

# RFC: Service catalogue service functional and technical design

## Introduction

The aim of the service catalogue service is first and foremost to enable institutions within national federations to learn about the properties of services. It is assumed that by providing information like for example attribute requirements, licensing, IPR and data protection and other legal information, such catalogue information may help institutions to make more informed decisions on allowing access to a service by their users.

On a technical level, the solution should be as simple as possible allowing federations with fewer resources to make use of the catalogue capabilities. At the same time, federations that already have a catalogue should also be supported.

The intent is to have a service acting as an 'exchange point' for federations to exchange curated information on services. Representing the information is considered the responsibility of a national federation, although a lightweight client implementation will be made available.

## Assumptions

- SAML metadata is considered leading, the catalogue does not in any way act as a replacement for the SAML metadata.
- The information in the catalogue is considered to be informational and is no replacement for proper checking of the licence and/or contractual obligations between parties.
- All information in the service catalogue is assumed to be public information.
- The service catalogue does not store personal data as part of the public information
- Services may either be national or international, where for the latter it is assumed that they are connected via eduGAIN.
- The catalogue will provide curated information beyond the technical data currently available via SAML metadata. A detailed list of what information may be provided is listed later on in this document.
- The catalogue supports multilingual information on a per-service basis. It is left up to the federation to mandate specific languages. It is strongly advised for services exported to eduGAIN to support at least English.

## Audience

Institutions  
SPs

## Curating service data

The catalogue will use information from SAML metadata from national federations and eduGAIN as its basis. It will however also provide additional information which is curated separately. One might be inclined to mandate adding such additional data to the 'regular' metadata. In reality, however, the capability of service owners to properly fill their SP metadata is limited and would require support from many, many vendors, both in and outside of the R&E community. That is not a feasible strategy if we want a working catalogue in the next 5 or so years.

The major challenge for a curated service catalogue is to collect the service information. It is proposed to make the service owners responsible for maintaining this information. Whether that is on a voluntary or mandatory basis is left up to the federation who is the primary registrar for a given service.

To support the federations in curating this information and to solicit the help of the service providers, the federation may make use of a web-based interface that enables a service provider to create and maintain the service information data. The aforementioned web interface makes sure only the owner of the service as registered in SAML metadata (either administrative contact or technical contact) is eligible to create and maintain this data. The initiative to invite Service owners to create and update information lies with the registrar.

If the federation already has an existing means for curating service provider information from the service owner, it may use the service catalogue service API to present this data to the service catalogue service.

All curated data is stored in a JSON format in a public Git repository, with entityID as key. For convenience, a copy of the information represented in a service SAML metadata is also provided in the JSON format, so both may be easily retrieved from the catalogue. The latter cannot be modified via the catalogue and is periodically refreshed in an automated fashion.

## Architecture

## Retrieving data of a catalogued service

The SAML entityID is the leading identifier for SAML based services. However, to use it conveniently in the catalogue, a derived identifier is created by creating a SHA1 hash over entityID, sha1EntityID. The catalogue service API may be queried for service information based on this identifier, by visiting:

- <https://catalogueURL/{sha1EntityID}>

The catalogue will respond to the request with per service information presented as a JSON array containing information on 1 entity.

If no entry exists for the requested entity identifier, the catalogue will respond with HTTP 404 "Not Found".

## Bulk access to data

To facilitate federations in retrieving all known additional information about the service in their national federation, they may request the catalogue to list all information on known entities as presented in an XML metadata URL.

- <https://catalogueURL/{metadataURL}>

This will request the catalogue to look up information for all known entityID values in the metadata URL and respond with a JSON array containing multiple entities.

As a request using metadata URL may be costly in terms of catalogue resources, the catalogue may cache its responses for up to 24 hours.

## Language support

Depending on published content  
UI in English