



Analysis and Comparison of Persistent Identifier Use and Integration across Disciplines and Sectors

Document Information

Date: 21/09/2016

Authors: Sünje Dallmeier-Tiessen (CERN), Robin Dasler (CERN)

Reviewers: Maaïke Duine (ORCID), Laura Rueda (DataCite)

Abstract: In order to evaluate the outcomes of THOR project and the continued sustainability of its outputs, it is necessary to continuously monitor the evolution of the PID ecosystem. We have developed systems and tools with which to conduct this monitoring. This document describes our initial observations, the challenges faced, and recommendations for how to solve them. This is an interim report. It will be updated with final results at the end of the project.

DOI 10.5281/zenodo.154592

This work was supported by the THOR Project. The THOR project is funded by the European Union under H2020-EINFRA-2014-2 (Grant Agreement number 654039). The following report is based on an internal deliverable submitted on 30/06/2016.

Visit <http://project-thor.eu> for more information.



Contents

Executive Summary	1
1 Introduction	1
2 Challenges of PID Evaluation	1
2.1 Defining Disciplines	1
2.2 Identifying Institutions	2
2.3 Incomplete Metadata	2
3 PID Trends	3
3.1 Quantifiable Metrics	3
3.1.1 THOR Dashboard	3
3.1.2 Partner Sources	5
3.2 User Perspective	6
3.2.1 ORCID Survey	6
3.2.2 Data Publishing Survey	6
3.2.3 Broader Trends in the PID Landscape	7
4 Alignment with Project Outreach	7
5 Implications for THOR	9
5.1 Interaction with Project Partners	9
5.2 Interim Recommendations: Research	9
5.3 Interim Recommendations: Development	10
5.4 Interim Recommendations: Outreach	10
6 Conclusions	11
7 References	11
Appendix A: Terminology	13
Appendix B: Project Summary	14



Executive Summary

In order to evaluate the outcomes of project THOR as well as the continued sustainability of its outputs, it is necessary to continuously monitor the evolution of the PID ecosystem. We have developed systems and tools with which to conduct this monitoring.

This document describes our initial observations, challenges and recommendations. It is an interim report outlining our progress so far, and will be updated with final results at the conclusion of the project.

1 Introduction

The THOR project has set out to address the gaps in the persistent identifier (PID) landscape identified as part of the project that directly preceded it, the ‘ORCID and DataCite Interoperability Network’ (ODIN). During the first year of THOR, our research, development and outreach has already addressed some of these gaps. To ensure that this work continues to address the perceived gaps appropriately, it is necessary to observe the PID landscape continuously to gauge adoption of PID systems and services. This will assist us with identifying remaining challenges, and allow us to evaluate whether THOR actions have been achieving the expected impact. The sustainability work of THOR has already presented – and will continue to present – observations and measurements, which will in turn continue to influence research, software and service developments within the project as well as our community engagement and outreach strategies.

In order to foster and leverage this observational work, we have developed a dashboard for monitoring the interconnected PID ecosystem (Dasler, 2016) as well as a set of key performance indicators (KPIs) to assist with strategic decision making. This document provides an interim assessment of these initial inroads as well as an interim analysis of PID use and integration in our stakeholder community based on these measures.

2 Challenges of PID Evaluation

In the course of developing tools to observe and evaluate adoption of PID systems across stakeholders and disciplines, we encountered a few challenges. These challenges have been presented to the THOR project team, feeding into different work streams. They are now lines of inquiry for the research arm of THOR.

2.1 Defining Disciplines

Open Science is a framework and philosophy that does not explicitly entail a one-size-fits-all solution. To better understand motivations and barriers for Open Science, it is important to have an understanding of differences in disciplinary practices. However, as is true in other areas that rely on specific classification, there are limitations in our current ability to distinguish between disciplines in a definitive and quantifiable way. When designing and discussing the requirements of the dashboard, it became evident that it is currently not possible to show such disciplinary patterns in a precise way, and at a level of granularity



that would be meaningful for informing discipline-specific outreach efforts. We can use readily available statistics as a proxy, such as overall counts from disciplinary data centres, but these are at best rough approximations, given that:

1. A data centre's affiliation may be too broad; for example, 'biology' is too general to be useful.
2. A data centre may represent more disciplines than its primary affiliation; for example, not everyone who publishes an article or deposits data at CERN is a physicist.

Even without these concerns, metadata about subject classification is often not provided by the data centres; DataCite does not require it to mint a DOI, for example. When it is provided, the information provided by multiple data centres may be at varying levels of granularity, or otherwise not well-aligned due to the use of different classification schemes. This makes it difficult to come to meaningful aggregate conclusions.

The challenge of defining disciplines stems largely from metadata issues, and as such is related to the more general challenge of incomplete metadata described below (see Heading 2.3). However, it is highlighted here as a specific class of metadata problem that has a direct effect on project goals and outcomes.

2.2 Identifying Institutions

During the focus groups conducted as part of the dashboard development (Dasler, 2016), it became evident that there is a need for institutions to observe Open Science developments among their constituents. This is not only insightful in terms of observing the spread of policy compliance, but it enables institutions to understand service gaps and determine where additional institutional support is needed. As with the challenge of defining disciplines, there is currently little information available related to the uptake of standardised institutional identifiers. Moreover, it is difficult to extract and compare affiliations from DOI metadata and ORCID metadata. This makes a case for pursuing the improvement and/or promotion of organisational identifiers as a means to identify specific institutions efficiently and in an unambiguous way.

Aside from the progress monitoring function, institutional IDs, once adopted, could also facilitate institutional impact assessment in an efficient and agile way; they could bring a new and comprehensive perspective to funding decision-making processes. These conversations have fed into the THOR research efforts to make this a higher priority in the project.

2.3 Incomplete Metadata

The services provided by DataCite and ORCID require minimal metadata prior to acquiring a PID (DataCite, 2011; ORCID, n.d. b). This is beneficial for reducing barriers to user adoption and fitting different use-cases. However, it makes it difficult to obtain a complete picture of the current state of PIDs. This is reflected in the THOR dashboard (see Figure 1), which was introduced in a previous report (Dasler, 2016) and briefly described in the next section. Object type definitions found in DataCite metadata include a significant number of 'unknowns' and 'others': 'unknown' refers to objects that did not indicate a resource type, while 'other' refers to objects that has a resource type other than the ones listed, out of the 13 allowed types.



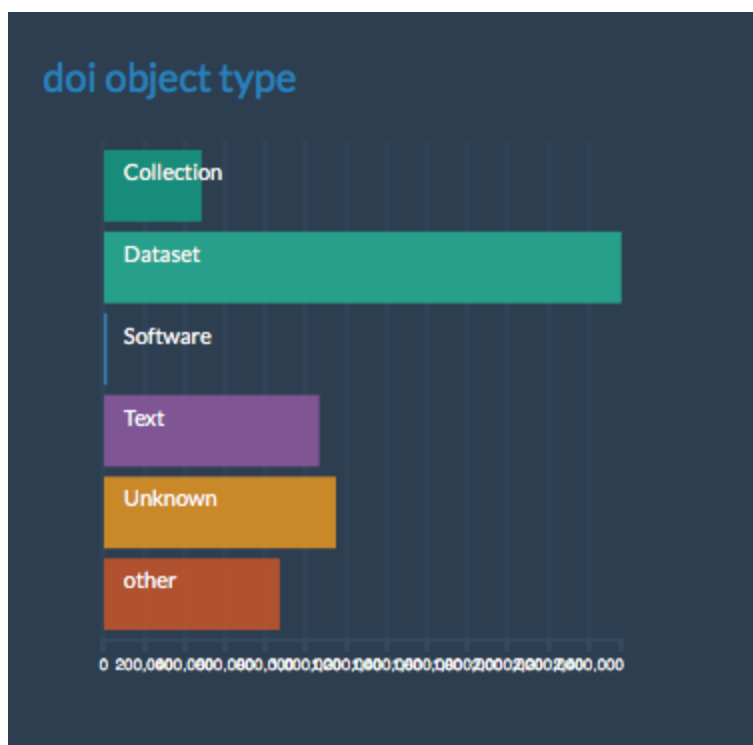


Figure 1: Object type definitions in DataCite metadata. Screenshot taken from the THOR dashboard.

This issue was also raised in the focus group conducted by the THOR sustainability team in September 2015, when attendees expressed an interest in viewing aggregate metadata for the purpose of assessing their own institution's compliance with, and contribution to clean and complete metadata. THOR's work on metadata standards and interoperability represent milestones in addressing this challenge, identifying gaps in existing PID infrastructures and potential strategies to overcome these gaps (Fenner et al., 2015). This is also being addressed in DataCite's forthcoming metadata schema update.

3 PID Trends

To build a complete picture of the adoption of PID systems, it is necessary to gather perspectives from multiple sources. Basic quantifiable metrics can be automatically harvested or gathered from partner analytical sources in order to provide a baseline of raw counts. This information is useful for visualising trends at a glance, but it does not give us a rounded understanding of the human motivations involved in PID trends. For this information, it is necessary to hear from users, through surveys or studies. Both quantifiable metrics and the human perspective on trends in PID uptake are presented below.

3.1 Quantifiable Metrics

3.1.1 THOR Dashboard

Basic metrics are harvested in an easy-to-use dashboard (see Figure 2). The dashboard was built as part of the sustainability and evaluation component of THOR to aggregate statistics on DOIs and ORCID iDs using the DataCite and ORCID APIs. Statistics are collected monthly from DataCite and ORCID. Different overlays allow users to dig down into object views (DataCite DOIs metrics) and researcher views (ORCID metrics).

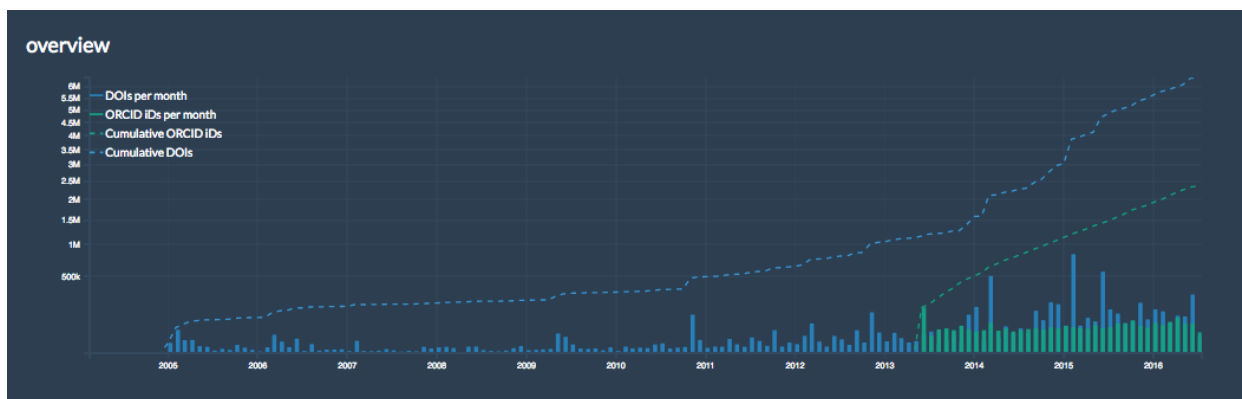


Figure 2: Portion of the dashboard showing statistics collected on DOIs from DataCite

A number of baseline statistics on ORCID and DataCite metadata and their interconnections with other services are monitored via the dashboard. These are used as key performance indicators in conjunction with metrics relevant to other aspects of the THOR project. The status of these PID indicators as of 1 June 2016 is shown below (see Table 1).

These indicators represent a steady upward trend in the amount of both ORCID iDs and DataCite DOIs, showing that uptake is still on the rise. For instance, between 1 June 2015 and 1 June 2016 (the life of THOR thus far) the total number of live ORCID iDs has increased from 1,423,886 to 2,297,632, an increase of approximately 60%, while the total number of DataCite DOIs minted has increased from 4,701,605 to 6,150,207 during the same period, an increase of approximately 31%. This rise is unlikely to be directly attributable to THOR, but the rise is an indication that demand in ORCID and DataCite services continues to grow.

Most DataCite DOIs that indicate a resource type are datasets, which is not surprising. Assigning collection-level DOIs is also a significant use case for DataCite. However, just under 19% of DataCite DOIs do not have an associated resource type. While this information is not required in order to mint a DOI, its inclusion is recommended by DataCite, and will become a mandatory field in the DataCite Metadata Schema 4.0 to accommodate the increasing diversity of the Open Science landscape. As this information can be useful to data centres (be it the original DOI-assigning data centre or another), additional or improved outreach efforts may be required in order to encourage data centres to collect it.

The indicators around ORCID iD uptake and usage paint an interesting picture. Out of a total of over 2 million ORCID iDs, only 20% include works (that is, references to research outputs such as publications, conference presentations, and data sets) (ORCID n.d. a). This shows that there is a significant portion of ORCID iD holders who do not maintain their ORCID profiles as a works list, whether public or not. While the reasons for this are not definitive, a majority of respondents in a recent ORCID survey (see Section 3.2.1) indicated their primary use of their ORCID iD was in manuscript submission, which is backed up by anecdotal evidence. From the institutional perspective, it is also evident from the statistics reported in the dashboard that few publishers or repositories are consistently adding ORCID iDs to the metadata they submit to DataCite when minting a DOI. The entire system would benefit if this reporting rate were improved.



Table 1: PID Indicators from the Dashboard

Metric	Value
Total number of live ORCID iDs	2,297,632
Total number of ORCID records that include works	467,587
Total number of works across all ORCID iDs	13,840,369
Total number of unique DOIs across all ORCID iDs	6,225,032
Total number of ORCID records that include employment information	822,024
Total number of ORCID records that include funding information	176,664
Total number of DataCite DOIs minted	6,150,207
Total number of DataCite DOIs whose records include at least one ORCID	255,822
DOIs by General Resource Type	
Total number of DataCite DOIs having the Dataset general resource type	2,563,708
Total number of DataCite DOIs having the Collection general resource type	484,467
Total number of DataCite DOIs having the Software general resource type	14,738
Total number of DataCite DOIs having the Text general resource type	1,067,138
Total number of DataCite DOIs having any other general resource type	871,223
Total number of DataCite DOIs having no associated general resource type	1,149,433

Table 2: PID types for works reported in ORCID author records

PID Type Reported in ORCID Record	Number of Records Containing This PID Type
DOI	9,133,100
Scopus EID	4,960,974
Web of Science UID	1,238,569
PubMed PMID	1,082,314
ISSN	1,075,876
PMC	247,127
ISBN	210,373
arXiv	118,759
BibCode	75,404
Handle	20,833

Even though only 20% of ORCID iDs include works, those iDs account for the over 13 million works claimed across the ORCID system. This means that those iDs that include works contain an average of just under 30 works each. Distribution of works per ORCID iD is not readily available from data presented on the dashboard, but it is not uncommon for some authors to have fuller ORCID profiles that list over a hundred works. This will be further investigated in the coming months.

3.1.2 Partner Sources

A variety of PID types for works are additionally reported in ORCID author records (see Table 2). This information was obtained through collaboration with ORCID, as this type of information is not currently available via the dashboard. Unsurprisingly, DOI is the primary PID type reported in ORCID author records, followed by popular database-specific identifiers.



3.2 User Perspective

The indicators we can observe via the THOR dashboard rely on information automatically harvested from partner services. While they present readily quantifiable data, they do not reflect the nuances that may factor into human decision-making. Insight into the thoughts and motivations of those who make use of partner services must therefore be gathered from elsewhere.

3.2.1 ORCID Survey

In 2015 ORCID conducted a survey (Armstrong, et al. 2015a) to gauge awareness of their services. The full details can be found in both the report and the dataset, which are available on Figshare (Armstrong, et al. 2015b). Those results most relevant to the work of THOR are summarised here.

56% of respondents indicated they were either ‘very familiar’ or ‘somewhat familiar’ with ORCID iDs, with a significantly higher level of awareness among publishers and librarians. 41% of respondents, including more than 70% of those who indicated they did not have an ORCID iD, were unaware that ORCID iDs are collected or required by many publishers, research funders, and institutions. As the report suggests, this shows that more work is needed to raise awareness of ORCID and its potential benefits to the researcher beyond name disambiguation. This is a point for the THOR outreach team to address, as it will have implications for the overall success of the project.

In particular, respondents indicated that colleagues were the most likely points of referral to ORCID (31%). Indications like these prompted the start of the THOR ambassador programme. In contrast, respondents were much less likely to have learned about ORCID via a funder or professional organisation (3% and 5%, respectively). This provides a good point of focus for THOR outreach efforts. By informing the professional organisations about PID services, we can leverage existing relationships to increase PID uptake.

Some of the findings supported observations made from the dashboard. A majority of respondents (56%) said that their primary use of ORCID iDs was in submitting manuscripts for publication, which could be a cause for the high percentage of ORCID iDs that exist but include no works information. Furthermore, 29% of respondents indicated they had not updated their ORCID profile at all.

3.2.2 Data Publishing Survey

A survey by Kratz and Strasser (2015) investigated researcher expectations around the concept of ‘data publishing.’ This work made a distinction between data publishing, and data that was merely shared. Kratz and Strasser found that data publishing was perceived to include a unique identifier and rich, formal metadata.

There are recommendations that arise from these findings. Data publishers such as disciplinary repositories should enable this formal citation approach by assigning PIDs, and specifying a preferred citation format. Article publishers should encourage authors to cite data formally in the reference list, thereby elevating and formalising the citation of data. There are now more data publishing solutions than ever, most of which include PIDs: DataCite’s re3data.org¹ even provides the means for researchers to identify appropriate data repositories, and now includes badges to highlight repository services, such as DOI assignment.

¹ <http://re3data.org>





THOR can assist in this formalisation of data citation by working with publishers, repositories and researchers to encourage formal citation practices and foster the development of additional or improved data publishing solutions that include easy-to-implement support for PIDs.

3.2.3 Broader Trends in the PID Landscape

Publishers increasingly require or support the use of ORCID iDs in the manuscript submission process (Haak, 2016), as do repositories for data submissions. While similar mandates from funders regarding funding applications are not yet as common, they are becoming more prevalent (Wellcome Trust, 2015; Kerridge, 2015). And yet the ORCID survey indicates that many in the scholarly research community are unaware of this trend – and even fewer are likely to have learned of ORCID through these entities. This disconnect in the scholarly ecosystem must be addressed both for the benefit of researchers, publishers and funders, and for the overall efficiency of the PID network. The human infrastructure arm of THOR is positioned to help connect these pieces, with both researchers and publisher/funders as target stakeholders.

In addition to a high-level trend for mandates among publishers and funders, several disciplines are beginning to extend the call for persistent identifiers for researchers and their outputs. The Coalition on Publishing Data in the Earth and Space Sciences (COPDESS) has published a Statement of Commitment that stipulates the referencing of datasets through registered DOIs, and identification of researchers with ‘relevant community persistent identifiers’ (COPDESS, 2015); funders and grants should also receive PIDs.

The COPDESS Statement further calls for uniquely identifying samples, which is itself a growing trend in several disciplines. In the field sciences (earth sciences and others), McNutt et al. (2016) endorse ‘liberating’ samples and data in part through the use of persistent identifiers for these materials and outputs. In nucleic acid research, the Encyclopaedia of DNA Elements (ENCODE) project assigns PIDs to experiments, files, antibody lots, and biosamples so that specific pieces of data can be uniquely identified (Sloan et al., 2016).

Other disciplines are also joining the call for promoting PID use. Researchers and librarians are now promoting the value and utility of PIDs for their various disciplines. Such calls-to-action have surfaced in materials science (Austin, 2016), medical sciences (Akers et al., 2016), crop sciences (Williams, 2016), and chemistry (Walter, 2016), among others.

The call to make better use of PIDs is not limited to the sciences. In the social sciences and humanities, IMPACT-EV is investigating how to measure the scientific, political, and social impact of EU research in those disciplines. PIDs, especially ORCID iDs, are a part of their strategy.

4 Alignment with Project Outreach

When monitoring broader trends in the PID landscape, human involvement cannot be ignored. It is for this reason that the ongoing outreach message of THOR is consistently aligned with the trends in PID uptake, and is poised to promote and/or influence these trends accordingly. Though a full analysis of project outputs falls under the purview of other ongoing work, an overview of trends in the THOR message is presented below to provide some framing at this interim phase.



The THOR blog has been active on the project website (<http://project-thor.eu>) since the beginning of the project. The THOR Twitter account tweeted its first tweet on 10 July 2015. These two media are our primary online outlets.

It is clear that the community is responding most actively to concrete outputs. On the THOR blog, practical information, and announcements about events and project outputs receive higher views and engagement than other content posted (see Table 3). This shows an appetite for meaningful personal engagement with topics related to PIDs.

Announcements of concrete practical outputs have also ranked highly on Twitter, as have announcements for the dashboard and the soft launch of the knowledge hub coordinated by the THOR outreach team (see Table 4). The dashboard in particular resulted in the most 'engagements' of any tweet (see Table 5), meaning that more users clicked through on information in that tweet.

Table 3: Top 10 Blog Posts by Views

Post Title	Total Views
The next step for open science: a state-of-the-art identifier network	637
Differences between ORCID and DataCite Metadata	528
July 7, 2016 THOR Workshop: Identifiers - Infrastructure, Impact and Innovation	399
THOR at a glance	289
Knowledge hub set to soft-launch, join our webinar on January 28th!	268
Monitor the Identifier Landscape with the THOR Dashboard	228
Organisation IDs for scholarly communications: where next?	222
Contributor Information in DataCite Metadata	208
The THOR Ambassador Programme is GO!	206
From Pilot to Service	179

Table 4: Top 10 Tweets by Impressions (Views)

Tweet Text	Impressions (Views)
Join our Webinar on Article - Data linking services, March 17! https://t.co/og5lydnMFG	6244
Monitor the Identifier Landscape with the THOR Dashboard https://t.co/RYnv93vnQI https://t.co/ASO4plwIO2	5985
Do you have a background in Library and Information Science? @CERN is hiring someone to work on THOR! See: https://t.co/gglqqQUvIN	5264
Interactive API docs for ORCID http://t.co/EuA4FuuzWA http://t.co/eaNQ0Ug8Ou	5052
We're hiring a THOR Events and Training Officer (ORCID). Great opportunity to get involved! https://t.co/FzVMqQTo3E	4995
Differences between ORCID and DataCite Metadata http://t.co/JQ9TkXJl9J	4566
The THOR Ambassador Programme is GO! http://t.co/eiE3RAXAKC	4307
#Data experts from @CERN @emblebi PANGAEA and many more speaking at THOR Workshop! Don't miss it! See: https://t.co/52PuoQVnz0	3244
July 7, 2016 THOR Workshop: Identifiers Infrastructure, Impact and Innovation https://t.co/EH8AiNtYPX	2312
Knowledge hub set to soft-launch, join our webinar on January 28th! https://t.co/Fhq1OqQtxs	2122





Table 5: Top 10 Tweets by Engagements (Clicks)

Tweet Text	Engagements (Clicks)
Monitor the Identifier Landscape with the THOR Dashboard https://t.co/RYnv93vnQI https://t.co/ASO4plwIO2	100
Differences between ORCID and DataCite Metadata http://t.co/JQ9TkXJI9J	80
Do you have a background in Library and Information Science? @CERN is hiring someone to work on THOR! See: https://t.co/gglqqQUvIN	64
Interactive API docs for ORCID http://t.co/EuA4FuuzWA http://t.co/eaNQ0Ug8Ou	52
We're hiring a THOR Events and Training Officer (ORCID). Great opportunity to get involved! https://t.co/FzVMqQTo3E	51
Reps from #EU projects @Eudat_eu @EdisonEU @OpenDreamKit and more working towards collaboration @DANSKNAW https://t.co/xoLt6COEKp	46
July 7, 2016 THOR Workshop: Identifiers Infrastructure, Impact and Innovation https://t.co/EH8AiNtYPX	42
DOI for geoscience data: how early practices shape present perceptions http://t.co/uWfTHkr2XN	38
The THOR Ambassador Programme is GO! http://t.co/eiE3RAXAKC	37
Join our Webinar on Article - Data linking services, March 17! https://t.co/og5lydnMFG	36

5 Implications for THOR

These trends in the adoption of PIDs have implications that can be fed back into the work of THOR in the sustainability and evaluation cycle built into the project. At this interim stage, we propose a number of recommendations for the research, service development and outreach components of THOR.

5.1 Interaction with Project Partners

The responsibilities of the THOR sustainability team entail a close collaboration with partners, to understand their ongoing developments while simultaneously feeding information back to them about general developments both inside and outside the project. This involves monitoring the PID landscape through the THOR dashboard as well as developing a set of key performance indicators (KPIs) to gauge project progress. Additionally, regular meetings are held to update the THOR team and to discuss specific sustainability activities, such as the focus groups, KPI development and sustainability models.

5.2 Interim Recommendations: Research

Integration with funders should be improved. As evidenced by the ORCID survey and the observed trends described in Section 3, there is policy support for the use of PIDs in funding agency requirements, and yet, as further evidenced by the ORCID survey, many researchers are unaware of this support.

The question of metadata completeness and quality, mentioned previously as a challenge in gathering PID trends and as a metric desired by focus group participants, remains an open question. Although it sits outside the original scope of THOR, metadata improvement could assist with PID adoption. The



extent to which partner services can assist in improving metadata, and whether this does indeed require a technical component instead of a purely human one, is still to be explored.

The inclusion of organisational persistent identifiers in repositories and publisher systems would help researchers to associate themselves with the correct institution in an efficient, precise and consistent way, which would, in turn, help data centres and publishers gather and maintain accurate metadata. It would also then be easier to gather aggregate statistical information on institutions and their outputs in an automated way, which would benefit evaluation services such as the THOR dashboard. The THOR research team has already begun work on exploring the inclusion and promotion of organisational identifiers as part of our work on interoperability of PIDs²; however, it is critical that this work is continued.

5.3 Interim Recommendations: Development

THOR partners are already making PID integration into repository software easier, as evidenced in a recent report (de Mello et al., 2016). To further support and continue this work, it would be beneficial to provide better documentation on liabilities and dependencies to support ORCID integration, and to develop modules for easy DOI minting as part of the publishing process on these platforms. Lee (2015) shows that many institutional repositories use the identifiers that came with the repository software, and those identifying a need to move to a different identifier schema must go through a planning and testing phase. This demonstrates a need to foster easier, more seamless integration into repository software.

Integration of PID services into publishing workflows needs to become a priority. We must facilitate the efforts of publishers to support ORCID integration, make data citation seamless, and provide bi-directional article-data linking. Discussions on the technical infrastructure necessary to support this work have already begun within THOR. Furthermore, now that the disciplinary partners have demonstrated the successful incorporation of ORCID authentication into existing services (de Mello et al., 2016), additional human infrastructure work is necessary to communicate the technical results to non-partner publishers appropriately.

5.4 Interim Recommendations: Outreach

Because there is a need for grassroots efforts to educate researchers about PIDs, and because colleagues are a primary influencer, as evidenced by the ORCID survey, the THOR outreach team should continue to facilitate and expand the project's ambassador network. Specifically, we should expand the network by reaching out to 'multipliers', to those potential ambassadors who already have their own established network that we can tap in to for expansion. We should also leverage the reach and influence of scholarly organisations: these are currently underutilised connections that have much to gain from increasing researcher awareness of PID services, particularly those services that the organisations themselves are working to implement. It could be helpful to encourage other organisations to follow the example of the American Geophysical Union (AGU), an early adopter of ORCID's peer review programme for acknowledging and giving credit for peer review activities.

In terms of message, the outreach team, through the ambassadors, should advocate the use of PIDs for both community improvement and individual convenience. Special emphasis should be placed on recommending implementations of PID services based on the work of the THOR technical infrastructure

² See: <https://doi.org/10.5281/zenodo.48705>





branch, such as article-data linking in repositories, and the outreach team should pay special attention to reaching out to those with the technical capacity to further this infrastructure. It is of special importance for THOR to provide direct, hands-on help to integrators to make PID integration in repositories intrinsic and expected. This is valuable justification for additional face-to-face workshops, especially the multi-day ‘bootcamps’ currently planned for autumn 2016 and 2017.

Given that there appears to be support for policies and mandates around PID use, it is important for THOR to engage with policy makers – a currently underutilised stakeholder group. In particular, THOR can serve as a central community support hub for initiatives around PIDs, assisting policy makers with implementing and communicating their PID-related efforts.

Collaboration with publishers is similarly important, as the ORCID survey shows that publisher requirements are a prime motivation for iD adoption. Using DOIs to reference data, as with articles, helps avoid link rot, and having both machine readable and human readable reference mechanisms helps to foster adoption by both repositories and researchers.

6 Conclusions

As part of THOR’s sustainability and evaluation work, we have developed the means to continuously monitor the PID landscape. Our initial investigations have shown that PID uptake is on the rise and support for the use of PIDs, both for researchers and data, is becoming more commonplace. THOR is well-placed to foster further uptake and use, so that unique and persistent identification and citation become mainstream. Through our work, we will make these capabilities and habits an expected component of the practice of Open Science.

7 References

- Akers, K. G., Sarkozy, A., Wu, W., & Slyman, A. (2016). ORCID Author Identifiers: A Primer for Librarians. *Medical Reference Services Quarterly*, 35 (2), 135–144. Retrieved from <http://doi.org/10.1080/02763869.2016.1152139>
- Armstrong, D., Haak, L., Meadows, A., & Stone, A. (2015)a. *ORCID 2015 Survey Report*. Retrieved from 10.6084/m9.figshare.2008206
- Armstrong, D., Haak, L., Meadows, A., & Stone, A. (2015)b. *ORCID 2015 Survey Report Data*. Retrieved from <http://doi.org/10.6084/m9.figshare.2007522>
- Austin, T. (2016). Towards a digital infrastructure for engineering materials data. *Materials Discovery*. Retrieved from <http://doi.org/10.1016/j.md.2015.12.003>
- COPDESS. (2015, 14 January). COPDESS Statement of Commitment – COPDESS. Retrieved from <http://www.copdess.org/statement-of-commitment/>
- DataCite. (2011). DataCite Metadata Schema for the Publication and Citation of Research Data. *DataCite*. Retrieved from https://datacite.commpla.com/sites/default/files/document/DataCite-MetadataKernel_v3.1.pdf



- Dasler, R. (2016). THOR: Metrics and Tools. *Zenodo*. Retrieved from <http://doi.org/10.5281/zenodo.46761>
- de Mello, Guilherme et al. (2016). Demonstration of Services to Integrate ORCIDs into Data Records and Database Systems. *Zenodo*. doi: [10.5281/zenodo.58971](https://doi.org/10.5281/zenodo.58971)
- Fenner, M., Consortium, T., Dasler, R., Vision, T., Rueda, L., Kotarski, R., ... Demeranville, T. (2015). D2.1: Artefact, Contributor, and Organisation Relationship Data Schema. *Zenodo*. Retrieved from <http://doi.org/10.5281/zenodo.30799>
- Fenner, Martin et al. (2016). THOR: Conceptual Model of Persistent Identifier Linking. *Zenodo*. doi: [10.5281/zenodo.48705](https://doi.org/10.5281/zenodo.48705)
- Haak, L. (2016). Publishers to Start Requiring ORCID iDs [Text]. Retrieved 5 May 2016, from <https://orcid.org/blog/2016/01/07/publishers-start-requiring-orcid-ids>
- IMPACT-EV. (n.d.). Retrieved 5 May 2016 from <http://impact-ev.eu/>
- Kerridge, S. (2015, 20 July). The UK Recommends ORCID [Text]. Retrieved 5 May 2016 from <http://orcid.org/blog/2015/07/16/uk-recommends-orcid>
- Kratz, J. E., & Strasser, C. (2015). Researcher Perspectives on Publication and Peer Review of Data. *PLOS ONE*, 10 (2), e0117619. Retrieved from <http://doi.org/10.1371/journal.pone.0117619>
- Lee, D. J. (2015). Research Data Curation Practices in Institutional Repositories and Data Identifiers. Retrieved from <http://fsu.digital.flvc.org/islandora/object/fsu%3A253104/>
- McNutt, M., Lehnert, K., Hanson, B., Nosek, B. A., Ellison, A. M., & King, J. L. (2016). Liberating field science samples and data. *Science*, 351 (6277), 1024–1026. Retrieved from <http://doi.org/10.1126/science.aad7048>
- ORCID (n.d.)a. Add works to your ORCID record. Retrieved 20 September 2016 from <http://support.orcid.org/knowledgebase/articles/462032-add-works-to-your-orcid-record>
- ORCID (n.d.)b. XML for orcid-bio. Retrieved 13 May 2016 from <http://members.orcid.org/api/xml-orcid-bio>
- Sloan, C. A., Chan, E. T., Davidson, J. M., Malladi, V. S., Strattan, J. S., Hitz, B. C., ... Cherry, J. M. (2016). ENCODE data at the ENCODE portal. *Nucleic Acids Research*, 44 (D1), D726–D732. Retrieved from <http://doi.org/10.1093/nar/gkv1160>
- Walter, P. (2016, 2 February). I am a number! Retrieved from <http://www.rsc.org/chemistryworld/2016/02/researcher-number-orcid-news-leader-editorial>
- Wellcome Trust. (2015, 30 June). Who are you? Recognising researchers with ORCID identifiers. Retrieved from <https://blog.wellcome.ac.uk/2015/06/30/who-are-you-recognising-researchers-with-orcid-identifiers/>
- Williams, S. C. (2016). Practices, Policies, and Persistence: A Study of Supplementary Materials in Crop Science Journals. *Journal of Agricultural & Food Information*, 17 (1), 11–22. <http://doi.org/10.1080/10496505.2015.1120213>





Appendix A: Terminology

Additional terms are defined below:

Term	Definition
API	Application programming interface
COPDESS	Coalition on Publishing Data in the Earth and Space Sciences
DataCite	An organisation that develops and supports methods to locate, identify and cite data and other research objects. Specifically, DataCite develops and supports the standards behind persistent identifiers for data, and the members assign them. See https://www.datacite.org
DOI	Digital Object Identifier
EC	European Commission
ENCODE	Encyclopaedia of DNA Elements
EU	European Union
KPI	key performance indicators
ODIN	ORCID and DataCite Interoperability Network
ORCID	An organisation that creates and maintains a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers. See http://orcid.org
PID	Persistent Identifier
THOR	Technical and Human Infrastructure for Open Research



Appendix B: Project Summary

The **THOR** project establishes a sustainable international e-infrastructure for persistent identifiers that enables long-term access to critical information about the life cycle of research projects. It enables seamless integration between articles, data, and researcher information creating a wealth of open resources. This will result in reduced duplication, economies of scale, richer research services, and opportunities for innovation.

The project has four concrete aims:

1. Establishing interoperability
2. Integrating services
3. Building capacity
4. Achieving sustainability

The project will meet these aims by defining relations between contributors, research artefacts (including data), and organisations. We will incorporate these relationships into the ORCID and DataCite systems. We will also expand existing linkages between different types of identifiers and versions of artefacts to improve interoperability across platforms and integrate ORCID iDs into production systems for article and data submission services in pilot communities and beyond.

The consortium will develop systems to embed new PID resolution techniques into existing services to support seamless direct access to artefacts, and in particular data. We will create services to allow associations between datasets, articles, contributors and organisations at the time of submission. Building on these, we will deliver the means to integrate trans-disciplinary PID services in community-specific platforms, focusing on cross-linking, claiming mechanisms and data citation (guided by the FORCE 11 data citation principles³).

For more information, visit <http://project-thor.eu> or contact info@project-thor.eu

³ <https://www.force11.org/group/joint-declaration-data-citation-principles-final>

