

Technical Considerations for an Organization Identifier Registry

Martin Fenner, Laura Paglione, Tom Demeranville, Geoff Bilder
Version 1.0, 31 October 2016

Summary

Organizational identifiers are needed to help solve the affiliation use case in scholarly communication, i.e., which research outputs are produced by researchers affiliated to a particular institution. Organizations need to be involved in changes to the organization identifier and associated metadata, including splitting or merging of organizations, and information about sub-organizations. Organizational identifiers must follow established best practices for persistent identifiers, including the linking to other organization identifiers and other resources in the metadata. This paper was prepared jointly by Crossref, DataCite, and ORCID and summarizes technical use cases for an organizational identifier system, and our understanding of priorities based on community consultations carried out over the course of the past year.

Approach

Given that organization identifiers for scholarly organizations have been around for some time and have arguably not reached critical traction in some important areas, we highly recommend an approach that keeps the initial implementation focussed by concentrating on the most important use case, on organizations relevant for currently active researchers, and on information about sub-organisations needed for the main use case.

Use Cases

We identified a number of use cases within our own organizations, and in conversations with various stakeholder groups over the past few months:

Membership and Credentialing: Organization identifiers are needed to be able to effectively manage membership accounts, correctly express the name of a member organization, merge accounts when an organization is acquired (or split when divested), and connect that identifier into both financial/invoicing systems and the API credentialing process.

Subscriptions and article-processing charges (APCs): Organization identifiers are needed to manage subscriptions of scholarly content, or content that otherwise is associated with payments, e.g. APCs.

Attribution: Organization identifiers should be able to unambiguously associate a research output with an institution, either directly or using the contributors to the work as proxy.

Attribution of publisher/repository: Organization identifiers should allow to unambiguously identify the publisher of a resource.

Assertions: Organization identifiers to be able to record the source of assertions, such as: this university with this name and organization identifier, asserts that this person with this ORCID identifier have a specific role relationship (e.g., employee or student).

Self assertions and usability: Individuals may also make self assertions, so the identifiers need to be associated with human-readable metadata including alternate names and abbreviations to enable user selection from a pre-populated list.

Matching Organization Identifiers: Organization identifiers need to allow cross-referencing to identifiers used in other instances to understand that it is the same organization.

Prioritization

The use cases for organization identifiers described above fall into these broad categories:

- Affiliation
- Authentication
- Internal

We have prioritized these use cases based on perceived need by the community, as well as anticipated difficulty in implementing them.

Affiliation

Unambiguous Affiliation information is the main use case for ORCID, Crossref and DataCite. This would enable proper description of relationships between contributors, contributions, funders, publishers, and funders. For ORCID “discovery” also includes disambiguation from alternative names.

The complexity of describing affiliation information for researchers increases dramatically when going further back in time, while at the same time the potential benefits of linking research outputs and their contributors to institutions decrease. For these reasons, we suggest that it makes sense to initially focus on current researchers and their affiliation at present and in the recent past.

Authentication

Authentication for access to research resources, and for tracking access by different parties, is an important use case for organizations providing subscription content, but is far less important for freely available research resources. Authentication use cases tend to have high risk profiles and tend to be some of the hardest and most expensive to meet. For these two reasons (highly relevant to only a subset of stakeholders, expensive to implement) authentication use cases are seen as lower priority use cases, and it is conceivable that authentication use cases are covered by another service. There is scope for existing federated identity infrastructure to adopt organisation identifiers in a way that meets their specific use cases in the future, which would enable cross-walking between the two systems.

Internal

Internal use cases such as administration of membership, or disambiguation of internal organization information, are important for many research organizations. Given that the numbers of organizations that are impacted by this use case is relatively small, the return on investment on internal applications of an organisational identifier likely would be modest. One possible scenario would be to use the organization identifier as linking identifier to other organization identifiers, such as that from GLEIF, would provide coordinated information to accommodate internal use cases.

Updating Organization Information

We expect that in most cases the organizational information will not come from the organization itself, in particular if the organizational identifier system is seeded with pre-existing organization information. At the same time we see it as an essential requirement that there should be a mechanism for organizations to add and change information about themselves in the system. How this is implemented can change over time as the service grows, but it should be clear from the start that this will be possible. An organization should always have the ability to control the information about itself associated with the organization identifier, including name variants.

We do not see that an organizational registry system must only support input and updates by the organization itself (which would be similar to how the ORCID registry operates), but rather suggest that information about organizations should be managed by a combination of updates by organizational identifier provider staff and self-updates by the organization. This will result in a dataset that is a hybrid of curated and self asserted information, which should be made distinct through provenance metadata.

Organizational Hierarchies

Most organizations consist of sub-organizations (departments, institutes, etc.), and managing these organizational hierarchies is seen as one of the biggest challenges for implementing an organization identifier. Three open questions are, (i) the depth of the organizational hierarchy that should be represented, (ii) who is the party that can best assert these relationships, and (iii) how to express where in the hierarchy the organization expects connections to be made. We assume that in general an organization should know this information best and can provide it to the organizational identifier provider, but more work needs to be done to clarify the needed workflows, including permissions.

Organization Identifier as Linking Identifier

While the initial focus is on solving the affiliation use case, it is important to keep the other use cases in mind going forward. One important feature of the proposed organization identifier is therefore the ability to link to other organization identifiers, such as those used for authentication.

General PID best practice

We expect organization identifiers to follow the persistent identifier best practices that have been established for other identifiers (e.g., DOI, ORCID ID), where appropriate. Organization identifiers should for example be:

- globally unique,
- resolvable,
- expressed as HTTP(S) URIs,
- associated with appropriate metadata, and
- durable.