PRIVACY PRINCIPLES FOR IDENTITY IN THE DIGITAL AGE

Draft for Comment - Version 1.4

Center for Democracy & Technology - December 2007

I. INTRODUCTION

Intersection of Identity and Technology

How to create and manage individual identity is becoming a central challenge of the digital age. Various identity-related initiatives are being developed and implemented in both the public and private sectors. A major goal of many of these programs is to prevent illegal activity or enhance security, whether it be the security of our national borders, airplanes, workplaces, health records, or online transactions. Identity-related technologies – such as databases, machine-readable identification cards, and online accounts – can help realize the potential of the digital age, whether by making e-commerce exchanges more seamless, tying together information on multiple devices, combating fraud, or enabling yet unimagined services. Increasingly, as individuals go about their lives online and off, they will be generating or disclosing personal information or will be asked to identify themselves in some way. Undoubtedly, the range of transactions and events that can be linked to individual identity will grow.

However, the collection, storage, and disclosure of identity information can create risks to personal privacy and security. Poorly implemented identity systems can unnecessarily invade the privacy of innocent Americans, and can actually contribute to identity theft or weaken security. The digitization of information – by facilitating the collection, storage and sharing of large amounts of data – can exacerbate the privacy and security risks inherent in identity systems.

To mitigate these risks while achieving the benefits of identity systems, it is essential that these systems be designed with effective privacy and security measures built in. Incorporating such protections at the very beginning will help achieve the goals of identity systems while maximizing user control and other elements of privacy.

Summary of the Principles

In this document CDT outlines the following 11 privacy and security principles to guide public and private sector entities in the development of programs or systems involving the collection, authentication, and use of identity information:

Overarching Principles	Principles Based on FIPs
• **Diversity & Decentralization** • **Proportionality** • **Privacy & Security By Design**	• **Purpose Specification** • **Limited Use** • **Notice** • **Individual Control and Choice** • **Security** • **Accountability** • **Access** • **Data Quality**

The first three principles are overarching guidelines that are particularly relevant to identity in the digital age. The remaining eight principles are adaptations of the widely recognized fair information practices (FIPs) to the identity context.

The principles focus on privacy but also address security in certain instances. This is because privacy and security are interrelated and often should be considered together. When privacy is compromised, security of the individual, the organization or even the country is also threatened. Conversely, security breaches can also lead to invasions of privacy.

This document is meant to apply only to systems that identify individuals rather than groups or other entities.

The remainder of this section provides a general framework for understanding these principles. Section II discusses the principles themselves. Section III contains a glossary of terms used throughout this document.

Are privacy and identification at odds with each other?

For those with limited exposure to the concepts of identity and privacy, it may seem as though identification and privacy are in contradiction to each other. In many cases, this is true: privacy can be served though the lack of identity. Anonymity is a constitutional right in some circumstances. As individuals are increasingly required to reveal more information about themselves and authenticate their identities more often, their privacy may be at greater risk.

However, the relationship between identity and privacy is a nuanced one. Consider a fraud detection system as an example. To determine if a particular transaction involves the fraudulent use of identity – and if an innocent individual's financial or other personal information has been compromised – it is useful to gather a good deal of information about the transaction and the identity claims of the individual seeking to engage in the transaction. This information can be compared with other identity information that has been compiled for fraud prevention purposes. Although gathering a lot of identity information may seem antithetical to privacy interests, in this case it may actually help to protect privacy by identifying an instance of identity theft. Thus, although in many cases less identification can mean more privacy, in this case the opposite may be true. Nevertheless, even an anti-fraud identity verification system can be designed in a pro-privacy fashion and should be guided by the principles set forth here.

The "less identification equals more privacy" idea also fails to take into account the type and sensitivity of the identity information involved. For example, a person's name, address, and telephone number may constitute a greater quantity of information than the person's fingerprint or DNA profile, but the latter reveal much more about the person. Likewise, a small amount of identity information that is shared with a multitude of parties or is not properly secured may put an individual's privacy at

greater risk than a large amount of information that is properly secured and accessed only by authorized parties.

These nuances ultimately lead to the conclusion that the evaluation of identity systems with respect to privacy must be done in context. Determining how to apply the principles set forth in the next section to a particular identity system will require a solid understanding of the environment in which the system operates and of all the risks and benefits that the system must balance.

Goals of the Principles

The purpose of this document is:

- To provide the public interest and privacy advocacy community with a general, high-level framework for evaluating the privacy and security of identity systems, and

- To provide policymakers, and identity system designers, implementers, and users – including those who may be unfamiliar with privacy concepts – with guidance on how to safeguard personal privacy and security in identity systems.

The ultimate goal of the principles is to help ensure that any given identity system maximizes personal privacy and security, or at the very least, minimizes invasions of privacy and threats to security. It should be noted, however, that applying all of the principles will not necessarily guarantee that a given identity system will be privacy-protective.. Nor are the principles intended to be a mere checklist. Rather, they are intended to spur the development of creative solutions.

The principles are interrelated and must be viewed as one overarching policy framework. All of the principles should be considered together when developing an identity system. However, while it is possible to apply each principle to an identity system, it may be that not all of the principles will apply to a given identity system with equal force. Policymakers and system designers must fully consider each principle and how it can be maximized within a given identity system, but may reasonably conclude that it is more appropriate or feasible to focus on some principles over others depending on the particular context or specified purpose of the identity system.

Next Steps

This Version 1.4 is a draft document that is open for comment. CDT continues to convene stakeholders with diverse perspectives on this issue with the hope of achieving a comprehensive set of guidelines that can be useful across the public and private sectors and in many different contexts.

II. PRINCIPLES

Preliminary Question: Is Individual Identification Necessary?

The first consideration for both governmental and commercial entities should always be whether an identity system is in fact necessary and effective for solving the problem at hand. Many goals can be accomplished without using any identity information at all.

Policymakers and system designers should not assume that adding an identification element to a system – an access system, payment system, communications system, or other transactional system – will make it more robust. The advantages of collecting and using identity, authentication, and linked information should be weighed against the risks to privacy and security.

Once a specific problem or goal is clearly articulated, the key question must be asked: Is individual identification necessary for solving the problem or accomplishing the goal? Developers should always be open to solutions that do not involve individual identification. However, if the answer is "yes," then the following 11 principles should be addressed during the development of the identity system.

Each principle below begins with a concise statement (in bold). A lengthier description and examples follow this statement.

Overarching Principles

The first three principles are overarching guidelines that are particularly relevant to identity in the digital age.

✦ Diversity and Decentralization

Rather than attempt to serve as a perfect single solution, enrollment and authentication options should function like keys on a key ring, with different identities for different purposes. They should allow individuals to choose the appropriate option to satisfy a specific need. On balance, it is not optimal to centralize identity information or use a single credential for a multitude of purposes. In cases where linking of identity systems and databases is deemed necessary, appropriate safeguards should be implemented to limit the associated privacy and security risks.

Using only one or a very small handful of centralized identity solutions for multiple purposes leaves individuals with few choices and diminishes the ability of identity systems to protect privacy and security. Requiring individuals to use a single identifier or credential for multiple purposes creates a single target for privacy and security abuses by identity thieves, terrorists, government, business, and others.

Using a single identity for multiple purposes may, however, offer convenience and efficiency benefits. These benefits should be weighed against the risk of concentrating identity information in a single location or credential.

EXAMPLE 1: Credit Cards

Individuals have the option of using merchant-specific credit cards or a single general-purpose credit card. Carrying a single card may be more convenient because it can be used in many different locations. However, this allows a single credit card company to maintain an individual's entire transaction history. Using multiple merchant-specific cards spreads this information among several parties.

It is important for both kinds of credit cards to exist so that individuals can weigh each option's benefits and drawbacks related to both privacy and convenience. Some may prefer the convenience of a single card, while others may prefer maintaining multiple separate transaction histories. Rather than requiring individuals to use one system or the other, both systems should be able to coexist.

Different government agencies, companies and organizations, and different types of functions within organizations, will likely need different types of identity systems. Identity systems should be designed to function in a marketplace offering multiple services that deliver varying degrees and kinds of enrollment, authentication, and use of identity information. This diversity of systems compliments the principle of Individual Control and Choice, which recommends that individuals be provided with options for expressing and authenticating their identities within a single system.

EXAMPLE 2: Diversity in Authentication Mechanisms

Consider the authentication mechanisms necessary for two different scenarios: accessing health records at a doctor's office, and accessing a Web-based email account.

At the doctor's office, a patient may be required to provide an identification card, such as a health insurance card, in order to access his or her own health records. The card might include information such as the patient's name and date of birth. Or a doctor or nurse may simply recognize a long-time patient and provide access to the appropriate records.

For Web-based email, a username and password combination is frequently used to authenticate the owner of an account. Some accounts may use two-factor authentication that combines knowledge of a password or PIN with possession of a security token or card. These authenticators may or may not reveal the account owner's name or other identity information.

Each of these authentication mechanisms is suitable to its own context. It would make little sense and may be harmful to privacy if individuals were required to login to their email accounts using a health insurance card – it is not necessary for most Web-based email providers to know the information on the card, and most cards are not remotely readable. Having a diversity of authentication mechanisms available is key to ensuring that suitable solutions exist for all kinds of authentication contexts.

The concept of decentralized storage and access to identity information closely parallels the idea of having a diversity of mechanisms for expressing and authenticating identity. As identity information becomes more centralized – whether through storage in a single physical location or linkage across disparate databases – there is increased likelihood for abuse.

In a networked world, the urge to link identity systems and databases together will always exist. Linking together disparate identity data may improve convenience,

efficiency, and even security (in cases such as fraud detection where linking information together can help to detect and deter fraudulent activity). Linking should occur in cases where its specific benefits exceed the associated privacy and security risks. When linking is deemed necessary, strong safeguards should be erected to ensure that unnecessary linkages do not occur. These safeguards should be addressed in the design phase of an identity system (consistent with the principle of Privacy & Security By Design) and not as an afterthought.

✦ Proportionality

The amount and type of identity information collected from individuals by an identity system should be proportional to the purpose for which it is collected. This means that the amount and sensitivity of identity information required for enrollment or participation in an identity system should be reasonable and appropriate in relation to the articulated purposes of the system.

Generally speaking, it is reasonable for an identity system to collect larger amounts and/or more sensitive identity information from individuals seeking to participate in transactions of higher significance. Similarly, it is generally not reasonable for an identity system to collect a multitude of attributes, or those that divulge substantial identity information, for transactions of lower significance.

> EXAMPLE 3: Gym ID Card
> An athletic club might print members' names and photos on club ID cards, but collecting biometrics exceeds what may reasonably be considered necessary to ensure that only club members have access.

For many transactions, it will never be appropriate to collect certain kinds of identity information. Only in the most select of situations is it ever appropriate to ask individuals about their race, ethnicity, or religious or political affiliation, and even then this information should anonymized to the greatest extent possible.

> EXAMPLE 4: College Applications
> College applicants are frequently asked for race, ethnicity, or religious information for admissions purposes, but it is generally anonymized and aggregated after it is collected.

Not all transactions need to be tied to identity. Identity systems relying on pseudonymous identifiers and authentication relying on anonymous attributes should be used whenever possible.

> EXAMPLE 5: IRS
> The IRS may require individuals to authenticate themselves by providing their previous year's total income and a PIN number of their choice, both of which are anonymous attributes.

One way for organizations to achieve proportionality in the collection of identity information is to use trusted networks that allow individuals to leverage secure identities created through other organizations. Trusted networks reduce the number of

organizations that need to collect identity information without reducing the variety of identity systems and options available to individuals.

> EXAMPLE 6: OpenID
> OpenID is one example of a system that provides a way for Web sites to leverage an identity created by a user through a separate "identity provider." Using this system, individuals can choose to share their identity information only with the identity provider and not with individual Web sites. When these individuals want to login to a particular blogging service, for example, the service contacts the identity provider to authenticate the individual, but the blogging service does not collect any identity information itself.

✦ Privacy and Security By Design

Privacy and security considerations should be incorporated into an identity system from the very outset of the design process. These include both safeguards for the physical system components and policies and procedures that guide the implementation of the system. Internal privacy and security practices should incorporate applicable regulatory and self-regulatory guidelines. Privacy impact assessments should be issued in conjunction with system design plans.

Identity systems should be designed with attention to human strengths and limitations that may impact the privacy and security of the systems. Knowledge of human behavior and how people will likely interact with an identity system should be incorporated from the first phases of a system's design.

> EXAMPLE 7: Forgetting Passwords
> People have difficulty remembering complicated passwords, so they choose passwords that are easy for others to guess. This human tendency should be central in deciding whether passwords are a strong enough authentication mechanism for the task at hand.

Consistent with the principle of Limited Use, identity systems should be designed to make secondary uses difficult. Incorporating technological and policy-based limits on the use of the system into its design will make "mission creep" – authorized but initially unintended uses – easier to avoid and less appealing later on.

Identity systems should have consistent, robust interfaces so that individuals can learn to trust legitimate systems and distinguish them from fraudulent ones.

Principles Based on Fair Information Practices

The remaining seven principles are adaptations of widely used fair information practices to the specific context of identity in the digital age.

✦ Purpose Specification

The first step in designing an identity system should be to specify the purpose of the system and the purposes for which identity information will be collected and used. The purposes for collecting and using identity information should be directly linked to the ultimate purpose of the system. Each purpose should have a clear and publicly communicated rationale behind it.

This specification should guide all further decisions about how identity systems will be designed, implemented, and used. Adhering to the principles of Proportionality, Limited Use, and Notice will require making decisions in accordance with the purpose specification.

✦ Limited Use

Identity, authentication, and linked information should be used and retained only for the specific purposes for which they were collected. Uses should be limited to those consistent with the identity system's purpose specification.

Secondary use, sharing, and sale of identifiers or credentials can compromise privacy and security. In particular, identification numbers can become open to privacy misuses and security threats if they are used for secondary purposes, especially in the case of authentication. Therefore, multiple uses of such identifiers should be avoided, particularly in the authentication context.

> EXAMPLE 8: Social Security Numbers
> The Social Security Number system was initially intended to be used for tracking income and issuing federal benefits. In the decades since it was introduced, however, the Social Security Number has been used across a whole range of other contexts, and it is now commonly used as an authenticator in setting up bank accounts, opening lines of credit, and obtaining loans. Because it is in such wide use as an authenticator, the Social Security Number has become a prime target for identity thieves and other criminals.

Use of identity, authentication, and linked information should be disclosed and minimized, and the information should only be stored until the purposes for which it was collected have been fulfilled. Identity, authentication and linked information should be shared with third parties – including data transfers between government and commercial entities – only when necessary, and should be stored by third parties only until the purpose for which it was shared has been completed.

EXAMPLE 9: Information Collection at a Bar

Consider a bar owner who decides to scan the barcodes on the backs of patrons' driver's licenses and store the names and addresses read from the barcodes in a database. The bar owner's purpose for doing this is to maintain a list of rowdy patrons who will not be allowed back to the bar. The bar owner discloses this to patrons before scanning their licenses, and turns away patrons who refuse to have their licenses scanned.

To conform with the principle of Limited Use, the bar owner should not later sell his or her database to a marketer. This would constitute a use that does not conform to the purpose for which the information was collected and was not disclosed to the individuals involved.

The amount and type of data linked to an identity should be limited, and linking should occur for specific, limited and disclosed purposes.

✦ Individual Control and Choice

Whenever possible, an identity system should offer individuals reasonable, granular control and choice over the attributes and identifiers needed to enroll in the system and the credentials that can subsequently be used within the system.

Individual controls help build trust in identity systems.

EXAMPLE 10: Choices in Air Travel

There are several examples of choice in the air travel context. At U.S. airports, individuals can choose among several different government-issued identification documents for use in authenticating their identities for check-in. When checking in for a flight online, many airlines will accept several different authenticators or combinations of authenticators that reveal different kinds of identity information (first name, last name, confirmation number, credit card number, airline member number, and others).

Individuals should have the option of using a single credential or form of authentication that always discloses the same information for all interactions, or employing a variety of authentication tools for different transactions. This principle is particularly important in a system designed for both authentication and authorization, which will likely be successful only if it balances added convenience with trust in the system.

EXAMPLE 11: Control in Online Accounts

Many online services allow a single individual to maintain multiple accounts. Consider the case of a social networking site. An individual might maintain different accounts to interact with family, friends, and colleagues. Each account might be associated with different contact details, photos, and other information. The ability to maintain multiple accounts gives individuals control over not only which information is used in each context, but also which sets of information are correlated with each other. An individual may choose to put different information in an account linked to his or her real name than in a pseudonymous account.

Individuals should be given the opportunity to consent to the terms of an identity system's notice (as described in the Notice principle) prior to enrollment, authentication, or use of identity or linked information. If an individual declines to accept the notice, no information should be collected. When possible, individuals should be able to consent to participation in an identity system but decline particular

terms of the notice. Should new uses of identity or linked information be developed after an identity is created within a system, individuals should be given the opportunity to consent to or decline such uses.

Individuals should not be required to accept the sharing of information for secondary uses as a condition of enrolling in an identity system.

✦ Notice

Individuals should be provided with a clear statement about the collection and use of identity, authentication, and linked information. Notice should be conspicuous and timely, and it should be provided in a manner appropriate to the technology being used. Notice provides a basis for accountability, in accordance with the Accountability principle.

> EXAMPLE 12: Cell Phone Notices
> Displaying a long, multi-paged notice on a small cell phone screen is an example of how notice could be inappropriate for the technology being used.

Individuals should be notified in situations where it may not otherwise be obvious that identities are being created for them.

Prior to enrollment, individuals should be notified of:

- The purposes for which their information is being collected (as developed based on the Purpose Specification principle);
- Who is managing the identity system and creating identities for individuals within the system;
- What information will be collected and how it will be used and secured;
- How long the identity information will be stored;
- Whether and how the identity and authentication information will be used by third parties;
- What other information will be linked to the identity and whether and how that information will be used;
- Whether individuals might need to authenticate themselves in the future and how to do so;
- How the individual will be able to access and correct information related to his or her identity within the system (consistent with the Access and Data Quality principle); and
- How the individual may decline to enroll in the system.

When identity systems make use of a technology that may be unfamiliar to participants in the system, notice should be provided about the presence of the technology and its privacy implications, in accordance with the items listed above.

> EXAMPLE 13: RFID

Many individuals may be unfamiliar with radio frequency identification (RFID), a technology that uses radio waves to identify and object. Individuals should be notified about how information about them – such as their location or items they have purchased – can be linked to their identities when RFID is used in an identity system.

Should new uses of identity or linked information be developed after enrollment in an identity system, individuals should be notified in accordance with the items listed above.

Individuals should be notified when other information is gathered about them and linked to their identity.

✦ Security

Organizations that handle identity, authentication, and linked information should provide reasonable technical, physical, and administrative safeguards to protect against loss or misuse of the information. Such measures should cover credentials, back-end systems that process and store information, personnel that handle the information, and physical facilities, among others.

In so doing, organizations should establish and maintain an information security program in keeping with industry standards and applicable laws, appropriate to the amount and sensitivity of the information stored in their systems. Such a security program should include processes to identify reasonably foreseeable internal and external risks to the security, confidentiality, and integrity of identity information; address those risks; and provide notice as appropriate for security breaches.

EXAMPLE 14: Industry Security Standards
ISO/IEC 17799 is one widely recognized international standard that provides best practice recommendations for information security management.

Identity systems that handle large amounts of identity information may be more vulnerable to tampering, loss, and unauthorized access (both internal and external). Adhering to strict, logical security procedures should be a top priority for such systems.

The authentication mechanism used for internal access to an identity system should be at least as strong or stronger than the mechanism for external access by participants in the system.

EXAMPLE 15: Internal Authentication
System administrators for a database of identity information may be required to provide two biometric credentials for authentication while participants in the system are required to provide only one biometric credential.

✦ Accountability

Organizations that handle identity, authentication, and linked information should be able to verify that they are complying with applicable privacy and security protections. Regular audits are necessary to ensure that reasonable technical, physical, and administrative privacy and security safeguards are being used. Personnel involved in handling identity information should be trained and educated about the privacy and security risks involved in dealing with identity and about applicable laws, guidelines, and procedures.

Any organization that handles identity information should include in its contracts provisions requiring that the entities with which identity, authentication, and linked information is shared will afford that shared data a level of protection consistent with or exceeding the organization's own standards, consistent with these principles and any industry standards that conform to these principles.

> EXAMPLE 16: Industry Standards for Shared Information
> The PCI Data Security Standard is an example of an industry standard that can be implemented via contracts between entities sharing identity information.

✦ Access

Individuals should be provided reasonable access to the identity, authentication, and linked information that organizations maintain about them and use in the ordinary course of business. This ability should be secured against unauthorized access.

The information should be easy for individuals to access, view, understand and change. Individuals should also be able to challenge conclusions drawn from identity and other information via structured and impartial processes. Whenever possible, individuals should be able to see when their identity information has been disclosed and to whom.

Depending on the context, access should either be provided by the organization that enrolls the individual or the organization interfacing with the individual, if they are different.

✦ Data Quality

Organizations should strive to ensure that the identity information they hold is timely, complete, and accurate.

Individuals should be able to correct inaccurate, out-of-date, and incomplete information. The data quality principle may thus be partly dependent on the access principle, since individuals will need to access their information in order to correct it.

III. GLOSSARY

Italicized definitions are from the National Research Council's *Who Goes There? Authentication Through the Lens of Privacy*.[1]

Attribute. An attribute describes a property associated with an individual.

Authentication. Authentication is the process of establishing confidence in the truth of some claim.

Authentication Information. One or more facts presented to support the authentication of an identity.

Authorization. Authorization is the process of deciding what an individual ought to be allowed to do.

Credential. Credentials are objects that are verified when presented to the verifier in an authentication transaction. Credentials may be bound in some way to the individual to whom they were issued, or they may be bearer credentials. The former are necessary for identification, while the latter may be acceptable for some forms of authorization.

Enrollment. Enrollment is the process by which an identity for individual X is created in identity system Y.

Identification. Identification is the process of using claimed or observed attributes of an individual to infer who the individual is.

Identifier. An identifier points to an individual. An identifier could be a name, a serial number, or some other pointer to the individual being identified.

Identity. The identity of X is the set of information about individual X, which is associated with that individual in a particular identity system Y. However, Y is not always named explicitly.

Identity Authentication. Identity authentication is the process of establishing an understood level of confidence that an identifier refers to an identity. It may or may not be possible to link the authenticated identity to an individual.

Identity Information: One or more attributes used to establish an identity.

Identity System: Any program or framework that involves the collection, authentication, or use of identity or linked information. Identity systems may be designed, implemented and used by government, businesses, or individuals.

Individual Authentication. Individual authentication is the process of establishing an understood level of confidence that an identifier refers to a specific individual.

[1]National Research Council of the National Academies. *Who Goes There? Authentication Through the Lens of Privacy*. Eds. Stephen T. Kent and Lynette I. Millett. Washington: The National Academies Press, 2003.

Linked Information: Other facts about an individual, such as transactional, shopping or travel behavior, tied to an identity.

Pseudonymous: Using a name or label that may identify an individual within a system but does not correlate to that individual outside of the system.

Secondary Use (of information): Any use of identity or linked information that is inconsistent with an identity system's purpose specification.

Use (of information): Any use of identity, authentication, or linked information other than for enrollment and authentication purposes. Use may follow either enrollment or authentication.