

Consultation: eduPersonAnalyticsID



This consultation is now closed.

Background

The REFEDS Schema Editorial Board's subcommittee, the eduPerson Analytics Code subcommittee, has proposed a new attribute to be added to the eduPerson schema that would provide a way for an institution to send through a set of reporting codes as part of the authentication transaction, which SP will then use to create segmented usage reports. The primary use case captures the need of a publisher/library scenario where data is needed to understand the use of a given resource and be able to classify that resource into buckets (such as internal billing codes). This happens outside the authentication/authorization transaction and so is not itself an entitlement.

Overview

This consultation is open from: 4 March 2021 15:00 CET to 5 April 2021 17:00 CET.

Participants are invited:

- to consider the proposed new attribute for the eduPerson schema; and
- to propose appropriate changes / challenges to the proposed document.



The specification is available as a [Google doc in Comment mode](#) or as a [PDF](#). Background on the [eduPerson Analytics Code Subcommittee](#) is available in the REFEDS Schema Editorial Board wiki space. All comments should be made on: consultations@lists.refeds.org or added to the change log below. Comments posted to other channels will not be included in the consultation review.

Following the consultation all comments will be taken back to the REFEDS Schema Editorial Board subcommittee for review and if appropriate the document will then be forwarded to the REFEDS Steering Committee for sign-off and publication on the REFEDS website as per the REFEDS [participants agreement](#).

Change Log

comment #	Line /Reference #	Proposed Change or Query	Proposer / Affiliation	Action / Decision (please leave blank)
1	n/a	I don't like the attribute name - eduPersonAnalyticsID - its the ID part that worries me, its not an 'Identity' per se. I can foresee confusion and people sending almost or real PII in that ID value.... eduPersonAnalyticsTag or eduPersonAnalyticsValue ?	Alan Buxey	The committee agrees to changing this to eduPersonAnalyticsTag
2	"represents the use of a service by a subject"	The spec seems to allow for use as a targeted, privacy preserving identifier. I don't have a strong opinion regarding the name, but as worded I think this attribute could be used to represent an identity. If this is not the intent then use of the term 'subject' may need to be clarified.	Mark Jones / UTHHealth	Language change: "aggregates the use of a service by a set of subjects"
3	8	From the definition I get the impression the identifier intends to identify a service from the perspective of the Institution. Perhaps the name of the attribute should reflect that. My initial association with Analytics is "learning analytics"	Niels van Dijk	We believe this is addressed by the renaming of the attribute to eduPersonAnalytics Tag
4	24	Might be worth clarifying whether it is expected that multiple subjects could have the same value, or giving an example. This paragraph is a little tricky to parse and fully understand	Hannah Short	We will add a sentence to this effect; that combined with the language change described in comment 2 we believe addresses this issue.
5	33	This format would not allow URI's to be used as value, as neither semi-colon nor back-slash are allowed. I think however the URL of the service, or a SAML entityID would actually make very good potential values?	Niels van Dijk	SAML entityID is not terribly useful to be reported in usage reporting systems as mostly all users from one institution will have the same. Also because this is for analytics purposes, values in the form of URIs may be more confusing than not. We will extend the example to offer some clarity on what to do when there are two values.

6	35	For the purposes of the spec it would be better to be specific; either the values are case-sensitive, or they are not.	Meshna Koren	Agree; will make a change to be clear. We do not want people using values that differ by case.
7	41	It should be possible for the home organization to correlate usage of various services. It will make analytics simpler if they don't have to decipher the meaning of different values.	Meshna Koren	These aren't personal identifiers, so correlation between IdPs by the SPs isn't relevant. If an IdP does want to make sure this isn't possible, the IdP needs to be the ones that use different values for each SP. Will clarify the text.
8	"It is RECOMMENDED, if stored in an LDAP or X.500 directory, ..."	After reading "use of a service by a subject" at the beginning of the spec i was thinking about ePTID, and this paragraph suggests to me that the value would represent a unique subject and the service a unique SP. If that is not the intent then some clarification may be needed.	Mark Jones	We will remove this text.
9	"It is RECOMMENDED, if stored in an LDAP or X.500 directory, ..."	There are at least 100 potential ways this data could be stored. Why a specific recommendation of 1 of these?	Niels van Dijk	We will remove this text.
10	"separated by a pipe (ASCII 124)"	I believe this is generally called "vertical bar" outside of *nix shells.	David Walker	Overtaken by events of removing the LDAP text.
11	"FOOBAR_ZORKMID"	Opaque?	Niels van Dijk	it is opaque in that it does not identify the individual. There is no reason it can't be human-readable.
12		Would recommending " " create a limitation in the use of LDAP products (i.e. as it is used as internal separator?)	Niels van Dijk	Overtaken by events of removing the LDAP text.