eduPerson Reporting Code Kick-off Call, 10 July 2020

Attendees:

Michael Gettes, Benn Oshrin, David St Pierre Bantz, Meshna Koren, Ralph Youngen, Chris Shillum, Heather Flanagan

Use case

Institutions ask publishers for a breakdown of their usage reporting in their invoicing. There is a standard called COUNTER that offers guidance on how to do this at an institution level, but customers are requesting that the reports be broken down even more granularly, sometimes to support internal cost allocations, or so they can understand who is using the resources. Different publishers have tried different schemes, including asking for charge back codes, asking the user to select from a pre-defined pick list, etc.
They want to provide a way for an institution to send through a set of reporting codes as part of the authentication transaction, which the publisher will then use to create segmented usage report. The publisher does not need to know what those codes mean or represent, they just need to break down those codes into reports. So, looking for something semantically opaque.

Here's a good example of how granular usage serves the purpose at the institutions, to watch later:https://www.youtube.com/watch?v=FD3wWtP59Bg&feature=emb_logo

Discussion:

Why can't we use entitlements for this? Entitlements are used to support authorization decision. That's a very different purpose than a reporting attribute which would be a pass-through value. Since Shibboleth and SAML are about authorization, it's a question as to whether a reporting code that does not inform authorization makes sense.

If we look at the definition of eduPersonEntitlement, it says it offers a set of rights. We maybe be overloading this attribute if we're not using this information for authorization. There is an assumption that the reporting attribute comes as part of the real-time transaction. Either we end up with a second attribute that looks like eduPersonEntitlement, or we restructure the definition of eduPersonEntitlement to encompass this requested functionality.

Different institutions are currently requesting this in different ways, sometimes by saying to report on email addresses.

What would the value space look like? The needs around the value space are very different. SPs look for a specific value form a known set, and then make an authorization decision. That. Means the SP and the IdP need to agree on a set of values that would note a user is entitled to something. In general, this is a very coarse-grained grouping. With the reporting code, the SP would NOT know what the value would be, and wouldn't care. There is no agreement required.

eduPerson is about defining something so that both parties understand what they're communicating. If it's local and arbitrary, it shouldn't be in a schema at all. We could, however, change the requirement to let us come to a value structure that would let share this information. We would have a URN that has a standard preface, with an arbitrary closing value.

How easy would this be to implement with things like AD, Okta, etc? Their attribute mapping capabilities may be limited. But since the eduPerson Object Class is already defined in those, then we just need the value space defined.

Biggest tension here is if this semantically overloads eduPersonEntitlement, and the workload for smaller institutions to create something entirely new.
University of Alaska worked with ACS to use the following:
name="urn:oid:2.5.4.11" multi-valued organizationalUnitName (local friendlyName "MAUaffiliations")
name="urn:oid:1.2.840.113556.1.4.261" single-valued Division (local friendlyName "primaryMAUaffiliation")

The problem is that would not provide enough granularity for larger institutions.

Using one attribute for two separate purposes does not work (at least in some common cases). It adds significant technical challenge.

If we talked about eduPersonSPData - guidance would be not to use this for authorization.
Value space could be something like "ePSPdata: entityid VALUE". IdP would send that, and then the SP could determine what to do with it. The VALUE can be any multi-valued, non-binary, ASCII data, constructed as the IdP chooses from LDAP or other source. The IdP will have to have internal discussion on if and how to populate that VALUE.

Could we solve this with ePTID? No, because the analytics are done on the SP side to generate the report.

One area this isn't solving for: how to prioritize the multi-values in a case where one side or the other wants to deal with just the primary choice. Michael and David will discuss and come back with a proposal.

Next steps:

- Heather to schedule a second call
- Heather to create a first draft for a new attribute
- Michael and David will discuss how to resolve the use case where either the IdP or SP request information on which value in a multi-valued attribute should be considered "primary" and report back to schema-discuss