**January 2015**

**Citizen-Centric Attributes:**

**A Report from the Internet2 NSTIC Pilot work in Scalable Privacy**

**Context:** The Internet2 Scalable Privacy Initiative, funded in the original NSTIC pilot round, is an attempt to distill and adapt the leading edge efforts in identity within the R&E community to a broader citizen centric context. The particular focus of the work has been privacy at scale. With complex online interactions, numerous interactions per user per day, and the user’s desire to “just have it do the right thing about privacy,” rolling out effective, resilient, scalable, and cost-­‐effective privacy infrastructure will be difficult. Privacy needs to scale with the growing number of users, scale with the number of services users need to manage, and types of applications in which privacy must be managed.  Moreover, there is a growing diversity of users and personal attitudes towards privacy that should be accommodated. These challenges have lead us to focus on foundational elements that are essential to scale and currently missing from the Identity Ecosystem. The Scalable Privacy work, led by Internet2 staff, draws heavily on the expertise of the Internet2 community in identity management, privacy and federation and examines a set of missing components that the identity ecosystem needs for deployment at scale. Many those involved in the work have had long and productive engagements with federal identity work over several decades. Deliverables have included critical software (MFA management for the leading open source federation software, certificate management tools, etc.), business and deployment plans, identification of key schema for inclusion in citizen-centric uses, evaluation of anonymous credential deployment at scale, and end-user consent management tools.

**Background and Methodology:**

The NIST-funded “Scalable Privacy” activity committed to looking at a set of related issues that address attributes and transport mechanisms for citizen to government activity. This space is characterized by a number of distinctive features, including multiple federations of identity providers, an emerging set of privacy requirements, and wide accessibility requirements. The members of the Internet2 community that contributed to this report have had long experience in federal identity approaches, working with GSA activities for over ten years, as well as engagement in current IDESG activities, design experience in broadly adopted community schema, and a commitment to privacy at scale. In addition, some members of the group have worked in some depth with constituent communities, such as the accessibility community[[1]](#footnote-1) and open access interest groups[[2]](#footnote-2).

In addition, the following Scalable Privacy deliverables played a key role in shaping this report:

* The citizen centric attribute registry[[3]](#footnote-3) that was established resulted in interactions with several sectors about their needs and approaches.
* The periodic table of trust elements[[4]](#footnote-4) was another effort that informed this work.
* And, most recently, the ongoing development of PrivacyLens[[5]](#footnote-5), a next-generation consent manager, has elevated a number of concerns about some of the attributes discussed below.
* All of these deliverables can be found on the project website at https://spaces.internet2.edu/display/scalepriv/Scalable+Privacy

The material below is distilled from those discussions and activities.

The work falls into three related categories:

1. Schema – discussions about relevant attributes – and important

 metadata for the C2G use cases

2. Transports – a set of specific flows among repositories for identity,

 attributes and authorization

3. Privacy and Accessibility by Design – a set of good practices and

 principles for developers

**Schema:**

There is a set of attributes that are of particular importance to the C2G relationship. These include a few attributes of multi-sector interest (e.g., name, DoB), as well as specific attributes for which governments are authoritative for and/or consumers of, such as citizenship and voting district. While this set of attributes are of interest across governments at many levels and internationally, there is little coordination on approaches. That may be because current use cases do not cross levels or countries, but it is likely that such cases will come forward and require coordination.[[6]](#footnote-6) Some existing standards bodies, such as IETF and ISO[[7]](#footnote-7), provide normative international codes. There will need to be mappings across such perimeters but there are many unaddressed issues associated with mappings (such as gateways versus distributed approaches, and differences in semantics or syntax.)

Two particular sets of schema were identified for their importance and amenability to federated approaches. These are the accessibility schema developed by GPII (gpii.net) and standardized by ITU (ISO 24751), and a “veteran’s schema”, as defined by DD214[[8]](#footnote-8). That latter schema may be developed as part of another NSTIC pilot activity.

One distinguishing characteristic of the C2G use cases is the need to accommodate multiple identity providers and multiple federations. This has many implications, including the value of an architecture that depends on rich metadata and mapping commonalities in attribute approaches across federations. The metadata enables multiple providers; the mappings handle a manageable set of perimeters between verticals or internationally.

It is likely that a relatively small number of attributes can address the large majority of use cases. While some, such as DisplayName and email address, are defined in normative schema, others, such as over legal age, lack adaptation to an on-line world. [[9]](#footnote-9)

We have created an attribute registry at [https://spaces.internet2.edu/display/scalepriv/Attribute+Registry+Overview](https://spaces.internet2.edu/display/scalepriv/Attribute%2BRegistry%2BOverview) to contain attributes from normative schema that are of relevance to C2G use cases. As noted in the current issues in attributes white paper, such registries are proliferating and, at a minimum, should be linked. See the IETF Draft I-D, on 'requirements on an attribute registry', <http://www.ietf.org/id/draft-johansson-areg-reqs-02.txt>

Conversations are just now starting on metadata about attributes. Examples of attribute metadata include level of assurance, provenance, consent settings, validity period, source of authority, and permissible use. See the accompanying paper on current issues in attributes for more discussion.

**Transports**

Broadly, there are two types of transaction-oriented flows for attributes – real time flows where the user and the RP are currently engaged, and flows where the user is not present. In addition there are “front channel” flows, typically as part of an authentication activity, and “back channel” flows where a RP can acquire additional attributes via a variety of local strategies.

While this discussion is intended to be largely protocol independent, it should be noted that each of the two primary real-time transport protocols – SAML and OpenId Connect – shades the approaches. For example, OpenId has the concept of scope, which roughly resembles the bundles being constructed in the SAML landscape. OpenId Connect has a natural integration with UMA[[10]](#footnote-10); integrating SAML and UMA takes more work but has been done.[[11]](#footnote-11)

It is important to distinguish management of the initial exchange of attributes between an IdP (or AA) and a RP from any subsequent downstream use, and perhaps retransport of those attributes. The technologies that might control downstream use do not yet exist; management of that flow is the space of policy.

Each of the flows has merit for certain use cases, and the project identified touch points where the different types of flows needed a common approach. These touch points include integration across protocols, consistent consent management, similar orchestration, agreements on common payloads and technology and creation of standards that foster identity portability.[[12]](#footnote-12)

UMA, user-managed access, provides a set of important capabilities to the ecosystem, when the user is not present and some downstream controls mechanisms. The project did develop an architectural approach to integrating UMA capabilities into its approaches, addressing which of the touch points merited leverage and how. As mentioned above, there is a working proof of concept that does integrate SAML 2.0 and UMA. It suggests that an IdP run a UMA resource server as part of its IdM infrastructure, and be able to integrate the consent mechanisms in the SAML release to also populate the settings of the UMA server. Thus, a single consistent user experience can manage consent for both real-time and asynchronous mediated release of attributes.

In the complex spaced of electronic health records (EHR), there are challenging needs for a large and diverse number of attributes and consent flows following those attributes across a complex mesh of independent health care providers. In the details, it seems a most daunting use case.[[13]](#footnote-13)

**Privacy and Accessibility by Design:**

There is an increasing body of materials that address privacy by design[[14]](#footnote-14). Few get into the specific topic of federated attribute release or what the transports permit and encourage. We list a number of specific design considerations for privacy.

1. Use of privacy-preserving techniques in federated application development .

Those classes of applications for which known user identities are essential are out of scope. (Most customer service applications are in this category.) However, in those classes of applications where user identity is not essential to the application, there are design alternatives depending on application’s requirements on issues such as the need to retain the users’ settings and the freshness of data. Such applications include content access, opinion polling and voting, marketing, and many others. Some privacy-preserving application approaches include using:

* Opaque user identifiers, when the app needs to provide a persistent user state without a mapping of that state to a known individual. For forensic purposes, the IdP may well have ways of knowing which opaque identifiers map to which user identities. Per-SP opaque identifiers offer a counter-measure against user activity correlation between different relying parties
* Entitlement or ticket style assertions by a trusted IdP, which can convey the right of the (otherwise unidentified) user to a certain level or type of application access.
* Affiliation or Group membership assertions by a trusted IdP may be mapped by the application to a certain class of service eligibility. No identifier for the individual is required in such cases.
* Differentiating required attributes from optional attributes. Applications should be aware of emerging approaches to federated consent infrastructure.[[15]](#footnote-15)

2. Recognize the complexity of consent management and take steps to minimize it. To the degree that bundles emerge as a scalable approach to attribute release, applications should fit their needs to existing bundles.

3. The vision for accessibility, and the architecture, within GPII is comprehensive but still implementable in piecewise fashion. Each of the components of the overall architecture – the schema, the ability to create user profiles by practitioners, the transport and others - are in deployment or development. It is the ability of applications to respond to these attributes with adaptations – in presentation, in cognition, and other dimensions that is the least far along. Tools to facilitate such adaptive application behavior should be encouraged. For example, screen readers and color-blindness adaption tools are available, but tools do not exist for restructuring content into linear flows to help those with cognitive disabilities. Adaptive application design is challenging but possible.

4. The capability to create bundles as filtered subsets within the GPII attribute set (ISO27341) would allow much easier use of the accessibility approach. It also provides an important new capacity to the ecosystem - The ability to couple accessibility with privacy has not been achievable before. A situation in which GPII-mediated accessibility can be coupled with user privacy include a library patron selecting a hardware token that fits their accessibility profile, plugging it into a library-hosted browser with GPII support and having the content accommodate their accessibility needs. The resource provider would receive no indication of the individual user’s identity. Alternatively, common accessibility profiles could be mapped to entitlement values (as discussed above) thus allowing anonymized access to federated content.

5. A prototype implementation that integrated real-time consent-release and also interacted with an UMA store to provide off-line capabilities would be helpful. The Scalable Privacy Project is hoping to deliver one.

1. Work funded by Department of Commerce NIST Cooperative Agreement 70ANB12D300 (CFDA 11.609)

 See, for example, Raising the Floor at <http://raisingthefloor.org/> [↑](#footnote-ref-1)
2. See, for example, the Council on Library and Information resources at http://www.clir.org/ [↑](#footnote-ref-2)
3. https://spaces.internet2.edu/display/scalepriv/Attribute+Registry+Overview [↑](#footnote-ref-3)
4. https://spaces.internet2.edu/download/attachments/33099874/PeriodicTable\_131108.pdf?version=1&modificationDate=1385562509851 [↑](#footnote-ref-4)
5. https://work.iamtestbed.internet2.edu/drupal/about [↑](#footnote-ref-5)
6. For example, regional government cooperation, particularly involving multiple states, and transnational activities in Europe. [↑](#footnote-ref-6)
7. See, for example, the X.500 attributes at http://www.ietf.org/rfc/rfc2256.txt and other attributes at inetorgperson at http://www.ietf.org/rfc/rfc2798.txt [↑](#footnote-ref-7)
8. http://www.archives.gov/veterans/ [↑](#footnote-ref-8)
9. In a user-IdP-RP relationship that spans multiple jurisdictions, who determines legal age? [↑](#footnote-ref-9)
10. https://kantarainitiative.org/groups/user-managed-access-work-group/ [↑](#footnote-ref-10)
11. https://events.nordu.net/display/NORDU2014/User+Managed+Access [↑](#footnote-ref-11)
12. See https://spaces.internet2.edu/display/scalepriv/Scalable+Privacy [↑](#footnote-ref-12)
13. Among the many issues here are how to present such complex consent to end users, how to bundle the large number of potential attributes into some user-manageable bundles, and how to translate the technical values of health records into human-understandable language. [↑](#footnote-ref-13)
14. See, for example, https://ico.org.uk/for-organisations/guide-to-data-protection/privacy-by-design/ and http://www.research.ibm.com/haifa/projects/imt/privacy/index.shtml [↑](#footnote-ref-14)
15. See the Refeds discussion list ([www.refeds.org](http://www.refeds.org)) and the R&E end-entity tag [↑](#footnote-ref-15)