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

D3.2 Requirements for Selected New PID Services

Deliverable type	Report
Dissemination level	Public
Due date	28 February 2019
Authors	Christine Ferguson (EMBL-EBI, orcid.org/0000-0002-9317-6819) Johanna McEntyre (EMBL-EBI, orcid.org/0000-0002-1611-6935) Ginny Hendricks (Crossref, orcid.org/0000-0002-0353-2702) Tina Dohna (PANGAEA, orcid.org/0000-0002-5948-0980) Ketil Koop-Jakobsen (PANGAEA, orcid.org/0000-0002-1540-6594) Frances Madden (British Library, orcid.org/0000-0002-5432-6116) Sünje Dallmeier-Tiessen (CERN, orcid.org/0000-0002-6137-2348) Stephanie van de Sandt (CERN, orcid.org/0000-0002-9576-1974) Artemis Lavasa (CERN, orcid.org/0000-0001-5633-2459) Simon Lambert (STFC, orcid.org/0000-0001-9570-8121) Vasily Bunakov (STFC, orcid.org/0000-0003-3467-5690) Robin Dasler (DataCite, orcid.org/0000-0002-4695-7874) Martin Fenner (DataCite, orcid.org/0000-0003-1419-2405)
Abstract	A comprehensive analysis of user stories relating to a range of entities needing persistent identifiers, with conclusions for further work in FREYA.
Status	Submitted to EC 11 March 2019 Revised version submitted to EC 15 November 2019 *Clarification of resolution model and relevance for EOSC of candidate services for prototyping by FREYA partners*

The FREYA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777523.

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.

Executive summary

FREYA deliverable D3.1 offered a comprehensive survey of the landscape of persistent identifiers (PIDs) across many disciplines, with assessments of maturity of different PID types and conclusions for the future. The present report follows on from that work. A large number of "user stories" have been collected, analysed and prioritized for their further use in the development of PIDs and services. The report documents the definitions, methodology, research and recommendations by the FREYA partners for potential new PID services that could be prototyped within the timeframe of the FREYA project.

All FREYA partners were invited to collect user stories from their organisations and the communities in which they are embedded, expressed in the form:

> *"As a <role>, I want <capability>, so that <benefit>".*

The user stories were a basis to prioritise the possible work that could be undertaken. In addition, outreach activities were conducted to gain an even broader perspective.

The following entities have been prioritised (with lead partner noted in parentheses):

- instruments (PANGAEA)
- facilities (STFC)
- grants (EMBL-EBI)
- organisations (DataCite)
- software (DataCite)
- research campaigns (PANGAEA)
- Data Management Plans (DataCite)
- physical samples and cultural artefacts (British Library)
- conferences (CERN)

Deep-dive analyses of these are presented, referring to the user stories relating to the entities, validation from outside the FREYA project, possible action by FREYA partners and the relationship with other FREYA Work Packages. The conclusions include candidate services for prototyping by FREYA partners.

Contents

1 Introduction

There are different ways to define and implement a "service". In contrast to public sector or corporate settings, the EU project environment necessitates a more agile approach to service development that is based on the expertise of dispersed project partners and takes into account a loosely coupled nature of their working relationships. Using terminology that is likely more familiar to government departments or corporations, the FREYA project's approach to service development comprises business analysis, IT architecture considerations, components development, components integration, and service validation. These elements are best thought of as interlinked project activities rather than distinct phases of service development. This deliverable is focussed on the first such activity, namely "business analysis" or research. The actual mechanism chosen by FREYA for its business analysis follows agile development methodology and is based on user stories that are collected, analysed and prioritized for their further use in the development of services.

This report documents the definitions, methodology, research and recommendations by the FREYA partners for potential new PID services that could be prototyped within the timeframe of the FREYA project (December 2017 to November 2020). We have used the user stories as a means to prioritise the possible work that could be undertaken. The report reflects both discussions and brainstorming that took place within the working group meetings, plus descriptions of potential pilot projects that could be undertaken by specific partners as well as any related work that is known to be taking place by communities outside of the FREYA consortium.

An important definition: in this report we use the word "entity"[1] to describe anything in the domain of research or scholarship that is assigned an identifier, including people and organisations, as well as resources like grants, instruments, samples, data, and scholarly outputs such as literature and conferences.

We agreed to investigate requirements for services that might implement *new* PID types (such as "ROR IDs" for organisations) in addition to services whereby entities previously without PIDs will be newly assigned *existing* PID types (such as DOIs for Data Management Plans). The research into possible services includes identification of relevant external working groups and initiatives across the globe currently involved in pursuing identifiers for these entities.

An underlying aim for the report is to provide a resource to stakeholders who share an interest in pursuing persistent identifiers for these various entities.

For this work package (WP3; see Table 1) FREYA partners undertook the research required to assess the state of the art for persistent identifiers (PIDs), to identify gaps and to specify use cases and requirements for potential new PID types and services.

Work Package (WP) title	Broad Aims
WP1 Project Management	
WP2 PID Core services	Improving what we have.
WP3 New PID types	Building what we don't have.
WP4 Integrating the PID Graph	Incorporating it.
WP5 Iterative Engagement	Sharing it.
WP6 Sustainability	Sustaining it.

Table 1 FREYA's Work Packages

[1] Oxford English Dictionary definition: 'A thing with distinct and independent existence'

The first deliverable (D3.1) drawn up in June 2018, described the evolving PID landscape and provided an assessment of the extent of PID usage, and maturity of PID services across research communities. The present deliverable (D3.2) focuses on the gaps in the PID landscape that might be filled and comprises three parts: (1) gathering use cases for new PIDs and PID services, (2) prioritizing/validating/mapping, (3) collecting requirements that are actionable. A subsequent task in the project will be to develop prototypes of selected new PID resources. Prototyped services may then be taken further by partners in WP2, WP4, or by an organisation external to FREYA. Alternatively, the prototypes may be sunsetted at the end of the FREYA project.

Update Q3 2019: The sections in chapter 3 that provide details of candidate services for prototyping by FREYA partners (Instruments, Facilities, Grants, Organisations and Research campaigns) have been updated to clarify the envisaged resolution model and the relevance for EOSC.

2 Methodology

2.1 User stories

The FREYA consortium is made up of partners who are invested in providing services for data management and embedded in their scholarly communities which collectively cover many research disciplines[2]. As such, partners are aware of the many varied PID service opportunities and are exposed to competing demands for such services. Therefore to help prioritise work that could be undertaken during the FREYA project, partners agreed to collect and be guided by "user stories" that reflect needs articulated by stakeholders, rather than "use cases" that describe functional solutions for perceived needs. Importantly, a user story describes something that the user needs to do in her/his day-to-day job; it is necessarily short and written in the language of the user, avoiding jargon and so easy for all to understand.

Most user-stories can be described as following the template:

> *"As a <role>, I want <capability>, so that <benefit>".*

All FREYA partners were invited to collect user stories with this format from their organisations and the communities in which they are embedded. Partners have also reached out to research communities at conferences and to the FREYA ambassadors[3] who are located across the globe. User stories collected since since Q3, 2018 have been collated in a Github repository set up for FREYA[4].

To encourage ongoing engagement of communities beyond those of FREYA partners, the user stories were replicated on *pidforum.org*, a community site made available to everyone with an interest in PIDs from Q1, 2019[5]. This is currently a living collection of user stories open to comment and addition and will be used in an ongoing way to inform future efforts of FREYA partners.

2.2 Prioritisation of user stories

At the end of Q3 2018, user stories were assessed for their relevance to WP3 (New PID types and services) by a small review committee representing three of the FREYA partners: Christine Ferguson (EMBL-EBI); Martin Fenner (DataCite) and Rachael Kotarski (British Library). The user stories that had been accumulated and are assessed in this report are available in table format in Annex A.

> Using Github, each user story has been entered as a "Github issue" accompanied by a short title and tagged using a controlled vocabulary of labels to allow for sorting (* the entity labels approximate the categories of entities described in earlier deliverable D3.1—see Annex B.

Table 2). Note that a single user story could be tagged with several different entity labels, e.g. user story #69[6]. The date of entry of the user story is also noted. The committee assigned a WP3 label where these had not been previously assigned by submitting partners. At the time of the exercise, 30 user stories in the collection were identified; and three further stories were added subsequently[7].

[2] https://www.project-freya.eu/en/about/partners
[3] https://www.project-freya.eu/en/ambassadors/our-ambassadors
[4] https://github.com/datacite/freya/issues
[5] https://www.pidforum.org/c/user-stories
[6] https://github.com/datacite/freya/issues/69
[7] The WP3 user stories can be seen here https://github.com/datacite/freya/issues?utf8=%E2%9C%93&q=is%3Aissue+is%3Aopen+label%3AWP3+

Category of label	Label applied using Github
Source of user story	FREYA partner organisation (eg British Library, ORCID, EMBL-EBI), ambassador, conference community
User/role in the user story	Library, facility, curator, funder, researcher, bibliometrician
Entity mentioned in the user story*	Article, data, grant, person, software, organisation, instrument, project, etc
FREYA Work package for which the user story is relevant	WP1-6, PID Graph
Additional labels used that other than the categories above	User story, geolocation, species, next

* the entity labels approximate the categories of entities described in earlier deliverable D3.1—see Annex B.

Table 2 Labels used to tag user stories in Github

We considered the following prioritisation possibilities for this subset identified by the WP3 label:

By entity: numbers of user stories per entity—this would reveal a sense of the demand to see this entity being linked.

By status quo: assess whether there is an external working group already working on this PID-need, perhaps an RDA working group; or a plan to take something forward (e.g. ROR.org).

By impact: assess the extent to which the PID/service will be used and will impact on workflow efficiency if implemented by a particular stakeholder group; assess whether there are barriers to implementation/adoption.

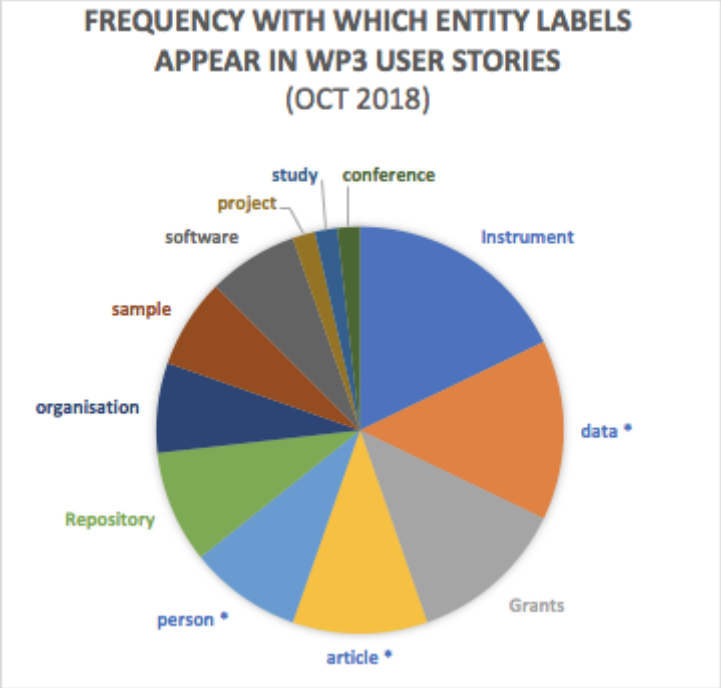
By partner expertise or special interest: (required for prototyping) assess whether any of the FREYA partners have the technical expertise and supporting infrastructure to build a prototype for a service; currently offer a similar service that could be extended to link a new entity; or have a specific interest in developing a specific service.

Partners focussed on prioritising by **entity** and **partner expertise** in the first instance. **Status quo** and **impact** were subsequently considered for groups of user stories when "validating and gather requirements", which is presented later in this report.

Priority entities for prototyping

Sorting the stories in Github using the entity labels provided an indication of the range and frequency of their mentions in the user stories. As can be seen from the collective of WP3-related user stories on Github, the entity labels applied to user stories can vary in number from one to as many as seven (these labels are shaded dark blue on Github). The pie chart (* entities with the PID infrastructures deemed most mature by FREYA partners—see initial maturity ranking in the landscape survey of PID services reported in a previous deliverable..

Figure 1 reflects the entity labels that were applied to the 30 user stories by submitting partners, indicating the relative frequency with which an entity label was applied across the user stories. Definitions of the labels become clear from the user stories, e.g. "study" refers to "a study registration record".



* entities with the PID infrastructures deemed most mature by FREYA partners—see initial maturity ranking in the landscape survey of PID services reported in a previous deliverable[8]..

Figure 1 Pie chart indicating the range and frequency of mentions of entities in user stories collected by FREYA partners

Observations and how these help to prioritise FREYA's research for new service prototyping:

- The size of the wedge indicates the number of user stories that mention the entity. Moving clockwise from 12, the chart shows that 'instruments' were mentioned most frequently (10 user stories) and 'conferences' mentioned in only one WP3 user story.
- The entities with higher numbers of user story mentions could reflect higher interest from FREYA partners and the communities they serve - and therefore be entities to prioritise for prototyping services. Note that "data", "article" and "person" labels are among the most frequent labels applied to user stories and these also have the most mature PID services in place. Closer scrutiny of the user stories that were assigned "data", "article", "person" and "repository" labels reveals that the user story is usually focussed on entity needing a PID that can subsequently be linked to mature PID infrastructures..
- "Conference", "study" and "project" entities were mentioned least and may reflect less stakeholder interest, or limited cross-discipline interest**.** Unless a FREYA partner could justify a specific interest in building a prototype to address these user stories, these would be excluded for research and prioritisation for this report.
- Entities not listed in the pie graph but for which some research was conducted for this report, include "data management plans (DMPs)" and "facilities" and "research cruises/campaigns". User stories for data management plans (DMPs) were contributed to the collective at a later date. "Facilities" was specified initially as a label for user role, but can also be seen as an "entity" that overlaps to a degree with "instruments". A user story mentioning "research cruises/campaigns" was considered belatedly for this report as a new entity requiring PIDs.

[8] https://doi.org/10.5281/zenodo.1324296

Partner Expertise

> The working group then conducted an exercise to match the PID-related expertise of each FREYA partner with these entities. The aims of the exercise were: to identify partners who are best placed to research specific user stories for this deliverable, and to share the research load meaningfully among the group. The exercise was carried out by representatives of partners attending a typical working group conference call. The group were asked to self-assess the maturity of the PID services for particular entities within their organisations or within the research disciplines they represent. The five-point scale employed for this purpose, was applied so that partners might identify who within a group had the expertise to conduct the research into a specific entity.

The partner organisations reflect specific disciplines and stakeholders. EMBL-EBI represents life sciences; and PANGAEA represents earth sciences. CERN and UKRI-STFC represent the high energy physics community; notably the UKRI-STFC also represent funders in that they award beam-time to the community. DANS represents the social sciences and humanities, which is also represented in part by the British Library.

The idea here is to reveal experienced FREYA partners who could take the lead or be consulted by less experienced partners in gathering requirements for user stories and identifying services that can possibly be prototyped.

	Maturity ranking of PIDs for disciplines **(1 = non-existent, 2 = nascent, 3 = emerging, 4 = in pilot, 5 = mature)**						
Entity (The range matches entity labels applied to user stories)	**EBI** Life sciences	**Datacite** Research data; PID provider	**BL** Cross disci-plinary	**DANS** Social sciences & human-ities	**CERN** High Energy Physics	**UKRI STFC** Facilities science; funder	**Pangaea** Earth sciences
Instrument	**2**	**1**	**1**	**1**	**1**	**1**	**2**
Data	**5**	**5**	**5**	**5**	**5**	**3**	**5**
Grants	**3**	**2**	**2**	**2**	**1**	**1**	**1**
Article	**5**	**5**	**5**	**5**	**5**	**5**	**5**
Person	**5**	**5**	**5**	**5**	**4-5**	**3**	**4-5**
Repository	**1**	**2**	**1**	**1**	**1**	**1**	**1**
Organisation	**2**	**2**	**2**	**2**	**1-2**	**1**	**2**

Sample	**5**	**2**	**1**	**1**	**1**	**1**	**2-3**
Software	**4**	**5**	**1**	**2**	**5**	**2**	**1**
Project	**2**	**1-2**	**1**	**2**	**1**	**1**	**1-2**
Study*	**2**	**1**	**1**	**2**	**1**	**5 ****	**1**
Conference	**1**	**2**	**1**	**1**	**1**	**1**	**1**

* For the purposes of the exercise, 'study' was deemed synonymous with 'investigation', 'experiment' or 'analysis'.

** Specifically, facility investigations

Table 3 Assessment of PID infrastructure maturity conducted by FREYA partners

Observations and how these help to prioritise FREYA's research for new service prototyping:

Where infrastructure was deemed mature by a partner in their organisation or community, it was an indication that that partner also had some expertise in building, hosting or maintaining the infrastructure. With this in mind:

- Varying levels of maturity (and expertise) across partner organisations can be seen for grants, samples and software, revealing obvious leads for the research into these user stories: thus EMBL-EBI could take the lead on grants, and DataCite (along with CERN) on software. For samples, the British Library has led in partnership with PANGAEA. See the notes provided by each partner in the next section.
- There are categories where the maturity of infrastructures was found to be more uniform across disciplines at the time—instruments, organisations, conferences. For these entities, the research was allocated to partners especially looking to prototype services for their communities.
- PID infrastructures for articles, people and data: as seen from the ranking (highlighted in pale green in the table above), PIDs for these entities are most broadly implemented across disciplines, and the infrastructures considered to be most mature. User stories relating specifically to extending the reach of PIDs for articles, data and people were deemed to be the focus of WP4 in FREYA.
- Note that three additional entities were prioritised for inclusion later in this report because of a special interest/ability from partners in the consortium, namely DMPs (by Datacite), facilities (by UKRI-STFC) and research cruises/campaigns (by PANGAEA).

2.3 Validating user stories and gathering requirements

Prioritisation. Handling the user stories according to entity, the partners prioritised the entities for which to collect requirements according to the following criteria:

- there are a number of user stories directly relevant to the entity;
- there is relevant partner expertise and/or partner interest in potentially developing services that address the user story.

The following entities were prioritised for requirements gathering (with lead partner noted in parentheses):

- instruments (PANGAEA)
- facilities (STFC)
- grants (EMBL-EBI)
- organisations (DataCite)
- software (DataCite)
- research campaigns (PANGAEA)
- Data Management Plans (DataCite)
- physical samples and cultural artefacts (British Library)
- conferences (CERN)

Two entities were excluded for requirements gathering for different reasons:

- PID infrastructures for the entities are mature and therefore deemed more relevant to WP4 than WP3. These include the entities: “article”, “data” and “person”
- No current partner expertise and capacity to drive services forward for the entity. This group comprises the entities “repository” and “project”.

Collecting requirements. The aim here was to undertake research that will allow partners to select the most promising candidate PID services for further development. Prototyping is the next task for this group. The requirements gathered here fill the gap between the use cases and what may be possible to be implemented within the timeframe of the FREYA project:

> *“In order to do X, this is what is needed.”*

For each entity, partners conducted the following analysis:

1. **Internal analysis of the user stories**: this provides a synopsis to summarise the essence of the relevant user stories.
2. **Validation**: this provides a summary of the research into the status/current interest in the scholarly community of the new PID or PID service, noting for example, whether there are external interest groups and the status of their work, or how many partner organisations had submitted user stories and whether there is interest from more than one scholarly discipline; whether there are any other relevant external groups/organisations with whom to collaborate.
3. **Potential action by FREYA partners:** this provides the possible action by FREYA partner(s) within the timescale of the FREYA project. If, during the research, the status of work is deemed too premature then a summary is provided of what could be contributed to existing external working groups. Timeframes are mentioned where relevant.
4. **Relationship with other FREYA work packages** : Here for clarity, to mention dependencies on, distinction from and alignment with other WPs where relevant.

This detailed analysis can be found in section 3 of this report.

Note: The sections in chapter 3 that provide details of candidate services for prototyping by FREYA partners in (Instruments, Facilities, Grants, Organisations and Research campaigns) have been updated to clarify the envisaged resolution model and the relevance for EOSC.

2.4 Outreach

Key to the current task is engagement of the wider user community: to inform them of the aims of this working group, to get an independent sense of their priorities and needs, to collect their user stories and identify any relevant external partners who could assist in moving forward the outcomes of this deliverable.

To this end, we list four events below that demonstrate specific outreach around user stories by members of the working group. Note that the outreach work was conducted in collaboration with FREYA partners working specifically on WP5 - iterative engagement of the community. The abstracts for these events can be found in Annex C.

- **A workshop at the Digital Infrastructures for Research 2018** (DI4R) conference in Lisbon, October 2018: in a World Café Session FREYA team members introduced the FREYA project and two of the then current research topics: the identifier landscape and the collection of user stories. In an interactive session on these two topics the community provided feedback on their priorities and challenges that might be addressed via PID services, and were able to contribute their own user stories[9]. The feedback was collected using Mentimeter[10] . Examples of questions asked and community feedback collated via Mentimeter can be seen in Figure 2.
- **A webinar with the FREYA ambassadors: (October 2018).** Early work conducted around user stories was presented, and exchanged information with members of the FREYA ambassador group[11] via Mentimeter and a Q&A session. See Figure 3 for a sample of the feedback received from participants.
- **Joint webinar FREYA and OpenAIRE: New developments in the field of Persistent Identifiers (January 2019):** This included a presentation on the user story approach to identify which PIDs are needed most by scholarly communities, what requirements and dependencies exist and thus which PID services can be developed by partners within the lifetime of the FREYA project.
- **A presentation at PIDapalooza 2019** (Dublin, January 2019): an interactive session with participants focussed on developments of the PID Graph, the collected user stories for new PID types, and the latest news on FREYA's PID Forum[12]. An interactive forum for the community, the *pidforum.org* was also launched at this meeting by FREYA partners in collaboration with members of ORCID, Datacite and Crossref. This forum is open for general dialogue to anyone interested in PIDs[13]. The user stories collected by FREYA partners have been uploaded to encourage feedback.

Figure 2, Figure 3 and the text boxes below offer examples of the feedback captured from community members who participated in outreach events.

[9] Link to a brief video recording summary of the session https://www.youtube.com/watch?v=NJsIc7jA7Hs
[10] https://www.mentimeter.com/features
[11] https://www.project-freya.eu/en/ambassadors/our-ambassadors
[12] https://www.project-freya.eu/en/engagement/pid-forum
[13] https://doi.org/10.5281/zenodo.2548636

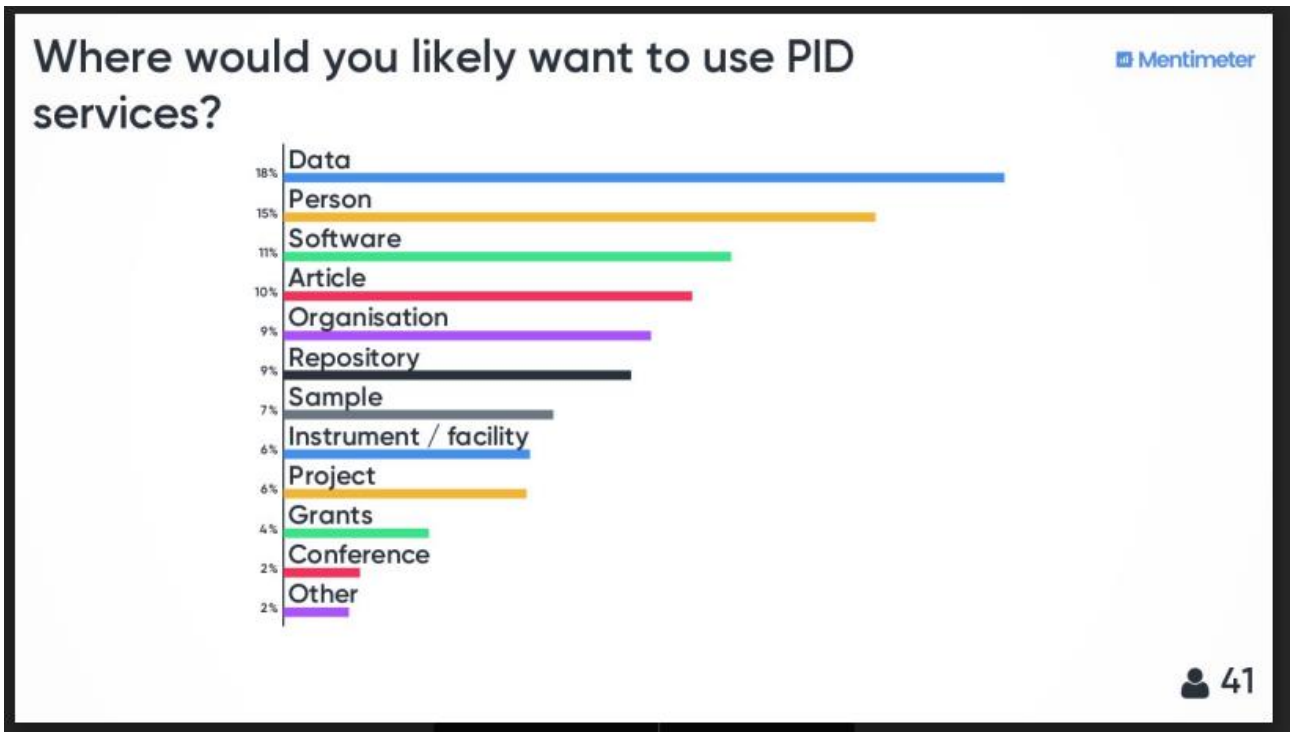


Figure 2 Voting-style feedback captured from participants during the DI4R 2018 interactive workshop

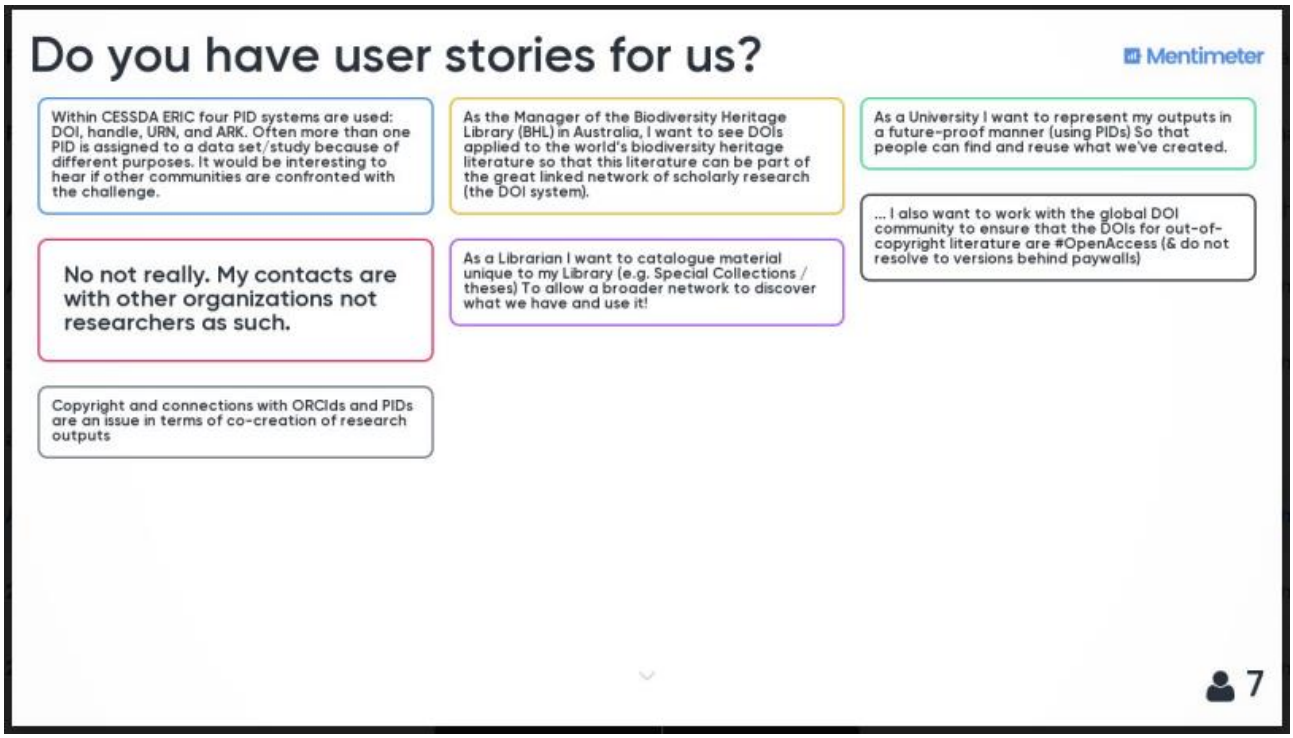


Figure 3 Free-text feedback during the FREYA Ambassadors webinar captured via Mentimeter. Seven answers were provided by five of the six ambassadors who were participating in this particular webinar.

DI4R attracted the data infrastructure community. 78 workshop delegates signed onto mentimeter, 41 of whom responded to the question shown in Figure 2. A single delegate could vote for PID services in more than one entity. The responders identified themselves as follows: PID newcomers (7), PID enthusiasts (10), PID users (6), PID producers (6), PID innovators (4), or did not identify themselves (8)

While the methods of measurement are not directly comparable, it is interesting to compare and contrast the entities that DI4R delegates wish to link going forwards, versus those indicated in the user stories collected by the FREYA partners.

Snapshot of general issues raised by Ambassadors during the October 2018 webinar (courtesy of Barbara Lemon, formerly at the British Library[14]**)**

- User stories raised more ethical and administrative issues than they did technological ones
- Issues raised to consider included:
    - Several PID types being assigned to the same dataset (where's the limit?)
    - DOIs being assigned to out-of-copyright materials by commercial entities;
    - Catering for creators or authors outside of the "researcher" mould (e.g. indigenous knowledge creators, computers, artists);
    - Possibility of incorporating existing globally recognised identification or numerical systems as part of PID systems (e.g. opus numbers for musical works)
    - Making non-persistent identifiers (such as those used in institutions) persistent
    - Is FREYA connecting with Wikipedia and their work with PIDs?
- This group uses primarily DOIs, also ISSNs and ORCID iDs, using them for articles, theses, research data repository, libraries, grey literature, in-house publications, monographs, museum publications, datasets.
- Interested in stronger links between, for example, specimens and organisations and data; codebooks and data.

[14] Currently working with the National and State Libraries Australia

3 Deep-dive analysis of user stories

Of the twenty-five entities identified in our previous PID landscape analysis[15], nine have user stories where PIDs need to be newly assigned. Those cases have been analysed further here by FREYA partners with the relevant expertise or specific interest in the entity. The following entities are discussed (with lead partner noted in parentheses):

- instruments (PANGAEA)
- facilities (STFC)
- grants (EMBL-EBI)
- organisations (DataCite)
- software (DataCite)
- research campaigns (PANGAEA)
- Data Management Plans (DataCite)
- physical samples and cultural artefacts (British Library)
- conferences (CERN)

3.1 Instruments

3.1.1 Synopsis

Instruments form a central connection node in field-based and lab-based quantitative research. Persistent identifiers for instruments are needed so that their identity can be included in metadata, letting data users decide on data compatibility, quality and measurement precision. The interest lies in connecting measurement data to the instrument with which it was taken, for the benefit of improved data provenance. Beyond that, many of the instrument-related user stories (see text box) focus on instrument owners and producers wanting to trace the use and output of their instruments beyond raw data, in the form of researcher careers and publications. Linking instrument-PIDs to other essential PIDs for example to DOIs for publication and data via a PID-graph, would significantly improve the amount for information available (Figure 4).

Furthermore, implementing PIDs for instruments in dataset metadata would greatly increase our ability to combine data from different sources and thereby the reusability of measurement data.

There are some first efforts for instrument PIDs in institutional contexts (e.g. *sensor.awi*), but no wider use throughout the community. Essential for wider implementation would be the willingness of instrument producers to commit to a registration process for their instruments and adherence to a metadata standard for instrument descriptions.

User stories requiring a new instrument PID

#55: “As a researcher, I would like to track other researchers who are using the same scientific instrument as me, and get access to their scientific outcome (data, publications, samples, genetic-markers etc).”

#57: “As a facility, I would like to track the published output related to the instruments provided. Also, i would like to be able to evaluate their impact.”

[15] https://doi.org/10.5281/zenodo.1324296

#64: "As a researcher, I want references to instruments in (meta)data published by data repositories to be actionable so that I am unambiguously redirected to metadata about the instruments which enables me to learn more about instruments and the context in which data were acquired."

#65: "As a researcher, I want to discover data by an instrument mentioned in a paper I just read because that data may be useful in my research."

#68: "As an infrastructure provider I want to be able to track people associated with my instruments, equipment and services so I can follow their careers."

#75: "I am a company producing scientific instruments and/or software. For a marketing analysis, I would like to trace the current use of our products (instrument/software-PID) across scientific disciplines and geographical areas by analyzing article and data publications (publication-PID/data-PID) specifically produced using our products (instrument/software-PID)."

#87: "As a researcher, I would like to find all data in the repository produced by a specific instrument/sensor on a research vessel . To decide whether data is compatible with other data from the same or similar instruments I would also need an actionable link to the measurement protocol or DOI of best practice document."

(The user story number refers to the github issue #).

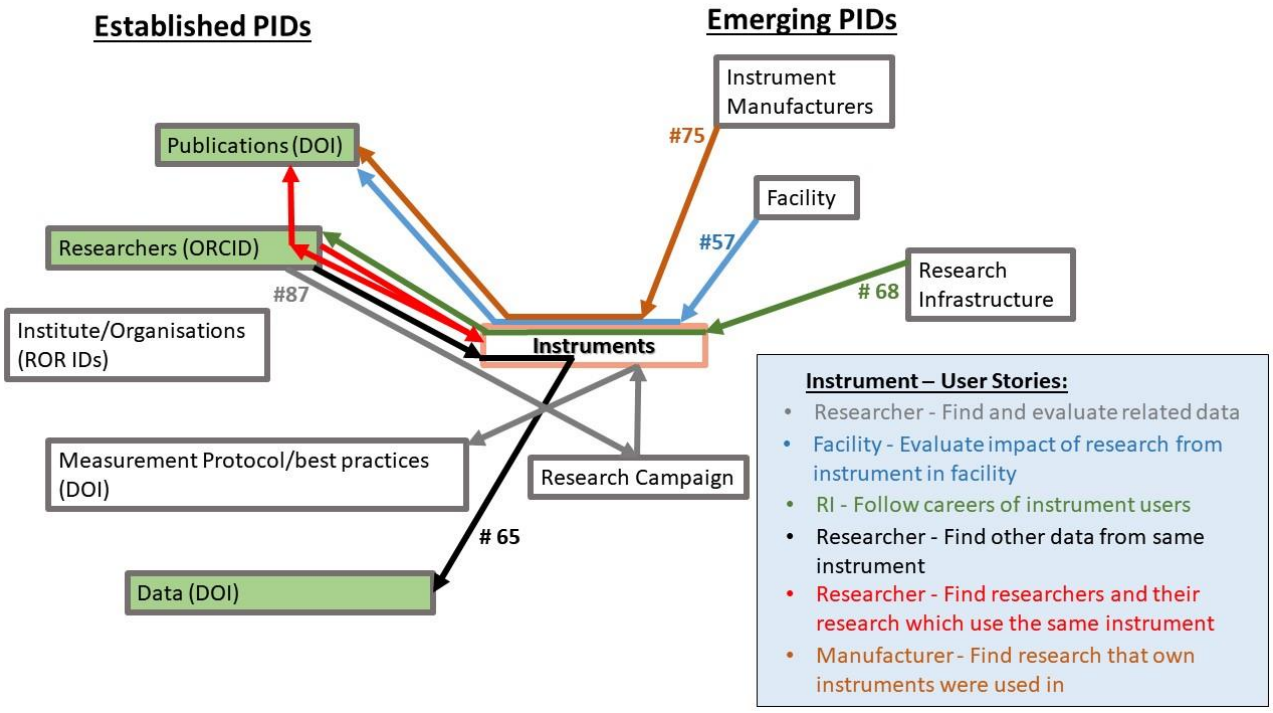


Figure 4 A PID-graph illustrating links needed between instrument-PIDs and other relevant PIDs to enable instrument-related user stories. Green metadata components indicate mature PIDs currently in use.

3.1.2 Validation

The RDA 'Persistent Identification of Instruments Working Group (WG)'[16] has recently completed a first metadata schema for instruments. The work is based on use cases from 13 different research organizations,

[16] https://www.rd-alliance.org/groups/persistent-identification-instruments-wg

that have described their instrument PID needs to the RDA WG. Submitters include research institutes and networks, EU research infrastructures, data repositories, library services, PID providers, and international initiatives. The collection of international use cases shows that the need for instrument PIDs from a user perspective is very high. Currently missing is any involvement from instrument producers. Initial contacts from the WG to industrial partners have indicated, that for the producers, the financial benefit of instrument PIDs need to be clearly demonstrated, to warrant any investment on their part.

The WG will expire in its current form in 2019 and a new WG or other more persistent format may need to replace it. A critical question remains of how to drive the adoption of instrument PIDs, if instrument producers, do not consent to 'PIDing' their products prior to distribution, despite the benefits of tracing usage and users for marketing purposes.

3.1.3 Possible action by FREYA partners

FREYA partners PANGAEA and ORCID are already part of the RDA WG for instrument PIDs and will contribute to finding the next suitable format for the WG initiative. To further support the group's output, PANGAEA will create a simple survey to test the applicability of the metadata schema to individual and complex systems of instruments. PANGAEA will use large aggregations of technology vendors alongside research conferences to perform these surveys. The surveys will evaluate two things: how well the metadata schema can be completed using the general specifications provided with the instrument; and how able and willing vendors are to describe their instrument using the suggested instrument metadata schema. These efforts will allow us to improve upon the schema, and to provide feedback on the general sentiment of industrial partners regarding PIDs for instruments. Overall, the work can give direction to future efforts, by reflecting stakeholder and community needs and working with and around obstacles which could slow the adoption.

PANGAEA also plans to include Instrument-IDs from *sensor.awi* for the fixed instruments on German research vessels in dataset metadata within PANGAEA. Building the PID-graph within PANGAEA, we will use the instruments as a node to connect to other similar data, researchers, cruises etc. in the database. This is an ambitious goal, but carries tangible progress if achieved. The RDA IG for Instrument PIDs is working on finding a PID provider for instrument PIDs and has developed draft metadata schema to go with the PID registration. As soon as this service is available, we will add these instrument PIDs to the persistent IDs given out by sensor.awi.in the dataset metadata.

3.1.4 Relationship with other FREYA work packages

Currently no dependencies.

3.1.5 Envisioned resolution model and the relevance for EOSC

A PID-system for scientific instrumentation is still at developmental stage, and there is no universal solution that is at a technological readiness level suitable for implementation in the EOSC services registry or any other EOSC platforms at present time.

The development is ongoing under the auspices of RDA. Currently, the scalability of an instrument PID system is being investigated through a mapping exercise, where institutional/organizational Instrument identifier systems are being sought mapped to a universal metadata-schema from DataCite. This is an important first step laying the grounds for a future universal scientific instrument PID system. The envisioned resolution model will expand Instrument PID usage to a universal level that encompasses the entire range of scientific disciplines generating data using scientific instruments. Only constraint is that the instrument shall be a "measuring instrument" definable as a "instrument providing an output signal carrying information about the value of the quantity being measured", in accordance with the guidelines provided by the Joint Committee for Guides in Metrology (JCGM). When such a PID system has been widely accepted, it will be a valuable tool for the curation of the outcome of scientific instruments and expand the Altmetrics of scientific resources, adding a new layer providing information about origin of scientific data

and embracing open science practices. This is important for the researcher services envisioned in EOSC, where PIDs for instruments will expand the search & browse functionally and recommender system, allowing research to identify scientific resources generated using the same equipment.

3.2 Facilities

3.2.1 Synopsis

Large-scale research facilities such as neutron sources or synchrotron radiation sources are natural hubs of multidisciplinary research, and also typically have an obligation of long-term preservation of experimental data and of records of science (structured descriptions of investigations performed by visitor scientists). This makes a facility an important element of research data sharing and research publication workflows. Having persistent identifiers for facilities will contribute to the Open Science agenda of facilities themselves, and of research institutions and publications that refer to data collected in facility experiments. It could also contribute to more structured provenance information for experimental data that is subsequently made public by DataCite, national and European funding portals, subject-specific databases and other research information systems.

The following user stories relate, either directly or indirectly, to facilities research and may benefit from having PIDs for facilities.

User stories requiring a Facility ID

#83 “As a user of [facility/resource/archive] I want to connect my use of the [facility/resource/archive] to outputs resulting from my work there.”

#70 “As a facility User Office interested in measuring the facility impact, or as a reviewer of research proposals (who evaluate applications for facility beamtime), or as a funder who supports the facility with public money or industrial contribution, I am interested in linking facility awards (beamtime) with structured records in renowned scientific databases, such as biomedical or crystallography databases. Ideally, these links and measures should be as granular as possible: what database record(s) resulted from what particular facility award(s).”

#65 “As a researcher, I want to discover data by an instrument mentioned in a paper I just read because that data may be useful in my research” (Facility as a whole can be considered an instrument, also facility beamlines are in fact large-scale instruments)

#57 “As a facility, I would like to track the published output related to the instruments provided. Also, I would like to be able to evaluate their impact.”

#55 “As a researcher, I would like to track other researchers who are using the same scientific instrument as me, and get access to their scientific outcome (data, publications, samples, genetic-markers etc). – for inspiration, validation and collaboration.”

#33 “As a core facility provider, I want to track usage of my facility so that I can demonstrate its value.”

#27 “As a staff member at STFC, I want to see all the publications based on raw data generated in our facilities, so that I can demonstrate the impact of the services provided by us.”

(The user story number refers to the github issue #).

3.2.2 Validation

Community interest in PIDs for facilities is evident in several joint initiatives:

A project lead by ORCID® on acknowledging research resources[17] in general, includes the requirement to attribute research support to facilities. There are ongoing discussions with UKRI-STFC and other facilities, who recognise the need for identifiers for facilities and their large-scale instruments (beamlines).

The PaNOSC project (Photon and Neutron Open Science Cloud)[18], recently funded under the Horizon 2020 e-infrastructure programme, aims to align the efforts of the existing and new photon and neutron sources to link up to the EOSC. It recognises that scientists are increasingly using several different infrastructures to perform their research, creating a need for unambiguous identification of facilities.

The RDA Interest Group on 'Research data needs of the Photon and Neutron Science community'[19] is an additional forum collecting feedback from and establishing best practices for obtaining research attribution for facilities. This Interest Group discusses data-related issues of science applications associated with large-scale research facilities that are shared and used by many research teams from different branches of science. The disciplinary diversity and the global character of facilities research makes standardization and interoperability even more challenging. A representative from UKRI-STFC co-chairs this RDA group, which provides further means to encourage elaboration of relevant user stories and promote best practices for the use of Facility PIDs.

3.2.3 Possible action by FREYA partners

A user story that underlines the need for a facilities identifier is #70 "Linking facilities research to records in scientific databases". Two of the FREYA partners, STFC and EMBL-EBI, have been working on requirements to address this user story which seeks to link facility time (awarded by the Diamond synchrotron facility), with resulting records in biomedical databases based at EMBL-EBI. In the first instance, the aim would be to link past facility awards to the data and published research articles that have resulted from those awards. STFC is considering PURLs (persistent URLs[20]) as a viable starting option for facility PIDs. While a solution for managing landing pages for facility PIDs is yet to be decided, it may either rely on using the existing facilities websites or using a reliable repository back-end. Bibliographic records for research articles that reference Diamond synchrotron data are currently made available in the Diamond publications database. These will be linked to Diamond synchrotron-derived records in the Protein Data Bank in Europe (PDBe) and in EuropePMC, both situated at EMBL-EBI, with the aim of enriching the records that stem from the Diamond synchrotron. UKRI-STFC and the EMBL-EBI aim to deliver a pilot demonstration of these links by the end of 2019. The work on facility PIDs relates closely to work on instrument PIDs (by PANGAEA) and therefore exchange of insights between these FREYA partners will be key to progress. Discussion of the metadata to be associated with a facility PID will require consultation with a few FREYA partners. This is because the multifaceted nature of facility operation means it can behave like an organization, a funder or large-scale Instrument depending on the information context. This discussion will also benefit from contributions from a large photon and neutron sources community, for which the aforementioned RDA Interest Group on research data needs of the Photon and Neutron Science community is a reasonable forum.

3.2.4 Relationship with other FREYA work packages

Improved provenance information that comes from linking UKRI-STFC to EMBL-EBI records will contribute to work on metadata recommendations and common APIs, to be conducted under WP2 and WP4. Extending PID graphs to include facility PIDs when available would be a focus of WP4.

[17] https://orcid.org/blog/2018/04/10/acknowledging-research-resources-new-orcid-data-model
[18] https://eoscpilot.eu/content/photon-and-neutron-open-science-cloud-part-european-open-science-cloud
[19] https://www.rd-alliance.org/groups/research-data-needs-photon-and-neutron-science-community.html
[20] https://archive.org/services/purl/help

3.2.5 Envisioned resolution model and the relevance for EOSC

As explained above, Facility may behave as an Organization or as a large-scale Instrument depending on the context where it is mentioned. In the user stories developed in FREYA, organizational aspect of facilities prevail, therefore organizational PIDs will be used to designate facilities, with the emphasis on RORs. The resolution of RORs will rely on the model offered by the ROR consortium: currently, this is a single provider model. Assigning RORs to facilities will allow circulation of unambiguous references to facilities in EOSC services, e.g. in OpenAIRE that will have a better opportunity to connect research outputs such as articles to facilities.

The instrumental aspect of facilities will be explored through collaboration with the RDA Instrument PIDs Working Group. STFC is going to run a small-scale pilot in order to adopt this Group recommendations; the principles of modelling Facility entities in the pilot will be decided upon later. If modelled as Instruments, the resolution model for Facilities is likely to be decentralized (federated) with individual organizations minting PIDs according to the RDA Group recommendations; using Handles or DOIs will ensure the uniqueness of the identifiers. Modelling of Facilities as Instruments will be important in EOSC services that use references to facilities in software, e.g. in workflow management platforms that handle data collection on facilities, and sharing this data (with the initial embargo period).

Overall, the services supported by facility PIDs will fall in the two major EOSC services categories: Resource Registries and APIs. Occasionally, they can also be used in other service categories, e.g. in AAI services to decide on the user permissions.

3.3 Grants

3.3.1 Synopsis

A large proportion of the user stories[21] collated relate to “grants” and to “funders”(# 39, 43, 45, 56, 62, 69, 70) . At the heart of these user stories is the need to link a grant (award) to its output, whether that be a physical sample collected, software; data, database record, thesis, publication. For the funder, this would allow a measure of usefulness of that allotted award and acknowledgement to the source of that award, providing a starting point for analysis of impact of funding. For the researcher it would allow a measure of productivity linked to that award.

While every funder has an internal identifier allocated to each of their grants, there is currently no global unique identifier system in place for grants which leads to many potential ambiguities across funders. A first step would be to set up a grant identifier.

User stories requiring a new Grant PID

As a funder in HEP, I would like to measure the impact of my grants, i.e. did the funding to specific software projects lead to more shared code and/or research outputs?

“As a funder, we want to be able to find all the outputs related to our awarded grants, including block grants such as doctoral training grants, for management info and looking at impact”

“As a funder, we want to be able to identify who (including orgs and individuals) benefitted from a given grant, for boosting management info and for looking at impact”

“As a researcher, I want to acknowledge a particular grant in funding the creation of my software.”

[21] See Appendix 1.

"As a funder, I want to know what software has been developed from a project I have funded."

'As a funder who supports the facility with public money or industrial contribution, I am interested in linking facility awards (beamtime) with structured records in renowned scientific databases, such as biomedical or crystallography databases.

As a funder, I want to track down the outcomes and beneficiaries of PhD studentship awards that I granted.

As a funding agency, I would like to trace the outcome of my financial contribution to a marine research cruise (Cruise ID) by tracking the data generated (data-PID) and articles (publication-PID) published, as well as physical samples taken (IGSN) and the repository (organization ID), where these samples are physically stored.

(The user story number refers to the github issue #).

3.3.2 Validation

Community interest in developing such a system is manifest in the Grant Identifier Initiative[22] which is being run by Crossref (a scholarly content registration agency) in consultation with funders such as the Wellcome Trust, JST, ERC, SNSF, DOE-OSTI, and NIH. Thus far, there has been agreement that the PID will be a DOI, which will be assigned for every grant awarded. It will also be accompanied by a dedicated widget for systems to integrate, an important requirement in helping researchers submit outputs, and in reducing errors. There are two working groups within the project: one to look into governance, participation, and fees for the introduction of grant IDs and the second more technical group to look into the schema for registering grant records and the metadata associated with each grant that should be made available. The project's proposed sustainability model—including fees and membership—was approved in November 2018, and the starting schema is to be finalised Q1 2019.

3.3.3 Possible action by FREYA partners

Two FREYA partners are involved in the community initiative mentioned above: Crossref and EMBL-EBI. By mid-2019, the expectation is that DOIs will be registered for an initial cohort of grants from early adopters. For funders of Europe PMC, EMBL-EBI will create these DOIs and add them to existing grants available in their Grants Finder Repository and integrate the DOIs within the links it already provides between grants and the biomedical literature aggregated in the repository.

This work could serve as a model for other partners to integrate Grant IDs in their workflows.

3.3.4 Relationship with other FREYA work packages

Since the work contributed by Crossref and EMBL-EBI around Grant IDs involve existing TRL8-level services, this relates to both WP2 and WP4. Many of the user stories that mention grants refer to impact assessment. By definition this means that these initial grant-output links need to connect to further research outputs in the PID Graph (WP4). The community engagement group (WP5) could have a role to communicate updates to potential stakeholders and promote adoption of Grant IDs.

[22] https://www.crossref.org/blog/global-persistent-identifiers-for-grants-awards-and-facilities/

3.3.5 Envisioned resolution model and the relevance for EOSC

There are two components to this initiative:

1. Grant DOI registration service: Crossref (an unfunded FREYA partner) has set out a funder membership scheme and developed a metadata schema for grant DOI registration. This will be maintained and governed by Crossref.
2. Early adoption of Grant DOIs into an existing grant finder service - Europe PMC hosts an existing grant finding tool and landing pages for grants awarded by its 29 funders. With the help of FREYA funding, Europe PMC is developing the means to add Grant DOIs and additional associated metadata to the existing records and tool. Initial implementation involves Wellcome Trust awarded grants for 2019. This data will be made freely available to all users in and beyond EOSC via the web-service (API) and Europe PMC webpages.

   While Crossref will provide the central registration service for Grant DOIs, a Grant DOI is envisaged to resolve to an openly available landing page of a specific service provider, Europe PMC being one such service provider. Going forwards it is envisaged there may be multiple service providers who will host the landing pages for grants from other funders.

Overall, the services developed above to implement Grant IDs are envisaged to be discoverable and accessible via EOSC hub's proposed Service Portfolio (the layer of services that are independently owned by their respective providers) and could fall in the two major EOSC services categories: "Resource Registries" and "APIs".

3.4 Organisations

3.4.1 Synopsis

PIDs for organisations are key to much of the interconnection work in FREYA. As such, they are implicit in addressing many of the other user stories collected by FREYA across all types of PIDs. A handful of users stories collected were specific to organizations, but other user stories elsewhere in this document will also rely heavily on the author affiliation connections provided by PIDs for organisations.

User stories related to organization PIDs

#45 "As a funder, we want to be able to find all the outputs related to our awarded grants, including block grants such as doctoral training grants, for management info and looking at impact" "As a funder, we want to be able to identify who (including orgs and individuals) benefitted from a given grant, for boosting management info and for looking at impact"

#68 As a funder, I want to track down the outcomes and beneficiaries of PhD studentship awards that I granted. There are many possible questions to be answered (with the help of PID graph): a. Whether the PhD studentship actually ended in a thesis; b. What organisations benefited from the PhD during or soon after the PhD research period, e.g. by hiring the PhD I sponsored; c. who co-funded or otherwise supported the PhD research; d. What artefacts (papers, data, software, samples, instruments,...) can be identified that either contributed to the PhD research or are the PhD research outcomes.

#71 As an owner or an operator of an institutional publications, data or software repository I am interested in gap analysis between my repository and other repositories of similar kinds. An example could be STFC ePubs repository (for publications) that has to operate in the diverse and ever changing world of information where other repositories potentially capturing STFC employees' publications exist: Zenodo, INSPIRE-HEP, preprint services. PIDs (for people, institutions, papers, potentially grants and projects, too) may help to identify gaps between what is captured by a local repository and what is captured elsewhere. The gap analysis can be the first step to further actions, such as: ingest records from

outer sources, or link to them, or merge with them, or simply disregard them (if they are somehow "out of scope").

(The user story number refers to the github issue #).

3.4.2 Validation

There is demonstrated interest in the topic of PIDs for organisations, as evidenced by the formation of the ROR (Research Organization Registry) initiative[23], which contains several FREYA partners as co-founders and contributors. While ROR is a separate initiative to FREYA, both sides can inform each other to improve the overall state of PID infrastructure.

A lean implementation group for launching the ROR identifier began in October 2018 and delivered a minimum viable registry at the ROR Stakeholder Meeting alongside the PIDapalooza conference in January 2019. ROR IDs are ready to be used and evaluated by the broader community, and the FREYA partners can provide a valuable testbed for exploring, improving, and expanding the capabilities of the ROR registry.

3.4.3 Possible action by FREYA partners

With ROR IDs ready and available to the broader PID community, the FREYA partners have the opportunity to build upon that work. A first step that would be useful to both the ROR initiative and to FREYA would be disambiguation projects on the part of the disciplinary partners. Such projects would simultaneously provide validation for the ROR registry and would hopefully solve the primary problem of successfully and accurately identifying research object creator affiliations in repositories.

For example, organisation identifiers would create a rich layer of linkages between datasets, theses and authors in the British Library's EThOS system, as well as stronger links to institutional repositories in the UK tertiary sector. Organisation identifiers would also assist in the development of the British Library's shared institutional repository, differentiating publishers from source institutions, holding institutions, author-affiliated institutions, etc. The British Library is also exploring this functionality with International Standard Name Identifiers (ISNI).

Following the validation provided by such disambiguation projects, FREYA partners could expand the metadata submitted to their repositories and subsequently registered with PID registration agencies like DataCite and Crossref. In turn these registration agencies could update their schemas to better accommodate the ROR ID as an accepted PID as part of a controlled vocabulary.

3.4.4 Envisioned resolution model and the relevance for EOSC

The ROR identifier will not be a DOI, but rather a distinct identifier of its own. Similar to a DOI, the ROR ID is backed by a registry, though ROR IDs resolve to the record in the registry for the entity that is being identified, rather than resolving to another URL provided by the registry.

ROR is currently run by a steering group comprised of Crossref, DataCite, Digital Science, and the California Digital Library, which is the same team behind the building and maintenance of the ROR registry. ROR is intended to be a community initiative, and as such the steering group is supported by a network of Community Advisors, Signatories, and Supporters who variously provide community input as well as in-kind and financial support.

[23] https://ror.org/

The ROR organisational identifier is itself a service that we will list in the EOSC services registry and that is available for providers of other downstream data services to make use of when enhancing their service offerings and when making connections between services. The particular tasks taken on by FREYA partners to disambiguate organisations with ROR and to expand metadata that is submitted to PID registration agencies will further enhance the information available through other EOSC-listed services. Specifically, the wider use of ROR as an organisational identifier in multiple data repositories and systems will make data that is associated with particular institutions more easily discoverable, both via EOSC tools and services and beyond.

3.5 Software

3.5.1 Synopsis

The user stories related to software PIDs that were gathered by the FREYA partners could be grouped into a few themes: software citation, software contribution and authorship, analysis of specific datasets, and aggregation of software versions.

User stories related to software PIDs

Software citation

#48 As a HEP researcher, I want to know who (author) used my software and for what purpose (their paper).

#49 As a HEP researcher, I want to know which results/paper has been produced with which software.

Contribution

#51 As an institution, I want to track the outputs of all affiliated researchers. This concerns papers, data, code and their impact (citations), but also contributions to specific conferences.

#61 As a young scientist I would like that my contribution to a publication is distinguishable from my co-writers' contributions, e.g. that it is clear who contributed to the code, data, analysis etc.

Analysis of specific datasets

#39 "As a researcher in the digital humanities, I want to analyse British Library datasets using software that has been developed specifically for those datasets and logged in GitHub." "As a researcher, I want to acknowledge a particular grant in funding the creation of my software." "As a funder, I want to know what software has been developed from a project I have funded."

Aggregating software versions

#63 As a software author, I want to able to see the citations of my software aggregated across all versions. so that I see a complete picture of reuse.

(The user story number refers to the github issue #.)

In general, these user stories were not requesting a new PID for software, but instead centred on making better use of existing PIDs for software. Collectively, these user stories operated on the assumption that software should be recognized and rewarded as a standard part of a researcher's outputs, in much the

same way as other research objects that currently have PIDs, such as articles or datasets. For example, user story number 61[24] is largely more relevant to use of the CreDiT taxonomy[25] than a specific PID, though assigning PIDs to software and interlinking PIDs describing multiple research object types is an assumed step toward making such microcontributions feasible. Though the user stories related to software citation and software contribution mentioned software specifically, this is more a matter of extending mechanisms of interconnection to software PIDs as well as other PID types, rather than inventing new mechanisms from scratch.

In contrast, the user stories about analysis of specific datasets and about aggregating reuse of software across versions each acknowledged a nuance of tracking citation and reuse that is specific to software.

The user story about analysis of specific datasets was contributed by the British Library from a digital humanities context. It is referring to the potential for linking research objects, such as datasets, with the software that was developed specifically to analyse that research object, as well as the documentation for said software. This user story presents a workflow challenge, as there could be significant intervals of time between the DOI for the dataset being created and a DOI being assigned to the interpreting software once deposited in Zenodo. The documentation may also only be published after the software itself has been archived.

The user story about aggregating reuse of software across versions addresses a disparity between software citation practices that are beneficial for a reuser and those that are beneficial for a software author. DOIs assigned to software typically point to a specific version of the software that was used, to avoid ambiguity when attempting to reproduce results. While this makes sense for a software reuser, individual DOIs for individual versions can quickly become cumbersome and undesirable for software authors wishing to concisely demonstrate their wider impacts to the public or to tenure committees. In this way, the concerns of aggregating and citing different versions of the same piece of software are not far removed from the concerns of citing dynamic datasets, so solving these challenges for software may be beneficial to other areas of data citation broadly. For software authors, it would be preferable to have a way to aggregate the reuse of all versions of their software, for purposes of either citation count or display.

3.5.2 Validation

The discussion of PIDs for software is still ongoing. While the FREYA partners did not identify a need for a new software PID type in their own institutions and workflows, PIDs other than the DOI for software are being actively discussed, such as git commit hashes. A relevant RDA working group will soon begin investigating this topic[26].

There are other active initiatives seeking to improve workflows around software citation, such as CodeMeta[27] and Citation File Format[28]. These initiatives demonstrate that there is interest in the topic of software citation in the broader community.

3.5.3 Possible action by FREYA partners

The FREYA disciplinary partners could take on a range of actions particular to their own needs, while being supported by relevant changes on the part of the infrastructural partners. A case in point is that of the bespoke digital humanities software user story (#39[29]) submitted by the British Library. The British Library is planning to promote the use of GitHub across the institution for managing software and will encourage researchers to deposit key versions of software in a repository such as Zenodo at particular milestones,

[24] https://github.com/datacite/freya/issues/61
[25] https://casrai.org/credit/
[26] https://www.rd-alliance.org/groups/software-source-code-identification-wg
[27] https://codemeta.github.io/
[28] https://citation-file-format.github.io/
[29] https://github.com/datacite/freya/issues/39

such as corresponding with a publication. In addition, addressing this user story will require a workflow to ensure that related objects are accurately connected via PIDs, which could be bolstered by improving backward linking capabilities between related PIDs. Besides technical mechanisms, appropriately addressing this user story may require improvements in metadata support for expanded software relation types, such as a research object to software relationship or a research object to documentation relationship.

Similarly, the FREYA project can benefit from and build on previous work undertaken by the project's partners. For instance, Zenodo, a multidisciplinary research data repository run by CERN, already mints DOIs for software records from GitHub. Recently, Zenodo started to implement software citation metrics[30], which is a first step towards a realization of the user stories related to software citation. The Zenodo citation metrics project primarily covers citations to astronomy and astrophysics resources, though citations to every record are displayed. Citations are fetched by the brokering software "Asclepias Broker"[31], which harvests data from DataCite, Crossref Event Data, and the NASA Astrophysics Data System. Zenodo by default rolls-up citations to all versions of software records in order to show its full impact. A filter is offered to show only citations to a specific version of a record. This approach to aggregation of software citations could be a good example for the other FREYA partners to follow, either in disciplinary services or as a model for extending this concept to services provided by the infrastructural partners.

However, the Zenodo software citation service is still in beta phase because of the difficulties of providing reliable citation data with high coverage. Right now, the citation coverage is quite low as formal software citations are not the norm, and the coverage relies on publishers to make citation data freely available, which is something that initiatives like the I4OC[32] are focused on. At launch (January 2019), around 2500 records have a minimum of one citation with a total of around 3500 total citations (about half from ADS and half from Crossref/DataCite Event Data). Only around 250 out of the 3500 citations are known by both systems.[33] Zenodo is already expanding its coverage to include further data resources. An early example is the addition of citations harvested from the literature repository, EuropePMC, data for which was presented at PIDapalooza 2019[34].

3.5.4 Relationship with other FREYA work packages

Because the use and citation of software comes with different needs and contexts depending on the discipline, possible actions to address software PID user stories may extend into WP4, which focuses on implementations in disciplinary contexts. Validation for any new developed services, whether disciplinary or otherwise, will benefit from collaboration with WP5 and the PID Forum.

3.6 Research campaigns

User story requiring a new research campaign PID

#62 As a funding agency, I would like to trace the outcome of my financial contribution to a marine research cruise (Cruise ID) by tracking the data generated (data-PID) and articles (publication-PID) published, as well as physical samples taken (IGSN) and the repository (organization ID), where these samples are physically stored. In this regard, I would also like to track the future data and publications generated from these samples. https://github.com/datacite/freya/issues/62

[30] Nielsen, L.H. (2019): Software citations now available in Zenodo. Zenodo Blog. URL: http://blog.zenodo.org/2019/01/10/2019-01-10-asclepias/
[31] https://asclepias-broker.readthedocs.io/en/latest/ (last checked on 11.01.2019)
[32] https://i4oc.org/#
[33] http://help.zenodo.org/#citations (last checked on 11.01.2019)
[34] https://zenodo.org/record/2548643#.XFG0_M_7TUJ See slides 37-43

(The user story number refers to the github issue #.)

3.6.1 Synopsis

A research campaign is unique research event that can be described with a specific purpose and which is set in time and space. The campaign can involve many researchers from various organization and involve multiple projects. In this deliverable, our use-case comes from marine science revolving around research cruise conducted with larger research-vessels.

One user story comes from marine science and involves tracing the outcome of research-cruises through linking cruise-IDs with PIDs for articles, data and samples. Although this PID type is only represented by one use-case, FREYA-partner, PANGAEA has experience in the implementation of identifiers for research-cruises in their data-sets from marine expeditions.

There is no universal PID for research cruises. However, given that larger research vessels comprise a substantial financial investment, identifiers for research cruises are often developed on a national level, and these IDs are to some extent implemented to track outcomes such as cruise-reports, publications and data-sets. While the experience drawn from working with national level identifiers would be highly relevant to generating a universal PID-system, there are pros and cons to consider: implementation will require replacement of existing identifier systems operated by many larger national fleets of research-vessels;the task would require significant time and funding. On the other hand, a universal persistent identifier for research cruises would significantly improve the cross-disciplinary discoverability and traceability of outcomes from what is commonly an international collaborative effort with participation and financing from multiple countries.

3.6.2 Validation

Countries with larger research fleets usually have their own “national” identifier systems. In Germany, The Federal Maritime and Hydrographic Agency (Bundesamt für Seeschifffahrt und Hydrographie, BSH) maintains a catalog of research cruises conducted by the fleet of major German research-vessels. Each cruise is assigned a cruise number, which is a alphanumeric identifier referring to the specific research vessel and the time-frame of the cruise. Furthermore, a separate identifier (a DOD-ref-No.) is added referring to the Inventory of the specific cruise. The latter ID is actionable in that it resolves to landing page of the cruise inventory, where additional information such as researcher, institute and data can be found. However, these entities are not represented by PIDs. PANGAEA (data publisher and FREYA partner) curates and publishes much of the data collected from these research cruises. As part of the curation process, the cruise-ID and DOD-ref-No. is implemented in the published data, making the outcome of specific cruises searchable in the PANGAEA database and linked to PIDs for Author (ORCID), data (DOI), articles (DOI) and samples (IGSN). However as the Cruise-IDs are not a universal PID, these links only exist within the domain of PANGAEA.

In the USA, the linking of cruise-IDs with other PIDs is further along. The Rolling Deck to Repository (R2R) program maintains a master catalog of research cruises conducted by the US fleet of research-vessels. The catalog currently has over 7,000 expeditions and continuously records cruises from 26 active vessels. In the R2R complete cruises are assigned a Digital Object Identifier (DOI), which are linked to PIDs such as DOIs for datasets, journal articles, participating researchers (ORCID), samples collected on the cruise (IGSN) as well as the funders (Crossref Funder Registry).

3.6.3 Possible action by FREYA partners

Within FREYA, PANGAEA is in the process of implementing PIDs for research campaigns primarily from the German fleet of Research Vessels and is working with the authorities assigning cruise identifiers. This activity can serve as a demonstor highlighting the benefits of linking sustainable cruise IDs to PIDs for articles (DOI), authors (ORCID), data (DOI) and samples (IGSN), while also exploring the challenges associated with implementing research cruise IDs.

In this report, the use case regarding research campaign comes from marine science. However, campaigns are not limited to marine research, but can refer to any research event with a specific purpose set in time and space, such as for example a research campaign conducted with the Hadron Collider at CERN. FREYA partners will explore the potential overlap between these scenarios in order to explore possible synergies which could be used in the generation of a universal research campaign-PID.

3.6.4 Relationship with other FREYA work packages

As Cruise-IDs represent a new and currently non-existing PID, the work is of primary consideration for WP3. There are currently no larger international initiatives working on a universal solution for Cruise-IDs.

3.6.5 Envisioned resolution model and the relevance for EOSC

A PID-system for research campaigns is very specific to marine research, and does not have a broader usage in the scientific community. In the context of FREYA, it serves as a demonstrator highlighting the benefits of linking metadata and data through PIDs. However, it is not universal enough to be of major importance in an EOSC context. Any developments on Cruise IDs that reaches a TLR of 8-9, will be made available through the EOSC services registry.

3.7 Data Management Plans

3.7.1 Synopsis

The FREYA partners, in particular DataCite, collected use cases around PIDs for data management plans (DMPs). These user stories are primarily concerned with automating processes surrounding the creation and use of DMPs, and the ability of these DMPs to be interconnected to other research materials by way of PIDs is central to accomplishing the types of automation required.

User stories related to PIDs for data management plans

#95 As a stakeholder in the research community, I want to uniquely identify a DMP the same way I can uniquely identify other research outputs, so that it can more easily be folded into the broader PID ecosystem.

#96 As a researcher, I want to automate the DMP creation process as much as possible (using other existing PIDs), so I can avoid extra effort or duplication of work.

#97 As a grant funder or institution, I want to readily see and link to other works related to a DMP, so that I can (e.g) follow up on data deposits post-award.

(The user story number refers the the github issue #.)

3.7.2 Validation

There has been demonstrated community interest in identifiers for DMPs, as evidenced by the RDA working groups on common standards for DMPs[35] and on exposing DMPs[36], which are both specifically concerned with machine-actionable DMPs. The information models and strategies that will be developed by these groups will inform the FREYA partners' implementation of infrastructure for machine-actionable DMPs. Further, machine-actionable DMPs have been prototyped by the California Digital Library[37], building on investigations conducted by students at the Technical University of Vienna[38]. Future work on machine-actionable DMPs will use DOIs as the PID to describe the DMPs, but will make use of other PIDs to link the DMP to other entities, such as funders, creators, or data[39].

3.7.3 Possible action by FREYA partners

DataCite will be taking on work over the course of 2019 and 2020 related to machine-actionable DMPs as part of an NSF EAGER grant in collaboration with the California Digital Library. This work can be complemented by FREYA by leveraging the feedback and outreach mechanisms of the PID Forum to build community consensus on machine-actionable DMPs and by introducing the resulting DMPs into the PID Graph.

3.7.4 Relationship with other FREYA work packages

This work will be related to WP5 via the PID Forum. While the mechanisms around machine-actionable DMPs are probably most relevant to WP3, the use and relevance of DMPs could be valuable to WP4's disciplinary development.

3.8 PIDs for physical samples and cultural artefacts

3.8.1 Synopsis

Seven of the user stories collected for this report pertained to identifiers for "samples" (#31, 36, 41, 42, 44, 46 and 91[40]). The primary difficulty here is that the definition of "a sample" can be very broad depending on scientific discipline, which the diversity of sample concepts in the collected use cases demonstrates: included were soil samples and field research sites, as well as cultural artefacts, geographical boundaries and historical personages. In practice, it would be difficult to accommodate this wide variety of samples using one generic sample PID. Given the current PID landscape, a series of different PID types will be required to address these user stories. At present these seem to fall broadly into three categories, a PID for physical samples, a PID for cultural artefacts and a PID with different metadata requirements for the conceptual entities described below.

Several user stories were for various types of conceptual things such as historical or mythical personage, historical timeframes and locations and historical geographical locations, which were also classified as PIDs for samples but it was noted from the outset of collection by the contributors that existing metadata standards would not meet these concepts easily.

[35] https://www.rd-alliance.org/groups/dmp-common-standards-wg
[36] https://rd-alliance.org/groups/exposing-data-management-plans-wg
[37] https://blog.dmptool.org/2018/07/09/scoping-machine-actionable-dmps/
[38] https://blog.dmptool.org/2018/08/20/machine-actionable-dmps-what-can-we-automate/
[39] https://blog.dmptool.org/2018/11/01/common-standards-and-pids-for-machine-actionable-dmps
[40] See user stories here
https://github.com/datacite/freya/issues?page=2&q=is%3Aissue+is%3Aclosed&utf8=%E2%9C%93
Or https://github.com/datacite/freya/issues?utf8=%E2%9C%93&q=is%3Aissue+is%3Aopen

User stories requiring PIDs for a sample or artefact

As a visitor of the Bremen Core Repository, I would like to get more information about a sediment core/sample in a repository with a smartphone.

As a museum curator, I would like to history of the placement and display of museum items, where it has been stored and the atmospheric conditions of those storage locations over time.

As a museum curator, I want to access accurate data about paint samples taken from artworks over time to assist in their conservation.

As a provenance researcher, I would like to be able to trace/relocate misplaced cultural artefacts.

As a researcher at scientific research facility, I would like to be able to identify the provenance of my soil sample by linking it to a specific research site.

3.8.2 Validation

The main sample identifier in use within earth, space and environmental sciences is the International Geo Standard Number (IGSN). IGSN has recently been awarded funding from the Sloan Foundation for a project to improve its underlying architecture and to expand its use beyond physical samples to include other types of materials[41]. The Steering Committee of Project IGSN 2040 contains representatives from geological sciences, life sciences and archaeology. The project will also focus on development and cross-disciplinary adoption of a common core metadata scheme for physical samples, which enables federated catalogues and cross-linking of digital sample representations with literature and data.

It should be noted that alternative PIDs for samples are employed in disciplines beyond the geosciences: Within the biosciences and life sciences, many collections use accession numbers to identify samples within their collections. Accession numbers, like LSIDs and GUIDs used for biodiversity data[42], are compact identifiers which comprise any local unique identifier with a prefix that is 'repository identifying'[43]. Using accession numbers, the BioSamples Database at EMBL-EBI for example, indexes more than 5 million biological samples, such as cell lines used in sequencing, gene expression and proteomics molecular experiments[44]. IGSNs have gained some traction within this space too. Another initiative, Research Resource Identifiers (RRIDs), have to date had limited uptake within life sciences[45]; that said, there is a subset of BioSamples that have recently also been assigned RRIDs[46].

An RDA Interest Group focusing on sample PIDs[47] is also working on facilitating a community around PIDs for samples and pursues the following agenda:

[41] Project IGSN 2040; see Lehnert (2018) IGSN: Toward a Mature and Generic Persistent Identifier for Samples, AGU Fall Meeting. https://www.slideshare.net/klehnert/igsn-toward-a-mature-and-generic-persistent-identifier-for-samples

[42] Life Science IDentifiers and Globally Unique IDentifiers https://www.gbif.org/document/80662/adoption-of-persistent-identifiers-for-biodiversity-informatics

[43] Wimalaratne SM , Juty N , Kunze J , Janée G , McMurry JA , Beard N , Jimenez R , Grethe JS ,Hermjakob H , Martone ME , Clark T . Uniform resolution of compact identifiers for biomedical data. Sci Data [08 May 2018, 5:180029]

[44] Nucleic Acids Research, Volume 47, Issue D1, 8 January 2019, Pages D1172–D1178,https://doi.org/10.1093/nar/gky1061

[45] D3.1, p.25, https://doi.org/10.5281/zenodo.1324296. D3.1 provides a more detailed overview of other PIDs in this area.

[46] https://scicrunch.org/scicrunch/about/blog/1132

[47] https://www.rd-alliance.org/ig-physical-samples-and-collections-research-data-ecosystem-rda-13-plenary

(RDA IG Statement excerpt:)

'*This group aims to facilitate cross-domain exchange and convergence on key issues related to the digital representation of physical samples and collections, including but not limited to:*

- *the use of globally unique and persistent identifiers for samples to support unambiguous citation and linking of information in distributed data systems and with publications;*
- *metadata standards for documenting a diverse range of samples and collections and for landing pages; access policies; and best practices for sample and collection cataloguing, including a broad range of issues from interoperability to persistence.'*[48]

In the current state, several of the user stories include a requirement for geolocation of samples and there is potential for this to be accommodated within IGSN, which supports this type of metadata but has minimal mandatory metadata requirements, enabling flexibility but also presenting a difficulty for cross-linking. The paint sample user story (#63) could also be met by IGSN[49].

There were several user stories which express a need for PIDs for a variety of cultural artefacts as well as the more conceptual artefacts noted above. D3.1 noted that there have been a number of initiatives around identifiers in this area but none of these have gained particular traction even though there is a recognised use case for PIDS within museum and cultural institutions systems[50]. Many institutions are using internal accession number or identifier systems, but none have received widespread adoption. There appear to be many reasons for this, including differences in requirements; varying approaches to the structuring of collections; and the sheer size of the legacy collections. All of these pose challenges not just between institutions but even in agreeing a common approach to identifiers internally, between departments and sub-collections. Museum professionals can also struggle to make the case for PIDs, as they do not relate closely to their internal drivers.

There are various projects and working groups developing this such as the International Standard Manuscript Identifier (ISMI)[51], and MuseumID[52] but as yet an agreed solution has not been found. Some organisations are using DOIs for cultural artefacts such as Rutgers University Libraries who create DOIs for Roman coins[53], however those institutions may not be able to meet all of their requirements as the metadata schema were not always designed for this purpose.

3.8.3 Possible action by FREYA partners

As part of the FREYA project, PANGAEA (University of Bremen) will expand its integration of IGSNs in the PID graph by implementing actionable IGSNs, linking to PIDs for Data (DOIs) and Authors (ORCID). For this task, PANGAEA is working with the Bremen Core repository, which hosts marine sediment cores from the Atlantic Ocean, Mediterranean, Black and Baltic Seas and Arctic Ocean for the International Ocean Drilling Program (IODP), on implementing IGSNs for specific samples that can be identified in the core repository with a barcode.

As PIDs for cultural artefacts remain relatively immature and consensus is some way off, The British Library has committed to explore this and is involved with initiatives such as ISMI and DISSCO[54] which is tasked with addressing this issue for natural science collections. It will also continue dialogue with other independent research organisations. The British Library will attempt to make its existing PIDs resolvable

[48] https://www.rd-alliance.org/ig-physical-samples-and-collections-research-data-ecosystem-rda-13-plenary
[49] D3.1, p. 24, https://doi.org/10.5281/zenodo.1324296
[50] D3.1, p. 27, https://doi.org/10.5281/zenodo.1324296
[51] https://www.irht.cnrs.fr/?q=fr/agenda/manuscript-ids-pour-un-identifiant-unique-des-manuscrits-2 accessed 14/01/19
[52] http://museumid.net/documentation accessed 14/01/19
[53] For example, https://rucore.libraries.rutgers.edu/rutgers-lib/41160/
[54] https://dissco.eu/

externally, specifically Archive Resource Keys (ARKs) for digital and digitised collection items and take steps to make them easier to cite by researchers.

The British Library will also maintain a watching brief on the conceptual user stories identified and will review the area regularly.

3.8.4 Relationship with other FREYA work packages

The integration of actionable IGSNs relates to the Bremen Core Repository which is a TRL8 service, which will relate to WP2. Many of the user stories mention relating to existing PIDs which will need to be explored with WP4, Integrating the PID Graph.

3.9 Conferences

3.9.1 Synopsis

There are many use cases concerning PIDs for conferences. Not all of those are reflected in the user stories that were collected, but PIDs for conferences could have a real impact on how "scholarly outputs" are trackable, therefore closing a huge gap that exists there. Other core user stories concern the tracking of conference proceedings with the purpose of observing the different publishing outlets with many new conferences emerging and others only being "one hit wonders". Another user story linked to this is being able to track early results, such as posters and early conference proceedings from researchers or institutions (again touching on user story #51). There also seems to be a need to be able to identify duplication, where journal articles and proceedings are being published with similar content. In some communities, (e.g. Computer Science) much is being published in conferences rather than journals, so the challenge and potential impact of PIDs for conferences should not be underestimated. For such communities, a connection to the PID Graph will be closing an essential gap in the scholarly record.

Example user story requiring a PID for conferences

#51 "As an institution, I want to track the outputs of all affiliated researchers. This concerns papers, data, code and their impact (citations), but also contributions to specific conferences."

(The user story number refers to the github issue #.)

The hope is that by the end of the project there will be a robust concept of PIDs for conferences and, perhaps, first prototypes available. A prototype would be a first pilot application. First use cases have been identified, e.g a workflow at Springer Nature as well as at CERN, CDS[55] and Indico[56]. One aim is to provide conference PIDs to Google Scholar.

3.9.2 Validation

Community interest in such activities has emerged over the last years. Work has been done by Springer Nature and DataCite in that regard[57],and has been presented at PIDapalooza[58] and Force2017/8. A good set of use cases and a better understanding of the boundary conditions have been derived.

[55] http://cds.cern.ch/ (last checked on 10.01.2019)
[56] https://indico.cern.ch/ (last checked on 10.01.2019)
[57] Birukou, A. (2018): PIDs for conferences - your comments are welcome!. DataCite Blog. URL: https://www.crossref.org/blog/pids-for-conferences-your-comments-are-welcome/
[58] https://pidapalooza18.sched.com/event/Cwmu/conferencepids (last checked on 10.01.2019)

Participation in the kick-off meeting at CERN (see more in the section below) is very broad, i.e. covers many TRL9-level services that have shown an interest to further the discussion and implementation. These partners come from various disciplines and industries.

3.9.3 Possible action by FREYA partners

In February 2019, FREYA and Springer Nature will co-host the first technical kick-off meeting[59] for conference PIDs. The aim is to refine the work plan and commit to concrete projects to seed first solutions. The draft work plan is currently focusing on a better definition of the use cases, e.g. registering a conference series PID. Following that, a roadmap for technical specifications and implementation should be set in place, including the expected timeframe for first assignment of conference PIDs. It should be noted that the meeting agenda is subject to change as the meeting will have taken place after the writing of the present document.

3.9.4 Relationship with other FREYA work packages

The communication work package (WP5) will be needed to get more input on the concrete action points and ensure its suitability for the wider community. If the FREYA time-frame allows, concrete implementations could be possible within the WP4 pilot applications.

3.10 Entities excluded for analysis

As noted earlier, several entities were excluded for requirements gathering for different reasons:

- PID infrastructures for these entities are mature and therefore deemed more relevant to WP4 than WP3. These include the entities: “**article**”, “**data**” and “**person**”
- No current partner expertise and capacity to drive services forward for the entity. This group comprises the entities “**repository**” and “**project**”.

While there are a number of WP-relevant user stories that have the label “repository” it was felt that many are secondary to the entity of focus in the user story (#65, #69, #87, #89) or lie beyond the expertise of the partners to prototype at this time (#73, #66). The project user story (#59) falls into the latter case.

An analysis was conducted for the user stories that mention “articles”—an entity for which PID services are deemed “mature”. A decision was taken to exclude these cases for further consideration for WP3. Rather than breaking new ground and filling a gap in the PID infrastructure landscape, the analysis revealed that these cases require extending existing article PIDs to new types of literary content; or cases where “articles” are not central to the user story and are mentioned due to the desire is to link a new PID type to articles within the PID graph.

[59] Agenda and notes: https://indico.cern.ch/event/780651/ (last checked on 10.01.2019)

4 Conclusions, including candidate PID services for prototyping by FREYA partners

This report documents the research undertaken by partners into new PID services that address needs captured in user stories from stakeholders and that could potentially be moved forward (prototyped) by one or more FREYA partners within the timespan of the FREYA project.

This report will be used as a basis for the subsequent task of prototyping select new PID services. During discussions, a small but significant point was raised about using the term “prototyping” or “pilot”: the latter term can be applied to services but we believe should be avoided for PIDs per se, since it is not possible to ensure persistence of a PID if the “pilot” is subsequently sunsetted.

FREYA partners agreed a working definition of a prototype for a demonstrator:

The service will focus on entities that are being newly assigned a persistent identifier, whether an identifier in current use such as a DOI[60] or a new identifier type such as a ROR ID[61]. Where a PID is not yet available for an entity, or in a case where continued technical discussion is required e.g. to agree the metadata that will be captured for the identifier, this would be deemed too immature to prototype. In such cases, a white paper around recommendations could be created but would be considered a lower priority and not to be pursued by FREYA partners for this task. FREYA partners do not have to deliver something for all new PID types or entities that are being newly assigned PIDs.

As a result of the discussions, outreach and requirements which we have captured in this report, we propose that the most promising candidate PID services for prototyping by FREYA partners involve PIDs for instruments, facilities, grants, organisations, research campaigns and cultural artefacts. See Table 4.

The deep-dive investigations into requirements for PID services for each entity has revealed relevant PID communities outside of FREYA (such as RDA working groups, the ROR community) with whom to collaborate and keep informed of any prototypes that are taken forward here.

Entity and PID (if known)	Possible for prototyping within the timeframe of FREYA	Lead partner for requirements gathering in this report
Instruments (on German research vessels)—Instrument IDs	yes	PANGAEA
Facilities (photon and neutron sources)—Facility IDs	yes	STFC
Grants—Grant IDs	yes	EMBL-EBI
Organisations—ROR-IDs	yes	DataCite
Software—DOIs	No (too mature)	DataCite
Research campaigns—Cruise IDs	yes	PANGAEA
Data Management Plans	No (too immature)	DataCite

[60] https://www.doi.org/faq.html
[61] https://ror.org/scope/

Samples—IGSNs, BioSample accession numbers; RRIDs; ARKs	No (too mature)	BL
Samples (Cultural Artefacts)	No (too immature)	BL
Conferences	No (too immature)	CERN

Table 4 Readiness for building service prototypes

At the end of the prototyping task (currently scheduled to end in approximately February 2020), demonstrator services will be available for take up by partners in other work packages that are focussed on improving existing PID infrastructures (WP2) or integrating disciplinary and EOSC contexts with the PID Graph (WP4). For the former one can envisage ORCID profiles integrating organization (ROR) identifiers. For work on the PID Graph, one can envisage scenarios such as those depicted in Figure 5: e.g. being able to query which datasets have been published by staff of a particular research organisation. In this scenario, not only are there APIs that identify links between researchers and their datasets, there are additional APIs that identify the research organisation to which the researcher-dataset links belong.

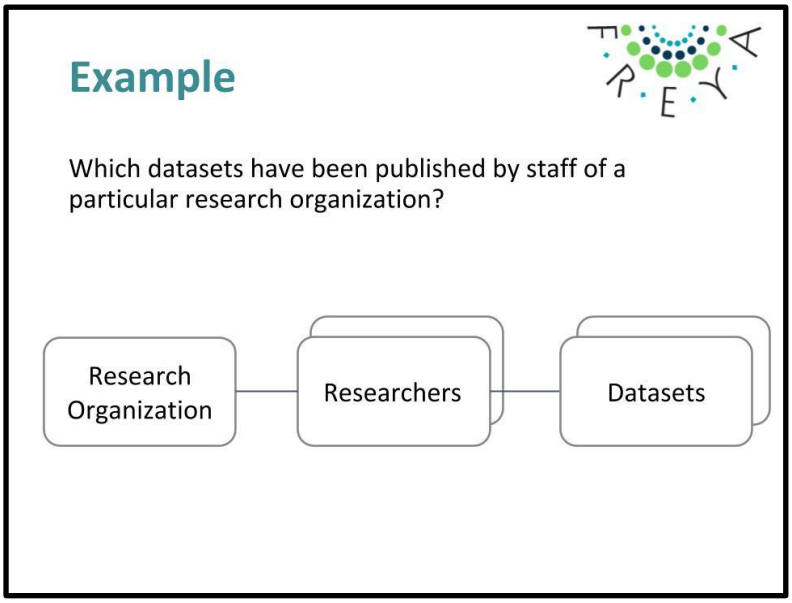


Figure 5 Expanding the PID Graph - an example (courtesy of Martin Fenner, DataCite).

Annex A: User stories collated by FREYA partners (as of January 2019)

The following table includes the 69 user stories that were collated in the FREYA Github repository[62] by FREYA partners and are referred to within this report. These were kindly collated into Excel format by Manuel Bernal Llinares (EMBL-EBI).

Github issue #	Title	User Story	Labels
97	use links between DMPs and PIDs to follow up on deposits	As a grant funder or institution, I want to readily see and link to other works related to a DMP, so that I can (e.g) follow up on data deposits post-award.	DMPs, WP3
96	automation of DMP creation	As a researcher, I want to automate the DMP creation process as much as possible (using other existing PIDs), so I can avoid extra effort or duplication of work.	DMPs, WP3
95	PIDs for DMPs	As a stakeholder in the research community, I want to uniquely identify a DMP the same way I can uniquely identify other research outputs, so that it can more easily be folded into the broader PID ecosystem.	DMPs, WP3, user story
94	PIDs for policies	As a research manager, I want to have policy IDs, so that I can easily identify relevant policies and assess the compatibility between different policies. *From DI4R workshop*	Community contribution, WP3
91	Tracing/relocating misplaced cultural artefacts	As a provenance researcher for Nazi-looted books, I wish for PIDs for stamps, exlibris, autographs, place of looting, place of storage. Could help to find objects of a former owner/collection which was torn apart by Nazi looting. *From DI4R workshop*	Community contribution, WP3, bibliometrician, geolocation, institution, national library, sample, user story

[62] See https://github.com/datacite/freya/issues;

90	Including heritage literature in repositories	As the Manager of the Biodiversity Heritage Library (BHL) in Australia, I want to see DOIs applied to the world's biodiversity heritage literature so that this literature can be part of the great linked network of scholarly research (the DOI system). ... I also want to work with the global DOI community to ensure that the DOIs for out-of-copyright literature are #OpenAccess (& do not resolve to versions behind paywalls) *From FREYA Ambassador Webinar 16 Oct 2018*	Ambassador, article, library, user story
89	University outputs for discovery and re-use	As a University I want to represent my outputs in a future-proof manner (using PIDs) so that people can find and reuse what we've created. *From FREYA Ambassadors Webinar 16 Oct 2018*	Ambassador, WP3, article, data, grant, organization, person, repository, software, user story
88	Adding content to repository	As a Librarian I want to catalogue material unique to my Library (e.g. Special Collections / theses) to allow a broader network to discover what we have and use it!	Ambassador, article, bibliometrician,data,funder, library,organization,user story
87	Linking data, instruments and protocols/best practices	As a researcher, I would like to find all data in the repository produced by a specific instrument/sensor on a research vessel . To decide wether data is compatible with other data from the same or similar instruments I would also need an actionable link to the measurement protocol or DOI of best practice document.	PANGAEA,PID Graph,STFC, WP3,data science, instrument,repository,user story
86	Reuse of my data	As a researcher, I want to be able to track the reuse of my data	DANS,PID Graph,data, person,researcher,user story
85	Retrieve from Research Graph a PID-Graph on basis of a certain PID	I would like to see an API (Maybe this is already existing) were I can search on a certain PID and retrieve all relation to this PID. Maybe set how many steps you want to go.	DANS,PID Graph,WP4, article, data,user story
83	acknowledging infrastructure use	As a user of [facility/resource/archive] I want to connect my use of the [facility/resource/ archive] to outputs resulting from my work there.	ORCID,PID Graph,STFC, WP3,facility,researcher,user story

82	Reducing administration	As a researcher I want the information about me to be readily available so I don't have to keep re-entering the same information into different systems Specialist versions of this include: - As an author, I want to be able to create simple funding acknowledgements while I submit articles. - As an applicant, I want to be able to auto populate my application for funding with information held by other systems	ORCID,PID Graph,funder, publisher,researcher,user story
81	Researcher outputs	As a researcher I want my contributions to be unambiguously associated with me so that others can discover my output. As a librarian/reviewer/administrator I want to unambiguously discover the output of a specific researcher	ORCID,PID Graph, bibliometrician,curator, researcher,user story
80	Representing multiple linked items and hierarchies	As a researcher, I wants to see/visualize how multiple linked outputs interact within a given collection or archive - and how these links and hierarchies have been changed and adjusted over time. This will help me understand the relationships and organisation of the collection/archive. This particularly comes from archival research, where structure may be as important as content.	British Library,PID Graph, WP4,curator,library, national library,researcher, user story
79	Enriching metadata and content	As a national library, we want to dynamically enrich the information we have about our collections based on information held by other organisations, without needing to hold the information separately. This will help our users in discovering and accessing our own items. This could be enriching information about: works, authors, publishers, related outputs among other things.	British Library,PID Graph, WP4,article,data,library, national library, organization, person,user story
78	Linking to unique cultural content	As a cultural institution or library, I want to enable links to our unique content from external sources where those are mentioned and discussed, e.g. wikipedia, news items, blogs, research outputs.	British Library,PID Graph, WP4,curator,library, national library,next,user story

76	PID endpoint metadata	Provide metadata at PID endpoints to support graph building.	EMBL-EBI,PID Graph,data, repository,user story
75	Industry/business sector and the commercial use of PIDs	Let's not forget the Industry/business sector and the commercial use of PIDs. I am a company producing scientific instruments and/or software. For a marketing analysis, I would like to trace the current use of our products (instrument/software-PID) across scientific disciplines and geographical areas by analyzing article and data publications (publication-PID/data-PID) specifically produced using our products (instrument/software-PID)	PANGAEA,PID Graph,WP3, instrument,user story
74	Linking preprints to their published article versions	As a data scientist (or researcher), I want to know whether any given preprint has subsequently been published. If yes, then for these to be linked reciprocally (from preprint to publication; from publication to preprint).	Crossref,EMBL-EBI,PID Graph,WP3,article,next,researcher,user story
73	A registry for preprint servers	As a literature repository wanting to aggregate preprints, I want to find all life sciences preprints in existence.	EMBL-EBI,PID Graph,WP3, repository,user story
72	Records enrichment in institutional repositories	As an owner or an operator of an institutional publications or data repository I am interested in records enrichment with links to identifiable chemical substances, biological objects, species, geolocations etc. that are mentioned in the record title, abstract, or in the associated full text/description. I would like to focus on text mining techniques first with possible extensions for the automated image or data characterization.	STFC,article,chemical,data, geolocation,species
71	Gap analysis for institutional repositories	As an owner or an operator of an institutional publications, data or software repository I am interested in gap analysis between my repository and other repositories of similar kinds. An example could be STFC ePubs repository (for publications) that has to operate in the diverse and ever changing world of information where other repositories potentially capturing STFC employees' publications exist: Zenodo, INSPIRE-HEP, preprint services. PIDs (for people, institutions, papers, potentially grants and projects, too) may help to identify gaps between what is captured by a local repository and what is captured elsewhere. The gap analysis can be	PID Graph,STFC,article,data, grant,organization,person, project,softwar

		the first step to further actions, such as: ingest records from outer sources, or link to them, or merge with them, or simply disregard them (if they are somehow "out of scope").	
70	Linking facilities research to records in scientific databases	As a facility User Office interested in measuring the facility impact, or as a reviewer of research proposals (who evaluate applications for facility beamtime), or as a funder who supports the facility with public money or industrial contribution, I am interested in linking facility awards (beamtime) with structured records in renowned scientific databases, such as biomedical or crystallography databases. Ideally, these links and measures should be as granular as possible: what database record(s) resulted from what particular facility award(s).	PID Graph,STFC,WP3,data, facility,funder,instrument, user story
69	Tracking down PhD studentship outcomes, beneficiaries, co-funders and supporters	As a funder, I want to track down the outcomes and beneficiaries of PhD studentship awards that I granted. There are many possible questions to be answered (with the help of PID graph): - whether the PhD studentship actually ended in thesis, how to find and how to cite this thesis, - what organizations benefited from PhD during or soon after the PhD research period, e.g. by hiring the PhD that I sponsored, - who co-funded or otherwise supported the PhD research, - what artefacts (papers, data, software, samples, instruments, ...) can be identified that either contributed to the PhD research or are the PhD research outcomes. This issue is related to #35 and in part to #68 (as facilities are frequent supporters of PhDs).	British Library,DataCite,PID Graph,STFC,WP3,article, data,funder,instrument, next,organization,person ,researcher,software,user story
68	Tracking researchers (facility/infrastructure)	As an infrastructure provider I want to be able to track people associated with my instruments, equipment and services so I can follow their careers.	ORCID,PID Graph,STFC, WP3, instrument,person, service_provider,user story
67	Data citations by repository	As a repository manager, I want to get notified of new citations of datasets hosted in my	CERN,Crossref,DataCite,PA NGAEA,PID Graph,WP3, article, data, data

		repository, so that I can inform the authors.	center,next,repository,user story
66	PIDs and registry for data repositories	As a researcher, I would like a central global list of research data repositories so that I can choose where best to deposit my data.	WP3,repository,researcher, user story
65	Cross-linking literature and data via instruments	As a researcher, I want to discover data by an instrument mentioned in a paper I just read because that data my be useful in my research.	PID Graph,WP3,article, data,instrument,publisher ,repository,researcher,user story
64	Linking published (meta)data with instrument metadata	As a researcher, I want references to instruments in (meta)data published by data repositories to be actionable so that I am unambiguously redirected to metadata about the instruments which enables me to learn more about instruments and the context in which data were acquired.	PANGAEA,PID Graph,WP3, data center,instrument, researcher,user story
63	Tracking reuse of software across versions	As a software author, I want to able to see the citations of my software aggregated across all versions. so that I see a complete picture of reuse.	CERN,DataCite,PID Graph, STFC,WP3,next,software, software author,user story
62	Tracing outcome of Research cruise (campaigns)	As a funding agency, I would like to trace the outcome of my financial contribution to a marine research cruise (Cruise ID) by tracking the data generated (data-PID) and articles (publication-PID) published, as well as physical samples taken (IGSN) and the repository (organization ID), where these samples are physically stored. In this regard, I would also like to track the future data and publications generated from these samples.	PANGAEA,PID Graph,WP3,a rticle,data,funder,organizati on,sample,user story
61	Microcontributions	As a young scientists I would like that my contribution to a publication is distinguishable from my cowriters contribution, e.g. that it is clear who contributed to the code, data, analysis etc. [general use case, not specific to CERN]	CERN,PID Graph,article, data,researcher,software, user story
60	Expose citation stats	As a information researcher, bibliometrician or ..., I want to have all the citation stats from trusted sources, independent of a service provider (e.g. google scholar, …) with high quality metadata.	CERN,PID Graph,article, bibliometrician,data, software,user story
59	Tracking groups/projects	As a service provider i would like to be table to track the outputs of the experimental	CERN,PID Graph,WP3, project, service_provider,

		collaborations of High-Energy Physics. That would include developing an identifier for (changing) groups.	user story
58	Reuse of links for services	As a service provider, I would like to be able to access/follow links between authors, papers, data, code, ….	CERN,PID Graph, article, data,next,person, service_provider, software, user story
57	Impact of instrument/equipment	As a facility, I would like to track the published output related to the instruments provided. Also, i would like to be able to evaluate their impact.	CERN,PID Graph,STFC,WP3, facility,instrument,user story
56	Impact of funding	As a funder in HEP, I would like to measure the impact of my grants, i.e. did the funding to specific software projects lead to more shared code and/or research outputs?	CERN,PID Graph,WP3, funder,grant,next,user story
55	Tracing outcome of scientific Instrument / Sensor	As a researcher, I would like to track other researchers who are using the same scientific instrument as me, and get access to their scientific outcome (data, publications, samples, genetic-markers etc). – for inspiration, validation and collaboration.	PANGAEA,PID Graph, STFC,WP3,instrument, researcher,user story
54	Metrics covering reproducibility	As a service provider, I would like to show the statistics of successful reuse/rerun of a physics analysis. This should also be included in the H index. Related to https://github.com/datacite/freya/issues/53	CERN,PID Graph, service_provider ,user story
53	Track reproduciblity	As a HEP scientist, I would like to be able to track whether and how often a physics analysis has been rerun/reproduced.	CERN,researcher,user story
52	Enable collaboration networks	As a service provider, I want to enable a visualization of the collaboration network, based on data, code, papers, …	CERN,PID Graph,article, data,next,service_provider, software,user story
51	Outputs by researchers from a specific institution	As an institution, I want to track the outputs of all affiliated researchers. This concerns papers, data, code and their impact (citations), but also contributions to specific conferences.	CERN,ORCID,PID Graph, WP3,article,conference, data,institution,person, software,user story
50	Impact of outputs	As a HEP researcher, I want to analyse the impact of my papers, data, code - how often have they been cited? This should be included in the H index.	CERN,PID Graph,article, data,researcher,software, user story

49	Tracking software use 2	As a HEP researcher, I want to know which results/paper has been produced with which software.	CERN,PID Graph, article,researcher,software, user story
48	Tracking software use	As a HEP researcher, I want to know who (author) used my software and for what purpose (their paper).	CERN,PID Graph,STFC, article,person,researcher, software,user story
47	Making easier to cite resources	As a HEP researcher, I want it to be easier to add correct citations to data and code in a specific state in order to give credit and to ensure the transparency of my analysis.	CERN,data,researcher, software,user story
46	Identifying historical timeframes	As an historian, I need to identify which definition of a particular historical timeframe I have adopted (e.g. the Elizabethan era) in my work, particularly where the dates and definitions of those eras are contested.	British Library,WP3, researcher,sample,user story
45	User stories for funding PIDs	As a funder, we want to be able to find all the outputs related to our awarded grants, including block grants such as doctoral training grants, for management info and looking at impact As a funder, we want to be able to identify who (including orgs and individuals) benefitted from a given grant, for boosting management info and for looking at impact	British Library,PID Graph, STFC,WP3,funder,grant, organization,person,user story
44	Identifying historical locations and geographical boundaries	As an archivist at a public record office, I need to be able to identify the precise geographical boundaries of wards and parishes at specific points in time in order to provide accurate information to researchers and legal professionals.	British Library,WP3,sample, user story
43	Funder needs for org IDs	As a funder, we want to be able to internally identify past and present affiliated institutes and their names over time, so that we can associate them with their related host institutions, staff, outputs and accolades.	British Library,WP3,funder, organization,user story
42	Identifying historical or mythical personages	As an historian of ancient civilisations, I need a means of identifying precisely which version of a fictional or mythical personage I am writing about, particularly when the existence of that individual and/or their surrounding circumstances are contested.	British Library,WP3,person ,researcher,sample,user story
41	User story for scientific research site PIDs	As a researcher at a scientific research facility, I want to authenticate the provenance of my soil	British Library,WP3,facility, researcher,sample,user

		sample by linking it to a specific research site, and to access information about the historic treatment of that site.	story
40	User story for study registration PIDs	As a peer reviewer for an article, I want to see the study registration record for the research paper that I'm reviewing so that I can assess the degree to which the researchers complied with their original proposal in obtaining their research results.	British Library,PID Graph, WP3,study,user story
39	Linking to software for analysis of specific research datasets	As a researcher in the digital humanities, I want to analyse British Library datasets using software that has been developed specifically for those datasets and logged in GitHub. As a researcher, I want to acknowledge a particular grant in funding the creation of my software. As a funder, I want to know what software has been developed from a project I have funded.	British Library,PID Graph, WP3,data,funder,grant, researcher,software,user story
38	Metrics for data with multiple PIDs (subsets)	As a longitudinal study, I want to be able to deduplicate the metrics/impact for our data, so that I can see the impact of our study's date as a whole. NOTE: Recommendations for dynamic data will lead to studies having multiple DOIs for single datasets, and multiple DOIs for the study may be used in any one given paper. So deduplication is needed to reduce double-counting.	British Library,DataCite, PANGAEA,PID Graph, STFC,WP4,data,next, researcher,study,user story
37	User stories for longitudinal study data PIDs	As a policy maker, I want to cite specific subsets and extractions of data from longitudinal studies as evidence for policy change As a researcher, I want to be able to find the survey instruments used to collect longitudinal data, so that I can collect my own data with the same survey, making the data comparable. As a policy maker, I want to know which longitudinal project collected which data, so that I can contact them for more information As an institution, I want to know which researcher collected which data from a longitudinal study, so that I can look at the contribution of our institution specifically to	British Library,PID Graph,WP4, data, institution,researcher, study, user story

		the research	
36	User stories for cultural artefact PIDs	As a museum curator, I want to track the history of the placement of an item including where exactly it has been displayed and in which exhibitions (including loans), where it has been stored, and the atmospheric conditions of those storage locations over time. OR As a museum curator, I want access to accurate data about paint samples taken from artworks over time to assist me in their conservation. I want this information to be publicly accessible.	British Library,WP3, sample,user story
35	Linking people and research outputs to theses	As a student using the British Library's EThOS database, I want to be able to see which students' theses a given researcher has supervised, and what those students went on to do. As a researcher, I want to see researcher family trees* between students, supervisors, grandparent supervisors and onwards. As a researcher I want easily to jump between the articles and datasets cited in thesis bibliographies.	British Library,ORCID,PID Graph,WP4,article,data, person,researcher,user story
34	More effective linking of data to publications	As a researcher I want (easy ways) to (more effectively) link all data to publications. As a reader I want to be able to easily find all data related to a publication.	CERN,DataCite,EMBL-EBI,ORCID,PANGAEA,PID Graph,WP2,article,data, next,organization,user story
33	Contribution of core facility to scientific discovery	As a core facility provider, I want to track usage of my facility so that I can demonstrate it's value.	EMBL-EBI,ORCID,PID Graph,STFC,WP3,facility, instrument,user story
32	Data Recommender	As a user of PANGAEA's data portal, I'd like to get dataset recommendations, so i can find related datasets, covering the the same area of interest. The current recommender on PANGAEA cannot present recommendations about datasets measured with similar instuments, but that's something I'd like to look	PANGAEA,PID Graph,WP4, data,user story

		into.	
31	Getting more information about a sediment core/sample in core repository with smartphone	As a visitor of the Bremen Core Repository, I'd like to use my smartphone with a barcode/QR code scanner to get more information about a core / sample. I am not only interested in metadata about the sample, but I'd also like to know more about the scientists doing measurements, what funding was used when sample was taken, related scientific articles, and finally which data was already gathered from it!	DataCite,PANGAEA,PID Graph,WP3,WP4,article, data,next,sample,user story
30	2nd degree citations	As a data centre, I want to see the citations of publications that use my repository for the underlying data, so that I can demonstrate the impact of our repository.	CERN,DataCite,PANGAEA, PID Graph,WP2,article, data,data center,next,user story
29	Find Data Citation for a given DOI	As a researcher or infrastructure provider, I want to see all citations and references to a given DOI including traditional citation, Twitter, blogs and grey literature.	ARDC,Crossref,DataCite,PID Graph,article,conference, data,next,user story
28	Expanding bi-directional links in HEP	As a High-Energy Physics researcher, I want to see bi-directional links between CERN Analysis Preservation records and INSPIRE or HEPData records for cases where a physics analysis on CAP has resulted in a published resource.	CERN,data,researcher,user story
27	Indirect citations of raw data	As a staff member at STFC, I want to see all the publications based on raw data generated in our facilities, so that I can demonstrate the impact of the services provided by us.	PID Graph,STFC,article, data,instrument,user story
26	Co-author graph	As a bibliometrician, I want to know all the co-authors of a particular researcher, so that I can do a network analysis of the researcher's collaborations.	ORCID,PID Graph,article, bibliometrician,data, person,user story

Annex B: Table comparing the categories of entities discussed in D3.1 vs D3.2

(D3.1) Research entity	(D3.1) Maturity of PID Infrastructure	(D3.2) WP3 User story and label applied	Requirements reported in D3.2
Publication (article)	**Mature**	Article	Annex D
Citation	Emerging	—	—
Conference	Emerging	Conference	yes
Researcher (or Scholar)	**Mature**	Person	—
Organization	Emerging	Organization	yes
Data	**Mature**	Data	—
Data repository	Immature	Repository	—
Grants	Emerging	Grants	yes
Project	Emerging	Project / Research campaign *	yes
Experiment	immature	—	—
Investigation	Emerging	—	—
Analysis	Immature	—	—
Software	Emerging	Software	yes
Computer Simulation	Emerging	—	—
Software License	Immature	—	—
Equipment			
Instrument, Device, Sensor, Platform, **Research Facility**	Emerging	Instrument	yes

		Facility**	yes
Archival/Storage facility	Emerging	—	—
Field Station	Immature	—	—
Sample			
Geological or Biological Sample	Emerging	sample	yes
Cultural artefact	Emerging	sample	yes
Historical or mythical person	Emerging	—	—
Temporal period & historical place	Immature	—	—
Study registration			
Clinical trial; non-clinical registration	Immature	Study	-
Data Management Plan	Immature	DMP	yes
Workflow	Immature	—	—
Protocol	Immature	—	—

* A “Research Campaign” user story (relevant to geosciences) was researched in this report whereas the “project” user story was not (“Project” was defined in D3.1 as a higher order entity “which aims to formalize the connectivity between research entities. It is also a term used by several research information systems. There is currently no widely-adopted standard for the identification of projects”.

** “Instruments” and “Facilities” were researched as separate entities for this report.

Annex C: Abstracts/emails for outreach programmes

Abstract for the WORLD Cafe Session accepted by DI4R 2018

https://indico.egi.eu/indico/event/3973/session/36/contribution/135;

Title: Persistent Identifiers in use: Exchanging ideas about new developments in the field of PID services.

Presenters: Eliane Fankhauser (DANS, for WP5), Simon Lambert and Brian Matthews (UKRI, STFC for WP6) and Christine Ferguson (EMBL-EBI for WP3).

Persistent identifiers (PIDs) like DOIs for articles or ORCiDs for researchers are a core component of open science as they improve discovery, navigation, retrieval, and access of research resources. FREYA, a 3-year EU-funded project, aims to extend the PID infrastructure by cross-linking PID services, facilitating the development of new PID types, and creating community of practice. The engagement with the stakeholders and the wider PID community is an important means with which to exchange knowledge and get feedback about the development of new PID types and services. Currently, FREYA is establishing the PID Forum consisting of a user community whose members collectively oversee the development and deployment of new services. Anyone with an interested in PIDs is invited to join this session, exchanging ideas and contributing to the discussions. At this World Café Session, the PID Forum will be introduced; some of the work that has been done in the first few months of this project will be presented and discussed with the audience in a workshop. The workshop will focus on two current FREYA activities: (i) mapping the identifier landscape and (ii) understanding how stakeholders operate within the landscape. Both of these activities we would like to discuss with and get feedback about from the user community. FREYA has recently surveyed the current identifier landscape and would like to share key findings with the user community. Moreover, FREYA would like feedback from the community on user stories that have already collected. Questions like “Is there broader value to be gained from addressing the user story?” or “What is needed to deliver the value identified in the user story?” will be addressed. Finally, FREYA is eager to connect with any stakeholders in the user community to learn about their user stories and identify gaps where research resources could be better connected and services extended or built.

The email brief for the ambassador webinar:

Presenters: Eliane Fankhauser (DANS, for WP5) and Christine Ferguson (EMBL-EBI for WP3)

Your user stories for FREYA WP3

Hello all,

The FREYA team is seeking your point of view, as our Ambassadors, for our next major piece of work. We are proposing an online meeting on Monday 15 October (a week or so after the next Ambassador webinar) and are looking for an early indication of your interest and availability.

This tranche of work involves gathering user stories to feed into our third work package. As you know, FREYA is building links where they are currently missing between research resources and research outputs. We are a consortium of cross-disciplinary service providers adept at linking research resources, but we

want to be sure that our projects reflect real life priorities - including those of you and your colleagues. We are looking for bite-sized user stories that describe unmet needs relating to PIDs. For example:

"As a [staff member at Institution X] I want to [see all publications that stem from raw data generated in our facilities] so that I can [demonstrate the impact of services provided by Institution X]."

"As a [biologist] I want to [reuse and remix data] so that [I can do my research]."

If possible we'd like some information about the size of the research community likely to be affected by the identifier type/service in your examples, and the volume of research data likely to become available as a result of its development. We'd like to add your stories to our growing collection and use them to help us prioritise our work for FREYA. It's a great opportunity to make sure your research community is represented.

If this is of interest and you think you can gather one or more examples to share by October 15, please do let me know and I will follow up with a calendar invitation.

Kind regards,

Barbara

Barbara Lemon (British Library, for WP5)

Event listing on the FREYA website for the Joint Webinar presented by FREYA and OpenAIRE

https://www.project-freya.eu/en/events/joint-webinar-freya-and-openaire-new-developments-in-the-field-of-persistent-identifiers

After all the festivities at the end of the year where family and friends connect, OpenAIRE together with FREYA will start off the new year with a webinar on digital connections: the Persistent Identifiers. The Science Europe Data Glossary defines the term Persistent Identifier (PID) as "a long-lasting reference to a digital object — a single file or set of files". As such, the importance of PIDs to build stable connections between research entities such as grants, projects, articles, or funders is recognized and addressed by several initiatives and projects.

FREYA is a 3-year project funded by the European Commission, aiming to extend the infrastructure for persistent identifiers (PIDs) as a core component of open research, in the EU and globally. FREYA will improve discovery, navigation, retrieval, and access to research resources. In so doing, FREYA has carried out a survey of the current PID landscape, collected a vast amount of user stories in order to identify needs of the community to expand existing and establish new PID services, and is currently working on building a PID Graph.

In the webinar, Ketil Koop-Jakobsen will talk about a report on requirements for new PID Services. To identify demands and requirements for emerging PIDs, FREYA collected user stories from their respective communities and networks. More than 70 user stories were compiled, each identifying a specific PID demand from the community. Koop-Jakobsen will introduce some of these stories and explain their influence on the development of new and emerging PID types. Amir Aryani, moreover, will shed light on FREYA's work on the PID Graph, talking about the discussion around the concept of the PID Graph itself and how FREYA partners are contributing to the actual setup of such a Graph.

Does this sound interesting to you? If so, sign up for this webinar and learn more about PIDs and why they are important for the research community.

Speakers: Iryna Kuchma (OpenAIRE), Ketil Koop-Jakobsen (PANGAEA for WP3, FREYA) and Amir Aryani (ARDC for WP4, FREYA)

Abstract for the presentation accepted by the PIDapalooza 2019 festival:

Title: FREYA proudly presents: the power of PIDs.

FREYA is a 3-year EU project that aims to build the infrastructure for persistent identifiers (PIDs) as a core component of Open Science. The work of FREYA will improve discovery, navigation, retrieval, and access of research resources. The project team is currently working on the establishment of new PID types, and on connecting existing and new PIDs into a PID Graph. PIDapalooza is the place to be to discuss our work with other PID-tellectuals and provide new directions to it.

Our session covers three parts which will be hosted by speakers from various FREYA partners: First, we will start with a discussion lead by Christine Ferguson (EBI) on the use cases for new PID types FREYA has recently collected. Then, Martin Fenner (Datacite) will present how new and existing PIDs can be integrated into the PID Graph that FREYA is developing. Lastly, Eliane Fankhauser (DANS) will encourage the PIDapaloozans to join the PID Forum to continue the discussion during and after the festival season.

Annex D: Analysis of user stories relating to "Articles"

D.1 Synopsis:

The user stories that relate to articles appear to outline three broad community needs:

1. More comprehensive indexing in literature repositories of "newer" literature types such as preprints, PhD theses, heritage literature (#74, 69 and 90 respectively)

2. Adding metadata to existing literature, such as more details about contributions made by "authors" (contributors) (#61)

3. More comprehensive linking of articles with other relevant research objects i.e.

    a. Entities with PIDs such as data, software, ORCIDs - see user story (#65, 67, 51, and 89)

    b. Entities using PIDs with limited uptake such as samples - (user story #31)

    c. Entities where new global PID systems are close to being implemented: such as organisations and grants (user stories #51 and 69)

    d. Entities that have no PIDs such as instruments, projects /ocean cruises (user stories #65 and 62)

D.2 Relationship with other FREYA work packages:

Published articles across disciplines are identifiable via several PID types and systems that are considered to be "mature" by FREYA partners. Further integration of literature is thus the focus of WP4. On scrutiny the user stories collated here largely focus on new PIDs /services for other entitie ; with the future aim that those entities will be linked up with articles PIDs.

The exceptions where user stories focus on articles per se, concern addition of alternative literature types to existing literature repositories. Is it feasible that these warrant action by this working group? Arguably not, if one considers the following initiatives to include these newer literature types:

- In the life sciences, since mid 2018, EuropePMC has been ingesting preprints from repositories that assign DOIs provided by Crossref. The preprints are ingested using an existing Crossref service (REST API). So although the content is 'new', neither the PID nor the API service is new.
- British Library holds records of all theses associated with PhD awards across disciplines in the UK. While many of these are not digital records, and nor are all associated with a PID (DOI, ISNI or ORCID), there is a move to digitize records and assign DOIs to these records. This constitutes expanding the reach of mature PIDs and incorporating these records into the PID graph which is the focus of WP4 (see the EThOS project described in D4.1)
- One of our FREYA ambassadors, Nicole Kearney, submitted a user story about registering PIDs for older content such as historic literature and out-of-copyright content[63]. Heritage literature records such as these are found within the Biodiversity Heritage Library. Currently these are assigned stable URLs with select data being assigned DOIs[64]. There are initiatives afoot to register PIDs (DOIs) for more of their records, but this is accompanied by logistical problems eg costs and agreement over who 'owns' the DOI for works belonging to long-deceased researchers. Current work to raise visibility over these issues is ongoing (see Nicole Kearney's presentation given at

[63] https://www.project-freya.eu/en/blogs/blogs/pidapalooza-competition-winner-1
[64] https://about.biodiversitylibrary.org/tools-and-services/

PIDapalooza2019[65]) and important to ensure this doesn't impede addition of heritage literature to the corpus of available records.

User stories alluding to new kinds of literatures

"As a data scientist (or researcher), I want to know whether any given preprint has subsequently been published. If yes, then for these to be linked reciprocally (from preprint to publication; from publication to preprint)."

"As a studentship funder I want to know - whether the PhD studentship actually ended in thesis, how to find and how to cite this thesis; what artefacts (papers, data, software, samples, instruments, ...) can be identified that either contributed to the PhD research or are the PhD research outcomes."

"As the Manager of the Biodiversity Heritage Library (BHL) in Australia, I want to see DOIs applied to the world's biodiversity heritage literature so that this literature can be part of the great linked network of scholarly research (the DOI system)."

[65] https://zenodo.org/record/2547570#.XG6OJJP7TUI