LIBERTY ALLIANCE PROJECT

Privacy Summit

hosted by the Public Policy Expert Group
of the Liberty Alliance
in conjunction with the Net-ID 2007 conference

Monday April 23rd 2007
Brussels

Meeting Summary

Foreword

This report summarises the second in the programme of Privacy Summits organised and hosted by the Public Policy Expert Group (PPEG) of the Liberty Alliance.

As for the first (Berlin) Summit, the aim of this meeting was to bring together a wide range of stakeholders, so as to stimulate a multi-disciplinary discussion of the technical, legal, social and other perspectives on identity and privacy. To that end, the participants included academics, lawyers, user organisations from the public and commercial sectors, technologists, industry analysts and so on.

The discussion was held under the 'Chatham House Rule'[1] – so the participants are free to repeat what was discussed, but may not reveal who said what, nor the affiliation of any of the participants.

This document is an attempt to summarise the topics discussed; it is not an exhaustive record of the meeting, and participants are encouraged to reply with any further notes or comments they would like to add to the output of the Summit.

The corresponding output report from the first Summit is available online from either of the two following URLs:

http://www.projectliberty.org/liberty/files/whitepapers/privacy_summit_meeting_at_net_id_berlin_summary_report

http://www.projectliberty.org/liberty/content/download/3114/20838/file/Privacy-Summit-Final.pdf

Anyone wishing to comment on, add to or correct a Summit report is warmly invited to do so by sending an email to robin.wilton@sun.com.

Many thanks for your participation

and your contribution to this ongoing programme.

Dr. Hellmuth Broda - Chairman

Robin Wilton - Moderator

Document date: 21st May 2007

Document reference: PS-Brussels-2007

[1] The Chatham House Rule

"When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed" (http://www.chathamhouse.org.uk/index.php?id=14).

Table of Contents

The Continuing Programme of Privacy Summits

The second, Brussels Summit represented an interesting phase in the programme, in that we approached it with no preconceptions as to whether the participants would want to simply pick up where the previous summit had left off (bearing in mind that several of the participants had not attended the Berlin meeting) or whether it was a case of starting from scratch.

In the event, an initial half-hour of 'brainstorming' generated a very broad scope of issues, questions and concerns which provided ample new material for the remainder of the discussion. In fact, the principal challenge (at least from a moderator's perspective) was to identify those topics which would generate the most useful output, and capture them in a form which would be both accessible and useful. That is what I have attempted to do in this document, which is intended as a summary of the principal themes of the discussion, rather than a point-by-point resumé of the meeting.

As before, if participants have other points which they wish to include or changes to recommend, an email setting these out would be most welcome.

Three High-level Models

This paper does not present a 'recipe for achieving privacy in any context'; however, it does give simple, systematic examples of how to identify the factors which can contribute to (or undermine) privacy in a distributed identity architecture.

From the opening half hour of the discussion, it was obvious that not only are 'digital identity' and privacy extremely wide-ranging topics, but that there is also a great deal of disparate thought, ambiguity and confusion over core concepts such as identity, identifiers, credentials and attributes.

Questions then arise as to how these relate to 'functional' topics such as privacy, data protection and identity management in all the various contexts in which they apply.

Finally, there are practical questions about how digital identities and their management can contribute to notions of trust in online systems.

While I cannot claim that the workshop definitively answered these questions, it certainly allowed us to formulate three high-level conceptual models which not only made some of the core concepts easier to understand, but also condensed them into a form which is relatively easy to reproduce and pass on to others who may be experiencing some of the same confusion and ambiguity.

The three models can be summarised as:

1. The “Onion” Model of Identity-related Data
2. A 3-Circle Model for 'Identity in Context'
3. A “Silo” model for identity across contexts

We start with a model which looks at different kinds of data which are often referred to as 'identity data'; from this, we extend to a model which considers where those kinds of data crop up in multi-application or multi-sector systems, and finally we take a wide perspective to look at how these data and architectural models relate to the technical and non-technical disciplines of identity management.

The next sections of this report will describe each of these three models in turn. In each case, I describe the basic model and the principles it illustrates, and then 'annotate' the basic model with additional points.

The “Onion” Model of Identity-related Data

Even among experts (and perhaps especially among experts), the phrase “identity data” can mean a wide range of things. How, for instance, do identity, identifiers or credentials, and personal attributes relate to one another to constitute a 'digital identity'? If digital identities are composed of multiple elements, do those elements or types of element need to be managed differently? Is my digital identity made up of all the digital facts that are associated with me, or is it important to be able to segregate some facts from others (and if so, how and why)?

One principle we noted was that there seems to be a strong theme of 'uniqueness' about digital identity. This has both philosophical and practical roots. Philosophically, Leibniz formulated (in the late 1600s) the principle of the 'Identity of Indiscernibles', which states that if two things have exactly the same set of properties then they are one and the same thing – they are identical. A relation of 'identity' obtains between them. Practically, we can determine that two things (or people) are not identical by looking for some attribute that they do not have in common – in other words, we look to prove identity through uniqueness.

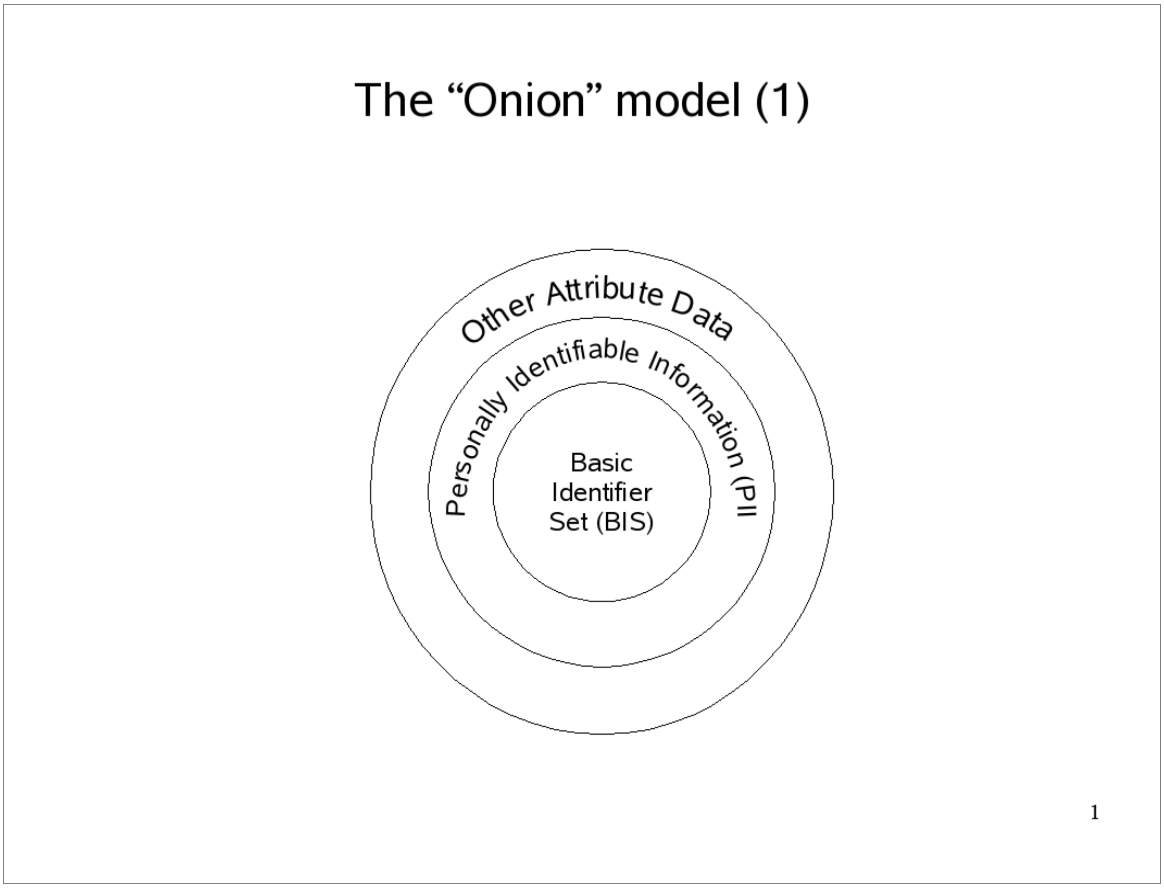
When it comes to identity-related data in practice, the discussion considered some national schemes which are based on a so-called “Basic Identifier Set” (BIS). These are the small set of data attributes which are generally considered, in such cases, sufficient to establish the uniqueness of a given individual. An example of a BIS might be:

- Name
- Date of Birth
- Place of Birth
- Gender

Examples have been given[2] of cases in which any or all of these attributes might not be immutable, but for practical purposes they form the core of most large-scale identity schemes such as passports, identity cards and so on. However, identity-related data clearly also encompasses a much wider range of data than just the BIS. The model we derived was a layered one, in which the BIS is the centre, surrounded by other Personally Identifiable Information (PII), which in turn is surrounded by other attributes and historical data relating to an individual. This is illustrated in the diagrams below.

[2] Gillian Ormiston, OECD Workshop, Trondheim May 2007: http://www.oecd.org/document/41/0,2340,en_2649_34255_38327849_1_1_1_1,00.html

Basic “Onion” model


The “Onion” model (1)

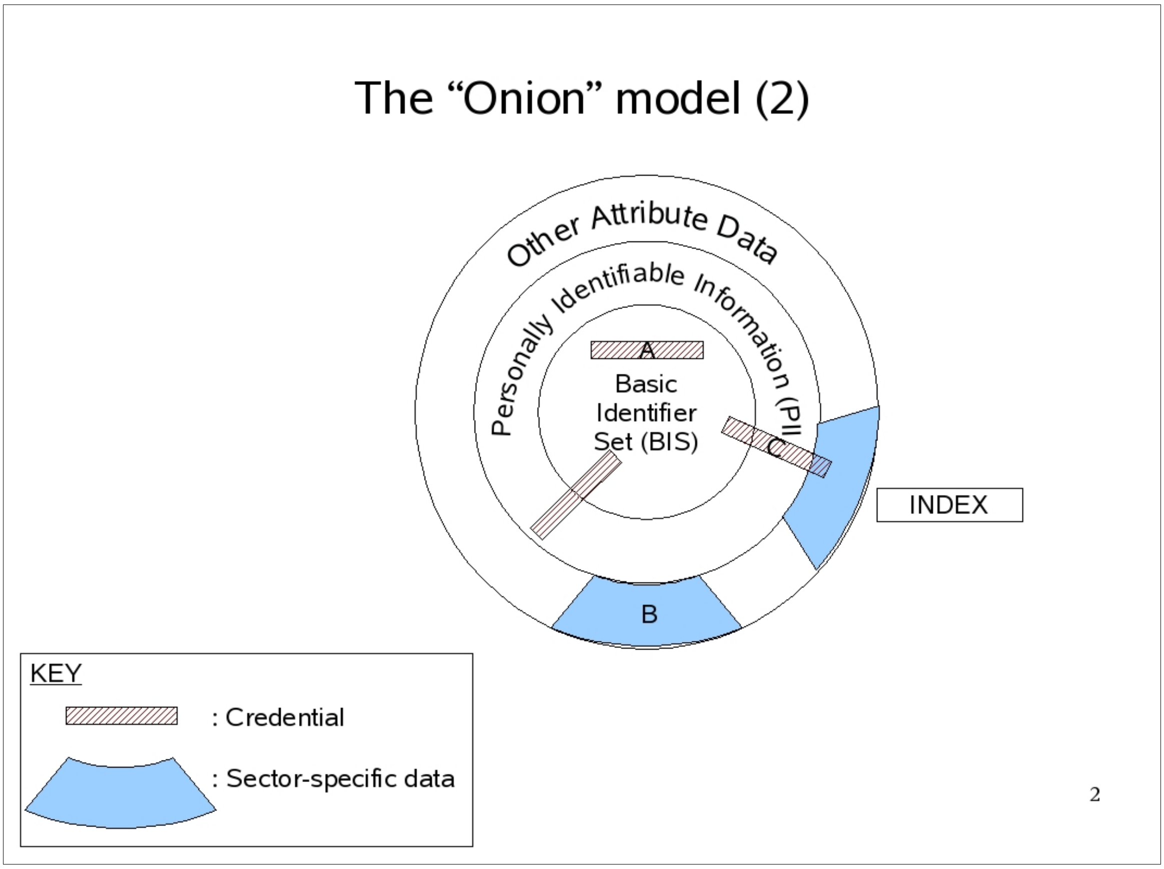
Other Attribute Data
Personally Identifiable Information (PII
Basic Identifier Set (BIS)
1

The first diagram, above, illustrates the basic principle of layers of identity data.

- The core BIS is the minimum set of attributes accepted as being sufficient to establish the uniqueness of a given individual.
- The next ring consists of the kind of personally identifiable data which might not meet the BIS criteria, but is probably covered by national data protection laws or their equivalent. An example of this might be “current address”.
- The outer ring consists of other attribute data associated with the individual, such as transaction histories. It also includes sector-specific data such as blood type, which might not in itself identify an individual, but is clearly useless unless correctly attributed to the right person.

It is interesting to note that one effect of increased computerisation (and increased computing power) is that data in the 'other attribute' category which might not previously have been sufficient to identify an individual might now be sufficient to do so. For instance, many web servers accumulate data about the browsing behaviour patterns of users; over time, interaction with a given website would allow the website owner to say, with reasonable certainty, whether a given user is the same one as visited the same website the previous day from the same IP address (as opposed to, say, a different member of the same household).

Annotated “Onion” model



The “Onion” model (2)

Other Attribute Data
Personally Identifiable Information (PII)
Basic Identifier Set (BIS)
A
B
C
INDEX

KEY

▨ : Credential

: Sector-specific data

2

This version of the diagram is annotated to illustrate some further useful points.

A represents a credential which contains only those data items about the user which form part of the BIS. By contrast, C illustrates a credential which contains some items from all three rings. An example of such a credential might be a driving licence, which could contain the following:

- BIS items such as name, date of birth, gender;
- PII items such as current address;
- Other attribute data such as 'entitlement to drive heavy goods vehicle'.

As B suggests, the 'onion' will often tend to be divided into sector-specific wedges, some of which may rely on their own sector-specific credentials (such as the driving licence).

The special case of the "index"

A further significant point, often subject to confusion in this area, is that whenever sector-specific data about an individual is stored (for instance, by tax authorities, driver/vehicle licensing agencies and so on), there is almost inevitably an index value which is used to identify each unique record in that store. The index value may or may not appear on a sector-specific credential issued by that agency.

The importance of this point is that such indices can be over-exposed and over-used, and this can undermine the integrity of the identity data in question. An example of this is the US Social Security Number. This fulfils the role of an index to each citizen's social security records, but over time (despite laws to the contrary) has come to be used as a credential. As a result, there is now widespread inappropriate reliance on Social Security Numbers, and their utility as an identifier is greatly compromised.

Where an index exists, it is important that it be appropriately managed (and if necessary, subjected to quite different management disciplines from, say, the credentials associated with it). An example of this is the Norwegian government's policy for national identity numbers. These are, by default, not to be revealed – and applications wishing to use the national identity number as a means of indexing an individual's sector-specific records may only do so with specific legal permission.

"To point or to store... that is the question"

A further note about credentials is that, for privacy reasons, some governments take the view that the closer a credential stays to the centre of this onion model, the better: that is, credentials should serve to identify the individual, but not necessarily be loaded with attributes and other personal data.

By analogy, imagine that, in order to establish your entitlement to buy alcohol in a bar, you show your driving licence. The bar staff only need to know that you are over the required age – but the credential might also reveal to them your date of birth, place of birth, current address, driver/licence index number, which types of vehicle you are entitled to drive, and possibly any endorsements you have.

In the online environment, where all that is needed is a pointer to the authoritative source of that information, it seems a sound principle that credentials should gravitate towards the centre of the onion, and point to, rather than hold, PII and attribute data.

By implication, this means that the more centralised a repository is in its design, the more attractive it is for the system to focus on 'proof of uniqueness' as opposed to broader sets of PII and sector-specific data. As the subsequent models will show, this is not a guarantee of 'unlinkability' (if that is an objective of the design), but may contribute towards it.

A 'silo' model for identity across contexts

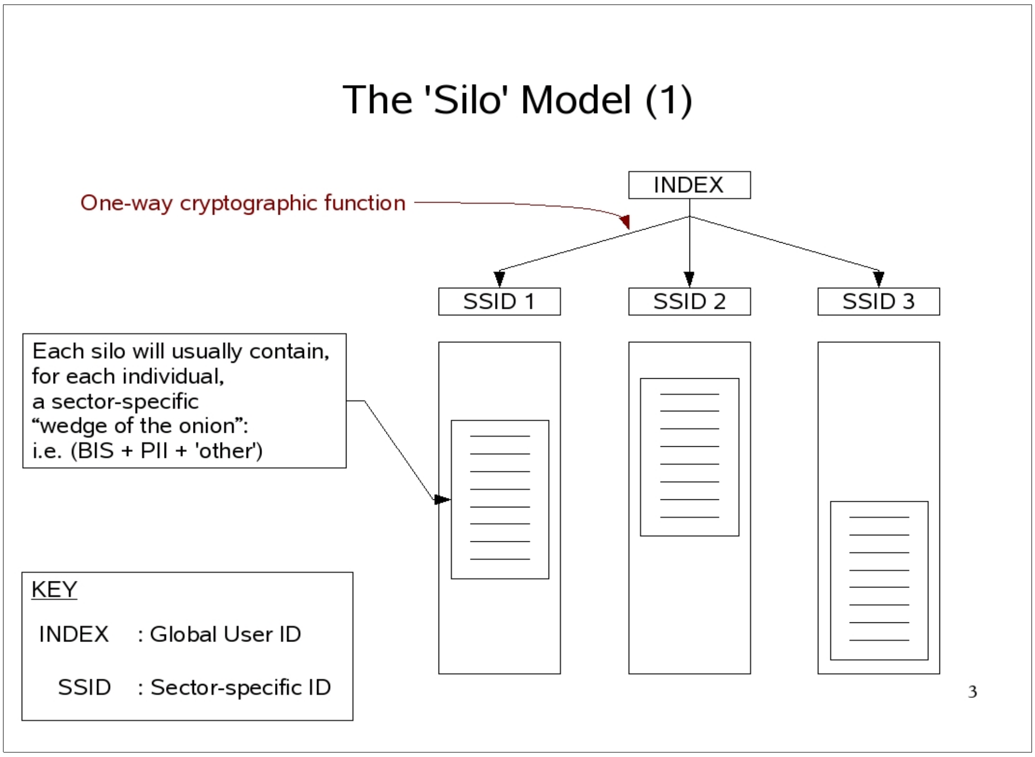
As noted in the summary document from the Berlin meeting, the Austrian ID Cards implementation can be used to illustrate a number of useful concepts. A concise description of the Austrian scheme can be found here[3] on the European Commission's IDABC website, and a web search using the argument "Austrian Citizen Card" will return a wide range of further documentation.

In the Austrian government example as described, a single state-issued identifier is used as the basis for multiple sector-specific identifiers – which can only be correlated by the Privacy Commissioner. The sector-specific identifiers are generated using one-way cryptographic functions, so that the sector-specific identifier can easily be derived from the original identifier on the citizen card, but conversely, the original identifier cannot be derived from the sector-specific identifiers. This gives rise to an architecture as illustrated in the diagrams below.

The first diagram allows us to relate this multi-sector view to the 'onion' model described above. Each silo will usually contain a sector-specific set of information about the individual, corresponding to a 'wedge of the onion' – that is, some or all of the Basic Identifier Set (BIS), some Personally Identifiable Information, and other data such as transaction history, entitlements and so on.

Note that this architectural model applies equally to public- and commercial-sector systems, and to intra-organisational as well as inter-organisational systems. Within a single organisation, this model illustrates a classic 'application silo' set-up. Between organisations, it illustrates a typical 'distributed identity' set-up. Each organisation (or application) stores the information it needs, regardless of whether this results in duplication. As far as the user is concerned, the applications appear disjunct.

Basic silo model



The 'Silo' Model (1)

Each silo will usually contain, for each individual, a sector-specific "wedge of the onion": i.e. (BIS + PII + 'other')

KEY
INDEX : Global User ID
SSID : Sector-specific ID

[3] http://ec.europa.eu/idabc/en/document/4486/5584

The second silo diagram illustrates some further points about identity in distributed applications. First, note that each silo can be regarded as a 'context', within which the user discloses some information to the service provider. For two silos to be federated, the appropriate technical and non-technical measures need to be in place to establish the context within which the user will disclose information to both service providers.

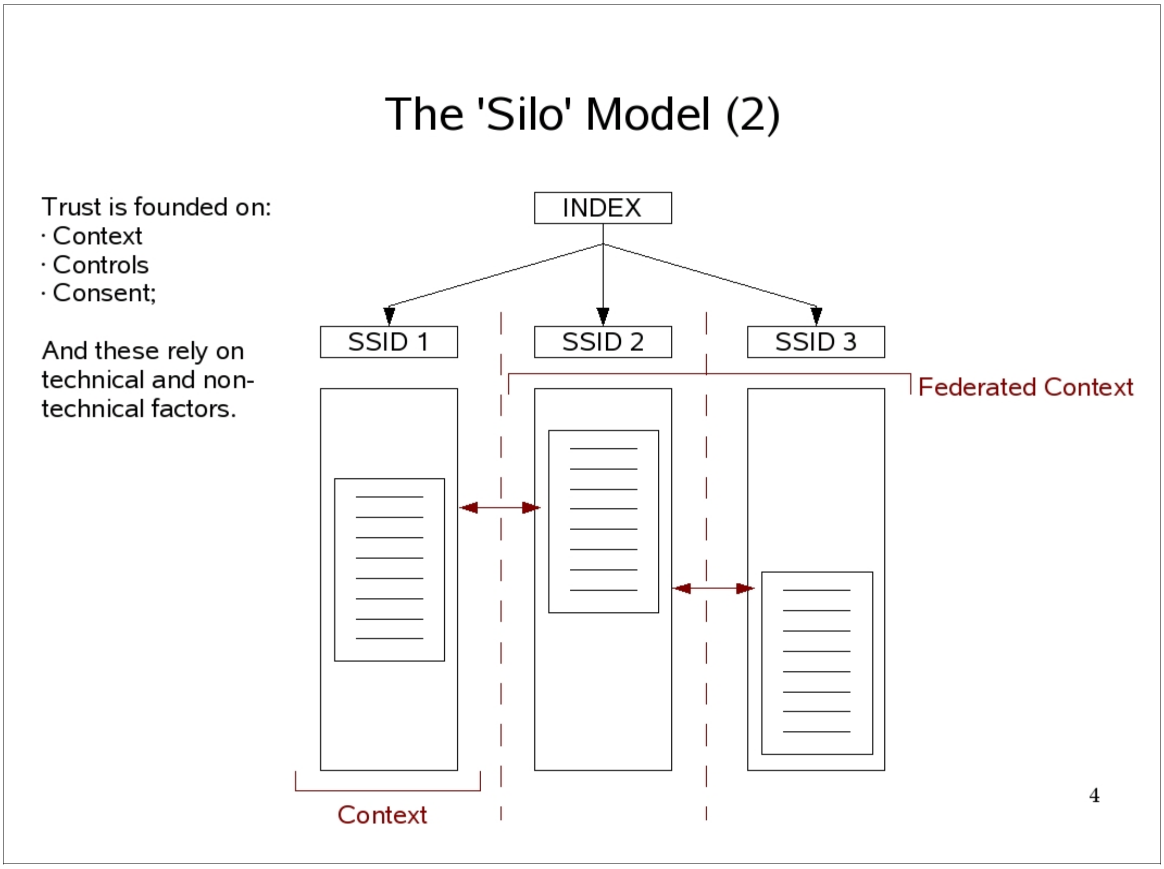
The Austrian example uses specific technical means to enforce a particular relationship between the INDEX and the SSIDs. In other cases, those relationships may be defined and enforced through policy measures or through different federation technologies. Similarly, data exchange between contexts may be controlled through technical or non-technical measures.

The second diagram also illustrates the principle that having a technically-enforced one-way relationship between the INDEX and the SSIDs does not, in itself, guarantee that data between contexts cannot be linked and attributed to the same user. For example, anyone able to search each database looking for a given BIS, or an attribute such as Postal Code, would quickly be able to find the sector-specific records relating to a given individual, even if they did not know the INDEX or SSID for that individual.

Thus, if 'unlinkability' is a requirement, it must be enforced through good data custody practices as much as through any technical means at the SSID level.

If the user is to trust a system such as this, the system must contain adequate (technical and non-technical) measures to deal with context, user consent, and controls over the exchange of data. These measures must be able to cope with different credentials, different levels of trust, and different control mechanisms between, for example, public sector and commercial sector contexts.

Annotated silo model



The 'Silo' Model (2)

Trust is founded on:
· Context
· Controls
· Consent;

And these rely on technical and non-technical factors.

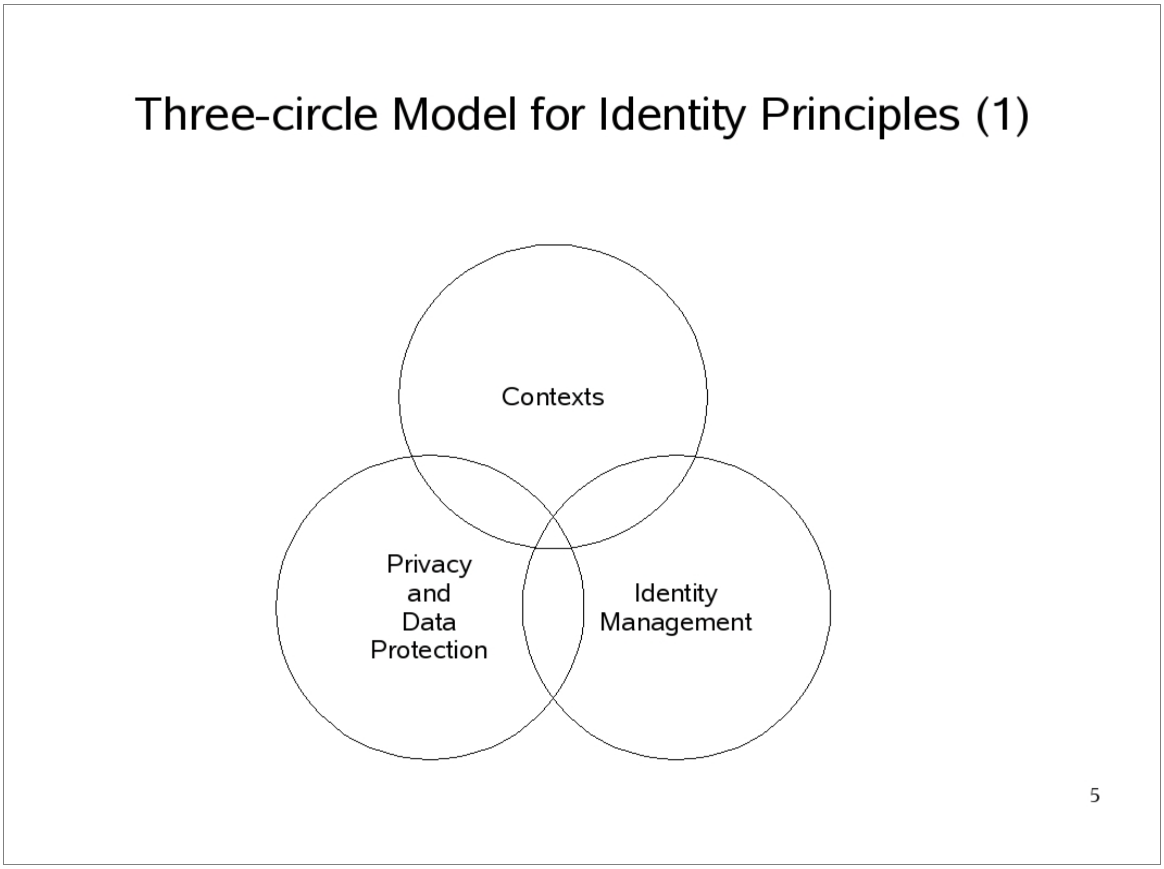
A three-circle model for 'identity in context'

Having used the two preceding models to analyse the characteristics of identity data and identity contexts, we can now take a step back and look at how these fit into a wider model – one which deals with the maturing disciplines of identity management, and relates the technical aspects to legal and social factors such as data protection law and privacy.

How does the technology of identity and identity management relate to non-technical measures such as legislation, regulation and codes of practice, and how do all of these contribute to privacy and data protection in a wide range of commercial, public sector and social-networking contexts.

We derived the following simple three-circle model as a way of situating these major concepts relative to one another.

Basic 3-circle model



Three-circle Model for Identity Principles (1)

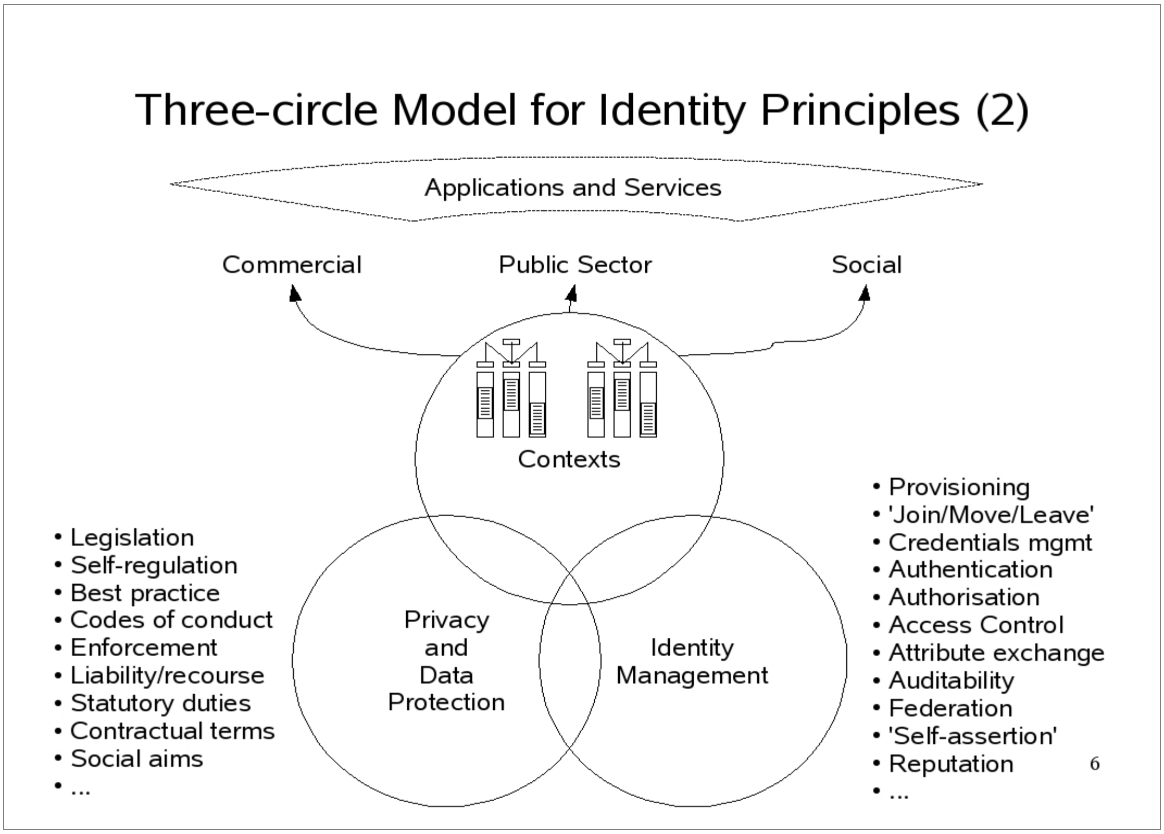
Contexts

Privacy
and
Data
Protection

Identity
Management

5

Annotated 3-circle model



Three-circle Model for Identity Principles (2)

- Legislation
- Self-regulation
- Best practice
- Codes of conduct
- Enforcement
- Liability/recourse
- Statutory duties
- Contractual terms
- Social aims
- ...

- Provisioning
- 'Join/Move/Leave'
- Credentials mgmt
- Authentication
- Authorisation
- Access Control
- Attribute exchange
- Auditability
- Federation
- 'Self-assertion'
- Reputation
- ...

As this version of the diagram illustrates, the 'contexts' circle provides a link back to the previous model describing how contexts and silos may correspond to one another. It also starts to show the key role that identity management disciplines have to play in enabling, managing and enforcing the contexts on which so many applications and services depend.

It is increasingly useful to consider commercial, public sector and social-networking uses of digital identity not just because they are all growing fields, but because there are visible trends towards service-delivery applications which span more than one of these sectors. One of the foreseeable challenges for the identity and privacy community is to design and implement systems which facilitate this while still dealing appropriately with the underlying issues of legal and regulatory compliance and liability.

For instance, in commercial sector use-cases, identity assertions are usually underpinned by some form of contractual liability provision: “if I rely on a credential issued by you, and my reliance turns out to be ill-founded, who is liable for any resulting damage?”. In the public sector, if there is such a liability underpinning it is generally more likely to arise out of statutory responsibilities than contractual terms, and the type and means of recourse are likely to differ considerably from the commercial sector case.

Note: Although not discussed in the workshop, an interesting example of cross-sector identity is to be found in the Swedish BankID[4] system, where bank-issued credentials have successfully been used in support of authentication for public-sector services. (Related link[5] – overview of Swedish e-Government IDM initiatives).

[4] http://www.bankid.com/BankidCom/Templates/StartPage.aspx?id=4&epslanguage=EN

[5] https://www.cosic.esat.kuleuven.be/modinis-idm/twiki/bin/view.cgi/Main/SwedishProfile