Identity, Identifiers and Attributes

"Meet Jean Blue, humanoid living in Centerville. Jean is a real person, an identity. Jean has many attributes, including gender, height, weight, preferred language, capabilities and disabilities, citizenship, voter registration, … Among these attributes are some identifiers – email address, ssn, passport number, membership club numbers, etc. Identifiers are attributes whose values are specific and unique to an individual, rather than attributes that multiple users might have the same value for. "

Within the Identity ecosystem are three important and related but distinct concepts:

Identity, roughly equivalent to a persona, vetted or self-created, that an individual may operate in, and the concept that most ties to the authentication aspects of the ecosystem.

Attributes, the virtually unlimited set of possible values that can be associated with identities, and transmitted via the protocols of the ecosystem. Attributes provide privacy-preserving and scalable access control within the ecosystem.

Identifiers, a particular category of attributes that can provide a variety of linkages to a specific identity. They are the part of the ecosystem that provide critical real-world functionally – uniqueness - to almost all applications and transactions. Identifiers are strings of bits that are used to manage almost every transaction on the Internet. Identifiers are used at all levels of the protocol stack, from the network layer to the application layer and the identity ecosystem interactions It is only through careful use of the proper identifiers that a functional and privacy-preserving ecosystem can be implemented.

Any "identity" (human, device, process) will have many attributes and many identifiers associated with it. The selection of the proper identifier, or set of identifiers, for a use case is one of the most critical decisions in implementations.

While more than a dozen distinct characteristics of identifiers can be discussed, in practice a smaller set of concepts are most relevant to the ecosystem:

Persistence
Reassignment
Privacy preserving with opacity
Human Palatability
Privacy preserving with non-correlating identifiers and unlinkability
Scope of Uniqueness

Persistence

Persistence is a measure of the length of time during which an identifier can be reliably associated with a particular subject. A very short-term identifier might be

associated with an application session. A permanent identifier is associated with its entry for its lifetime (which is not necessarily "forever", so permanence is just a relative notion). Examples of persistent identifiers include social security numbers, passport numbers, eppn, etc.

Reassignment

Many identifiers do **not** specifically guarantee that a given value will always refer to a single subject forever. Reassignment means the association of an identifier value to one subject, and then assigning the same value to a different subject at some point in the (possibly distant) future. Examples of non-reassignable identifiers include social security numbers, passport numbers, the orcid identifier,

Privacy-preserving with Opacity

Some identifiers are designed to preserve a subject's privacy and limit the ability of unrelated applications from correlating activity by comparing values they receive. Such identifiers are therefore required by design to be opaque, and to have no particular relationship to a subject's legal identity or other identifiers. Note that this definition still permits sharing/commonality of the identifier among multiple applications if they are deemed to be equivalent to a single application for privacy purposes. Examples include epTID,

Human Palatability

An identifier that is human-palatable is intended to be rememberable and reproducible by typical human users, in contrast to identifiers that are, for example, the randomly generated sequences of bits in opaque identifiers. There is a natural tension between palatability and both privacy and non-reassignment and they are often in opposition. The world does not have a popular solution to all three problems at once today, which feeds into the oft-noted recommendation that many applications really need to use multiple identifiers for different purposes. Examples of human palatable identifiers includes Display Name,

Privacy preserving with Non-Correlating identifiers and unlinkability

Transactional identifiers can be handed out by an identifier provider in several ways that can provide additional privacy protections. This includes

handing out a different short-term identifier to each web site that an authenticated user visits. This allows the many benefits of a stateful user experience but stops correlation attacks. It does not address liability concerns. Examples include epTID from eduperson schema.

Handing out a different short-term identifier *each time* a authenticated user to a web site. This provides ultimate unlinkability of users actions, but at some impact on user experience. Examples include

The scope of uniqueness

All identifiers must have some degree of uniqueness, within a particular "namespace" in which the identifiers are being created and managed. Sometimes this namespace is explicitly made part of the identifier (as in the case of a "scoped" identifier, see below), in which case the identifier is globally unique. In other cases, the namespace may be implicit, in which case the identifier may not stand alone without the namespace being articulated and stored in some form. This becomes particularly relevant when applications are truly federated (supporting multiple Identity Providers accessing the same data), or otherwise store identifiers in a common way. Because of DNS, email addresses are considered to be globally unique. Generally human names are not globally unique. Display names may not be unique even within a given domain.