduction

In order for government to fulfill its critical functions, it must be able to authenticate its citizens' claims about their own identities and characteristics. As digital government becomes a reality, the need for reliable digital identifiers becomes increasingly urgent. At the same time, digital government identifiers create unique threats to privacy as current practices of using personal information break down. The wide availability of information through electronic networks has the potential to erode privacy at an unprecedented rate, as well as making authentication based on personal "secret but shared" information increasingly untenable.

The Digital Government Civic Scenario Workshop convened to address the wide range of issues surrounding digital identity, and to plot a course to better understand the concept of digital identity through further research.

Objectives

Identity management is an emerging field. Yet there is an immediate need for a functioning digital authentication infrastructure. There is no single path forward. The Digital Government Civic Scenario Workshop brought together individuals from different domains and disciplines to develop a shared vision of future identification systems, and determine the key questions that need to be answered in this arena. This workshop produced four distinct results, for the nation and the participants.

Best Practices: The workshop participants identified a series of principles that any identity system must employ. For example, flexibility and adaptability are primary design values for a system that inevitably will not be flawless. Other best practices are also highlighted: in technical implementation, privacy practices and processes for risk management must receive increased attention.

The Liberty Alliance and the Microsoft .Net initiatives both promise to bring uniform management to all web sites. In addition there is an open source alternative to the .net project being led Ximian, called Mono. Because the "single sign-on" technologies are still emerging and quite competitive, clearly articulated e-government requirements could guide the development and adoption of these technologies in e-government.

Research Agenda: To help guide future decisions, the workshop continually noted open questions, important areas of research and points of contention in questions of identity. This report identifies seven major topic areas in the realm of identity management, and six broadly defined academic

fields under whose purview these problems might fall. Qualitative and quantitative research questions were explicitly identified when possible.

Community Building: The participants in the workshop came from a diverse set of backgrounds across the public, private and academic sectors. Pre-conference communication allowed experts to share their field of expertise with others, while the presence of practitioners-- especially from the state government level-- offered critical real-world insights in what was occasionally a model-driven discussion. The effects of this network-building event are expected to extend beyond the event and individuals to the institutions and professional organizations of the participants.

Shared vision: As is discussed in the second section of this report, identity management is a very complex problem. The assemblage of technologists, practitioners and policy academics allowed each to share their perspective and highlight the issues they felt important. The participants came away with a more complete understanding of how to approach identity in a networked world. The participants left with a common understanding of the difficulties of governing in a networked world.

Relevant Values

Technical systems have values embedded in their design in that their functionality makes some actions easier and some more difficult. (Friedman, 1997) Workshop participants came from many different backgrounds, personally and professionally, and each brought a unique set of values to the discussion. Bringing these values out to the forefront ensures that the research will address all pertinent issues. Not all the values identified below are shared by all participants, or by the authors of this report, but they all were present in the workshop.

Autonomy*:* Personal freedom is a foundation of American life. Individuals often expect to control the important aspects of a given system as it relates to them. Any new infrastructure should not impede personal liberties, but make it easier for individuals to interact with the government.

Privacy*:* Privacy means many things to many people. For some it is a measure of personal dignity: they do not like to be treated in a large system as only a number. Yet others may reject personalization and object to being 'watched' via data compilations. In terms of autonomy, the fear of having one's actions recorded and reported may prevent individuals from actualizing themselves. Privacy invasion often brings with it concerns of direct control rather than indirect incentive, not only the totalitarian concerns of "Big Brother," but also coercion by private actors.

Internal Government Efficiency*:* An important e-government goal is to enable government to do more for less. Faster, better and cheaper service is a key goal of digital government. Different data systems, and organizational plans makes it difficult, expensive and time consuming to work across bureaucratic boundaries.

Free Flow of Information*:* Impediments on the flow of information, whether legal or technical, can become troubling obstacles. A faster flow of information creates more efficient workplaces and markets, and saves expenses related to regulation. Information flows across national borders are also of concern, as the ongoing controversy over Safe Harbor with Europe shows. Access to information allows for more effective fraud detection and prevention.

Responsiveness and governance*:* Democracy is predicated on the interaction between citizen and state. Many theorists posit that improving citizens' ability to communicate with leaders and access government resources strengthens democracy. Governance may also be improved by faster service, less government paperwork and greater citizen understanding of the processes of government.

National Security and Law Enforcement*:* When an individual is a threat to society, then society should be able to protect itself by identifying and apprehending that individual. Crime using the information infrastructure, including identity theft, fraudulent spam, and malicious code, has grown dramatically recently. Identity theft is reported to have harmed 57 million Americans, and is the fastest growing crime in the country. Violations of privacy, such as identity theft, are used by terrorists and common criminals. false identities are used to create confusion, and perpetrators can rage from the attackers of September 11 to simple shoplifters. Potential victims deserve as much protection as possible. A new identity infrastructure should not create opportunities for increased criminal activity.

Definition of Terms

In any discussion, the participants must have a common vocabulary to ensure productive and meaningful communication. This is particularly true in interdisciplinary discussions, which include technical content. Computer science and information science frequently redefine common words, and it is easy to confuse their more colloquial meanings and connotations with specific information science concepts. Even across disciplines terms may have distinct meanings, so that even scholars may use these terms differently. Harmonization and clarity are the goals of these definitions, not canonical determination of meaning in all contexts. These terms are defined to allow readers to make full use of this report.

Attribute. A characteristic associated with an entity, such as an individual. Examples of persistent attributes include height, eye color and date of birth. Examples of temporary attributes include address, employer and organizational role. A Social Security Number is an example of a long-lived attribute. Some biometrics are persistent (e.g. fingerprints), some change over time or can be changed (e.g. hair color).

Identifier. An identifier identifies a distinct person, place or thing within the context of a specific namespace. For example, an automobile, a bank account and person each have identifiers. The automobile has a license plate and the bank account has a number. The person may be associated with either the auto or the account through additional information, such as a certificate or ownership, or a social security number. One identity can have multiple identifiers: a car has a permanent serial number and a temporary license place. Each identifier is meaningful only in a specific context, or namespace, and can reasonably be thought of as having a <thing identified, identifier> pair.

Personal identifier. Persistent identifiers associated with individual human and attributes that are difficult or impossible to change, such as human date of birth, height and genetic code.

Identity. The set of permanent or long-lived temporal attributes associated with an entity.

Anonym (as in anonymous). An identifier associated with no personal identifier, but only with a single-use attestation of an attribute. An anonymous identifier ascertains an attribute, once. An anonymous attribute used more than once becomes a pseudonym.

Pseudonym An identifier associated with attributes or sets of transactions, but with no permanent or personal identifier.

Identification Association of a personal identifier with an individual presenting certain attributes. For example, accepting the association between a physical person and claimed name, or determining the association with a medical record and a patient using physical attributes.

Authentication. Proving an association between an identifier or attribute, and the relevant entity. For example, an automobile is identified by its license place, and that is authenticated as legitimate by the database of cars that are not being sought for enforcement purposes.

Identity Authentication. Proving an association between an entity and an identity. For example, the association of a person with a credit or educational record.

Attribute Authentication. Proving an association between an entity and an attribute. Confirming some one's age is an example. This is usually a two-step process, where the association between an entity and an identifier is established, and then a link between identifier and attribute is established.

Authorization. A decision to allow a particular action based on an identifier or attribute. Examples include the ability of a person to make claims on lines of credit, the right of an emergency vehicle to pass through a red light or a certification of a radiation-hardened device to be attached to a satellite.

Defining the problem

The concept of identity is a sticky one, even in the sole context of a bureaucracy. How an individual is represented in a system and how that individual can prove that the information in the system refers back to him has a profound impact on system use. Identity management doesn't refer only to ID cards, or databases but a large infrastructure of personal information and authentication mechanisms, spanning public and private sectors and touching many aspects of modern life.

Evolution of Identity

As many have observed, a pre-bureaucratic society has little need of identification management, since social interactions were on a small enough scale to rest on trust and personal recognition. (Fukuyama 1995) Even well into the modern bureaucratic paper-driven era, access could be physically limited and institutions protected access to records and files; transactions were made within the confines of a physical locale and relatively closed networks. Computerization has made transactional histories more detailed and networks have made them available to many. As more data have become available, data have been integrated into available systems, enabling services never before possible. Information in the digital world can flow freely, and be copied and stored at almost no expense, and it is on this information that our transactions have become more dependent. In the increasingly digital realm, trust depends on transactional history-credit reports, educational history, employment history, and even criminal or medical history. This extension of trust is based on transactional histories associated with some common identifier. Across administrative domains, that identifier is often the Social Security Number (SSN).

The problem arises when an individual presents herself and requests permission for some transaction: how does the system, in the form of a bureaucrat or, increasingly, an automatic machine, confirm that the person is, in fact, who she says she is. The individual presents her identifier and the system must somehow confirm that the identifier refers back to the individual rather than some third party. This is the heart of identification, and a surprisingly complex problem.

A common solution today is to assume that there is some information that only that individual will have access to, and that presentation of that information, such as a mother's maiden name confirms the individual's identity. Today, a SSN and a mother’s maiden name can prove worth for creditor, employer or, increasingly, authorization for a one time purchase of

discount goods from the web. The problem, of course, is that the system verifying identity must also have access to the "secret" information. Such widespread distribution of information casts doubt on the necessary assumption of this scheme: that only the individual will have access to this information. The number is on a huge percentage of documents and files referring back to the individual, and the family name is often easily traceable using publicly-available documents. Using more personal information, correlated from different sources and compiled without the awareness of the subject is a profound privacy violation. Furthermore, personal information that is widely available in a networked world can hardly be considered secure, making it easier for malicious actors to subvert the system with identity theft.

As more identifiers are linked to one identity, the threat to privacy and data integrity increases, and the security of the data decreases. Absent substantial controls on how this information can be used, shared and stored, there are wildly varying management practices for the same data. Moreover, very high-value transactions and decisions -- employment, professionals managing large transactions -- use the same identity-specific data as very small transactions (Odlyzko, 2002). Because the risk in low value transactions can be decreased using personally identifiable information at the most detailed level (e.g., social security numbers, universal identifiers, credit information) these managers keep data long term. Identifiers simplify price discrimination. Yet because the value of the transactional records is low the level of protection is low. Use of this data resembles use of the proverbial tragedy of the commons -- all parties have incentive to use the data but only one has incentive to protect it according to the highest value. Any party looking to subvert data will seek data or systems at the lowest level of protection and then use the data for authorization to subvert the security surrounding high value uses.

Current identification systems rest on confirmation of personal information, yet that information is not uniformly secure or protected. The declining value of this verification poses a growing threat to the validity of these systems, yet any future verification must depend on currently used documents.

Identity in the Public Sector

Personal information is not only collected in the private commercial sector, of course. The modern state could not exist without a functioning identity system to determine just allotment of rights and privileges. Federal agencies collect information about those they serve and those they monitor. The growing trend to capitalize on information technology has increased reliance on identity systems, as citizens interact with their government in the online world, where information is required to verify identity. Government agencies are harmonizing their databases to allow agencies to collaborate more effectively, making it easier for personal information to flow

from where it was originally collected to other parts of the government. Individuals can grow weary of having to give the government their information repeatedly and may encourage inter-agency information sharing.

The state not only cares who an individual is; the state is also concerned about who individuals are not. In the interest of protecting society, access to anything from entry to federal buildings to moving into a neighborhood might be denied to a suspected terrorist or a convicted sex offender. Recent attention to national security has brought these issues to the forefront. Yet the more mundane practices of government in daily lives, the provision of services, the collection of taxes, and the distribution of benefits requires determining that the individual is in fact who he claims to be with some degree of confidence.

The state also has a central role to play in the administration of identity management systems. Many important tools in the identity process, from birth certificates to drivers' licenses, have shifted from their original purposes and become trusted identifiers. Often these serve as "breeder" documents with which one can generate other systemically valid identifiers. Since a birth certificate is allegedly hard to forge and can thus be trusted, the bearer is assumed to be the have the name on the certificate, thus authenticating that name. Few other institutions have the universality to issue documents that can be accepted on a wide enough scale. Even if one did, the issuer would be exposed to tremendous potential liability were the system to fail. Sovereign immunity hands the job of trusted root and issuer to the government.

The digital government movement also depends on identity systems, for much the same reason as the private sector. In a digitally networked environment the functioning of services requires some mechanism for identification. The President's Management Agenda for E-government proudly touts their achievements and goals, but everything from commenting on a proposed rulemaking to obtaining security clearance requires some amount of identity information. In the public sector, as in the private sector, our identities are practically defined as sorted and correlated personal information. With a focus on information flows throughout the federal administrative system, e-government is closely tied to identity management issues.

At the same time, a public sector solution would not be an easy sell for the American public. Apart from general skepticism in many government programs, many see any attempt to systematize national identification as the first step to a national ID card. (Safire 2001) After September 11 2001, initial studies showed that many Americans may support such a scheme, but it remains an incredibly controversial topic, even touching on individuals' religious beliefs. (DeLotto Behrens & Baum 2002)

Identity Breaks Down

Weaknesses in identity management systems discussed above, combined with the increasing availability of personal information, have led to the rise of "identity theft." This slightly vague term refers to any number of crimes and misdeeds perpetrated using the personal information of another. Government efforts to quantify the magnitude of the problem have been complicated by the challenges of data collection efforts, but recent government estimates are between one quarter and three quarters of a million victims annually. (GAO 2002) A private research company estimates that seven million Americans were victims this past year. (Litan 2003) Absence of a central reporting agency makes estimating the actual numbers difficult but there is consensus that identity theft is a large and growing problem.

Individual cases range from the issuance of credit to individuals under the victim's identity to committing crimes using their identifiers. While the victim is not legally liable for the money spent or crimes committed, clearing their credit record or asserting their good name can be frustrating, time consuming, expensive, and sometimes impossible to do completely. When major identifying documents have been compromised, proving ones identity is incredibly challenging. Many victims report long-lasting damage to their credit reports, livelihoods and quality of life. Information can linger in computers until manually removed, and many decisions are made by silently and automatically consulting databases. An individual may never know whether they have completely secured their identity. As is discussed more below, were it easier to identify and correct identity theft, false claims of theft could itself open a new avenue of fraud.

Fraud against private firms has concrete economic damage, but much of it is based on business practices from the firms themselves, making credit easily and quickly available, and the firms absorb the cost of fraud as a necessary business expense. Government, on the other hand, does not always have that luxury, and is often forced to pay closer attention to its bottom line. Most importantly, the government is not in a position to deny rights and privileges to an actor based on probabilities and efficiencies. Democracy is predicated on equality before the law, and efficacy must be balanced with legal priorities. The cost of fraud must thus include those denied access to what is due. The third cost of identity theft, and perhaps ultimately the most damaging, is a loss of trust.

Trust

Recent scholarship, particularly in the fields of sociology, anthropology and organization theory, has drawn attention to the important role trust plays in society, especially in an online world (Camp, McGrath &

Nissenbaum 2001). Whether between individuals, in commercial interactions or with government institutions, trust is critical for any relationship. Trust also requires a reliable identity framework. An individual must be confident in the relevant attributes of other parties in any relationship. One vision of trust is predictability, manifest in reputation. A reputation must be reliably tied to an identity, and this connection must be durable over a long period of time. Reputations require identifiers, and these too must be bound to the individual. Reputation is possible in pseudonymous or even anonymous systems but these systems are identity management systems nonetheless. They simply manage attribute information in a way that protects all personal identifiers. In an online world, where all identifiers and attributes are in the form of personal or public information, proper management of this information is critical. Too little information accumulated precludes enough trust to build a reputation. Too much information can have a chilling effect on behavior, as people fear commingling of too much identifying information.

A Multi-Dimensional Problem

The issues surrounding identity management are complex in part because the problem is so hard to bound. The set of risks and required analysis are completely different depending on the apparent crisis one is setting out to solve. Identity theft has very different causes than other serious crimes, and wildly divergent risk analyses in terms of costs, probabilities and viable alternatives. Standardizing drivers' licenses to make it harder for teenagers to buy beer would do little to solve either problem if the ease of obtaining fraudulent documents makes them cost-effective. Meanwhile, the enormous expense of developing an identifier system appropriate for Homeland Security may not prevent a doctor from falsifying treatments for Medicare fraud.

Some shift must be made in current identity management. Every scenario drafted for this workshop extrapolated current trends of identity theft, terror and loss of privacy to eventually drive policy shifts in some direction. Every scenario illustrated distinct social, technical, and political unknowns.

The first step is to gain a greater awareness of the scope of the problem, and acknowledge what identity must, can and cannot do. Every scheme will have different abilities, and knowing which issues are of foremost importance is critical in designing the best system. The workshop identified seven critical problem areas that any full discussion of identity management must consider:

Information architecture and management strategy

Privacy and personal information protection

Governmental policies

Accountability inside and outside the system

Metrics for design and evaluation

Implementation of the infrastructure

Roll-out and enrollment phase

These problem areas make up a research space. Each problem area offers many individual questions to be addressed, both qualitatively and quantitatively. In the research agenda section of this report, this range of research topics is presented, along with a means of viewing the whole problem space.

Best Practices in Managing Identity in Digital Government

In developing a compelling vision of the role of identity in digital government, the workshop identified a series of principles that any identity system must employ. Since no one identity management system exists as the right one, best practices can guide development discussion to ensure optimal design. Flexibility and adaptability is identified as a primary design values for a system that inevitably will not be flawless. Other best practices are highlighted, in technical implementation, privacy practices and processes for risk management.

Identity management is an infant science, with very real uncertainty. Therefore the most important recognition is that no initial implementation will be flawless at first implementation. As one participant commented, "There is no magic bullet to solve problems in identity management." Flexibility and the capacity for incremental change are therefore the lynchpins of a successful identity management plan for e-government. The uncertainty in identity management exists on three dimensions: technology, privacy, and processes:

Technologically there are key areas of development relevant to identity, including emerging cryptographic methods, biometrics, mobile devices, and secure hardware.

Emerging cryptographic methods include threshold systems that can be secured according to the different needs of distinct authentication providers. Many of these are options unimaginable in the paper realm. Group cryptography, for example, allows for the proof that a person or device is part of an authenticated group without providing unique identifying information.

Biometrics offer much promise, yet there are significant risks that biometric system design will be based on misperceptions.

The processing power and mobility of devices also changes issues of authentication. Communications and computing devices can be associated with a specific person, a particular role of that person, or may be shared by multiple people who all fulfill the same role. Devices also change hands, often with personal authenticating data remaining on the devices.

Secure hardware solves technical problems but creates policy problems. The most secure hardware is special purpose, difficult to change and often with limited interoperability.

Privacy constraints are not yet determined for provision of services on-line. Citizen expectations of privacy may be in conflict with

citizens' desire for efficient on-line service. Risk perception by citizens, and the associated policy responses connected with those perceptions, don't always accurately reflect or respond to the actual level of risk. For example the Personal Earnings and Benefits Systems offered personal reports on-line using slightly more authentication than the long-practiced phone version. In response to the privacy concerns (Garfinkel 1997) the system was suspended, then canceled, and then replaced with a less secure method using only a purchased mailing list with no user authentication. The Social Security Administration was trapped between two conflicting dimensions of privacy: making information available to the data subject and ensuring that data about one person are not released to another.

The **processes** for security risk management are not defined in terms of process across digital government. The economics of security are uncertain, and is in fact only an emerging area of research. There are not formal quantifiable metrics that are useful for comparing conflicting goals; for example, how would one empirically contrast the risk of information disclosure and the risk of denial of service to an authorized user.

Finally, good definitions are critical. Identity as a solution cannot solve an under-specified problem.

Technical Best Practices

A key element of authentication is that the *authenticator must also be authenticated*. If the system is not configured to authenticate itself to the citizen then effective attacks can be used to misdirect the citizen into disclosing their own authentication information inappropriately. Public key cryptography is the easiest and most powerful way of doing this, yet there are few widely available applications taking advantage of this capability for non-specific uses. Currently the only method to implement this is to institute an SSL connection upon the first request from the browser. SSL is widespread but inadequate. Other existing applications are difficult for average users to master, and many public key infrastructures suffer from their own weaknesses. (Kent and Millet, 2003)

Digital government services may choose to develop their own key hierarchies. However, this may decrease trust in the system if the result is a warning that the key is not from a pre-installed root. Digital government practitioners may choose to purchase a verified key; however, this results in a situation where the citizen trusts the government because the government has paid a company to extend trust.

Technology neutral specifications are optimal for two reasons. First, risks cross technological boundaries. Loss of data is loss of data regardless of platform. Authenticating information may be lost from physical

devices or software failures. Second, focusing on a single technology may result in a myopic concern for a set of particularly well-understood risks or may result in unnecessary bundling of functionality based on an assumed implementation. Moreover, system-neutral standards such as the IETF's standards can promote competition in performance across systems while ensuring interoperability and system-wide quality.

Biometrics are often touted as the solution to the authentication problem as a consistently reliable personal identifier, but *biometrics do not necessarily provide unique universal identifiers*. Biometrics may not identify individuals uniquely; for example facial recognition. Some biometrics (e.g., handprint geometry) are useful only to verify a claimed identity. Some biometrics can be used to identify anyone enrolled in the system (e.g., fingerprints and iris scans) yet there will be some who cannot enroll.

In order to minimize the risk of loss do *not store raw biometric data for authentication.* When biometric data are used as pass phrases, the security of the data is critical. Once biometric data are compiled into a database or accessible over a network, biometric information is simply data and data can be stolen. For example, if the connection to a fingerprint reader over the network is not completely secure false data may be fed into the connection.

Biometric authenticators pose particular problems once subverted. Thus any design based on biometrics must include the possibility that there is a loss of control over the authenticating data. *Biometric systems require measures of loss recovery.*

Finally, biometrics are available to any entity with a reader. *Therefore it is impossible to control the security of a raw biometric.* The authenticating entity can control the template, and the encryption method of the biometric but never the raw authenticating data.

In any authentication system, including biometric systems, the temptation is to manage for the false positive rate. The false positive rate is the rate at which impostors are allowed into the system. Conversely the false negative rate is the rate at which authorized users fail to authenticate. In all systems it is critical to *be as rigorous with the acceptance of false positives as with false negatives*. Biometrics systems in digital government pose a particular challenge as the more under-represented the population the more likely is the false negative.

Privacy Best Practices

Privacy is a problem that is easier to solve with consideration beforehand. Privacy by design is better than post-hoc liability. The phrase from the workshop is that *privacy is better built-in than bolted on.*

Privacy enhancing technologies can resolve the conflict between citizens' desires for efficient service and an expectation of privacy. *Privacy enhancing technologies are most effective when integrated in the design stage.*

Privacy has many dimensions. Some people may want to be left alone; and a simple lack of follow-up contact is adequate. Others are concerned about their autonomy and fear a digital Big Brother. In order to address different concepts of privacy, *use the principles of data protection.* In most cases, if the data are protected then privacy is inherently addressed.

The essential principles of data protection can be summarized as:

No unauthorized sharing

Data collection requires advanced permission

Justification required, including a clear specification and minimum use of data

User review and correction of data

This requires, above all, *knowing what information is needed for any given task.* Note that the requirement for justification of data compilations will be echoed in the process best practices for managing security risks. Limiting data compilations decreases security as well as privacy risks.

For personal information, and particularly for authentication information, *be aware of the life-time of the data.* Data that are no longer useful may become a liability. Keeping data with no specification for use is hazardous to privacy and risky in terms of security management.

Data protection provides a minimal threshold for protecting privacy. Anonymity provides the highest degree of privacy. Thus implementing data privacy does not remove the need to create anonymous alternatives to services when possible.

Process Best Practices

The most critical element of implementing an identity management system is to *have an exit strategy.* Given the range of uncertainty, there will be some strategic failure. Even the most perfect plan can be improved and must be upgraded over time. Even a perfect, flawless strategy that predicts exactly citizen response, diffusion, and integration with current systems will require upgrades as processing power and thus key lengths are altered. Absent an exit strategy, replacement and upgrade costs can be very expensive, and a publicly failed strategy may mar future attempts.

An exit strategy requires avoiding lock-in. Lock-in can result when any part of the technology forces backwards compatibility to the existing

system and thus limiting future choices. Examples of problematic lock-in within computer networks range from the centuries old lock-in by knowledge externality of the QWERTY keyboard to the modern shortage of Internet Protocol numbers created by IPv4. Lock-in can result from the technical implementation, the user base, or the protocol.

Data formats will be an increasing cause of lock-in. The use of digital rights management mechanisms for protecting data formats may offer improved security for the user. However, given the Digital Millennium Copyright Act, creating interoperable software with a format protected by encryption is a felony. Therefore, *selection of open formats is critical*.

Open code provides the greatest flexibility and prevents lock-in. Open systems, when available, prevent forced upgrades, prevent loss of control over data, and enhances long term strategic flexibility. If open code is absolutely not possible, open standards should be used to increase trust in the system. Secret specifications as security measures expose the system to break once, run anywhere attacks.

A critical part of any strategy addresses *initial roll-out and the diffusion of upgrades*. The ability to change or grow in an organic or by degrees will complement any exit strategy. The ability to have a limited roll-out distinguishes the failed X.509 and successful Pretty Good Privacy methods for key distribution. Both systems use public key encryption to provide confidential email. X.509 requires a centralized directory for all users who wish to communicate. PGP allows users to assert their association wit their own keys, and then that assertion can be validated by others. PGP was designed to enable organic patterns of diffusion while X.509 required simultaneous adoption.

An understanding of both an initial roll-out strategy and the issue of upgrades creates the ability to *plan on post-production changes. Pilots and gradual phase-in is risk-averse and allows for institutional learning.*

Risks can be evaluated against an ideal or a historic baseline. The historic baseline can be misleading as with the PEBES case mentioned above. In order to evaluate risks in digital government the Standards for Security in Federal Information Processing Standard (FIPS), which were released in draft form by NIST in May 2003, provide a baseline for evaluating risks in security. The FIPS proposes three questions that must be answered:

What security controls are necessary?

Are the security controls properly implemented?

What is the desired level of assurance that the implementation is functioning as designed?

Within this framework NIST recommends particular attention on confidentiality, availability, and integrity. However, as noted in the false

positive and false negative recommendation above there can be a conflict between assuring access to the authorized (availability) and preventing access by the unauthorized (confidentiality).

One way of avoiding unanticipated risks is to *actively pursue early engagement with as many stakeholders as possible*. In digital government the privacy community will be one of those stakeholders.

In addition to understanding risks, *digital government has a unique burden requiring it to communicate risks to citizens*. Of course, the focus should not be entirely on the risks.

Communicate the value of the service. Do not hesitate to advocate a service that has been examined and found to be likely to serve the community. Digital government can both provide services, and provide information about the availability and value of services. Risk and benefit communication should be one element of the necessary trust-building in digital government. Digital government websites should build trust through communicating the value added to citizens in every digital interaction. Illustrating the value of the organization is critical in building trust. Agencies as well as companies need to seek citizen trust. Trust can be built by advertising services, reminding users of the value of the services, and then building more services on the basis of trust.

Digital government brings the potential to transform the citizen/government relationship. Dramatic and effective transformations require citizens to opt-in, involve themselves in the new process, and leverage new ways of interacting. Citizen opt-in requires evolving technology, consistent privacy, and continuous information flow on the investments and services provided by government - on and off line.

Balancing Competing Interests

Building an identity system is full of trade-offs, some obvious, others hidden, and some misleading. Understanding competing forces brings about better design for several reasons. First, it helps force realistic assumptions into a field that has vague definitions and has been touted as a solution to a wide variety of problems. Knowing what goals may be mutually exclusive will help prevent unrealistic or contradictory objectives in the initial stage of development. Equally important, many aspects of an identity system have their champions, whether its privacy, national security or efficient bureaucracy. Seeing how these forces line up against each other theoretically can lead to a more complete list of parties involved in the planning process, and help make sure that no one aspect is poorly defended in the design process. Few of these trade-offs are absolute, of course. A third way is often available, but that frequently involves considerable effort and expense to handle matters manually and personally.

It must be noted, however, that the commonly perceived tradeoff of security versus privacy sets up a false dichotomy. We have seen numerous systems where a lack of one leads to a decline in the other: poorly protected personal information undermines access and authentication mechanisms for the entire system. Reliance on personal information for security systems may even decrease their original efficacy. Recent critics of the CAPPS II airline screening system, for example, have pointed out that using deterministic profiles based on personal information will allow malicious individuals to ascertain whether they are under suspicion. Personal information *can* provide security if one knows exactly what to look for, but this is often much clearer in retrospect. Ensuring that the necessary information is available requires massive data availability, and the ability to identify critical data. Of course, massive data compilations may also be useful to attackers who, by definition, have prior knowledge of information that would be useful during any assault. Thus large scale data repositories providing wide dissemination for law enforcement may create harms which outweigh the perceived benefits of these approaches.

Security vs. Flexibility

A central tenet in the principles of data protection is that a database containing mistakes should be correctable. The misrepresentation of an individual in a database, could cause a person to suffer harm from that misinformation, but also can even alter that individual's identity by reifying an incorrect identifier or attribute.

On the other hand, identifiers should be impossible to change without proper authorization. No one should be able to alter part of another person's identity without the express or state-derived permission to do so.

In the simple case where only some particular attribute is mistaken, it is merely a matter of process of authenticating the individual to correct it. The problem arises when major identifiers are incorrect. The easier it is to use secondary identifiers to correct mistaken identifiers, the more vulnerable a system is to malicious actors attempting to create false identifiers. In a highly responsive system, malicious actors could use publicly available sources, social engineering, or rely on personal connections to obtain identifying information about an individual, and then subvert the identity system. On the other hand, a security system that prevents some one with only a subset of identifiers to prove their identity would make it incredibly frustrating to set things right; one can imagine a Kafka-esque bureaucratic nightmare ensuing. This problem is further compounded by a feedback effect between security and target attractiveness. The better the security of a system, the more trust that system generates. A system that is highly trusted becomes a very valuable target for fraud. One direction to ameliorate the difficult trade off between flexibility and security is to employ a range of techniques for system security accompanied

by strict punishments for fraudulent users, rather than a "once in-always safe" strategy.

Security vs. Usability

Closely related to the idea of system flexibility is the idea of usability. As many IT professionals can attest, users are interested in using the system for their specific goal. Complex security procedures are often seen as barriers. This problem is compounded by security policies that are difficult and built for the machine rather than the user. For example requiring people to use multiple, random, and frequently changing passwords is a requirement that is in opposition to natural human practice. Individuals are much more interested in using the system to meet their needs than managing security. As a result, many seek out shortcuts to save time and hassle, or to make things easier. Parents give children their credit cards; people share PINs casually, based on (sometimes ill-founded) trust.

Anyone locked out of their own system from loss of a password knows that security can be an obstacle to the authorized user as well as the attacker.

Microsoft's Passport system capitalized on the popular desire for ease of use, offering individuals a single place for personal information, protected by a single password. A central point can create at least one secure juncture, but it also creates a single point of failure and, as above, a very attractive target for attack.

Security does not have to be an obstacle to usability. Rather, a combination of user education and proper security design is needed. System users must be made aware of the consequences of poor security behavior. Incentives are required for user behavior and for system design, and should align with the overall system purpose. From the systems design perspective, human-computer interaction scholars are beginning to address security issues, and guidelines are emerging for making security accessible.

The conflicts between security and usability often result from the nature of computer security as being an option, or add-on, rather than having security as a core design principle.

Accountability vs. Privacy

Any decision to log transactions and keep histories must be made in the larger framework of risk analysis, based on the specification of the problem the identity system is to solve. Often logging is used as a the ubiquitous cure for all unspecified system ills.

Transaction histories can be misused. Anonymous information systems do not keep personally-identifiable logs precisely because of the risk of misuse. In addition to the direct privacy threat of detailed compilation of information, there is the potential for abuse if security or policy protections are undermined. A shift in administrative polices, a failure to maintain updated software, or an abuse of subpoena-like powers could result in large amounts of data being exposed. Both risk management and organizational management indicate that information protection policies should be embedded in code whenever possible, rather than relying on administrative consistency. An example if this is automatic log dumping after a fixed period of time

Accountability versus privacy is often a false trade-off. The existence of records does not ensure that individuals are accountable. Indeed, access to logging records may create a hazard in and of itself.

A related problem lies in user review of data. Users should have the right to review their data, say standard privacy tenets, to ascertain that it is correct. But if it easier for an individual to access their data, it is easier for others to access it as well. This relates back to problems of availability, usability and security.

Bootstrapping

A critical issue is the problem of deploying an identity management system on top of an existing, flawed system. A reliable system must correct for all the errors of the previous system. Of course, the expenses and difficulties of correcting the previous system are among the reasons to alter the nature of the conception of identity for a conception of proof of specific attributes. Consider the difficulty of proving Jane Doe is indeed Jane Doe, and that this Jane Doe has been the carpenter paying Social Security taxes. Alternatively, considering proving that Social Security taxes have been paid by some entity now claiming benefits. The first requires bringing together a number of discrete documents to prove identity. The second could be proven by documents generated by the employer, the Social Security Administration, and the employee that prove a unique relationship between taxes paid and benefits claimed. The process of "bootstrapping," which derives from the mental image of pulling oneself up by one's bootstraps, presents a tricky dilemma: how do you get a secure, trusted ID in the first place in a world with fraudulent identifiers already existing? When are choices necessary between ease of implementation and the reliability of the system to be implemented?

Research Agenda

The workshop participants covered a wide range of expertise and professional interests, and the resulting discussion of open questions and issue to be addressed is similarly expansive. This section divides important research questions about identity systems into seven problem areas that must be addressed in any discussion of identity. Each problem area is then examined in light of six broadly defined disciplines. In each problem and discipline, critical questions are identified for the design of any identity system, and research questions are proposed that will bring a better understanding of what successful identity management must entail. We mention specific research projects and general areas of contention in a topical framework, and then mention several specific pilot programs and case studies.

Laying out an Agenda

Like many aspects of digital government and other instances of applying technology to policy, common themes run through different problems, and what might address one concern could exacerbate another. On the other hand, a topic such as identity is so broad that some division is necessary to gain a better understanding of the overall shape of the research agenda, and then to isolate what might be considered a research priority. We believe the latter argument is stronger in this case, and present the seven problem areas described below to serve as an organizational rather than methodological guide. While all problems should be considered concurrently, the ordering follows the logical flow of the development process, from concept to implementation.

Information architecture and management strategy. The overall shape of a system is an important point to isolate from lower level questions. The "back end" must be thought through, as must an understanding of how different entities, public and private, will interact with the system. What identifiers will be used, where will they be stored and how will information flow throughout the system? How can secure identifiers be created from the range of failed or fragile systems now in places?

Privacy and personal information protection. Privacy concerns were a primary theme throughout the workshop. Apart from its priority as an important and endangered social value, control over personal information is necessary for a good identification system.

Governmental policies. Assuming an ID system is at least tied to government programs, the federal administration will play a strong roll in dictating how a program will and will not be used. Inter- and intra-agency

policy will need to be defined in addition to regulation of commercial actors and citizens.

Accountability inside and outside the system. Abuse and fraud prevention is necessary to make sure that the problems the identity was designed to solve are not duplicated and new ones to not arise. Among other things, this means having the system capacity for due responsibility.

Metrics for design and evaluation. In order to design a successful system that can be judged to be an improvement, measures must be developed to evaluate what success would look like. Identity deals with risk, and proper risk analysis requires metrics to coordinate management.

Implementation of the infrastructure. Any system will be used by individuals, and those individuals must interface with the system with a minimum of difficulty and a maximum of efficacy, equity and comfort. The user end of a system, however, is often one of the largest security liabilities. Again diffusion and initialization are issues.

Roll-out and enrollment phase. Systems do not magically spring into implementation, and converting from an old system to a new system always has kinks. Identifying obstacles ahead of time helps smooth the transition process. All previous steps in the planning process require consideration of roll-out. A perfect system is useless without diffusion, as shown by the example of perfect X.509 and pretty good PGP.

Each problem area has many independent and related research topics. This report subdivides these topics into six academic disciplines. As above, an inter-disciplinary world advises caution, and many of these issues can fall between disciplines, or across multiple disciplines. This is noted where possible. Each discipline brings unique qualitative and quantitative tools to the study of identity, and this report seeks to highlight the value of each discipline in addressing identity issues. Qualitative research focuses on design principles; quantitative topics test implementations and prescribe standards.

Computer Science and Engineering: Hardware, biometric tools, cryptography and the technical side of human-computer are placed interaction are core issues in the development of an identity system. Of those, the cryptographic and hardware research is well supported, but there is little analysis of the policy implications of different technical choices. Qualitative research focuses on design principles; quantitative topics test implementations and prescribe standards.

Management Information Science. MIS focuses on information systems, and how they are shaped. Building on computer science, it focuses more on the structure of a system, and the impact of that structure. Research areas focus on both the design aspect and evaluation of performance.

Organizational Science. People can behave predictably in structured environments, and optimal identity systems require understanding these structures and behaviors. Public management scholarship has much theory on organizational dynamics and how technology can best be implemented inside an organization.

Economics. Identity affects how resources are distributed and is, in turn, affected by these resources and their distribution. The field comprises economic models, game theory and business issues on the qualitative side. There are many economic tools to evaluate the quantitative impact of policies on a large scale and on individual decision making.

Social Sciences. Understanding how society behaves is critical to properly evaluating policy. A critical topic trust, which spans traditional disciplines such as philosophy, psychology, sociology and political science. Qualitatively, the social sciences offer the ability to experiment with computer-human trust interaction. Quantitatively, we can better comprehend how the population might respond, or how segments are likely to use technologies and systems they are exposed to.

Law. Significant changes to current authentication practices will implicate current legislation on individual rights, administrative responsibilities and organizational liability burdens.

A Topography of Research Interests

The following more detailed explanation discusses the key research questions that were raised over workshop discussion. That some fields have more research questions than others does not make them less valid, or even less critical to understanding the problem. The number research projects or their important to understanding identity is, of course, fairly subjective. There are many research questions in management information science and organization science, and a clustering of economics, social science and law in certain problem areas.

Information architecture and management strategy

An identification system is defined by the shape of information flows inside the system. The user interfaces enable or preclude certain capabilities, but the power of a system relies in who gets what information, when and how. A thorough understanding of information architecture will rely on information systems design tools and knowledge of organizations, but still will draw on the technical and legal to determine what is feasible.

Computer Science and Engineering: Understanding the structure and design of an identity system requires understanding the technical options and networking context. What are the technical and computer security

limitations on how users can interact with the system? A system with authentication hardware that needs a security guard to watch it will be very different from a system that only uses a smart card reader that can be attached to a user's computer. Similarly, encryption systems that rely on public/private key pairs need a user-based feature for handling the private keys. How the security flaws in smart cards and other data devices is resolved will affect the how where identifying information flows from.

Management Information Studies: What are the management requirements in having data centrally located, spread across multiple databases or resident on user-controlled devices? How much information should be required to authenticate an attribute, and can this process be done without establishing the identity of the entity? How can data be transmitted and backed up without increasing disclosure risks? Government agencies and private databases are all in very different formats; how can they be harmonized and made interoperable to increase efficiency and take full advantage of the system's connectivity? Who will have access to information in databases, and how can a data subject access her information without fear exposing that information to others? The scale of any system will be very large, so models should be empirically tested to highlight stress points in large-scale implementations.

Organizational Science: The question for the administrators of this system is who gets what, and what will they do with it? The IRS needs different information than the Social Security Administration. The Department of Homeland Security needs a larger range of information, but also requires greater civil protections. How will information be segregated, and when it needs to flow between agencies, how can the process be streamlined on the administrative side? Personal information needs to flow not only up to national level databases, but down to state and local authorities, and possibly out to the individuals themselves. How can this process be made secure and efficient? Will private businesses be allowed access to this information? Empirically, scholars can examine current information flows through bureaucracies to model improvements under better identity management capabilities.

Economics: What are the current market demands for better identity? Where are the loss leaders in fraud, and how much more can they absorb? This information can help predict the key private sector players in building identity, and thus more about the eventual structure.

Social Sciences: More work is needed in how and why individuals interact with information infrastructures, apart from obvious need. Why do people trust certain institutions, and not others? How much of this relationship is shaped by reputation, by information or by interface? Would a better infrastructure, in fact, increase trust in a system, and over what kind of timeframe?

Law: Legal measures may be required to ensure seamless integration of new identity systems into everyday life. Apart from shifting liability, what other legal tools exist to ensure that the capabilities of good identification management are used wherever beneficial?

Privacy and Personal Information

Through the workshop a consensus emerged that data protection provided a fruitful framework for evaluating the impact of potential innovations in identity system. The framework of notice, consent and justification provides a comprehensive and consistent set mechanism for evaluation of alternatives. Privacy is an essential value, yet the continuing discourse about the underlying meaning of privacy in a democracy creates opportunities for misunderstanding when trying to communicate risks, and benefits. Data protection cannot entirely encompass the subtleties of privacy in public discourse, but it provides a powerful tool for analysis and design.

Computer Science and Engineering: Trust enhancing technologies including reputation, cryptographic and authenticating technologies. Privacy-enhancing technologies have experienced growth yet remain an active area of research. End to end encryption is not always used where applicable, and it is not always easy to use. Should these systems be solved at the applications level, the issue of human computer interaction remains. Personal data does not need to be a part of many transactions, although the shift away from cash in electronic transactions leaves more details from any interaction. While electronic cash schemes never became as ubiquitous as other methods of payment, encryption can be used to provide the requisite amount of information and trust for a specific transaction (e.g. a campsite reservation) without using personal identifiers. How can such systems be made more uniform and applicable?

Management Information Science: Temporal and location-specific information is highly sensitive from a privacy standpoint, enabling invasive behavior. Tied to identities, temporal and location-specific information allow detailed surveillance. Even when distinct from a defined identity, pattern matching and data correlating can yield far more about individual behavior than the data target may suspect. How can this information be protected against unnecessary collation? What unique technical or systemic protections would preclude the need for having to rely on policy to prevent that kind of monitoring?

Organization Science: Theoretical research should be conducted on the implications of the shift from paper to digital administration, especially when looking at privacy. What information is out there, and what is important for administration? What information is of particular use for fraud detection, or for those attempting fraud? If individuals have a right to inspect their own data,

how can we guarantee that right and still protect unauthorized access? With digital information, the path the data takes is only as secure as its least trustworthy link. Organizational scholars need to study the patterns of access to pinpoint potential trouble points. Wherever possible, individuals should have accountability mechanisms. Is privacy feasible with the interoperability organizations will demand?

Economics: Only recently have scholars begun to examine the economics of privacy. In multiple-identifier system, the strongest identifier tends to be overloaded in a tragedy of the commons scenario: a frivolous actor will gravitate toward using a simple but strong identity mechanism. The more this identifier it is used and diluted with poor management, the greater the chance of corruption. Researchers should investigate whether there any way to break this direction?

Social sciences: Research on the social implications of new services can help both with design and diffusion. If trust in a system can be conceived as a willingness to expose personal information, then there exist a range of potential real world experiments to determine how humans trust in computer-mediated interactions. To measure the value of personal information to other actors fighting fraud or defending national security, empirical work can be done to show how useful information is for any given data analysis system. Finally, access to important features of the electronic infrastructure is not ubiquitous: over 5% of the population does not have home telephone service, and that figure can be ten times higher in certain populations. (NTIA 1999) Even if these populations could be integrated into a digital identity system, would have the same privacy protections of review, opt-in and opt-out as others with more technical access?

Law: What controls exist over the private sector use of government-collected information? What assumptions about identity and anonymity are embedded in the law? What assumptions does the law make about "publicly available" information that supposedly anyone should be able to obtain? How much control should an individual have over information that refers back to her? This question abuts the quagmire of intellectual property, since the database owner may claim protection of the database, admittedly assembled at his expense. The public sector, for its part, relies increasingly on private sector information, for fraud protection and terrorism, for example. (Mack et al, 2002)

Governmental Policies

Since an identity system needs to have some coordination, standardization and centrality to be effective, and no other institution has the institutional trust (to be distinguished from personal trust) or liability protection

of the federal government, it is likely that the federal government will play a strong role in the administration of any changes in identity management. As one of the largest handlers of identifying information, many changes will also reflect its own role.

Management Information Systems: What improvements in government will a better identification system enable? How will better record systems and more efficient and reliable identifiers change government? The large scale of a government-wide system must also be considered. Many security flaws are greatly exacerbated by their ubiquity.

Organizational Science: Any major changes in identity systems will have to be made over a period of time that would likely include shifting government agendas, priorities and philosophies. Even if the basic implementation were to remain the same, "scope creep" might set in. Agencies and actors are likely to seek to do more with identifiers than was originally planned, to the possible detriment of privacy or system efficacy. How do human-scale political institutions plan to be adaptable but prevent corruption of the original function? Shifts in technology further confound this problem by altering costs and feasibilities. As Professor Jane Fountain has noted in her work, The Virtual State, Cooperation between agencies has improved recently, but systems are still stove-piped, and it is difficult to secure funding and attention for projects in shared jurisdictions, making harmonization less simple. One question, then, is whether or not a new government agency would be required to manage the identity system. A universal identifier system might, whereas a revised multiple identifier system may be claimed or fall under the purview of an existing bureaucracy. As interoperability increases, control over data may decrease, leading to deteriorated protection of personal information.

Social Sciences: The government has a very different role from private sector actors in terms of its audience. The state cannot always rely on a simple cost-benefit decision process, since certain services and rights must be protected, despite inefficient circumstances. The digital divide deserves special attention in the realm of e-government and identity management. How will less feedback ability and service shape the ability of less privileged groups to successfully represent themselves in a high tech identity system? If such groups do not have adequate access to computer technologies or do not understand the system, do they literally lose their identity? If those with access to technology push for improved online services, will what is left over be of decidedly poorer quality? How can the impact of inequality be minimized in implementing a new identity system?

Law: The government must set policies that govern distinctions between the state and private businesses. Traditionally, there have been strong political and legal protections against undue regulation of businesses. Keeping

the public and private sphere under the same identity system but under two very different legal regulation models will not be easy.

Accountability Inside and Outside

Oversight is critical to ensure proper system management of any large-scale centralized system,. Citizens have a right to know what personal information is kept in government databases, and many people have advocated that this right be extended to private databases as well. Citizen requests for transparency, a concept that is highly valued in a democracy, should be balanced with the state's ability to perform necessary security and enforcement functions, which also involves information monitoring and oversight.

Computer Science and Engineering: The issues with record keeping discussed above are compounded by the fact that accountability means tracking data usage. Thus, records must contain relevant meta-data that allows various actors to see to varying degrees who had access to information. Managing large quantities of metadata have been a topic in areas such as semantic systems, although further exploration in terms of access to the metadata itself is needed.

Management Information Science: Accountability in an information system is an information management problem. Transfers across administrative domains need to be authenticated ahead of time, and tracked post-facto to provide records. Audit trails can be used to protect personal information by revealing inappropriate use, but are themselves a form of personal information. How can accountability information be protected from misuse? Another key aspect of information management in identity systems is "exception handling." What happens when an identity has been compromised? No system is perfect, and errors are bound to happen; how can reliable identifiers be revoked? PKI research considers this an open question, and solutions should be explored that reflect administrative realities.

New records need to replace old records without making life too difficult for the individuals involved. Sometimes, the government may seek to change records and *not* leave an auditable trail, with the most obvious example being the witness protection program. An identity system should not expose to a casual user that an intentional shift has been made if it is properly authorized, but certain administrators may need to know. Creating the capability for authorized record-free changes opens another hole for fraud, and systematic protections are needed above the human-level safeguards. Automatic algorithms for predicting and detecting fraud will be necessary to keep system confidence high without a massive hands-on labor force. The current system is purported to have up to 10 million duplicate Social Security 004 Numbers. (SSA 1997) To

what extent does the idea of a unique, consistent <name SSN> pair fall short in current government databases, and what are the origins of these errors?

Organizational Science: Accountability is a key value in responsible e-government, and a fair amount of literature has been devoted to this topic as an essential feature of good governance. Should organizations strive to use a "need to know" approach personal data? What permissions structures and principles will prevent misuse but maximize organizational efficiency? Existing inter-agency information policies can be studied to make predictions about future data sharing security. When an error is made in identification, public and private bureaucracies must be equipped to handle it. False identification leading to fraud or crime should leave a trail and allow for follow-up investigations to determine its cause, and hopefully prevent future errors. False negative errors, where an individual incorrectly triggers suspicion, will be more likely given statistical rates of crime. In these cases, inconveniences and rights violations should be minimized, and data records should be structured to help an individual bolster his claimed identity.

Economics: Game theory offers important tools for studying systems vulnerability. It predicts that the more a system is used and trusted, the greater the incentive will be to subvert this system. Attacks will go up, so overall security may decline. Models should be developed that can mitigate this apparent paradox, and economic studies can guide thresholds of probabilities, expenses and rewards for system subversion.

Social Sciences: Accountability is about trust. All the checks and balances will not increase system functionality if users are not aware of them, or do not have faith that they will work. A larger social science issue tied to this is trust in the government. Can the government prove its trustworthiness? How do you make people trust an institution? Several participants in the public sector noted that few Americans see their government as a provider of good things, and that a positive relationship with the state would be possible if the government could effectively communicate its uses to citizens. How does government public relations relate to democracy? Is there any way to separate government service from domestic politics? Other mechanisms to increase trust in government should be explored.

Law: In order for accountability mechanisms to work, penalties have to be threatened and enforced. Inside the federal government, misuse can be punished easily enough. How does the state go about pursuing private parties who have misused others' personal information, or flaunted information rules? Not all misuse is malicious, of course. If harms result due to mistakes, should there be negligence penalties, especially if the harms are difficult to concretely measure? Moreover, individuals can be irresponsible with their own data. Will businesses still be able to trade in personal data by offering discounts? Limitations could be placed on that which could disrupt the system and allow

individuals to contractually give up their privacy in other arenas. Finally, if an individual accidentally compromises the security of the system, should the legal system pursue this? It may be a wise design choice to minimize the number of ways an average user can inadvertently cause any harms to avoid making this a legal issue.

Metrics for Design and Evaluation

In order to design a successful system that can be judged to be an improvement, measures must be developed to evaluate what success would look like. Identity deals with risk, and proper risk analysis requires metrics to coordinate management.

Computer Science and Engineering: Risk management depends on accurate understandings of the computer security issues. The most secure systems are limited in function, yet the most accessible systems are flexible and adaptable. As a balance is sought, engineers will have to come up with reliable metrics to compare multiple systems, and a wide range of standards to cover the many contexts of identity management systems. Accurate estimates of costs to perform cost-benefit analyses demand complete projections of technical requirements and the expenses involved. Biometrics are carefully tested individually, but testing multiple integrated systems, especially with respect to accessibility, fairness and false positive rates is critical for identity applications.

Management Information Systems: A model of identity management depends on accurate measurements of how much information is out there. MIS can offer the tools to determine how many different data sources and nexuses are in government and private sector systems. Where are the bottlenecks preventing information flow through systems when free flow is desirable? Where should gates be placed when such open data movements are less appealing?

Organizational Science: The discussion of whether a system will do more harm than good will benefit from novel advances in understanding how strong identity management will improve institutions. Scholars of organizations and bureaucracies can envision how their objects of study would change given certain levels of certainty in identity to help identify the winners and losers under new systems, as well as suggest means of quantifying those changes.

Economics: One of the more common metric systems for policy analysis compares the harms and expenses of a system with the good created. However, cost-benefit analysis is very difficult when terms are hard to define, and even harder to quantify, particularly when contested social values are at stake. Social scientists frequently observe that what is defined as a "cost" or a "benefit" can involve a deeply contested debate regarding social values. Many of the costs are not directly financially tangible, including the value of privacy

and data integrity. Care must be taken to properly integrate these ideas into any analysis, and strategies must be developed to not over-emphasize that which is easily quantifiable. Understanding the cost of system failure can help design a system that can minimize the damage caused by a compromising instance. Finally, we are just starting to understand the economics of privacy (see, e.g. Acquisti 2002), but far more work needs to be done in understanding it empirically. Having economic impact data will make it far easier to sell the importance of personal data protection to skeptical policy-makers.

Social Sciences: Much of e-government rhetoric is based on improved governance. How do we actually know when we have better governance and a healthier democracy? What are the appropriate indices of measurement?

Law: A new system will alter the legal landscape and scholars must identify ahead of time where new legal burdens may lie, and who would bear responsibility for liability. In addition to comparing different systems, this would help predict how various actors would respond to the new system. It would also help discern a clearer picture of what reactions would result from a given identity management scheme, allowing a better understanding of costs and benefits.

Implementation of the Infrastructure

The physical implementation will reflect the information architecture, but will rely heavily on technical expertise to ensure that the system functions as intended. Understanding the bounds of what is possible and what will never work is as important as maximizing the efficacy of a given device. Similarly, thinking about issues of an identity system in interactions with extant institutions can provide valuable insights to take back to the original system design.

Computer Science and Engineering: Among the technologies discussed at the workshop, cryptography and biometrics were mentioned repeatedly, yet neither is ready for wide-scale deployment with any reliable standards. The mathematics of cryptography have thus far outpaced the technical ability to attack them cryptographically, but it remains very easy to circumvent code-based protections by attacking weaker links, using social engineering. Public key systems have been understood for 25 years, yet they play relatively minor roles and are complex enough that they must be completely automated for the average user. One study also highlights the specific difficulty of using public key infrastructures for situations that are very wide in scope. (Kent and Millet 2003) Biometric error rates remain far too high for society-wide implementation: a 0.1% error rate would result in 350,000 people misidentified out of the national population. Recent attacks using

relatively unsophisticated methods have thwarted previously promising fingerprint and iris-scan technology. (Matsumoto et al 2002; Thalheim et al 2003) To prevent a single unexpected weakness from destroying all biometric validity, researchers must look into better coordinating multiple systems together. While completely tamper-proof hardware is essentially impossible, the trustworthiness and reliability of PCs, smart cards and other tools must be improved if hardware-based authentication is ever to enter the home.

Management Information Science: Legal persons (corporations) may wish to participate in the identity system, as the ability to authenticate would be a boon to corporations dealing with the government. A system should be able to handle legal persons, and make distinctions between humans and corporations in their interactions and permissions.

Organizational science: All systems will make mistakes, and thus secondary and tertiary identification mechanisms must be tested. A malicious actor whose actions are prevented from the primary identity management system will of course try to subvert error handling systems as well. Back up schemes must thus not be a weaker system of defense, but rather extend the number of ways an individual can reliable prove things about their identity. Thus, a mistakenly-identified party should be able to prove themselves with only marginal difficulty, while miscreants are denied an easy back door. If multiple biometrics are to be used, they must be selected with care, and tested as a combination.

Economics: Much of the cost of a system such as those discussed in the workshop is in the deployment phase. What are ways to minimize these costs? Outside of those benefits directly related to improved identity management, are there any tangible benefits of a new infrastructure? Computer simulation might be an appropriate mechanism to search for emergent effects.

Social Sciences: The social sciences need to inform views of how identity management plays out in society. Identity is a cultural idea, and the identification process is burdened with cultural values in a pluralist society. A biometric could raise issues about touch and hygiene. Mandatory photo IDs may raise modesty concerns, particularly in ethnic populations which eschew such depictions. Proposals for a universal identification number have drawn flack from powerful religious groups (Moore 1997). Is it possible to have a biometric that everyone can use, and if it were, would the very idea trigger too many social associations with criminals or oppressive regimes of the past? Electronic systems raise the problem of accessibility not only in terms of the price of the technology, but physical access. Disabilities may prevent some from using biometric devices.

Biometric systems must have an initial set of data from which to develop the ability to distinguish between individuals. The necessary

initialization and configuration of a biometric system is called "training". When a biometric system is to be used, it is important that the training population is similar to the population that will be authenticated using the system. Diverse populations will therefore require diverse training sets. Issues of diversity in authenticating systems include accents, for voice recognition systems, or saturation level for face recognition. Ensuring usability for a diverse population is a non-trivial issue for all authentication systems.

How can we build an infrastructure that's open to everyone? Even if certain goals were not met by an identity management system—lacking, for instance, the capability to effectively identify terrorists—who would continue to benefit? In this case, law enforcement might still have an easier job, and information infrastructure companies would still be paid. Thus, they might both prefer an imperfect system to no system at all. Being aware of these interest group pressures will help understand the political economy of implementation.

Law: As technology developed, more decisions are made online by intelligent agents. If agents can authenticate themselves online, new laws will have to be developed to handle responsibility for online transactions.

Roll-out and Enrollment

It is difficult to research the roll out of a system that has not yet been conceptualized, let alone designed. Nonetheless, there are critical issues that must be addressed before many systems can be widely implemented, and much of that research can be done now. The case studies and pilots mentioned below will also add much depth to this problem area.

Computer Science and Engineering: Enrollment into a biometric system-getting users' data into the system for the first time-is one of the stickiest parts of managing a biometric system. Human-computer interaction needs to inform the design of these machines. Studies have shown that even something as simple as voting with a digital interface can cause difficulties in large segments of the population, so such systems must be designed with care if they are to be automated.(Mercuri 2002) Early technologies must also be phased in with upgrades in mind, as new generations of technology will likely be available by the time implementation is complete.

Management and Information Science: As mentioned above, the information system must be designed with the capacity for field-updating. Distributing unique identifiers to a large population is a non-trivial problem, given estimates of up to 10 million duplicate social security numbers. How can such a distribution happen in a decentralized fashion that allows thousands of administrative centers to operate simultaneously?

Organizational Science: Perhaps one of the hardest problems in deploying a trusted identification system is determining on what to base trust in an individual's identity. So-called "breeder documents" such as birth certificates, passports and social security cards may already be based on false identification information. If they are incorporated into the new system, the integrity of that system is undermined. Figuring out how to "bootstrap" an identity management system should be a key research priority.

Economics: The costs of a national identification system will be enormous, and the rollout phase will make up much of that expense. How much will getting 290 million Americans enrolled into the system cost? Beyond enrollment, the transition costs to a new system will be substantial. Of note are the expenses of a partly-deployed system during the middle of the implementation process. Many institutions will have to simultaneously employ multiple systems. While federal agencies can manage this with incentives and regulation, private firms will have to contend with the early adopter problem. How will the traditional S-curve shape costs and business reactions?

Social Sciences: Any predictions or prescriptions about deployment will be valuable. How will the population and various social actors react to a new identification system? Should the government or private actors employ public relations or propaganda tactics to encourage support for a new system, or to guide its acceptance in society? How should the government deal with a vocal minority that might oppose a shift in the status quo?

Law: Some people may decline or refuse to use a new system. Can private institutions deny dissenting individuals service, much as air passengers who refuse to show ID are denied air travel? The government is faced with a problem when not dealing with services offered to citizens. If it just seeks more control over citizens for security or management purposes, will it be forced to implement mandatory carry rules, or are there other legal solutions?

Pilot Programs

One of the hardest aspects of studying large programs and systems is visualizing multiple components coming together and conceiving of how complex subsystems interact. Implementation on some level is necessary to discover emergent behavior. Simulations can yield useful results, but seldom can they reveal unexpected weaknesses; a pilot program can expose real-world considerations that may have been overlooked in planning. They can be used to prove specific aspects of the system, such as fraud prevention or user interface success, without the expense of a full range system. Finally, pilots serve as proof-of-concept vehicles to advance the validity of a given solution.

Given the wide range of applications of identity, there is no shortage of potential pilot ideas: they must be selected to maximize what can be

learned. Any pilot must examine rates of fraud, misuse, user satisfaction and back end costs. Persons directing pilots should keep careful and protected records of usage patterns, both of individuals and resources. Perhaps most importantly, pilots should be scalable. Any identity system is going to have to expand and adapt, so a pilot program should start small and record its success in growing and upgrading while maintaining original functionality. For example, programs with strategies and architectures that cannot develop and grow larger without reissuing identifiers should probably be revised before seeking a more public role.

Some pilot program ideas put forth by participants:

University library system: A library offers the right combination of various amounts of personal information, regular authentication to access privileges, and noticeable harms if fraud gets too bad. Library pilots were first proposed as a proof-of-concept for anonymous ID systems, as many participants were skeptical that it could work, or saw it as too complex to be understood. Implementation of an identity system without personal information that involved regular student interaction with the system would allow proponents to address complexity issues, show that personal information is not necessary to protect access permissions and iron out bugs in the program. Alternatively, such a program could be used to test a more conventional identifier-based system and develop an information architecture that incorporated principles of personal information protection and bureaucratic efficiency.

Campsite reservations: To secure a scarce resource remotely, personally identifying information is not necessary. All that needs to be shown is that there is a resource (a campsite) that exists, that some one has reserved it, and that the individual present is that some one. In the simplest case, at the time of reservation, the party making the reservation could be given a secret token, which she could present at the campsite to secure it. More complex scenarios involving anonymous deposit payments or cryptography can grow from there. It should be remembered that pilots should not attempt to have more functionality than their application recommends, since that may reduce the quality of final analysis.

Anonymous parking ticket payment: A more complex pilot might involve law enforcement directly. There is no need for a parking ticket to initially be linked to an individual. License number does not have to be correlated to an individual until the account is clearly delinquent. A ticket paid promptly should have all records destroyed, and anonymous payment mechanisms should be in place. Repeat offenders who do not pay can be identified by car, rather than driver, and only in the most egregious of cases should an actual identity be established. This pilot can, among other things, make the case that personal information is not always critical to law enforcement.

Hospital Records. Personal verification and auditing are an important part of any system that tries to protect personal information. A hospital or medical system can, in addition to collecting and protecting information as deemed fit by law and medical practice, use proper data protection practices. Patients should see their own records, which include a list of everyone who has looked at it. Requests to see it should include justification, also kept in the access log. Streamlining such a system will be valuable.

Case Studies

Often a pilot is not necessary to learn a specific lesson if an idea has already been implemented. In that instance, a case study is the perfect method to glean what lessons can be learned from others design decisions and implementation choices. The case study method is well-developed, and there are already numerous examples of modern identity management systems that have been attempted in recent years.

Many other nations have tried something like a national ID card, with a wide range of purposes, structures and success rates. Some of these are well-integrated into data systems, others are not. Are fraud rates lower with a national ID system? Can security measures be compared? How universal is the use of these cards inside the country, and for what range of purposes? The reaction and support of citizens for such a system should be gauged. Of particular note should be the story of implementation, and how personal information is handled by the system.

The Department of Defense Common Access card is a smart card technology using a public key infrastructure developed to serve as an identity management system for all DOD employees and contractors. Including reserve military personal, this system covers approximately 4 million people. This is a large enough scope that many of the size-related issued can be observed in the planning and deployment process. A complete study should note differences between the Defense environment and larger society, as well as the tools to evaluate reported metrics of security.

An identity management system will be a large-scale essential infrastructure that will need occasional technical upgrades. In order to learn more about updating a physical and information-driven infrastructure, similar large-scale operations that depend on inter-connectivity should be studied. The telecommunication industry's may not be a fair analog to an identity system, since the common carrier principle means that the edges of the network will behave the same while the center is being upgraded. The US Postal Service might serve as a better subject as a centrally-run infrastructure. How is functionality maintained with partially upgraded systems? How are employees

and clients equipped to deal with change, and what are strategies to shorten the transition period in a complex bureaucracy?

Errors exist in current information systems. Social security numbers are duplicated, or paired with the wrong names. Addresses are not up to date, or incorrectly entered. Legal status, family status, income, all have error rates. To understand how error can enter a system, a typology of error sources and rates needs to be compiled, and why not use existing data? Of particular interest will be the extent to which data is inconsistent across databases, and how it can be easily and automatically verified. Sampling techniques can be pioneered for eventual measurement of new programs.

Method

The method used in the Digital Government Civic Scenario Workshop was the civic scenario process. The scenario method was developed in business schools to address severe problems where participants disputed both the order of magnitude of the critical variables as well as the definition of critical variables and areas of concern. The use of the method in the public realm is usually dated to the *Mont Fleur* event in South Africa in 1991-92. Other notable examples include Colombia where in 1997-98 *Destino Colombia* aided in the leaders in coming to a common vision of peace, and Guatemala, where 1998-2000 *Vision Guatemala* offered a path out of the nation's nightmare.

Each of these projects focused on the construction of the future, and resulted in a shared vision. The outcome of this workshop was a shared vision about the pitfalls of an ill-designed information system and an understanding that business as usual will no longer suffice. While the passionate divides that exist between civil libertarians and law enforcement in the United States cannot be compared to the distances that had to be overcome in Mont Fleur or Guatemala, the scenario approach again enabled the construction of common ground. The result was an understanding of a compelling vision for digital government where accountability prevents exploitation of government resources, and loss of privacy is not the cost of participation.

The projects mentioned above are famed because of their success. For each project that concluded successfully however, many more projects ended as failures. Successful and unsuccessful projects can be differentiated by the quality and activity of the participants, and the intensity and adequacy of the preparation. Thus early preparation is critical. To ensure success, this proposal was broken into two discrete proposals, the first addressing the technological uncertainty began May 2002. The second proposal addressed the second set of activities. Before the first round of invitations was sent, a set of technological descriptions had been written. These descriptions lay the groundwork of what is and is not technically feasible in information management in 2003.

The identity scenario required some core agreements about the realm of technical possibility. The other civic projects required only statements about the obvious political realities.

Technological Descriptions

The investigation began with an underlying understanding of the technologies. Carefully constructed groups of technologists wrote simplified descriptions of authentication and risk management technologies. There were

five technology description groups: 1) Biometrics; 2) Cryptography with a focus on digital signatures; 3) Secure processing/computation; 4) Secure routing/communication and 5) Reputation systems. These descriptions were the basic building blocks that enabled the construction of scenarios.

Scenario groups were developed a month before the event itself. The scenario groups were formed with one representative from each of the technology description groups. Thus, the shared technological basis could be integrated into the scenarios.

This was the first application of the civic scenario task group, which combined government, academic, and business sectors with technical and policy representatives for all three. Previous civic scenario events were not based on technology. Previous technologically-based events were for the private sector. The heterogeneity of the participants combined with the technical complexity of the topic made this event a first. Each technology description group provided both coherent descriptions and technical individuals for the scenario groups.

Notice that the technology descriptions also discuss how objects will increasingly need identity. Technology description groups addressed the issue that hardware as well as people will require authentication as well as code that must be authenticated to be trustworthy. The concepts of identification are being mapped onto code and devices. Domain "names" are mnemonics that are used for branding and thus create trust. Mobile code is said to have a "true name" which verifies integrity of code and data. (This true name consists of the code and data hashed, then digitally signed by a trusted party.)

The various groups each developed a 25-page technical background paper. Each of these documents was completed in time to feed them into the scenario generation process. These technology descriptions included academics, representatives from competing commercial identity vendors, experts from government, and ethicist from the not-for-profit sector. The development of the technical reports assured that the scenario development process would not result debate of or differences of technical opinion.

The technological descriptions each included a brief technology background: what is it, how does it work. Using this, the scenarios were able to identify technological roads not taken. Each scenario and technology group was organized by a group leader. No person led both a scenario and a technology description team.

Development and Uses of Scenarios

The civic scenario process typically requires a two-day meeting concurrent with the construction of a set of scenarios. Usually groups at the

workshop first construct and then critique the scenarios under the guidance of the sponsoring organization (for example, Global Business Network). However, in this case a team constructed the scenarios in advance.

Each technology description task groups contained a minimum of five technical experts. The participants of each technology description group were divided across the scenarios. Therefore each scenario group had one representative from each of the technology description groups. The scenarios were completed shortly before the workshop. The lead-time for scenario construction was enabled by simple list technology, and allows for a shorter and more focused Harvard event. By making the event shorter a higher level of invitee was able to participate.

A long scenario construction time enabled more detailed analysis of the technological assumptions in the scenario. Building two scenarios on differing technological assumptions in real time risks a workshop blown off-course by those well-established technological debates closer to religious than scientific argument. Advance scenario development improved the probability of remarkable success in the workshop. As the workshop method was high-risk, the advance planning was time consuming, but worthwhile. The construction of the scenarios would have prevented the meeting of the minds.

There were four initial candidate scenarios. During the invitation process ideas for additional scenarios were requested, and the candidate scenarios were polished. The resulting five scenarios were as follows:

1. Single national identifier

The idea of a national identifier gained popularity in the wake of 9/11. The national identifier program is moving forward through the coordination of the fifty state drivers' licenses' authorities. A similar implementation can be seen in some identity management systems, which concentrate all data in a single account. Currently the Social Security Number is widely used as an identifier but it cannot be said to be ubiquitous and universal. This proposal will draw heavily on the secure hardware technology group.

2. Sets of identifiers

The national identifier scenario offers a single credential. In this proposal each person has a set of identifiers stored in secure hardware or in a series of devices. If the single credential is analogous to a signature, then the set of attributes is analogous to the key ring. In this case the multiple PKIs and devices will have some limited interoperability and potentially complex risk cascading issues. This scenario will draw heavily on the reputation technologies work.

3. Business as usual

In this scenario there will be a continuing growth of ad-hoc identifiers in the business world. The identifiers and practices in the business world are adopted unaltered for e-government. Such adoption is most likely in the form of closed code.

4. Ubiquitous anonymity

Under this scenario the tools of crypto-anarchy serve the ends of e-government. The most effective tools for ensuring anonymity are linked with particular assertions, for example, the assertion of Veteran status. Yet financial transactions and information requests can be made entirely anonymously.

5. Ubiquitous identity theft

The motivation for the ubiquitous identity theft scenario was difficult, until the Recording Industry Association of America began releasing waves of subpoenas. The file-swapping model became the intellectual basis for the identity-swapping model. Identities became so fluid and the personal information so badly protected that no reasonable person would want to expose their own personally identifiable information. Thus a single shopper might have a set of Social Security Numbers he or she uses, just as today we have multiple passwords.

The Harvard Meeting

At the invitation-only workshop the participants explored the set of technologically-based scenarios. These scenarios reflected possible visions of the interaction of citizens and government in the digital age and the digital marketplace.

The development of the breakout groups was a difficult scheduling process. Each breakout group needed to have representatives from each scenario group, each sector (academic, federal, state and private) and each technology description group. Yet the quality of discussion and results illustrated that the effort was a profitable investment.

The agenda in the appendix illustrates the intensity of the day. After each breakout group a representative from a discussion group brought forward the best practices and the research questions that arose from the discussion of the particular scenario.

The day began with greetings from the sponsor, the head of the local Center for Digital Government, and the workshop organizer, Lawrence Brandt, Jane Fountain and Jean Camp, respectively. Then the lead authors of the technology descriptions spoke, and the leads of the scenario groups spoke. Following that was lunch, with a keynote by Jeroen van den Hoven on the

differences between data protection and privacy. Then began a series of intense breakout groups to determine the best practices and the research agenda.

Designing break out groups is not trivial. Each person was to discuss every scenario except the one in which they participated. (There were four breakout groups and five scenarios.) Every group needed one participant from each technology description group for clarifying questions. Every group had one representative from each of the scenarios, excluding the one under discussion.

Every group had a rapporteur as well. Rapporteurs attended sections, discussing the scenarios that they helped develop. Therefore, they were both scribes and sources of clarifying information.

Works Cited

Acquisti, Alessandro. "Security of Personal Information and Privacy: Economic Incentives and Technological Solutions" *Workshop on Economics and Information Security* Berkeley 2002 http://www.cl.cam.ac.uk/users/rja14/econws/36.doc

Camp L Jean, Cathleen McGrath & Helen Nissenbaum. "Trust: A Collision of Paradigms", *Proceedings of Financial Cryptography*, 2001

De Lotto, Richard, Laura Behrens, Christopher Baum. "Soft Factors Will Impede Acceptance of US National ID" *Gartner Inc Market Analysis*, February 15, 2002.

Friedman, Batya. (Ed.) Human values and the design of computer technology. New York: Cambridge University Press and CSLI, Stanford University. 1997

Fukuyama, F. Trust: The Social Virtues and the Creation of Prosperity, Free Press, New York. 1995

General Accounting Office. "Identity Theft: Prevalence and Cost Appear to Be Growing" Report to Congressional Requesters GAO-02-363 March 2002

Garfinkel, Simson. "Few key bits of info open Social Security records" *USA Today*, 07 Apr 1997

Kent, Stephen T. and Lynette I. Millett, Eds *Who Goes There?: Authentication Through the Lens of Privacy on Authentication Technologies and Their Privacy Implications*, National Research Council 2003

Litan, Avivah. "Underreporting of Identity Theft Rewards the Thieves" *Gartner Inc Market Analysis*. July 7, 2003

Mack, Gregory, B. Bebee and G. Wenzel. "Total Information Awareness Program System Description Document v1.1" DARPA information Awareness Office, July 19, 2002

T. Matsumoto, H. Matsumoto, K. Yamada, S. Hoshino, "Impact of Artificial Gummy Fingers on Fingerprint Systems," *Proceedings of SPIE Vol. #4677, Optical Security and Counterfeit Deterrence Techniques IV*, 2002.

Mercuri, Rebecca. Humanizing Voting Interfaces," Rebecca, *Usability Professionals Association Conference*, Orlando, FL, July 11, 2002. http://www.notablesoftware.com/../Papers/UPAPaper.html

Moore, Stephen "The National ID Card: It's baaack!" Cato Institute Daily Report, September 23, 1997. http://www.cato.org/dailys/9-23-97.html

National Telecommunications and Information Administration (NTIA), *Falling Through the Net: Defining the Digital Divide*, (July, 8 1999) http://www.ntia.doc.gov/ ntiahome/press/fttn070899.htm

Safire, William. "Op Ed: Threat of National ID", *New York Times*, December 24, 2001

Social Security Administration "Report to Congress on Options for Enhancing the Social Security Card" SSA Publication No. 12-002-- September 1997

Thalheim, L., Jan Krissler, and Peter-Michael Ziegler. "Body Check: Biometric Access Protection Devices and their Programs Put to the Test" *c't* May 21, 2002 http://heise.de/ct/english/02/11/114/

Appendix: Workshop Agenda

Workshop on Identity in Digital Government Agenda

April 28, 2002

8:00 – 8:30 Breakfast and registration
8:30 - 9:00 **Introduction and Greeting** TB A& B

Jane Fountain - National Center for Digital Government
Valerie Gregg – the Digital Government Program of NSF

9:00 – 9:30 **Defining a Research Agenda** TB A& B
Jean Camp
workshop goals, the interaction of process & final product

9:30 – 10:00 **Session 1: Technology Description** TB A& B
a few slides each on definitions, and conclusions of technology descriptions

10:00-10:30 coffee break

10:30 – 12:00 **Session 2: The scenarios** TB A& B

- Single national
- Sets of attributes or special-purpose identifiers
- Business as usual
- Ubiquitous anonymity
- Ubiquitous identity theft

12:00 – 2:00 **Lunch**: Privacy vs. Data Protection

2:00 – 3:45 **Breakout 1** : TB A, B, C, foyer TB401

3:45 – 4:00 break Allison Dining Room

4:00 – 5:00 **Presentations and discussion**

April 28, 2002

8:00 – 8:30 Breakfast and registration

8:30 - 10:00 **Breakout 2**

list research questions arising from scenario discussions
eliminate scenario or identify one yet uncreated

10:00-10:30 coffee break

10:30– 12:00 **Breakout 3: Eliminate Scenarios**

Are there technological and organizational innovations that could collapse one scenario into another?

12:30 – 2:00 **Lunch in Allison Dining Room**

2:00 –3:00 **Discussion**

Evaluate remaining scenario (s)

3:00 – 4:00 **Breakout 4: Security Research**

4:00 – 5:30 **Concluding Session**
Confirming Consensus
Reviewing the major points in the research agenda
Process for completing and reviewing research report

For details and updates check:
http://www.ksg.harvard.edu/digitalcenter/conference/

Attendee list available upon request.

Workshop on Digital Identity
L 213
Kennedy School of Government
79 John F Kennedy Street
Cambridge, MA o2138