Organization Information gone Wild: ROR, Entity IDs and The Organization Ontology

Marius Politze[1*]
[1] RWTH Aachen University, Germany
politze@itc.rwth-aachen.de

Abstract

While building services for individuals from academia, uniquely identifying a person is a challenge that was widely addressed in several contexts like eduGAIN. Sometimes, alongside the "who?" information systems also need reliable information about the "from where?". During the past years several alternative standards came up to tackle that problem from different directions. In this paper we would like to introduce some of them: Research Organization Registry (ROR), eduGAIN Entities and The Organization Ontology and give an opinionated overview of how they can work together.

1 Motivation

For several years we have developed IT systems that support business processes. In many cases these systems would support a set of processes in a single facility like a university or other research institution. Only rare rarely allowing access across organizational boundaries and if only based on existing cooperative projects. With the dawn of new initiatives spanning across organizations like EOSC on international or NFDI and NHR on national level more and more facilities are becoming a service provider on a national or international level. The imminent question is how to provide authentication of external users. Within the academic context, Géant established eduGAIN, a multi federation approach that provides the necessary interfaces, infrastructure and implementations for decentral authentication.

Apart from user authentication projects may also require to obtain affiliation information from the users. For federated information systems we were naturally looking for a persistent and globally unique identifier that can be used to refer to an organization. In the project Coscine, we utilized DFN-AAI (the German sub-federation of eduGAIN) for Authentication and ROR for identification of research organizations and faced exactly the problem of synchronization of both catalogues.

[*] https://orcid.org/0000-0003-3175-0659

2 A Wild Organization ID appeared!

While there is no claim in that direction, the eduGAIN “Entities Database” (Géant, 2020) fulfils a lot of criteria for a catalogue of academic organizations: machine readable, curated in a distributed manner, provide a unique, relatively persistent ID and a reference to their logon interface definitions. For most IdPs there are additional information like organizations’ names or websites. Mapping users’ origin to Entity IDs is obviously embedded into the logon flow. With about 4200 entries worldwide the catalogue is also quite extensive. What else would we ask for? By design the “Entities Database” only contains currently active IdPs: Organizations that do not have an IdP are not listed, also eduGAIN is somewhat limited to public academic organizations. For example the “German National Library of Science and Technology” (TIB) is not available even though they certainly are a valuable part of the academic landscape, so are many other international private or commercial research bodies.

This is where the Research Organization Registry (ROR) comes into play as it claims to be “a community-led registry of open, sustainable, usable, and unique identifiers for every research organization in the world” (ROR, 2021). ROR offers a machine readable data dumps and an API for almost 100.000 research organizations including commercial and academia related nonprofit organizations like DFN or Géant themselves. Hence ROR offers the more extensive catalogue and additionally allows cross linking with identifiers e.g. to GRID or Wikidata. However, at the time of writing is missing the cross link to the Entity IDs from the eduGAIN “Entities Database”.

3 It's dangerous to go alone!

For our current project at hand, we decided to use ROR for top level research organizations while still using eduGAIN during the authentication process. Currently, this requires mapping between ROR IDs and Entity IDs to get the membership relation of a user to an organization. Both approaches also fall short modeling the internal structures like departments. Having, Entity IDs and ROR IDs, in our project we need a way of aligning the organizational information from both.

The data model of Coscine is based on RDF therefore using W3C standard “Organization Ontology” (ORG) was a natural choice. ORG is designed “to enable publication of information on organizations and organizational structures” (W3C, 2014). Every organization from ROR is modeled as a “org:FormalOrganization” using the ROR ID as the primary ID and assign the “org:identifier” for Entity IDs. Using ORG provided the additional benefit that it also allows modeling internal structures using organizational units and memberships. This essentially allows the data model of Coscine to provide a W3C standard compliant interface for internal structures while retaining compatibility with the existing catalogues from ROR and eduGAIN.

4 Just the Begin of the Journey

Bringing together ROR and EDUGAIN Entity IDs seems like a viable choice for connecting both directories. An initial brief test showed that aligning both directories using names and institutional websites yield high quality results. For authoritative records this would, however, need a more decent quality control.

Within Coscine we were able to bring both systems together and add internal organizational structures and organization memberships in a W3C conformant information model using ORG. Our future work in this area will entail to build a repository of machine readable organization information. Allowing academic and research organizations to maintain their own internal structure and reuse information from others using an open source inspired, collaborative model.

5 References

Géant. (2020). *eduGAIN Entities Database*. Retrieved 03 10, 2021, from https://technical.edugain.org/entities

ROR. (2021). *ROR*. Retrieved 03 10, 2021, from https://ror.org/

W3C. (2014). *The Organization Ontology*. Retrieved 03 10, 2021, from https://www.w3.org/TR/vocab-org/

6 Author biographies

Dr. Marius Politze is head of the group "Process and Application Development for Research" at the IT Center of RWTH Aachen University. Before that he held various posts at the IT Center as a software developer, software architect and as a teacher for scripting and programming languages. His research focusses on Semantic Web and Linked Data architectures for distributed and service-oriented systems in the area of research data management.