

Request for Comments

eduGAIN Catalogue Service

Introduction

Our Work

Proposed Curated Metadata

Engaging Service Provider Operators

Demonstration Version

Introduction

The eduGAIN Support Team, Secretariat and Business Development groups have over time received several requests from Service Provider Operators, Identity Provider Operators and NRENs for a Service Catalogue.

The aim of such a Catalogue service is primarily to enable institutions within national federations to learn about the properties of services offered via eduGAIN. It is assumed that by providing information such as attribute requirements, licensing, IPR and data protection, a Catalogue may help institutions to make more informed decisions on allowing access to a service by their users.

Currently, the eduGAIN Entities Database tool and REFEDS Metadata Explorer Tool (MET) are the two services closest to a service Catalogue, but there are inconsistencies in the information presented by the two tools. In addition, the tools take a rather technical view. Research was conducted in Q3 2017 and was built upon within the Trust and Identity Incubator from early 2021.

To provide this enhanced view of services within eduGAIN it was decided to augment the information provided by the SAML metadata with a set of "Curated Metadata". This Curated Metadata would be stored in a central location, updated (curated) by the operators of the services themselves, and be accessed through an API or a simple and lightweight webpage.

Catalogue Service

Basic research on this topic was carried out previously, and in September 2017 the REFEDS Working Group published its paper, describing examples of service Catalogues, technical requirements, best practices, and challenges. It was suggested that an eduGAIN Service Catalogue could be based on the Entities Database and MET.

The Trust and Identity Incubator took the previous work and extended it to take the latest state of eduGAIN; the number of services, increased size of federations and the work some federations had already undertaken to create service Catalogues.

From this research, a knowledge base of data was created, consisting of attributes being used around federations, both within eduGAIN SAML metadata and beyond. A great deal of work has been undertaken to reduce this superset down to a manageable quantity of attributes the Incubator believes is reasonable to ask a Service Provider operator to maintain.

Based on work conducted and discussions with stakeholders, a catalogue service is proposed that holds two sets of related data on a service:

- A JSON representation of data as found in each service's SAML metadata
- A JSON representation of 'curated' data provided through an alternative source

The catalogue service combines these sets on a per-service basis to provide the following capabilities:

- A public, read-only catalogue API presenting an aggregate of metadata and curated data on a per-service basis.
- A shielded catalogue API with read/write capabilities to allow curated data to be provided and made available on the public API.
- A lightweight, white label GUI which may be leveraged by federation operators to present the data from the catalogue service in their national context to create national level catalogues. The (JavaScript based) GUI allows for full branding and deep integration into existing websites by federation operators.
- An instance of the above white label GUI representing eduGAIN services, and branded as such.
- A GUI to allow a service owner to provide curated data.

Proposed Catalogue Fields

The following is taken from the list at

<https://wiki.geant.org/display/gn43wp5/Curated+Metadata+-+the+Proposed+List>.

- **Service name** - Short localised service name; sourced from mdui:DisplayName; in the user's language if available, otherwise, default to English if available, but also offer other available languages
- **Description** - Localised one-paragraph description; from mdui:Description
- **Registrar** - The available registrar; sourced from mdrpi:RegistrationInfo

- **Logo** - Sourced from mdui:Logo; can also be multilingual (if there are some cultural adjustments text in the graphics)
- **Access URL*** - Direct link where the end user may use the service. Source: curated from the service
- Service provider (name)- Sourced from md:OrganizationDisplayName, expandable to md:OrganizationName
- **Service category*** - One or more values from a limited vocabulary list of categories as established by the eduGAIN community. In case category entity-categories are provided, these will be provided additionally on a national level
- **Contacts** - Available values of md:ContactPerson
- **More service info** - Sourced from (localised) mdui:InformationURL
- **Privacy policy** - Sourced from mdui:PrivacyStatementURLSAML attributes - Requested and required attributes from sourced from md:RequestedAttribute and md:RequiredAttribute
- **License info*** - One or more values from a limited vocabulary list of (high-level) licences limited vocabulary as established by the eduGAIN community. Probably best implemented as a binary selector of, e.g., Commercial/Non-commercial or Restricted/Public or Subscription-based/Free type
- **Since** (first seen) - From the available metadata history
- **Last updated** - From the available metadata history
- **CatalogID** - A unique, URL usable, reference to the entity created as a sha1 hash over the SAML entityID

The items marked with * are not readily available from SAML metadata and in some cases their values need to be agreed upon by the community.

Engaging Service Provider Operators

To collect curated data a GUI is provided centrally where federation operators may point their SPs. This GUI will:

1. suggest SAML metadata improvements based on evaluating current SAML metadata of the SP and comparing that to current best practices and
2. allow a service to create and maintain additional curated data fields as listed above.

In addition, federations that currently already operate a catalogue containing curated data, may provide that data on behalf of the services via the catalogue r/w API.

Proposed engagement model:

- Any federation that exports SPs to eduGAIN will point the SP to the GUI to test existing SAML metadata for completeness and ask these SPs to provide the agreed-upon curated data.
- This activity provides a flow and form for that and will record the curated data in the Catalogue.
- Federations collect curated data via the API for their own catalog. If so required, a federation can make use of the white label Catalogue to present this data.

White Label Catalogue (Demonstration Version)

We have put together a demonstration system hosted at GitHub, making use of the built-in “GitHub pages” and CI services, which can be found here:

<https://github.com/surfnet-niels/surfnet-niels.github.io>

Put simply, the system takes the SAML metadata for each entity in eduGAIN, creates a SHA1 hash of the entity_id and builds a precise copy of the metadata as a JSON block. Then, should curated metadata exist for the same entity_id - through matching SHA1 hashes - the two datasets are combined and made available to anyone in the `output` directory of GitHub, above. The re-loading of eduGAIN and curated metadata is controlled by a CI flow which can be found in `.github/workflows`.

With metadata created, it is possible to view it using a variety of methods. For the Incubator, an implementation of the lightweight, single-page JavaScript application we propose can be found here: <https://surfnet-niels.github.io/>

The HTML is built such that it is easily styleable (CSS) and functionality could be extended by any group wishing to do so.

Hello Catalog!

The screenshot shows a web interface for a service catalogue. At the top right, there is a search bar with the text 'cam.ac.uk'. Below the search bar, there are two columns: 'Service Name' and 'Service Provider'. The main content area displays details for a service with the URL 'https://www-falcon.csx.cam.ac.uk/shibboleth'. The details are organized into sections: 'catalog ID:' (4cb3e5373ea9d2e0c57468d7e3d26803b5702697), 'entityID:' (https://www-falcon.csx.cam.ac.uk/shibboleth), 'Service Name:' (https://www-falcon.csx.cam.ac.uk/shibboleth), 'Service Provider:' (University of Cambridge: Falcon CMS), 'Personal Data:', 'Description:', 'Contact information:' (Support contact: Falcon Support (falcon-support@uis.cam.ac.uk), Technical contact: Falcon Support (falcon-support@uis.cam.ac.uk)), 'More information:' (with external link icons), 'Privacy policy:' (with external link icon), and 'Entity categories:'. A vertical scrollbar is visible on the right side of the details panel. The text '(ANAM' is partially visible at the bottom right of the interface.

For an end-of-sprint demonstration in early June 2021, the system was demonstrated and you can find a recording of that demonstration here :

https://wiki.geant.org/download/attachments/247431187/SC_demo.mp4