Project Name	**FREYA**
Project Title	**Connected Open Identifiers for Discovery, Access and Use of Research Resources**
EC Grant Agreement No	**777523**

D4.4 Organizational IDs in Practice

Deliverable type	Report
Dissemination level	Public
Due date	30 November 2019
Authors	Christine Ferguson (EMBL-EBI, orcid.org/0000-0002-9317-6819)
	Simon Lambert (STFC, orcid.org/0000-0001-9570-8121)
	Manuel Bernal Llinares (EMBL-EBI, orcid.org/0000-0002-7368-180X)
	Frances Madden (British Library, orcid.org/0000-0002-5432-6116)
	Robin Dasler (DataCite, orcid.org/0000-0002-4695-7874)
	Martin Fenner (DataCite, orcid.org/0000-0003-1419-2405)
	Artemis Lavasa (CERN, orcid.org/0000-0001-5633-2459)
	Chris Baars (DANS, orcid.org/0000-0002-5228-1970)
	Tina Dohna (PANGAEA, orcid.org/0000-0002-5948-0980)
	Ketil Koop-Jacobsen (PANGAEA, orcid.org/0000-0002-1540-6594)
	Dan Morgan (PLOS, orcid.org/0000-0003-0625-2770)

Abstract	This deliverable demonstrates how the various partners in FREYA are taking up persistent identifiers for organizations as part of the Work Package 4 work on integrating emerging PID types into disciplinary contexts. A range of organization IDs are discussed with special focus on the ROR ID as a community-led initiative with open infrastructure and data that is well suited for use in an open science environment.

Status	Submitted to EC 12 December 2019

FREYA project summary

The FREYA project iteratively extends a robust environment for Persistent Identifiers (PIDs) into a core component of European and global research e-infrastructures. The resulting FREYA services will cover a wide range of resources in the research and innovation landscape and enhance the links between them so that they can be exploited in many disciplines and research processes. This will provide an essential building block of the European Open Science Cloud (EOSC). Moreover, the FREYA project will establish an open, sustainable, and trusted framework for collaborative self-governance of PIDs and services built on them.

The vision of FREYA is built on three key ideas: the **PID Graph**, **PID Forum** and **PID Commons**. The PID Graph connects and integrates PID systems to create an information map of relationships across PIDs that provides a basis for new services. The PID Forum is a stakeholder community, whose members collectively oversee the development and deployment of new PID types; it will be strongly linked to the Research Data Alliance (RDA). The sustainability of the PID infrastructure resulting from FREYA beyond the lifetime of the project itself is the concern of the PID Commons, defining the roles, responsibilities and structures for good self-governance based on consensual decision-making.

The FREYA project builds on the success of the preceding THOR project and involves twelve partner organisations from across the globe, representing PID infrastructure providers and developers, users of PIDs in a wide range of research fields, and publishers.

For more information, visit www.project-freya.eu or email info@project-freya.eu.

Disclaimer

This document represents the views of the authors, and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright Notice

Executive summary

This deliverable corresponds to a specific task in FREYA's Work Package 4 concerned with integrating emerging persistent identifier (PID) types into disciplinary contexts. It builds on approaches developed in WP3's work on new and emerging PID types and shows how various FREYA stakeholders (disciplinary partners, PID service providers, etc.) are taking up organization identifiers and integrating them into their existing services or developing pilot applications that take advantage of organization IDs.

A variety of new and mature organization identifiers are presented with a focus on one consolidating organization PID, the ROR ID (Research Organization Registry identifier). This deliverable reports on the work and current status of integrations of organization Identifiers by FREYA partners and underlines advantages and challenges that the partners encountered through this process. These lessons learned and overall considerations regarding the implementation of organization PIDs, both in similar scenarios (such as those described in this report) but also in the context of the EOSC, are presented at the end.

As a result of this work, more FREYA partners have organization IDs integrated as part of their infrastructure (that previously had no PIDs of this type), while others (that already had other organization IDs) chose to also include ROR IDs; ROR IDs are now being the predominant type of organization ID in partner implementations. The work on integrating new and emerging PIDs is essential for creating more connections and thus adding more value to the disciplinary PID graphs that are developed as part of this Work Package, and will continue on future deliverables in order to further build on this with more PID types.

Contents

1 Introduction

The need for organization identifiers—that is, persistent identifiers that uniquely identify formally organized and recognized institutional bodies—has long been acknowledged[1]. Clearly the organizations are real-world entities rather than digital objects, but that does not diminish the importance of identifiers. An especially important use case is that of reliably capturing *affiliations*: being able to relate individual researchers (with their ORCID iDs) and research outputs (with their DOIs or other identifiers) to the organization (university, research institute or whatever it might be) that employed them and under whose auspices the outputs were produced[2].

Organization identifiers were already discussed in FREYA's deliverable D3.1, as part of the survey of the PID landscape[3]. The discussion began with this succinct account of the situation:

> "The community currently has the ability to assign identifiers to published content and to individuals, however, a missing piece is a comprehensive, open, and accessible organization identifier infrastructure that identifies affiliations and is independent of a particular organization identifier service provider's business needs. While there are many examples in this space there is no single solution that meets the broad needs of the research community".

In this context, organization identifiers were judged to be "emerging" in terms of maturity.

The development and widespread adoption of organization identifiers would allow the creation of rich PID graphs (one of the fundamental motivators of FREYA) by filling in the layer of affiliation and enabling applications built on it. A university might wish for a reliable way of identifying its research output for purposes of assessment, and a funding body might require the same information. If the process relies on free-text fields filled in by individuals at some stage in grant application or publication workflow, the process is so much more difficult and error-prone.

Organization identifiers are analogous to identifiers for persons: rather than resolve to a digital object or landing page, they point to a centralized registry. FREYA's deliverable D2.1[4] discussed practices for resolution of persistent identifiers; in those terms, organization identifiers require a "single-service resolver". The creation, maintenance and sustainability of such a service is then a key question for any approach to organization identifiers.

However, some particular difficulties arise with organization identifiers that are not found with identifiers for individuals. Organizations evolve over time: they dissolve, new ones are created, they merge or divide, they change their names, etc. Maintenance is a problem: both how to make updates, and who has the authority to do so. What of historical records - what is the correct way to handle historical affiliations to an organization that no longer exists?

This deliverable corresponds to a specific task in FREYA's Work Package 4 concerned with integrating emerging PID types into disciplinary contexts. In essence, it shows how the various partners in FREYA are taking up persistent identifiers for organizations and integrating them into their existing services; or developing pilot applications that take advantage of organization IDs, showing how these can enrich and extend the utility of PID graphs.

---

[1] Fenner, M., Demeranville, T., Kotarski, R., Vision, T., Rueda, L., Dasler, R., … Cruse, P. (2015). D2.1: Artefact, Contributor, And Organisation Relationship Data Schema. Zenodo. https://doi.org/10.5281/ZENODO.30799

[2] Laurel, H., Bilder, G., Brown, C., Cruse, P., Devenport, T., Fenner, M., … Smith, A. (2017). ORG ID WG Product Principles and Recommendations. Figshare. https://doi.org/10.23640/07243.5402047

[3] https://www.project-freya.eu/en/resources/project-output

[4] Wimalaratne, S., & Fenner, M. (2018). D2.1 Pid Resolution Services Best Practices. Zenodo. https://doi.org/10.5281/ZENODO.1324300

Before describing these applications, however, this deliverable will present a concise overview of a variety of already-existing organizational identifiers, and focus on one currently under development: the ROR ID, Research Organization Registry identifier, which has emerged as a result of community consultation (2016–2018) alluded to in this project's description of work (Annex A) and holds great promise as a consolidating PID.

2 Varieties of organizational identifiers

2.1 ROR

ROR is the Research Organization Registry[56]. It is an open identifier for organizations (Creative Commons Public Domain 1.0 International Licence for ROR IDs and metadata). It is meant as a "linking identifier," meaning that it does not aim to replace GRID, Ringgold, ISNI or others. An organization in the ROR registry can have any and all of those other IDs included, depending on the organization's function.

A ROR ID is an identifier with metadata. There is a registry of these identifiers, and it is possible to look up organizations and their identifiers in the registry. ROR does not link to other resources (researchers, publications, datasets, etc.). Thus, the crucial point is what other organizations do with this information.

ROR released an "MVR" (minimum viable registry) in January 2019, with all data coming from the regular GRID data dump, available under a CC0 waiver. The MVR includes an API and a search web interface. Essentially, since January ROR has been ready for use, and subsequent improvements have been concerned with improving the registry and developing conveniences. This includes streamlining the import of GRID data releases, and ROR has since been updated multiple times with the latest GRID data dump.

From a policy/community standpoint, ROR has also soft launched a statement of support[7] that organizations can sign.

In terms of including ROR IDs in the metadata for DOIs that are registered, DataCite and Crossref already provide support or have concrete plans. DataCite members or repositories can add ROR identifiers to the affiliation information of individual authors, or to organization authors in DataCite DOI metadata. This is possible since DataCite Metadata Schema release 4.3[8], released in August 2019. Crossref will include ROR in the next version of their metadata input schema which is scheduled for release in 2020. Crossref members will be able to attach ROR IDs to affiliation information, as well as to organizations (such as publisher names) wherever they appear in Crossref metadata. The urgency for this work was underlined at Crossref's annual meeting in November 2019, where a diverse group of Crossref members and metadata users recommended that ROR top the list of Crossref's areas of strategic focus. Support for ROR IDs as affiliation identifier in DOI metadata of other DOI registration agencies is not available yet.

Note that ROR has been developed during the time of the FREYA project and with the participation of FREYA partners DataCite and Crossref. Specifically, FREYA funds have supported DataCite's activities as a member of the lean implementation group that launched the "MVR" and in subsequent work as a member of the ROR steering committee to ready ROR IDs for adoption by community stakeholders (see FREYA deliverable D3.2, chapter 3.4)[9].

---

[5] ror.org

[6] www.github.com/ror-community

[7] http://tinyurl.com/ror-supporters

[8] DataCite Metadata Working Group. (2019). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data v4.3 (Version 4.3). DataCite. https://doi.org/10.14454/7XQ3-ZF69

[9] https://doi.org/10.5281/zenodo.3554196

2.2 GRID

GRID (Global Researcher Identifier Database) is a service launched by Digital Science in 2015. The service is available via web interface[10], API, and regular data dump[11]. The database is available with a Creative Commons Zero (CC0) waiver, allowing reuse without restrictions. There is no fee to register or update an organization. The latest GRID release describes 97,900 organizations.

GRID was initially developed internally to serve the use cases of Digital Science organizations such as Altmetric or Figshare. The main use cases focus on affiliation information, and the organizations included in GRID reflect this use case. GRID data are used in the Research Organization Registry (ROR), and the latest GRID release includes the corresponding ROR identifier in the metadata. The main reasons GRID was picked as the foundation for ROR include the open license, the high matching rate in disambiguation testing (better than organization identifiers that also include other organizations besides research organizations such as LEI), and the support by Digital Science to help with this work. ORCID has integrated GRID as supported identifier for affiliation information in the ORCID registry.

2.3 ISNI

The International Standard Name Identifier (ISNI) is an ISO standard conceived to identify the public identity of persons and organizations. Its scope is to identify entities in the production lifecycle of intellectual and artistic content including authors, cartographers, composers, performers and researchers as well as organizations. ISNI was conceived from the library community. The Governing Board of Directors is made up of founding members of the ISNI International Agency. ISNIs are used not only by many libraries but other organizations such as YouTube use them to reliably identify persons and organizations represented within them. There are approximately 700,000 organizations in the ISNI database. GRID/ROR databases have about 100,000 organizations, indicating that ISNI is broader in scope.

ISNIs are primarily created by registration agencies and the database is maintained by the ISNI Quality Team at the British Library and the Bibliothèque Nationale de France. The database is free to search and read. At present there are several methods for assigning ISNIs, contact can be made via email to an ISNI registration agency. A portal is under development which will allow individuals or organizations to request ISNIs for a minimal fee. This portal will also allow holders of ISNIs to add information to their records but changes will be managed via an email request. In order to assign an ISNI to an organization its name must be unique and not only consist of initials and certain essential information must be available including organization type, a full UN/LOCODE[12] a URI or URL referring to more information about the organization. In early 2020 a data dump of the database for people and organizations will be made available as RDF/XML and JSON-LD under CC0 as well as individual record download.

The quality level of ISNIs is very high due to the detailed requirements for assignment, the high level of curation records receive and their accuracy. They have also recently convened two consultation groups representing the music industry and research organizations to better represent the needs of these areas. However as an organization ISNI can be somewhat opaque in relation to its governance and it can be challenging for users to understand how to have ISNIs assigned apart from becoming a member or registration agency which requires a financial commitment.

---

[10] https://grid.ac

[11] The latest release from 6 October 2019 is available at https://10.6084/m9.figshare.9963719

[12] http://www.unece.org/cefact/locode/service/location

2.4 Ringgold

Ringgold is a registration agency for ISNI (International Standard Name Identifier), which means that it is possible to obtain an ISNI through Ringgold. It is possible for individuals or organizations to request ISNI for a fee of £25; the fee per ISNI is reduced when a batch of ISNIs is required.

Beside ISNIs, Ringgold manages Ringgold Identifier (Ringgold ID or RIN)[13]. This is a persistent numeric unique identifier for organizations in the publishing industry supply chain. Ringgold's Identifier Database includes almost 500,000 Ringgold IDs representing organizations and consortia who acquire scholarly publications and content. The Ringgold ID was introduced in 2003. Ringgold developed it in response to an issue raised by Oxford University Press, namely how to identify institutional subscribers unambiguously. The system is owned and administered by Ringgold, Inc. and Ringgold, Ltd., who also publish a taxonomy for classifying the subject interests of the listed organizations.

Ringgold services, including the lookup service, are not open. Ringgold does permit limited access after simple registration whereby the lookup service is available for seven days, with a maximum of 10 searches a day. Ringgold does not operate with an open license nor do they offer a data dump of Ringgold IDs, and thus are not amenable for use in Open Science.

2.5 LEI

The Legal Entity Identifier is a 20 character alpha-numeric code based on an ISO standard that was proposed by the G20 after the global financial crisis in 2008, to provide clear and unique identification of legal entities participating in financial transactions around the world. The scope of organizations indexed is thus much broader than scholarly organizations and to date LEI records number over 1.5 million. LEIs and their database, the Global LEI index, are maintained and governed by the Global Legal Identifier Foundation. GLEIF provides a free search tool (no registration required) for looking up LEIs[14].

While ORCID supports the use of LEIs to assert affiliation in researchers records[15], the LEI is not broadly used as an organization identifier in scholarly research. The reason for this is that many research organizations do not fulfil the criteria for a legal identity. For example DANS, the FREYA partner, is not a legal identity but part of the Royal Netherlands Academy of Arts and Sciences.

2.6 Crossref Funder ID

Crossref's funder ID is a DOI assigned to individual organizations that provide grants or awards. Organizations range from government agencies and large international funders to small single mission charities or private agencies. The publisher Elsevier donated the funder registry to Crossref in 2012 and continues to update it monthly with new funders. The registry, containing over 21 000 funders[16], is maintained and curated by Crossref who makes the identifiers and associated core metadata available under a CC0 waiver and provides an API for the retrieval and resolution of Funder IDs.

In the long term, the Crossref funder registry corpus will also be assigned ROR IDs and included in the Research Organization Registry.

---

[13] https://en.wikipedia.org/wiki/Ringgold_identifier

[14] https://search.gleif.org/#/search/

[15] https://members.orcid.org/api/resources/orgids-in-orcid

[16] https://search.crossref.org/funding

2.7 PSI OrgID

This is the organization ID used to identify academic institutions within The IPregistry.org[17]. The IP registry was developed by the independent company PSI IPV Ltd., to address the concerns of librarians and publishers over incorrect or misused IP addresses that can lead to use of content without the publisher/library's knowledge. The registry contains 1.5 billion verified IP addresses for more than 60 000 academic institutions. The PSI OrgID is made up of 5 components: 3-digit country code (GBR = Great Britain) category (A = academic), size (S0), type (3 letters: UNI = university or INS = institute) and ends with a unique 6 digit ID#. The IP registry serves to manage and communicate the IP ranges used to authenticate access to content registered by publishers and libraries.

In response to the 2017 Open organization Registry RFI, PSI RPV Ltd was considered as a prospective host for the open Persistent Institution Identifier Registry (openPIIR)[18] since developments made to establish the IPregistry.org meant many of the desired requirements for an openPIIR were met. Relevant features included the PSI orgID, a set of naming standards, curation facilities allowing Libraries to amend their own records following free registration, and take-up by key publisher and library stakeholders[19].

Ultimately, ROR emerged as a result of the work of the openPIIR group[20].

2.8 Wikidata identifier

Wikidata[21] is a project, driven by Wikimedia Deutschland, that was launched in October 2012 within the Wikimedia Foundation.

It's an open knowledge database focused on entities, e.g. topics, concepts, objects, people, etc. Every item in this document oriented database has an associated wikidata identifier (or QID) that is unique, and makes the item language independent, as well as not imposing any restrictions on any property of the items, e.g. their name. It is possible for two objects in the knowledge base to have the same value for this property, but refer to different elements.

Wikidata is a common source of structured data for Wikimedia projects like Wikipedia, released under a public domain license, and it also offers a query service based on SPARQL that was rolled out in September 2015.

Before this query interface, a statement-based mechanism allowed not only data access but also the interconnection of item properties within the knowledge base.

Their curation model is community-based and it can be made either anonymously or via a platform account.

---

[17] https://www.psiregistry.org/theipregistry

[18] https://orcid.org/content/2018-org-id-meeting

[19] https://orcid.org/sites/default/files/ckfinder/userfiles/files/PSI-Organisation%20Identifier%20Project%20-%20RFI%20Presentation%20v3.pdf

[20] https://orcid.org/content/organization-identifier-working-group

[21] https://www.wikidata.org/wiki/Wikidata:Main_Page

3 The FREYA position

It is not the role of FREYA to insist on the adoption of any particular persistent identifier for a given entity. FREYA is concerned with best practice, of course, but in this area the project's influence is exerted indirectly, through its work on PID policies, providing a scope and starting point for the EOSC governance bodies to develop and promulgate policies. It is likely that the policies that emerge will concentrate on principles, desired results and governance, though they might make recommendations for the use of mature technologies.

Nonetheless the FREYA partners consider that they can endorse and recommend ROR as a solution for organization identifiers. ROR has several advantages over other identifiers:

- it is underpinned by open principles
- it is a community-led project[22]
- its scope is appropriate for the type of organizations that will require identifiers in the PID Graph (ISNI for example has a much broader scope, while LEI requires that the organization be a legal entity in itself)
- it can subsume other more specialized registries such as those of funding bodies

The FREYA consortium (through the coordinator) has signed the public ROR statement of support undertaking to "support the ROR identifier and its future" and pledging "to adopt and use ROR in our applications, tools, workflows, and communications".

Several FREYA partners are involved in the development of ROR, but this does not mean that the project as a whole adopts an uncritical view of ROR. This deliverable is not intended to be an "evaluation" of ROR—it is less than 12 months since its launch and a definitive evaluation would be premature—but the diversity of applications does give some indication of ROR's strengths and weaknesses. Specific lessons learnt are given in the conclusions at the end of the document.

---

[22] These two points are compatible with the ideals of the PID Commons envisioned by FREYA, summarised in the project's Description of the Action as "defining the roles, responsibilities and structures for good self-governance based on consensual decision-making"

4 FREYA activity with organization identifiers

4.1 EMBL-EBI

4.1.1 Europe PMC

- *Background*

Prior to initiating the pilot integration described here, Europe PMC (a database of biomedical literature) included no organization IDs in its records[23]. The author affiliations are included as free text, which can result in misspellings and ambiguity. Figure 1 shows an example of the disparate results obtained when searching for Europe PMC records affiliated to a specific organization.

Figure 1 Results of the affiliation matches for EMBL's European Bioinformatics Institute (EMBL-EBI) captured in Europe PMC's records. Highlighted in red are the affiliations that are recorded for the EBI, and in blue are other EMBL related records that increase the ambiguity when searching for publications by affiliation.

---

[23] A small exception to this is that we store Crossref Funder IDs in our GRIST database for the 29 funders of Europe PMC: http://europepmc.org/grantfinder

● *Quick read summary of planned work for this deliverable*

To decrease the ambiguity of search results retrieved for EMBL-EBI, a minimal viable implementation was planned for Europe PMC whereby the ROR ID for EMBL-EBI is added retrospectively to a corpus of publications produced by authors at EMBL-EBI from 2016-2018.

The first step was to add an "org-ID" field to author records for publications indexed in Europe PMC's database to allow for input of persistent identifiers of any type for an organization, e.g. ROR IDs, GRIDs, ISNIs, Wikidata. Next, it was necessary to Identify EMBL-EBI-affiliated publications in Europe PMC: to this end, we retrieved a list of DOIs for EMBL-EBI publications (2016-2018) manually curated at EMBL-EBI with the aid of a CRIS system. This list was matched against DOIs held by Europe PMC. Where publications listed authors with an EMBL-EBI affiliation via free text, the ROR ID for EMBL-EBI (ror.org/02catss52) was included in the "AffiliationOrgID" index field associated with the author. Finally, this field was exposed to users of Europe PMC's APIs or website.

● *Timeline*

The index field (affiliationOrgId) & the data (ROR ID linked EMBL-EBI affiliations on 737 publications) is now stored in Europe PMC's articles database. This was exposed for users of the webservice in October 2019 via the core response of the APIs: SOAP searchPublications and REST search methods. Changes to the database search functionality on the website are estimated for December 2019.

● *Use case to be addressed*

ROR IDs have been developed to address the affiliation use case: "to unambiguously identify which organizations are affiliated with which research outputs"[24]. A stakeholder in this use case is the Communications Team at EMBL-EBI who have to manually curate a list of publications produced by staff each year.

To begin to address this use case, Europe PMC sought to assign the EMBL-EBI ROR ID to (the staff members who are authors on) the publications collated for 2016-2018. For relevant records with an unambiguous EMBL-EBI affiliation, the ROR ID has been assigned as part of the author record and exposed to users as demonstrated in Figure 2 (API core response) and Figure 3 (search results using the Europe PMC website).

```
<author>
        <fullName>Lafita A</fullName>
        <firstName>Aleix</firstName>
        <lastName>Lafita</lastName>
        <initials>A</initials>
        <authorId type="ORCID">0000-0003-1549-3162</authorId>
        <affiliation>
        European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome
Genome Campus, Hinxton, Cambridge, UK.
        </affiliation>
        <affiliationOrgId>ror.org/02catss52</affiliationOrgId>
</author>
```

---

[24] https://doi.org/10.6084/m9.figshare.8137961.v1; also https://github.com/datacite/freya/issues/51

Figure 2 Exposure of the affiliation information to users of the API: An example of how the information will appear in the core response of the API. Note that the <affiliation> field containing free text is now accompanied by <affiliationOrgId> field containing the ROR ID for EMBL-EBI.

Figure 3 Searching the Europe PMC database for articles authored unambiguously by EMBL-EBI researchers, reveals 737 results (red lozenges). The publications that can be unambiguously attributed to EMBL-EBI authors per year are highlighted by the blue lozenge and thus serve to address the use case for the EMBL-EBI communications team.

- *Questions arising*

It was not possible to assign ROR IDs to all articles on the lists that were curated for EMBL-EBI from 2016-2018—a proportion of articles list a non-EMBL-EBI affiliation for staff on the publication indicating that staff members having conducted their research at EMBL-EBI, have moved on to another organization by the time the journal article is published. So from the affiliations listed in the publication(s), we can't identify the EMBL-EBI author to whom to assign the ROR ID.

EuropePMC indexes publications from multitudes of journals and publishers in the biomedical sciences that capture affiliation information according to their individual publication needs and policies; these may include affiliations for where the work was conducted, affiliations indicating contact details for an

author that are current at the time of publication to aid publisher billing processes or to provide contact details for readers.

In the case of the current EMBL-EBI effort, the aim is to identify the affiliation where the work was done, not the current affiliation of an author. Deducing the affiliation via the researcher as intermediary means that matching issues such as the one above will arise when retrospectively adding ROR IDs to Europe PMC records. Therefore, in the case of retrospective addition of organization IDs to literature records, it is worth investigating how to provide a link between publication and research organization(s), rather than publication and researcher.

● *Lessons learned through implementation*

Europe PMC now has a database index field in which to store affiliation information in the form of organization IDs for machine reading. This represents the next step in expanding the PID Graph built around EuropePMC's literature-indexing services. Retrospective incorporation of ROR IDs for a subset of EMBL-EBI publications means that a search is now possible using a single ROR ID for publications with authors affiliated to EMBL-EBI.

Ideally ROR IDs are added prospectively to research articles by authors ahead of publication, so that the verified information can subsequently be incorporated in databases like EuropePMC. However, neither Europe PMC nor ROR can dictate which research affiliations will be captured and included in literature records by journal publishers. It is worth highlighting to the publishing community the importance of capturing the affiliation that shows where the work was done, to be able to address the use case of a research institute wanting to generate a list of publications produced by staff for a particular year. However, it will be a challenge to align publishers on this so that they ensure consistent practice around affiliations. This issue is discussed in other sections of this report, notably the sections from the publisher, PLOS, and from DataCite.

4.1.2 Identifiers.org

● *Background*

Identifiers.org holds information on the different resources that are registered as providers of information for the different ID spaces listed in its central registry.

Prior to 2019 this information used to be a simple line of free text, that was supposed to capture the most relevant details regarding an institution, where a resource was affiliated. No organization IDs were part of the registry.

● *Quick read summary of planned work for this deliverable*

Upon the release of the new Identifiers.org cloud native platform, many changes have been made to include information like organizational affiliation of resources, and geographical location (following ISO 3166-1).

As a consequence, we now have a more structured view of the metadata regarding affiliation for resources, and this will trigger a curation update registry-wide, to complete organization information with their corresponding ROR IDs, where possible (Figure 4).
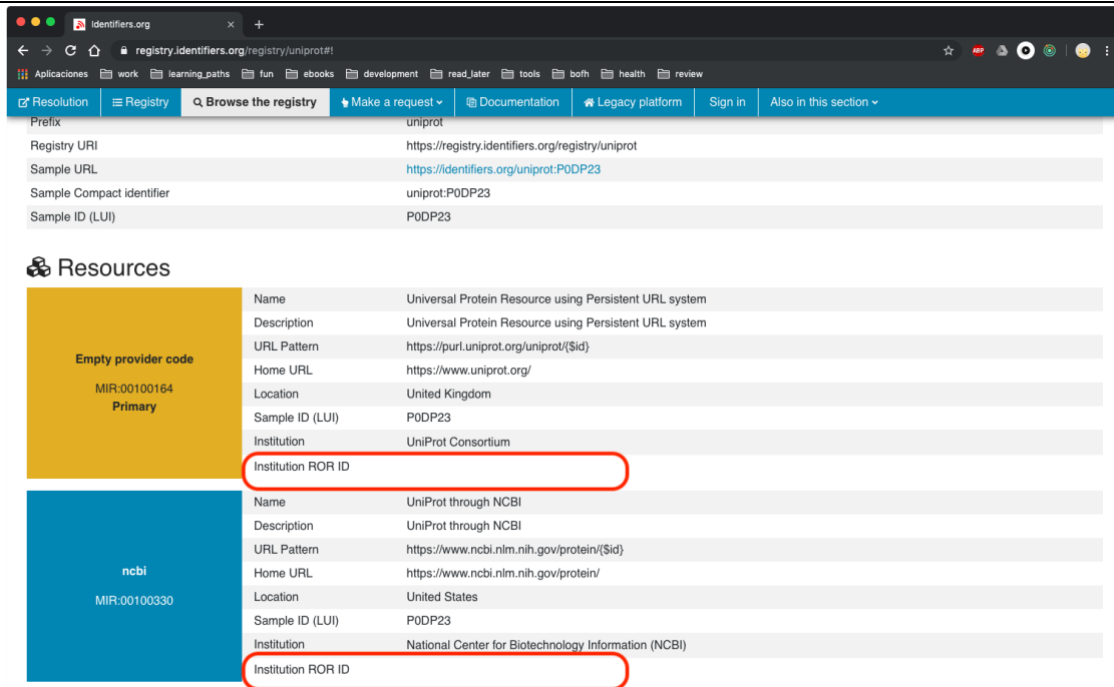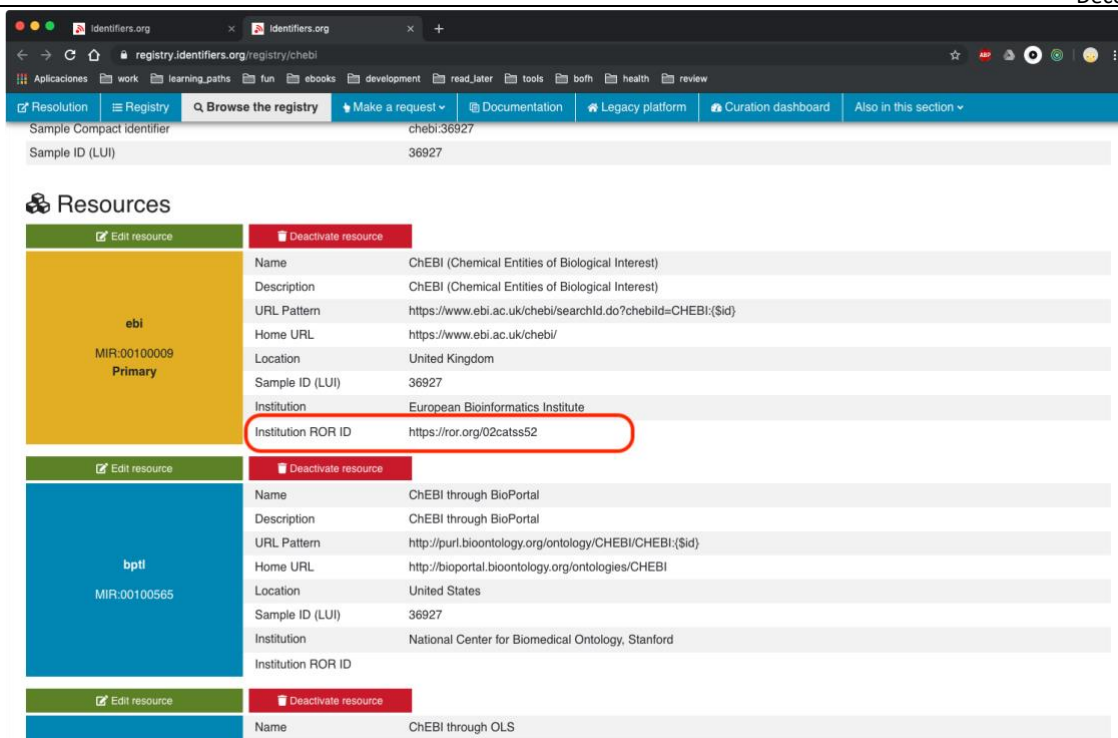
Figure 4 This is an example of a registry entry where we need to complete the ROR ID information for the institution associated with the provider.

In addition, ROR ID metadata model transformation and access at UI level will be made available to our community to aid in the interaction with Identifiers.org services, especially when submitting prefix or resource registration requests.

- *Timeline*

API communication and data model transformations are already implemented in the backend of Identifiers.org, as well as the needed elements on the Web frontend for handling ROR IDs in prefix and resource registration requests.

ROR IDs are also shown, when available (Figure 5), as part of the resources information when browsing the registry, and they are available through the REST API.

Figure 5 This is an example on how ROR IDs are shown to the users when present in the registry.

The necessary integration tests for these new features were completed by the end of September 2019.

- *Use case to be addressed*

Organizational affiliation information is required when submitting prefix and resource registration requests. By using ROR IDs, we are able to aid the user in providing this information to us (Figure 6).

The user provides a ROR ID in the affiliation section in the registration request form, and the Identifiers.org organization data model will automatically populate this information from either Identifiers.org Central Registry, or ROR API.

Figure 6 This is an example of how the ROR ID, provided by the user while filling in a prefix registration request, has triggered an auto-fill of the rest of the institutional fields in the form.

● *Questions arising*

In Identifiers.org we recognize the value of ROR IDs within FREYA as part of the federating mechanism, that aims for a seamless journey through the PID Graph by navigating between the different providers' APIs.

Because of the nature of Identifiers.org mission, the rationale for integrating ROR IDs has been clear, especially spotting where it would be of immediate benefit for our community, but also as it enhances interoperability ("I" in a FAIRness context).

However, the recent emergence of ROR IDs, compared with the pre-existing ISNIs, Wikidata and GRID, leads to concerns over their adoption, long term.

● *Lessons learned through implementation*

Integrating ROR IDs into Identifiers.org was reasonably easy: we keep information about organizations in our registry, to which the different resources are affiliated. From the point of view of our data models, a ROR ID is just another property of an institution. The interesting interoperational question was how to use that new attribute, and the ROR API.

When looking at the ROR API documentation and the "how to" guidelines[25], there was initial confusion about which part of the URL corresponds to the actual ROR ID. According to the API instructions, if we wanted to retrieve information about ROR ID "https://ror.org/02catss52" (The European Bioinformatics Institute), we would make an HTTP GET request to the API to the URL: *"https://api.ror.org/organizations/https://ror.org/02catss52"*. However, trying other variants that do not follow this instruction give the same results: *""https://api.ror.org/organizations/ror.org/02catss52"*

---

[25] https://github.com/ror-community/ror-api

and *"https://api.ror.org/organizations/02catss52"*. It turns out that the ROR API understands the ROR ID in different formats for convenience.

We encountered a URL redirection issue during our implementation: when we retrieve information from the API for a specific ROR ID, we implement a data model transformation that allows us to use that information within our backend and frontend. One of the attributes included in the ROR ID associated information, is the home URL for the organization, in the case of the EBI, it is listed as "http://www.ebi.ac.uk". This URL is flagged as invalid by Identifiers.org, yielding an HTTP Response Code 301 "Moved Permanently" due to the EBI web services being moved to HTTPS. Note that using this URL in a browser works, since the browser will follow the redirection. However, the type of redirection is key for Identifiers.org for programmatic processing and validation of URLs: Identifiers.org accepts HTTP 302 Found, as it is a common mechanism for URL beautifying, but when a URL returns HTTP 301, it signifies that the resource has permanently moved elsewhere (indicated by the "Location" in the header), that the URL is invalid and should no longer be used. The URL redirection issue is not unique to ROR IDs but also encountered elsewhere.

Currently there is no curation mechanism built into ROR itself. The advice is to feed corrections back to ROR via email[26] to have these added manually. ROR plans to establish a curation board in 2020 to consider requests and prioritize them for implementation.

In terms of the data modeling for institutional information, we found ROR granularity more or less aligned with the interests of Identifiers.org. However, by only modeling top level institutional information it is currently superficial and may not suit stakeholders requiring greater granularity to identify organizational subunits like departments, teams, etc.

Although ROR IDs have emerged very recently relative to existing organization IDs, feedback from early implementations such as those described in this report, are being shared in the ROR community calls. The functional advantage of ROR is therefore currently manifest in its community governance which enables feedback to be heard and acted upon with the aim to build a high performing organization ID that best serves the community.

4.2 STFC

- *Background*

STFC has integrated GRID identifiers in the PID graph built around PhD theses and related entities that is one of the pilot applications[27], and is working on incorporating ROR identifiers in the same graph, too. This relates to PhD researchers who either conduct experiments on STFC large-scale facilities or are sponsored by STFC in monetary form with a studentship grant, and identifies the university that hosts the PhD student.

Other applications of organization PIDs have been discussed in STFC, specifically for indexing bibliographic records with unambiguous authors' affiliations, but there is currently no demand from the institutional repositories to adopt such a practice.

- *Quick read summary of planned work for this deliverable*

Organization identifiers play a key role in the STFC pilot applications, enabling representation of affiliation of researchers and indirectly of research outputs. GRID identifiers are currently being used, with work in progress on incorporating ROR identifiers.

---

[26] info@ROR.org

[27] Description in FREYA Deliverable 4.1 "Integration of Mature PID Types" (https://doi.org/10.5281/zenodo.2414838)

- *Timeline*

It is expected that development of the PhD case will be completed by the end of 2019. It is the intention to take forward this pilot application of the PID graph in conjunction with the stakeholders in STFC.

- *Use case to be addressed*

The main use case addressed is affiliation of researchers and research outputs - providing the means to reliably identify the institution with which the researcher is employed or associated and to which the outputs should be attributed. The second use case concerns provenance of the records of science.

- *Questions arising*

The issue of changes over time is a particularly problematic one for organization identifiers. In particular, it is unclear whether and how to propagate new practices that involve organization identifiers back in time to cover the existing records of science created years ago.

- *Lessons learned through implementation*

Organization identifiers enable enrichment of PID graphs with promising possibilities; however potential users might not immediately perceive the benefits compared with apparent costs (e.g. of introducing new processes to ensure organization identifiers are captured).

4.3 PANGAEA

- *Background*

The implementation of ROR PIDs as metadata components of PANGAEA datasets is the next step in our expansion of the PID graph built around the PANGAEA data publication services. The expansion provides new standards for PID usage in research data handling. PIDs are envisaged to play a primary role in the EOSC and the PID Handler is playing a key role in EOSC as a class of services to generate and resolve PIDs. PANGAEA's ROR implementation may be used as a demonstrator for best practices in an EOSC perspective exemplifying the ROR implementation for the EOSC PID Handler services.

- *Quick read summary of planned work for this deliverable*

PANGAEA will implement ROR organization IDs with their published datasets. ROR IDs will be used to identify organizations which are directly linked to datasets and therefore reported with organization name in the metadata. In these cases, the metadata will be upgraded to display organization name and ROR ID. This implementation will not include linking organizations to the authors of the dataset, as the affiliation between author and organization may change over time and we maintain this as an updated list. Furthermore, for funder identification, PANGAEA will continue to use the Crossref funder registry and the associated IDs. Consequently, ROR IDs will not be associated with funders. Entities which are funders and research organizations in one, have a double entry in PANGAEA, one with the ROR Id and another with the funder ID.

Figure 7 A PANGAEA dataset with a Research Organization included in the title of the data publication, in this case the Alfred Wegener Institut (AWI). In the implemented version the "AWI" in the title will be a link resolving to the record in the ROR registry. Also visible in the ROR registry are other identifiers linked to the research organization.

For this deliverable, PANGAEA reports on the ROR ID implementation. This will demonstrate how PANGAEA will take advantage of the consolidated PID infrastructure that ROR is providing to expand the PANGAEA PIDgraph.

The ongoing work has was started by query and affiliation matching between the PANGAEA and ROR registries (Figure 7). The created file also contains columns like "score", "matching type", "chosen", "substring" which all relate to affiliation matching. Chosen (True or False) is an indicator of whether the score is high enough to consider the organization correctly matched by affiliation matching. This indicator is provided by the ROR API.

In our case, only 43% of id's received by affiliation matching are classified as "True" by the ROR API. However, several instances were discovered, where the ROR was labeled as false, even though it was indeed correct (Figure 8). The experience drawn from our ROR implementation will be provided as feedback for the ROR community.

	A	B	C	D	E	F	G	H
		qROR_id	aROR_id	chosen	score	matching_type	substring	ROR_name
	0		https://ror.org/032e6b942	True	1.0	ACRONYM	AWI	Alfred Wegener Institute for
	1	https://ror.org/04ers2y35	https://ror.org/04ers2y35	True	0.92	COMMON TERMS	Bremen University	University of Bremen
	2	https://ror.org/01xtthb56	https://ror.org/01xtthb56	True	1.0	PHRASE	University of Oslo	University of Oslo
		false negative h2x0161	https://ror.org/02h2x0161	True	1.0	COMMON TERMS	IFM-GEOMAR	GEOMAR Helmholtz Centre f
	4	https://ror.org/01y9bpm73	https://ror.org/01y9bpm73	False	0.86	PHRASE	UniversitÃ¤t GÃ¶ttinge	University of GÃ¶ttingen
	5	https://ror.org/04v76ef78	https://ror.org/04v76ef78	True	1.0	PHRASE	Christian-Albrechts-Un	Kiel University
	6	https://ror.org/01se4f844	https://ror.org/002gsek34	False	1.0	ACRONYM	IES	Institute for Employment Stu
	7		https://ror.org/00wkygr69	False	0.62	FUZZY	Marine Geology Depar	Maine Department of Marine
	8	https://ror.org/02dvf9b44	https://ror.org/02dvf9b44	True	1.0	PHRASE	Heidelberger Akademi	Heidelberg Academy of Scien
	9	TypeError occured while read	https://ror.org/050draa26	False	0.48	FUZZY	n_g not_given	NRG Oncology
	10	KeyError - query API error	https://ror.org/05g3dte14	True	1.0	PHRASE	Florida State University	Florida State University
	11	https://ror.org/013meh722	https://ror.org/013meh722	True	1.0	PHRASE	University of Cambridg	University of Cambridge
	12		https://ror.org/01nrxwf90	True	1.0	PHRASE	University of Edinburgl	University of Edinburgh
	13	https://ror.org/00hqnxt08	https://ror.org/00hqnxt08	True	1.0	PHRASE	Murmansk Marine Bio	Murmansk Marine Biological
	14		https://ror.org/05j7rz205	False	0.59	COMMON TERMS	Halifax Halifax	Halifax Health Medical Cente
	15	KeyError - query API error	https://ror.org/000pdyj29	False	1.0	HEURISTICS	Wisconsin-Madison Ur	Madison University
	16	KeyError - query API error	https://ror.org/00hj8s172	False	1.0	COMMON TERMS	Lamont-Doherty Earth	Columbia University
	17	https://ror.org/03wt3br81	https://ror.org/05f0yaq80	True	0.93	COMMON TERMS	University of Stockholn	Stockholm University
	18	https://ror.org/00wge5k78	https://ror.org/00wge5k78	False	0.82	HEURISTICS	TromsÃ¸ University	The Arctic University of Norw
	19		https://ror.org/03zbnzt98	True	1.0	PHRASE	Woods Hole Oceanogr	Woods Hole Oceanographic
	20	KeyError - query API error	https://ror.org/035a68863	True	1.0	PHRASE	Geological Survey	United States Geological Surv

institutions_Affil_match

Figure 8 Example excerpt from registry matching results via API. An example of false negative scoring is highlighted as this was encountered in several instances.

Considering that less than half of the organizations in our registry could be matched to ROR IDs via API, more work will flow into manual matching between registries.

4.3.1.1 Timeline

This implementation of ROR IDs in PANGAEA datasets was started at the beginning of September 2019 and is ongoing. Work, including mapping ROR IDs against the list of organizations that are part of metadata records, should be completed by the end of November. This will be a first step at PANGAEA to include organizational identifiers in the front end metadata records. We are currently discussing whether to include ROR IDs in other parts of the metadata records (e.g. "Coordinating Institution" in the "Project" metadata—see Figure 9).

Figure 9 PANGAEA editorial system "Project" input mask. To link organizations to data publications resulting from research projects, we are considering adding a ROR identifier field to the input mask and consequently allow for search filters based on this.

4.3.1.2   Use case to be addressed

ROR IDs have been developed to address the affiliation use case: "to unambiguously identify which organizations are affiliated with which research outputs". Stakeholders in this use case are organizations or individual users trying to filter their dataset search in PANGAEA according to organizations.

This is a very general use case and we focus most on populating the PID landscape with organization identifiers so that any use case involving organizations as an endpoint or intermediate identifier can be realized via Research Graph APIs involving PANGAEA records (via DataCite or schema.org).

4.3.1.3   Questions arising

Many questions remain in regards to how best to include the ROR identifiers in relation to published datasets. Since PANGAEA applies the DataCite metadata schema, which already provides a field for ROR identifiers, we could theoretically include our organizations there. However, research output is often a complex interplay of individual researchers doing research, while unrelated organizations may provide infrastructure, and still others may provide support as coordinating organizations. Simply including a ROR ID in the Datacite metadata schema cannot map the complexity of these relationships.

Mapping our registry against the ROR registry was successful but uncovered some issues that may affect infrastructures who may also wish to map long-term curated organization registries. Matching success rate was less than 50%, producing the need for more work on our side in manual matching and result checking. Due to the ROR approach of confining identifiers to entities residing on the higher levels of organization architecture (e.g. identifying institutes, but exclude departments within organizations) we are also consolidating some of the granularity of our registry if applying ROR IDs only. It is questionable whether a clear and consistent policy can even be applied in this regard considering the very fluid and complex structures within research organizations.

4.3.1.4   Lessons learned through implementation

ROR IDs have emerged from a community of PID providers who would like to consolidate the many identifiers within their records. Providing ROR IDs for the larger community, is a very important step in trying to consolidate the heterogeneous landscape of organization PIDs and enabling resource holders to link these resources to the responsible organizations. However, how to apply PIDs in the context of dataset publications or journal publications, or more generally devising rules for how organizations should be credited and referenced in the context of other linked resources is an additional aspect that will take more work. These are aspects of PID assignments that will gain traction with increasing uptake and also in the context of research graphs and EOSC services related to PIDs for organizations.

4.4 British Library

- *Background*

The British Library's repository has been developed with support for ISNIs as organizational creators. This had a beta launch in November 2019 with a second phase of development work in early 2020.

- *Quick read summary of planned work for this deliverable*

The intention is to integrate ROR, ISNI and possibly other organization identifiers into various fields in the newly developed repository platform based on Samvera Hyku. The repository is multi-tenanted with several other UK Independent Research Organizations, including the British Museum and Tate, availing of the repository system. The fields which have been identified are: Organizational Creator and Contributor, an affiliation identifier for Personal Creators supporting DataCite schema v4.3, and a lookup to the Fundref database will be included to populate the funder field described in Figure 10 for an article record and Figure 11 for a dataset. The lookup for the affiliation identifier will be managed through the depositor selecting the institution with which the output is affiliated.
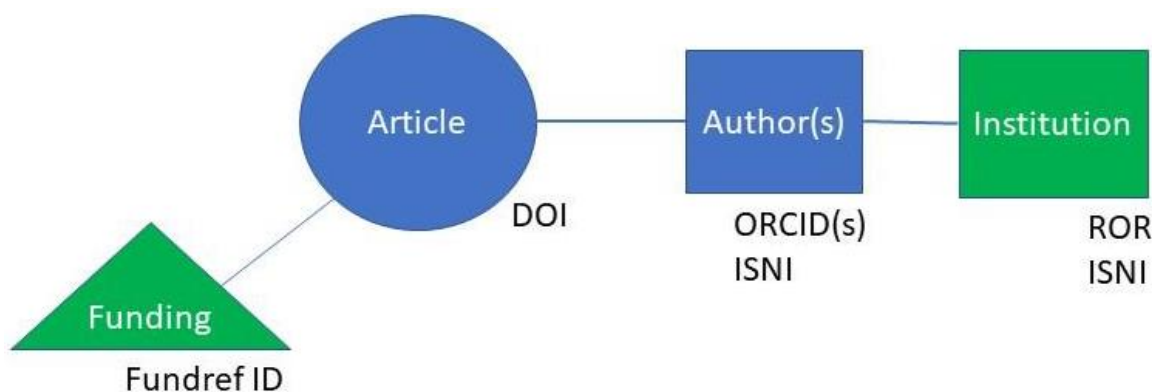

Figure 10 A diagram indicating how identifiers will be integrated into the metadata of an article in the British Library's repository. The items in green indicate where the new identifiers will be added.

The fields have been identified for matching to the repository but the actual implementation has not been fully scoped with a development partner. There are a number of possibilities for how this could be implemented e.g. a picklist allowing selection of an ID type containing multiple options or a specific ROR field. The plan is to display these identifiers as hyperlinked icons on record front end, currently there is an ORCID bubble on the end user interface where actionable links to ORCIDs profiles are provided for personal creators and contributors, it is hoped to do something similar for ROR IDs.

These developments will be carried over to be implemented within the EThOS (UK Theses) collection at the point of that collection's migration to the new repository platform.[28] Work is underway to enhance

---

[28] https://ethos.bl.uk/

the metadata of the EThOS records with organization identifiers for the current and awarding institution and these will be incorporated during the migration process.

Figure 11 A graph indicating the PID Graph for a dataset in the British Library's repository. The developments to the repository's metadata schema will allow PIDs to be assigned to the elements in green.

● *Timeline*

This will be completed following tender exercise in January 2020 where a development partner will be identified for the next three years. Developments to the repository metadata schema will be prioritized for the first phase of work. The EThOS migration is scheduled to take place beyond the lifespan of FREYA but the initial data matching can take place sooner and will be made available as a downloadable csv file through the repository.

● *Use case to be addressed*

To allow affiliation to be fully identified with PIDs on the BL's repository. This work also addresses user story #35 (from the FREYA user stories collected on GitHub[29]) - to understand the background of a piece of research, especially theses and British Library research collaborations etc.[30]

● *Questions arising*

The main questions which have arisen from scoping this piece of work are, how many types of organization IDs can or should be accommodated and displayed in a single repository record? There is a concern it could get very long and possibly confusing for a user. It is recommended that ROR IDs be included as URLs, and some of the display issue can hopefully be accommodated by the use of icons which link to the ID record. Should ISNIs, ROR IDs, GRIDs all be displayed for every type of organization or should it be restricted to one for ease of human readability? This will probably be addressed by cataloguing guidelines rather than any technical implementation.

How will the multi tenancy repository accommodate affiliation identifiers? The multi tenancy nature of this repository implementation means that affiliation is addressed in a fixed way through selecting the

---

[29] https://github.com/datacite/freya/issues?q=is%3Aopen+is%3Aissue+label%3A%22user+story%22
[30] https://www.pidforum.org/t/linking-people-and-research-outputs-to-theses/113

repository's institution. It is not known if this has been dealt with anywhere else and if it will present issues or not and which identifier should be used for this field.

Another question presented by the EThOS metadata is how to manage defunct former named institutions within the awarding institution and sponsor fields which would require ISNI or another organization identifier rather than ROR identifiers as ROR does not contain former names.

● *Lessons learned through implementation*

Lessons learned from the scoping exercise have highlighted the complexity of managing identifiers within historic data. As ROR develops its curation processes, these will need to be addressed in the future as organizations develop over time.

This use case has also highlighted the issue of the number of identifiers which can be managed within human readable records and how to present them in a way which is both human and machine readable. Some of the early adopters of ROR, such as Dryad[31], have not used ROR on the front end of the service so this will be an example of a different application.

4.5 CERN

● *Background*

ROR implementations for CERN concern the INSPIRE[32] and CERN Analysis Preservation (CAP)[33] services. The former is a core High-Energy Physics (HEP) information system (predominantly literature) and the latter is a closed-access preservation tool for physics analyses (currently in beta).

The only CERN service that has been using organization identifiers is INSPIRE, as GRID IDs have been included into their Institutions database (for some of the records). GRID IDs were matched to approximately 2K records (out of 11.5K in total in the database). Attempts to match all the records have not been made, as it is too much work and it is possible that the INSPIRE Institutions database is simplified in the future to allow for lower granularity and, as a result, better mapping to the GRID (and ROR) database.

● *Quick read summary of planned work for this deliverable*

In CAP, users can enter their affiliation and get back a ROR ID. This is in place for cases where basic metadata (such as authors/persons involved) is not fetched automatically from internal databases, so the user would input such information manually. The user can search for their organization in the ROR database (search and autocomplete) through the CAP user interface. Then, they can also fetch the ROR ID for that organization which is accompanied by other relevant metadata. See Figure 12.

---

[31] UC3 published a blog post In July 2019 that has a good overview of the Dryad implementation: https://uc3.cdlib.org/2019/07/11/ror-ing-together-implementing-organization-ids-in-dryad/
[32] inspirehep.net/
[33] github.com/cernanalysispreservation/analysispreservation.cern.ch

Figure 12 ROR implementation in CERN Analysis Preservation: searching in the ROR database, fetching the ROR ID for an organization and relevant metadata.

In terms of the INSPIRE implementation, INSPIRE has an Institutions database[34] with more than 11500 curated records. INSPIRE will soon be migrated to a new version, but this database is planned to be retained in the new INSPIRE as well. The ROR implementation concerns the current platform. Development for INSPIRE is much simpler, as there is no need for an autocomplete function or search interface. An institution record provides simple metadata about an institution/organization[35]. There,

---

[34] https://inspirehep.net/collection/Institutions

[35] E.g. as in https://inspirehep.net/record/902725?ln=en

ROR IDs were included along with the already-existing GRID IDs (which could potentially be entirely replaced by ROR IDs in the future). See Figure 13.

Figure 13 INSPIRE institution record for CERN that shows the ROR and GRID IDs

Since INSPIRE has various other databases, for any given institution in the database, it is also possible to check how many and which papers come from that institution (literature database), a list of people that are affiliated with the institution ("HEPNames" people database), as well as which experiments are conducted at the institution (experiments database).

It is expected that these implementations can become more concrete in the future (especially in the case of INSPIRE) and it is possible that ROR IDs will be integrated into other CERN services at some point (e.g. Zenodo).

- *Timeline*

Development is complete and the implementations are being tested. Screenshots are not necessarily from the production systems and might change.

- *Use case to be addressed*

Tracking literature and other outputs produced by a specific organization (that operates in HEP). Enhancing metadata quality and enriching the CERN PID graph.

- *Questions arising*

The main issue to take note of in the case of INSPIRE is granularity, which was mentioned above. The INSPIRE Institutions database is more granular (department level) that what is provided by GRID (and ROR), which is more high-level. Also the parent-child and other relationship types is another important feature for INSPIRE, which will need to be addressed by ROR to enhance its usefulness. See Figure 14.

Figure 14 Example INSPIRE institution record indicating various relationships with other institutions (other records): related, successive and parent institution

● *Lessons learned through implementation*

It is often the case that the value of a PID is not entirely obvious to the users/clients in the beginning. Once the potential of organization IDs is realized through addressing concrete use cases and additional integrations in more services, it will become easier to convince the HEP community to further adopt or use them when possible (e.g. making them available in metadata, adding them to author profiles such as the ones on INSPIRE, etc.).

4.6 PLOS

● *Background*

PLOS is a nonprofit, Open Access publisher with a portfolio of seven journals in 2019 - *PLOS Biology, PLOS Medicine, PLOS ONE, PLOS Computational Biology, PLOS Genetics, PLOS Pathogens*, and *PLOS Neglected Tropical Diseases*. Historically, PLOS's primary revenue model has been Article Processing Charges (APCs) coordinated by authors, per article, after acceptance.

The PLOS suite of journals used the Editorial Manager (™) journals editorial system owned by Aries Systems Corporation. Editorial Manager provides access to the database of institutions provided by Ringgold.

● *Quick read summary of planned work for this deliverable*

PLOS is currently planning:

● How best to capture institutions (of all authors of/contributors to a manuscript) in the most accurate fashion possible, at the front end during manuscript submission.
● A project to clean up historical data on institutions.
● How to create tooling to keep the data in reasonable order over time as new records come in.
● Reporting on all the above information (see Use Case section below for full information).

ROR has been identified as the strong candidate as an open, robust, and community-supported list of institutions that PLOS could use to normalize this data, and use going forward, pending a technical review to help understand the available options.

- *Timeline*

- Internal documentation (engineering sprint) by December 1, 2019
- Decision to use ROR or other organization ID database by the end of 2019
- Data cleanup Q1 2020

- *Use case to be addressed*

Beyond simply keeping accurate records for their own sake, PLOS needs full clarity regarding institutions associated with submitted and published PLOS articles for 2 important functions:

1. To thoroughly and automatically identify manuscripts that should be billed to institutional or consortial accounts
2. To provide accurate reporting as needed to those institutions and consortia

While article-by-article billing for APCs can overlook ad hoc institutional nomenclature, lack of normalization in author institutions and the resultant need to do cleanup on a use-by-use basis has been an ongoing pain point for investigation into such business models.

A few use cases that illustrate uses of accurate article institution data:

- Internal reporting, e.g. generating accurate accounts with an internal Business Intelligence tool of how often a given institution or consortium publishes with PLOS, for the purpose of proposing and negotiating batch publishing deals.
- External reporting, e.g. producing rollup reports of institutional publishing for librarians.
- Accounting, e.g. producing accounts of publications with PLOS across a time period for the purpose of debiting an institutional account.

The ultimate solution to this issue is to capture institutions in the most accurate fashion possible, at the front end during manuscript submission. Unfortunately, while Editorial Manager does provide access to a the Ringgold list of institutions, its UI makes finding and selecting the right institution by authors very difficult, which pushes users to resort to free text input too easily (which would ideally only feel necessary/possible if no match exists in the master list). See Figure 15.
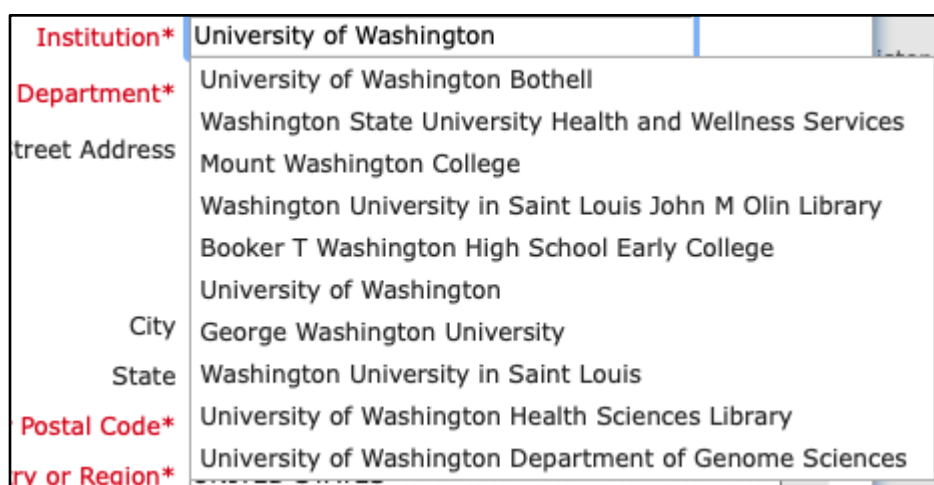

Figure 15 Screenshot of Editorial Manager institution drop-down menu of the Ringgold database

As a user types, the list of type-ahead matches seems to be sorted either randomly or invisible factors. For some cases it seems possible that an exact match may not even make the drop-down list. The type-ahead match list is limited to 10 items and does not scroll—in combination with the apparent randomness this can make finding, for example, "University of Texas at Dallas" difficult. Also, when the user mouses over any selection in the list, any item the pointer touches is populated into the field without a click, obliterating what has been typed.

PLOS understand Aries is considering changes to the UI, and is taking input on changes to the list(s) they offer, but there has been no official roadmapping for these things.

Regardless, improving the front-end of the process would not fix old data, and any cleanup of data can be revisited later if there was a need to change approaches. PLOS seeks to explore options for cleaning up the data it has on institutions and for creating tooling to keep the data in reasonable order over time as new records come in. As stated above, ROR has been identified as a strong candidate as an open list of institutions, from a like-minded organization, that could be used to normalize this data. It provides some metadata for each institution as well as ID links to other open databases that could be of assistance down the road toward identifying consortia memberships, via API. PLOS is also investigating the ROR reconciler[36] which has also been developed using OpenRefine.

- *Questions arising*

As has been noted elsewhere in this report (in particular, Section 4.1 EMBL-EBI) PLOS understands and notes that the research and PID community needs to investigate and discuss what the "affiliation" field in a published article (and, therefore, submission system) actually represents. Currently it could represent:

- the current or most recent "address" of authors (and therefore primarily an author-related field, for the main purpose of identifying or contacting authors).
- an unequivocal designation of "the institution at which, or under the auspices of which, the work was carried out".

In most published items, these two fields would be the same, but since most use cases seem to suggest that we are using organizational IDs for accurate representation of notion 2, then there is perhaps an open question for the PID and research community as to whether the affiliation field in scholarly communications is the correct place for this information, and thus this particular PID? It has been suggested that perhaps articles need a new field for institutions representing notion 2, which would operate in a similar way to a Funder field for an article.

- *Lessons learned through implementation*

While PLOS's implementations are all plans for future work, our experience with the Ringgold lookup in Editorial Manager highlights the need for UI/UX considerations in any tools requiring users to enter or select information, in addition to accurate databases in the back-end.

4.7 DataCite

- *Background*

DataCite is one of four partners on the steering committee at ROR and helped to develop the minimal viable repository for ROR as part of FREYA's Task 3.3.

Prior to 2019, the affiliation information for authors in DataCite DOI metadata was a text string. This information is optional, from the 3,561,969 DOIs registered in 2018, 799,651 (22.4%) had affiliation information. For organizational authors, a *nameIdentifier* has been supported for many years.
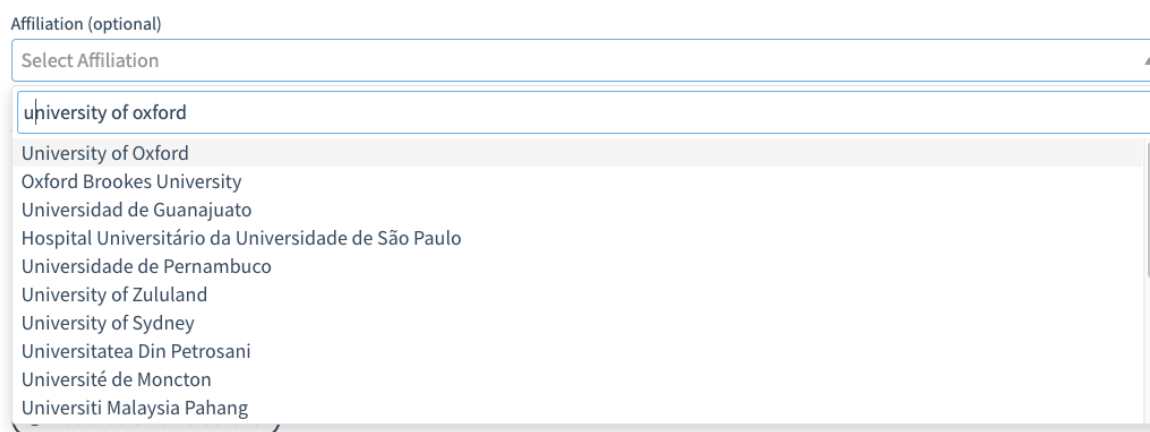
---

[36] https://github.com/ror-community/ror-reconciler

Contributor metadata hold exactly the same information, but also include a *contributorRole*, including for example *hostingInstitution*.

- *Quick read summary of planned work for this deliverable*

In April 2019 DataCite release an updated DOI registration form in the Fabrica DOI registration service where affiliation information was added via a lookup via the ROR API (Figure 16). The same functionality was added for organizational authors[37].

In August 2019, DataCite Metadata Schema 4.3 with support for affiliation identifiers was released[38], and promoted via blog post, newsletter, webinar, etc. Via the API, the type of affiliation identifier needs to be specified, but is not restricted to ROR. In the Fabrica DOI registration form only ROR identifiers are supported.

Figure 16 Affiliation lookup via ROR API in Fabrica DOI registration service

In August 2019, organizations and their connections to publications, datasets, software and other research outputs were added to the Crossref/DataCite Event Data service, and the DataCite GraphQL API for the PID Graph, allowing complex queries starting with a ROR identifier, e.g. listing all datasets associated with a particular organization (Figure 17).

---

[37] Dasler, R. (2019, April 25). DOI Fabrica Form Gets a Facelift. https://doi.org/10.5438/RZB6-G292
[38] DataCite Metadata Working Group. (2019). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data v4.3 (Version 4.3). DataCite. https://doi.org/10.14454/7XQ3-ZF69
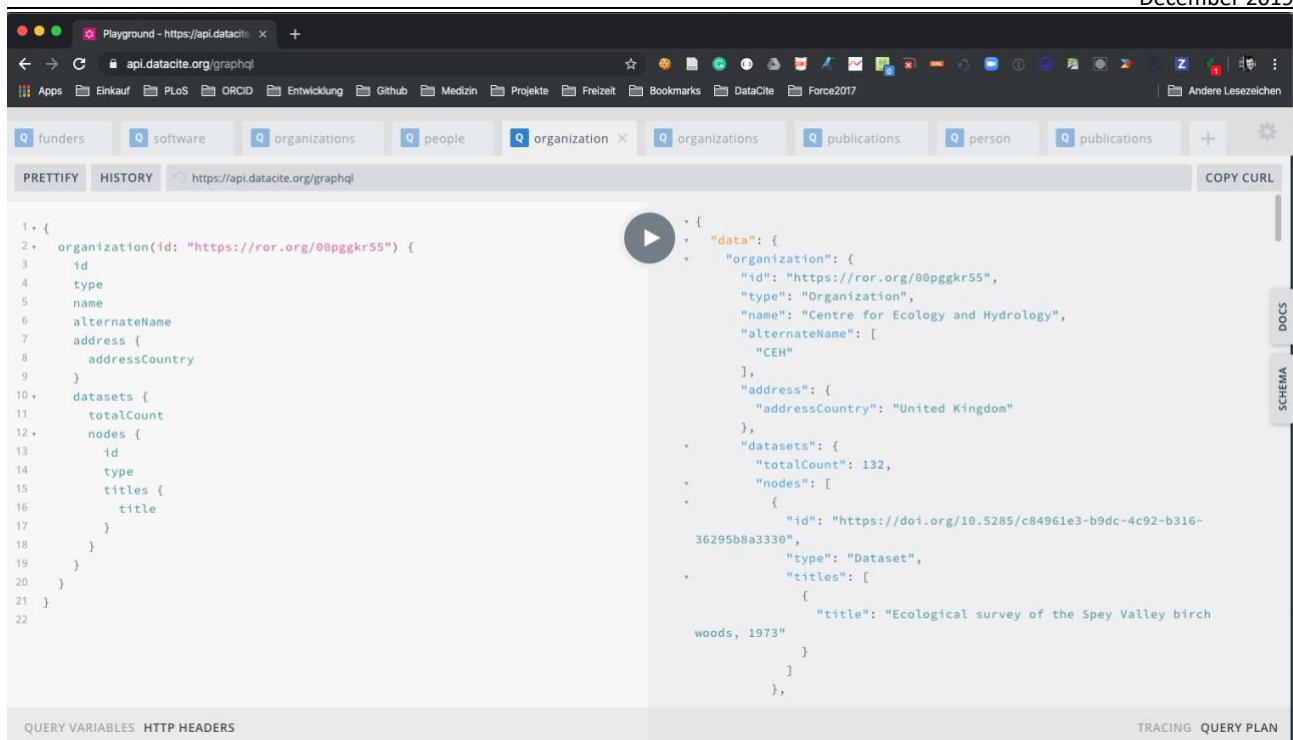
Figure 17 Query by ROR identifier in DataCite GraphQL API

In October 2019, after registration of more than 2,900 DOIs with affiliation identifier, a search facet allowing the filtering by institution was added to the Fabrica DOI registration system and the public DataCite Search (Figure 18).
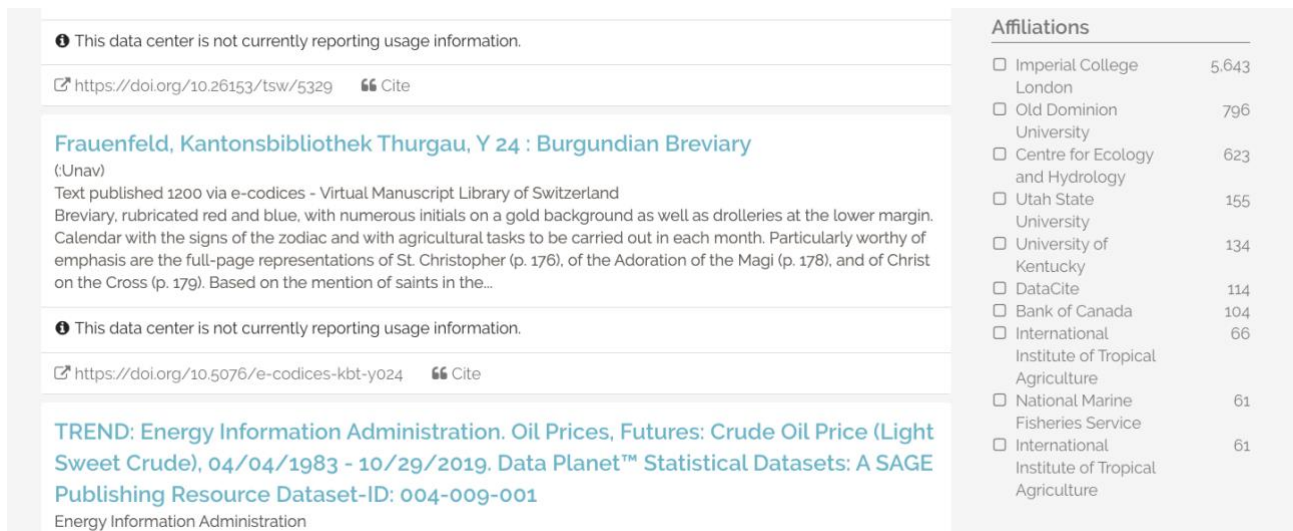

Figure 18 Filtering by affiliation in DataCite Search

As DataCite does not change the DOI metadata registered by its members, DataCite is not going to add affiliation identifiers to the many affiliations already registered with DataCite. Rather, DataCite is working with its members and individual repositories to update this information with the help of the ROR reconcile API.

- *Timeline*

- April 2019: Use of ROR database to provide affiliation names in DOI registration form
- August 2019: Support for organization identifiers in DataCite Metadata Schema
- October 2019: Facet for affiliations in the DataCite Search web interface

- ● *Use case to be addressed*

- Affiliation information for creators for scholarly outputs
- Uniquely identify institutional authors

- ● *Lessons learned through implementation*

The overall experience implementing organization identifier support using ROR at DataCite has been positive. ROR provides an easy to use and open JSON API, and the database with close to 100,000 affiliations has very good coverage of the affiliations needed for DataCite DOI registration. As predicted, most of the ROR IDs added since August 2019 use the web form, because updating API integrations will take more time. Several DataCite members indicated that they plan to update their DOIs with affiliation identifiers before the end of 2019, and we expect a significant number of DOIs with affiliation identifiers by August 2020, 12 months after the launch of the affiliation identifier support. A common use case of the early adopters is DOIs for dissertations, which facilitates the discovery of these works associated with particular institutions (**Text box 1**). A second second use case from the Centre for Ecology and Hydrology (CEH) is presented in **Text box 2**.

Text box 1: Dissertation Use Case

Several of the early adopters of affiliation identifiers in DataCite DOI metadata used them in metadata describing dissertations. One example is the Old Dominion University in Virginia, USA, who by October 2019 had registered more than 500 dissertations with affiliation identifiers. Also, an increasing number of universities in the UK are assigning DOIs to theses and sending the DOI metadata to EThOS, the UK's thesis index administered by the British Library.[39] The total number of DOIs in EThOS at end September 2019 was 10,472, up from 5019 in September 2018. The British Library is a DataCite member and FREYA partner, and using affiliation identifiers facilitates the discovery of theses from a particular institution, even if in a repository located elsewhere.

Two challenges were identified during the ROR implementation work at DataCite:

1. No ROR identifier for the institution
2. Complexity of the connection between institution, author and publication/dataset

The DataCite Fabrica implementation does not allow the manual entry of affiliation names, depending instead on information from ROR. We did this based on the experience at ORCID which had implemented an affiliation name lookup much earlier. ORCID helps users to select organizations via an ID powered Organization name lookup, and supplements this with the ability to customize the name of an organization once chosen. This provides a more detailed/specific affiliation name, but does lead to alternative names being linked to a single organization ID. Users of the DataCite Fabrica service face another problem: until ROR takes over curation from GRID and builds an interface that makes it easy to submit curation requests, information about affiliations might get lost if these affiliations are not in the ROR database.

As also discussed elsewhere in this document, the relationship between institution, author and publication can be complex, as authors change institutions, or work at more than one institution in parallel. This is confounded by the delay in publication: work done at institution x might not be published before the author moves to institution y. This makes it sometimes hard to unequivocally associate an institution with a publication using the author has the connection between the two. For this

---

[39] https://ethos.bl.uk/

reason it is essential to have a direct link between publication and institution, using for example DataCite (or Crossref) DOI metadata.

Text Box 2: Affiliation Identifier Use Case from The Centre for Ecology and Hydrology

The Centre for Ecology and Hydrology (CEH)[40], who use DataCite DOI services through a consortium led by the British Library, has adopted affiliation identifiers for authors using ROR IDs in the metadata for datasets deposited in the Environmental Information Data Centre (EIDC).[41] At present this is only available for researchers affiliated with the CEH and the ROR ID is hard coded to be added to the metadata sent to DataCite when a DOI is created. CEH will soon enable the capability for authors from elsewhere to have affiliation identifiers in the DOI metadata. As this will be a manual entry process, how best to incorporate this into record creation process is being determined. CEH decided to start using affiliation identifiers to remove the existing ambiguity with identifying organizations over time, to improve the machine readability of their metadata and so the EIDC's datasets will be incorporated fully into the PID Graph. ROR was chosen as the preferred identifier given it is community led and incorporates many of the affiliations CEH would wish to use.

4.8 DANS

● *Background*

Organization PIDs are of great relevance for the DANS services EASY[42], its online archiving system, and NARCIS[43], the Dutch national portal for scientific information, with many possible use cases were these PIDs could and should play a vital role. There is a clear need to identify an organization unambiguously and in a computer readable way, especially in the exchange of information to/from other services, in provenance, dissemination of datasets and as part of the PID Graph. For this WP4 task, DANS made a start with the implementation of organization PIDs in EASY and NARCIS.

● *Quick read summary of planned work for this deliverable*

DANS implemented organization PIDs in EASY and NARCIS. This means that EASY and NARCIS will support the use of PIDs, and NARCIS will automatically relay information relating to an organization PID. This implementation is done in a generic way, so other organization PID types can be added should there be a demand from users or stakeholders.

Both EASY and NARCIS have several types of users and stakeholders. First, there are the users who provide metadata and data, these can be individual researchers or research organizations. Secondly, there are users that reuse metadata and data, such as researchers, or other users/stakeholders, such as service and content providers, including OpenAIRE, Google Scholar, and Thematic Portals.

DANS chose to support different organization PIDs because the communities DANS serves are not homogeneous and use different identifiers. DANS is also not in a position to impose the use of a certain organization PID type. Although we recognize that the different organization PID types do not have the same maturity level, we are able to use them for connecting content to the NARCIS PID Graph. In the process of implementation, we found that some organization PIDs types are difficult to obtain, or even

---

[40] https://www.ceh.ac.uk/
[41] http://eidc.ceh.ac.uk/
[42] https://easy.dans.knaw.nl/ui/home
[43] https://www.narcis.nl/?Language=en

need to be bought, which is not in the best interest of Open Science. DANS did a small experiment and bought some ISNI, but this will not be scalable. For DANS, organization PIDs must be easy to obtain, free, globally unique, resolvable and preferably have useful services built around the organization ID.

- *Timeline*

The first phase of organization PID implementation was done in 2019. Future work will include outreach to different communities and users of EASY and NARCIS to adapt the usage of organization PIDs in their workflow and to include them in the metadata. NARCIS already supports organization PIDs fully and DANS aims to support the use of organization PIDs in EASY in 2020. In EASY is it already possible to include organization PIDs in the metadata, but the relevant information needs to be added to the deposit form for data.

- *Use case to be addressed*

This section describes the completed and future work in the DANS services EASY and NARCIS.

EASY - Electronic Archiving System

EASY, the Electronic Archiving SYstem of DANS, contains more than 100.000 dataset, with almost 50% from the archeological community. The life sciences collection is the second largest, with more than 30.000 datasets. These communities have different needs and it is not expected that they will use the same organization PID types. In fact, most smaller research organizations in the archeological field will not be found in the GRID or ROR registry. The first step in implementing the organization PIDs in EASY was to make the internal metadata schema ready to receive the following organization PID types: ROR, GRID or ISNI. The second step was to include the organization PIDs in the DataCite metadata schema in the process of obtaining a DOI for a certain dataset. EASY provides two different ways to deposit metadata and data in the digital archive: manually through a web-form, or automatically through the use of the SWORD-protocol (so-called machine-to-machine depositing). For machine-to-machine depositing organization PIDs are fully supported. Future work will include facilitating the entry of organization PIDs in the online submission forms for manual deposits, but since automated depositing is outnumbering manual depositing, we prioritized automated depositing.

To fully understand what is needed to introduce an organization PID in a certain field within the social sciences and humanities disciplines, it was decided to introduce an organization PID for all archeological research institutes. About 60 Dutch archaeological research institutes use EASY as Trusted Repository for the preservation of their research materials and reports. Most of the needs of these organizations are similar. An important requirement is the possibility to create an overview of all their datasets in EASY or outside our system, for example in services such as Europeana Collections and the ARIADNE e-infrastructure[44] (Enabling European Archaeological Research). For the latter, ACDM and Carare are used as metadata formats and EASY is intermediary in providing archaeological metadata to these services. To accomplish this requirement the use of an organization PID is requisite. Some archaeological organizations make an overview of their datasets manually for presentation of the web, but this is not sustainable and exchangeable. In this use case DANS assigned an ISNI to all these archeological research institutes and plans are made to assign them retrospectively to more than 10 000 archeological datasets (this work will be done in 2020). By assigning organization PIDs during the ingest process, datasets are assigned automatically to the right organization and overviews per organization are automatically generated, in the EASY archive, as well in any other service mentioned in Figure 19. The organization PIDs will be included in the metadata and in the PID Graph and information can be connected on basis of PIDs any time, at any stage, needed.

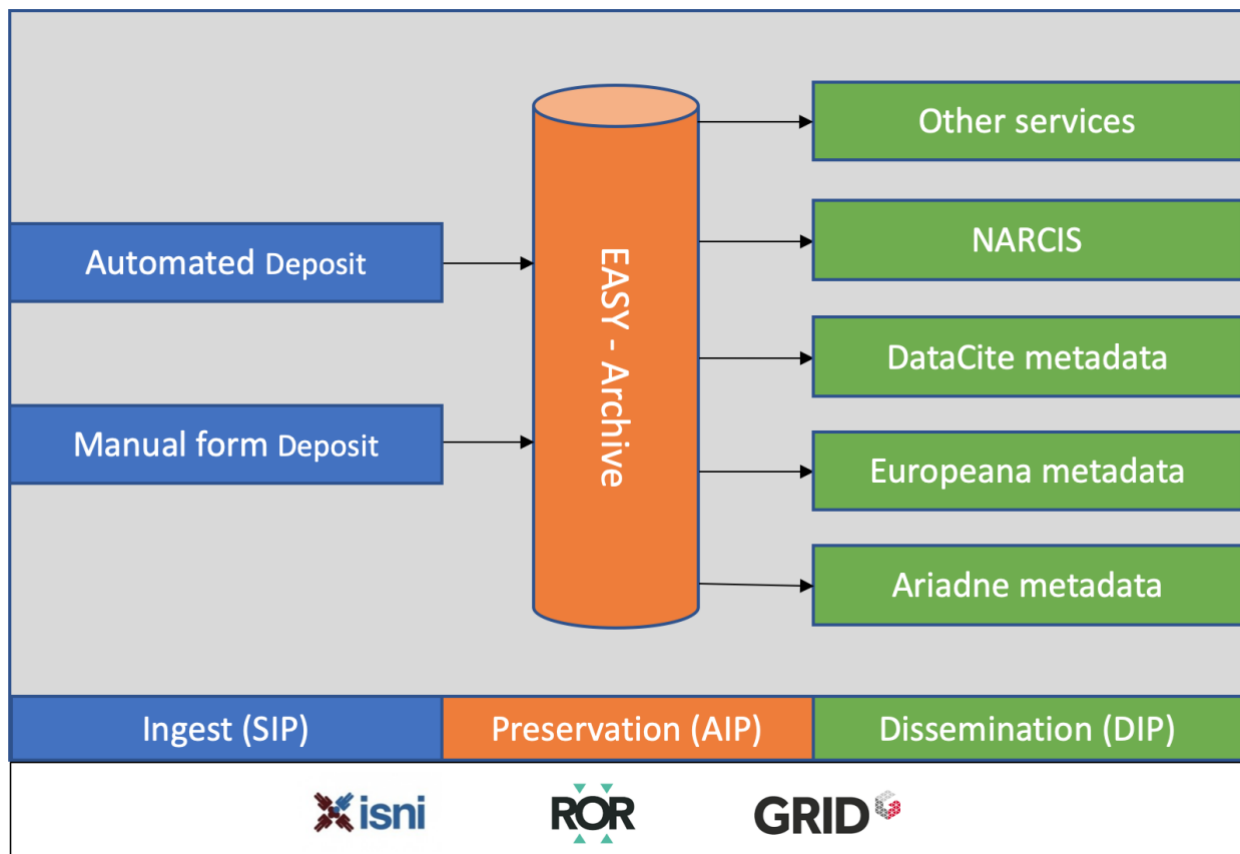[44] https://ariadne-infrastructure.eu/

Figure 19 Use of organization PIDs during ingest, preservation and dissemination process in EASY (present and future work)

Work done is creating the possibility to add ISNI, ROR or GRID to the metadata for the automated deposits. As mentioned before, there are plans to add this functionality to the deposit form for manual deposits, but this still needs to be done. EASY is able to include the organization PIDs in the DataCite metadata schema, which is used by DataCite and NARCIS and therefore will be part in those PID Graphs. Future work will be to include organization PIDs in other metadata formats (like ACDM, Carare, MODS) for the use of services like Europeana Collections and The ARIADNE E-Infrastructure (Enabling European Archaeological Research).

NARCIS - National Academic Research and Collaborations Information System

NARCIS aggregates information on publications and datasets from more than 40 different Dutch repositories and archiving systems and contains only metadata. In addition to the aggregated information from these repositories, NARCIS contains information about all Dutch research organization, researchers and research projects. The organizations that provide information to NARCIS make use of different systems, workflows, and technical solutions, and have different needs. Therefore, they also make different decisions on which kind of PIDs they use to identify different content types. Some even use two or more different PID systems for the same object, just because they serve different purposes, or there is an obligation to use one of them. To meet all these different requirements, DANS added functionality to NARCIS to support different organization PIDs. Currently NARCIS supports ROR, GRID, and ISNI.

One of the aims of NARCIS is to connect the various information types and present research information in its context. For example, funding is connected to projects, projects to data and publications, and all these types of research outputs to researchers and organizations in its different roles as funders, secretariat, etc. The aggregated metadata originates from multiple sources, and relations can only be

made through the use of PIDs in the metadata or in the PID Graph. In addition to other PIDs, organization PIDs play a role in identifying organizations unambiguously. Another aim is to make this "connected information" available for different communities: to information services such as Google Scholar, OpenAIRE, DART-Europe and so on, and to NARCIS' data providers. But also these PIDs are made available for reuse in Linked Open Data and schema.org, in different formats through API's, and through the NARCIS portal to the general public.

For most content types there were already PIDs available: publications and datasets can contain DOI, Handle or an URN, persons can contain PIDs such as ORCID, ISNI or DAI (Digital Author Identifier). As a result of the work in this task, ROR, GRID or ISNI can now also be used in NARCIS to make connections of digital content to organizations.

 The following five sections will give an overview of the work done and our future plans regarding organization PIDs.

1.   Assign organizations PIDs to organization records

NARCIS contains records of all Dutch research organizations, with their physical address, web address, phone, email, missions and other information. The portal contains information on nearly 3.000 research organizations, including the departments of universities. Most of these organizations are not included in the GRID, ROR or ISNI database yet.

 As first step we assigned a ROR, GRID and ISNI to several organizations that were already present in the different registries which resulted in almost 200 Dutch research organizations with one or more PIDs, see Figure 20.

Figure 20 University of Amsterdam, with GRID, ISNI, ROR

Another step was applying for ISNIs from Ringgold for 60 archeological research organizations to experiment with the process of obtaining organization PIDs. For the pilot with archaeological organizations, as described before, we needed these identifiers anyway. Future work will be a full coverage of organization PIDs for all Dutch research organizations, including faculties and departments. All stakeholders and users would greatly benefit from this work, for example by relating datasets or publications to organizations, generate overviews for each organization, provide organizations with usage statistics.

2.   Uptake of organization PIDs in metadata

A second step is to include the organization PIDs in the metadata and in the exchange of metadata or in Linked Open Data. By doing so, PIDs become part of the NARCIS PID Graph and will be openly available to the rest of the world.

For the archaeological use case DANS is testing including organization PIDs in the metadata, so that NARCIS can presents overviews for all participating organizations and the PIDs will become available for other (archaeological) services. This result is not completed yet, but will become available in 2020.

Furthermore, DANS is planning to engage with the different communities and stakeholders in the Dutch national research information infrastructure to stimulate the update and use of organization PIDs.

3.   Uptake of organization PIDs from other services

For deliverable 4.1, DANS developed a micro-service[45] which collects PIDs from other services and integrates them into the NARCIS PID Graph. The main purpose of this micro-service is to collect PIDs and relations between PIDs and add them to NARCIS for enrichment of relations between digital objects. This functionality has already been added to the NARCIS portal and DANS wants to extend this functionality in the near future to organization PIDs. By doing so, it would be possible to retrieve organization PIDs from ORCID.org, match them with NARCIS and enrich the relationship within NARCIS and the NARCIS PID Graph, schema.org and JSON-LD.

4.   NARCIS relates information on basis of organization PIDs

NARCIS is ready to receive metadata which contain organization PIDs. As soon the records flow into the NARCIS metadata store, the portal will automatically attach the publications or datasets to the right organization of basis of these PIDs. Figure 21 gives an example of how NARCIS presents the information in the web portal. The organization record and research is already in production, the functionally to add publications and datasets is ready, but adding the organization PIDs to the metadata of publications and datasets is work that still needs to be done. Organization PIDs are also indexed and searchable in the NARCIS portal.

---

[45] https://github.com/DANS-KNAW/narcis-pid-aggregator

Figure 21 Mock-up of organization record with publications and datasets attached to it

5. Publishing of organizations PIDs

All available organization PIDs will be published in the different metadata formats. This is already accomplished in JSON-LD, schema.org, CERIF and the DataCite format. Other formats still have to follow. They also will become available through the NARCIS PID Graph.

● *Questions arising*

In one of the use cases new organization PIDS were needed. Actually it was rather difficult to obtain an organization PID for an organization that is not in a certain registry yet. For the Netherlands is seemed that obtaining an ISNI required payment, however it was unclear how we could add organizations to registries like GRID or ROR in order to obtain persistent identifiers for organizations that don't yet have IDs assigned. Related to this is the issue of the governance of the PID system. It is not always clear who "owns" the organization PID and who had the right to edit the organizational information.

Concerning the structure of universities (faculties, departments, etc.), it is not clear how this structure can be maintained and what will happen when departments merge, for example.

● *Lessons learned through implementation*

Technical implementation and metadata is work that can be planned, but to get all stakeholders using an organization PID as part of their workflow is rather complicated and it will take time before adoption is a general practice.

5 Conclusions

5.1 FREYA partners' activity with organizational identifiers

This deliverable reports on the current status of integrations of organization Identifiers into resources run by FREYA partners. The FREYA partners here represent a broad range of resource/service providers including, a PID provider (DataCite), a PID resolver (Identifiers.org), disciplinary partners (EMBL-EBI, PANGAEA, CERN, UKRI-STFC, British Library and DANS) and a publisher (PLOS). See Table 1.

FREYA Partner	Resource/Service	Organization IDs implemented prior to 2019	Organization IDs being implemented from 2019	Status of new implementations
EMBL-EBI	Europe PMC	Crossref Funder IDs	ROR	Pilot implementation for ROR IDs; limited in scope
EMBL-EBI	Identifiers.org	-	ROR	Platform-wide implementation complete
STFC	PhD thesis pilot	-	GRIDs, ROR	In progress
PANGAEA	PANGAEA published dataset repository	-	ROR	Platform-wide implementation; in progress
British Library	BL repository & EThOS	ISNI	ISNI, ROR	Scoping exercise only (for ROR IDs)
CERN	INSPIRE-HEP	GRID	GRID, ROR	Complete
CERN	CERN analysis preservation (CAP)	-	ROR	Complete in testing system
PLOS	PLOS editorial management system (powered by ARIES); internal BI systems	RINGGOLD	Possibility to include ROR IDs going forward	Scoping exercise only
DataCite	DOI registration service (API and web frontend)	-	ROR	Platform-wide implementation complete
DataCite	DataCite public search	-	ROR	Implementation complete
DataCite/Crossref	Event Data service	-	ROR	Implementation complete
DANS	EASY Archive	-	ROR, GRID, ISNI	In progress
DANS	NARCIS	-	ROR, GRID, ISNI	Complete

Table 1 A summary table showing FREYA partner resources that are currently working with organization IDs, the status of the implementations and the organization ID types being incorporated

The range of partners and implementations demonstrates the wealth of use cases for organization IDs across disciplines and services.

Before 2019, very few FREYA partners incorporated organization IDs in their metadata. The organization IDs types that were included varied from partner to partner showing next to no market dominance by one type.

Of the organization IDs available for integration, functionality overlaps to the extent that they all identify an organization of some sort, yet each identifier has been developed to address a specific community need, and each has its pros and cons relative to other identifiers: ISNIs can be issued for organizations or individuals (both living and dead) and are important for social science arts and humanities communities served by the British library and DANS. However, ISNI does not yet operate with a CC0 license. LEIs serve the financial sector, yet won't serve the research community. Ringgolds are not open, yet the organization ID of choice by large service providers in the publishing community.

ROR is a newly emerging, open, community-governed registry and has emerged as the "consolidated infrastructure" alluded to in the projects description of work (see Annex 1). As such FREYA partners have actively sought to support the uptake of these identifiers in their work. Importantly these are linking identifiers that will enable connections to alternative open organization ID types. As can be seen in **Table 1**, ROR IDs are now the predominant type of organization ID in partner implementations from 2019.

5.2 Take-home messages from FREYA partner implementations

Partners have mentioned a range of challenges whilst implementing ROR IDs into their metadata. These are listed here and grouped where related. To give a sense of how often a particular issue arose, we have included a number in square brackets for each issue that indicates the number of partners who mentioned the issue.

5.2.1 Issues regarding Implementation of ROR IDs:

- Limited granularity of ROR records [3]. ROR's metadata model for organizational information is too superficial for some users. ROR is aware of this challenge and may address this, although not as an immediate priority.
- The number of organization IDs that can be accumulated per affiliation [1]. There are not yet functional standards or guidelines for this. The implementation in the DataCite Metadata Schema 4.3 allows only one organizational identifier per affiliation, but multiple organizational identifiers for institutional authors. Affiliation identifiers were added in schema 4.3, and allowing multiple identifiers would have led to breaking changes.

5.2.2 Issues regarding current functionality of the Research Organization Registry:

- It is currently unclear how to go about getting an ID created, if no ID exists for an organization [3]; ROR's current policy is not to add/change any records but this is high on the priority list for ROR.
- Curation is not possible by community stakeholders [3] - curation mechanisms are high in the priority list for ROR, as they will allow the community, among other things, to report amendments in existing entries as well as to report missing organizational data.

5.2.3 Issues highlighting the need to build community consensus:

- Uncertainty over what the affiliation field in a published article/dataset represents [3]? Currently, the affiliation details recorded for a researcher may address one of several use cases: - e.g. for PID providers, it might capture the current address of a member (and therefore require updating to keep it current); for publishers, it might record the billing address of an

author; whereas for discipline specific resources, the affiliation field would be the place to record where the associated research was conducted.

- Need for a set of best working practices for how best to credit organizations on published research items [4]: Authorship complexity on a research item may be a consideration in determining whether it is best to assign organization IDs at the level of the research entity or to the author. The solution might vary for research entities: e.g. for journal articles in the life sciences, authors and credit can be attributed to a number of different organizations; however, datasets in the lifesciences can be smaller entities than journal articles and attributed to one author or a set from one affiliation. Assigning organization IDs at the level of the author might work best for the former scenario, whereas assigning organization IDs to the research entity may work better for datasets.
- Getting stakeholders to adopt organization IDs as part of their workflows will take time [2].
- Need for guidelines on how best to display multiple organization IDs in a human-readable way on the front end of a resource [1]: This is a user interface issue rather than an issue with organization IDs themselves. There are several options to address this, for example using ORCID iD-like icons in the author list. If more than one type of organization ID is recorded per organization, does one show icons for all of these organization IDs? Does one record this information only for disambiguation and linking purposes but not expose this to users on the front end? The solution here likely resides in the use case being addressed by the resource.

5.2.4 Factors to consider when selecting an organization ID type for implementation

As stated earlier in this deliverable (section 3), it is not the role of FREYA to insist on the adoption of any particular persistent identifier for a given entity. Given the lessons learnt (above), FREYA partners suggest that the following factors are weighed up when determining which organization ID type is best for further implementations:

- Use case(s) to be addressed by the implementation
- Scope of organizations included by the repository
- Whether the organization ID resource operates with an Open license?
- APIs to query the organization identifier service
- Granularity of information provided on landing page (name variants, URL, country, parent-child institutes)
- Extent of curation offered to stakeholders by the identifier resource.
- Governance of the resource—this will determine who makes decisions about all of the above

5.3 Considerations for EOSC integration

Although we wish to underline the value to EOSC of organization identifiers in general, the FREYA partners feel they can endorse the ROR ID as an open, community-governed organization ID. This report demonstrates cases of early adoption of ROR IDs into metadata and although services that provide integration are just emerging, the capacity to accommodate these and other organization identifiers could be integrated into any new EOSC services. For ROR IDs to get a foothold as an organization ID used across research disciplines, implementation needs to be taken up by a wider group of stakeholders. EOSC can play a central role for the wider acceptance of ROR IDs, as many of the services that make use of ROR IDs (or other organization IDs) are candidates for the EOSC Service portfolio and federating core. EOSC possesses the necessary scale to facilitate the cross-cutting implementation of ROR IDs.

In April 2019, EOSC pilot partners released a description of the proposed EOSC Service Architecture that identifies 47 classes of services that will be considered for incorporation in the "Minimal Viable Product" able to match EOSC's overall service goal[46]. These services are grouped according to the user they intend

---

[46] D5.4 FInal EOSC Service Architecture https://eoscpilot.eu/sites/default/files/eoscpilot_d5.4-v1.1.pdf

to serve and include services for researchers, for research administrators, for EOSC system managers (including service providers), for EOSC suppliers, or for third-party service providers; finally there are services that are envisaged to provide for or link any of these stakeholders, the cross-cutting services such as AAI (Authentication and Authorization Infrastructure), a class of services that the EOSC System puts in place for authorization and access control. ROR IDs have the potential to assist many of the services categories in EOSC: from the pilot implementations and emerging services mentioned in this report, where linked to a research entity, the PID could benefit "EOSC attribution and reward services". Similarly, it assists with "repository services" as it allows for the tracking of affiliation identifiers. We envisage that organization identifiers will augment "peer review" and "search and browse" services as well as "social networking services" for researchers. Ultimately ROR could be considered for EOSC's cross-cutting services such as AAI.

EOSC encompasses many scientific disciplines and accommodates a diverse stakeholder group including both researchers, administrators and managers. Hence, EOSC holds the potential to promote and advance usage of ROR IDs as a universal PID for organizations through both recommending and facilitating ROR implementation by its services.

Annex A: FREYA project's description of work for deliverable D4.4

D4.4 : Organizational IDs in Practice

Report on integrating Organizational IDs in pilot applications.

Report due: end Nov 2019

T4.2.2 Organisation identifiers

There has been significant progress with regards to consolidation of organization identifier infrastructure in 2016 and early 2017, with significant involvement by members of the FREYA project. It can be assumed that by late 2017 there will be a need to start work on disciplinary pilot applications that can take advantage of this consolidated PID infrastructure for organizations.

CERN will standardize organization information in its community databases using organization identifiers. This will involve disambiguation of organization information and the enhancement of connections to local databases/directories.

BL will develop links to organization identifiers within research services, to help manage and support the provenance of content and related use cases.

PANGAEA will integrate organization identifiers into data publication metadata.

DANS will pilot organization PIDs in their social sciences data archive.

STFC will test the use of organization identifiers for facility users from different institutes for reporting of experiments from large-scale facilities.

PLOS will integrate organization identifiers into their author database to facilitate reporting of publications by institution.

EMBL will test integration of PIDs for organizations within EuropePMC and/or BioStudies, and Identifiers.org will integrate PIDs for organizations within their registry.

Annex B: ROR resources

Partners were pointed to the following resources in order to build customized integrations of ROR IDs.

- ROR GitHub organization: http://www.github.com/ror-community
    - Everything about ROR is open source. Feel free to contribute.
- API
    - The primary endpoint is https://api.ror.org/organizations
    - Documentation is in the ROR GitHub repo, but more specifically here: https://github.com/ror-community/ror-api
- OpenRefine reconciler: https://github.com/ror-community/ror-reconciler
    - This can help map internal identifiers to ROR identifiers.
- ROR community Slack workspace: https://ror-community.slack.com/
- Created for the ROR community - here the community are notified of updates to the ROR repository, relevant blog posts, links to relevant slide presentations, conferences, workshops, and advisory group calls.