



Business Plan for Sustaining the THOR Federated PID Infrastructure and Services

Document Information

Date: 28/04/2017

Authors: Xiaoli Chen (CERN, orcid.org/0000-0003-0207-2705), Sünje Dallmeier-Tiessen (CERN, orcid.org/0000-0002-6137-2348), Robin Dasler (CERN, orcid.org/0000-0002-4695-7874), Artemis Lavasa (CERN, orcid.org/0000-0001-5633-2459)

Reviewer: Jo McEntyre (EMBL-EBI)

Abstract: This interim report outlines the considerations of sustainability of the THOR project – an EU-funded project seeking to establish a sustainable international e-infrastructure for persistent identifiers. It reviews the initial business plan from the project’s proposal for continued relevance, considers the sustainability activities of DataCite and ORCID as key THOR infrastructural partners, and provides a set of open questions to guide the project’s future work. The final report will ultimately result in a business strategy for THOR outputs.

DOI 10.5281/zenodo.569862

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 interim deliverable produced on 31 August 2016.

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



Executive Summary

The THOR project aims to improve the overall interoperability of persistent identifiers (PIDs) by developing connected services. The THOR project is not creating new stand-alone tools, nor are we assembling an overarching organisation that seeks to be permanent. Instead, our partners are collaborating to build services that they can integrate into their existing offerings, thus absorbing and sustaining THOR outputs as part of their normal operations.

One of the goals of the sustainability branch of THOR, therefore, is to formulate a sustainable business plan for the project's outputs. The THOR project has considered sustainability from the outset, as evidenced by an initial project business plan included in the project proposal (THOR Consortium, 2014). At this interim stage, this report reviews that initial business plan for continued relevance, takes an initial look at approaches to sustainability within THOR, and presents a slate of open questions to pursue during the second half of the project.

The THOR approach to sustainability relies on THOR outputs being absorbed by our partner organisations as part of their regular operations. This, in turn, means that the sustainability of THOR's outputs relies on the continued sustainability of those partner organisations. While the continued sustainability of all THOR partners is relevant to the sustainability of the work of THOR, this document will focus on the sustainability activities of DataCite and ORCID: as the primary fundamental infrastructure providers on the project, their success is the linchpin of the disciplinary partners. Other relevant organisations and identifier systems are included for comparison.

As part of their involvement in THOR, both DataCite and ORCID have been expanding their services to make information sharing activities more straightforward, facilitating bi-directional communication links between their services and the services of other THOR partners through PIDs, thus providing direct incentives for the research community. The other THOR partners are actively contributing to the sustainability of THOR outputs by integrating PID services into production services within their own institutions and research communities.

As we pursue sustainability within and between the partners, some open questions have arisen. These questions revolve primarily around four areas: the need for openness in PID services, dependencies between PID infrastructural components, the role of centralisation in PID service provision, and persistence. These questions will be explored throughout the remainder of the THOR project, culminating in the final report at project's end.



Contents

1	Introduction	1
2	Creating Sustainable Value	1
2.1	THOR's Approach to Sustainability	1
2.2	THOR Outputs	2
2.3	Revisiting the Initial Business Plan	2
2.3.1	Goal and Community Needs	3
2.3.2	Business Model	3
2.3.3	Assumptions and Risks	3
3	Factors that Influence Sustainability	4
3.1	THOR Partners	4
3.1.1	DataCite	4
3.1.2	ORCID	6
3.2	Other PID Service Providers	7
3.2.1	Crossref	8
3.2.2	ISNI	8
3.2.3	Database-Specific IDs	9
4	Open Questions	9
4.1	Openness	9
4.2	Dependencies	10
4.3	Centralisation	10
4.4	Persistence	10
4.5	Service Adopter Sustainability	11
5	References	12
	Appendix A: Pricing Models	13
	Appendix B: Terminology	15



1 Introduction

The THOR project aims to improve the overall interoperability of persistent identifier (PID) services. This will enable researchers in Europe and beyond to identify themselves and their work, no matter the format, in a seamless network of attribution and access. In order for the project to truly make improvements, to truly effect change, the outcomes of the project must be everlasting.

One of the goals of the sustainability branch of THOR is to formulate a sustainable business plan for the project's outputs. In this first phase of development, we will document what consideration our partners have given to sustainability thus far, which in turn will determine our work until the end of the project. Specifically, we will:

- Outline the project's approach to sustainability
- Re-examine our initial sustainability assumptions
- Explore the commitment to sustainability by our key partners
- Raise open questions to further guide work for the remainder of the project period and beyond

2 Creating Sustainable Value

2.1 THOR's Approach to Sustainability

THOR has set out to make fundamental changes and improvements to the PID infrastructure. These changes and improvements ultimately belong to the research community at large, melding disparate tools and influencing everyday practice. Because of this, the THOR project is not creating new stand-alone tools, nor are we assembling an overarching organisation that seeks to be permanent. Instead, our partners are collaborating to build services that they can integrate into their existing offerings, thus absorbing and sustaining THOR outputs as part of their normal operations.

For the primary persistent identifier service providing partners, DataCite and ORCID, whose services form the backbone of our overall interoperability efforts, this is especially true. Both organisations have taken steps to ensure THOR outputs are brought into their business operations beyond the project term. This includes aligning THOR objectives with their business goals and addressing crucial organisational sustainability concerns in order to build trust among stakeholder communities. Thus, at this interim stage, the focus of this report is placed on the approaches taken by DataCite and ORCID regarding sustainability, given that they are pillars of PID infrastructure and that the sustainability of their services underpins the persistence of other THOR impacts.

The other THOR partners are actively contributing to the sustainability of THOR outputs by integrating PID services into production services within their own institutions and research communities. This means that THOR is directly resulting in usable community-embedded services, rather than pilots or proofs-of-concept. Though valuable, these activities are more appropriately monitored and expanded upon in other THOR reports designated for that purpose. The focus of the present business planning is on those THOR partners who provide core PID infrastructure, as their continued success is the linchpin of the success of the community-embedded services.



2.2 THOR Outputs

The THOR project is producing work that influences a range of stakeholders and impacts PID infrastructure from both a technical and human perspective. While a full evaluation of the project outputs after the first year has been undertaken in other reports (Dallmeier-Tiessen & Dasler, 2016; and Dallmeier-Tiessen, Dasler & Lavasa, forthcoming), a summary will be provided here.

THOR partners are building technical capacity from two fronts: service development by the PID issuing agencies; and service integration by PID service adopters, for example, data centres and publishers. In the meantime, all partners build ‘human capacity’ jointly by contributing to community engagement efforts.

Table 1: Sustainability factors relevant to the THOR project and its constituents

	PID issuing agency	PID service adopter	PID user
THOR project output	Community insights Technical capacity Core services	Service integration Community-specific development experience	Awareness
Factors that influence sustainability	Trust Financial stability Open governance Operational transparency Degree of collaboration	Trust Level of integration Investment of resources Value of service to end users	Trust Adoption Attitude
Means of evaluation	Business plan	Commitment	Community engagement

This variety of outputs poses a challenge to sustainability, particularly given that no single entity will ultimately be responsible for maintenance. As Table 1 shows, in order to present a comprehensive picture of the current – and future – state of PID use, different means of evaluation are needed to investigate sustainability. This is particularly relevant as multiple partners are involved in tasks and outputs. Given the central role of ORCID and DataCite as PID issuing agencies, particular emphasis will be given to their business plans in this report.

2.3 Revisiting the Initial Business Plan

The THOR project has considered sustainability from the outset, as evidenced by an initial project business plan included in the project proposal (THOR Consortium, 2014). The plan included an outline of business goals and community needs, a proposed business model, and a consideration of the assumptions and risks regarding project output and community-wide impact. A year into the project, it is worthwhile to reflect on the original business plan with an eye to the current situation, to see whether the assumptions present in that document still hold true.



2.3.1 Goal and Community Needs

The overall goal of THOR, as stated in the initial business plan, is ‘to act as a catalyst for the development of a widespread, interoperable PID infrastructure by addressing the need to unambiguously identify researchers and the digital objects they create and rely on’. This goal has remained constant since THOR’s inception. It is still our aim to support scholarly communication by building and reinforcing the PID infrastructure from both human and technical perspectives. By collaborating to enhance services, and forming partnerships with key community stakeholders, THOR project partners are creating concrete PID services that meet community needs, building capacity within different scholarly communities, and making sure these services persist through PID issuing organisations.

Furthermore, the original overarching project objectives continue to be relevant to addressing community needs for interoperable solutions, easily integrated services, training and support, and analysis.

2.3.2 Business Model

Our approach to sustainability has not changed since the inception of the project, and so the core tenets of the business model as proposed for THOR remain consistent with our current approach. Namely:

- The results of the project will persist in the day-to-day activities of the partners
- The intellectual property created by the project will be placed in the public domain
- Open Source software and training material will support PID integration

This commitment to openness is central and critical to THOR’s approach to sustainability, and indeed plays a role in the overall sustainability of open, neutral, and fair research infrastructure (Bilder et al., 2015). This commitment is shared by the THOR partner organisations, and will be specifically expanded upon in a later section.

2.3.3 Assumptions and Risks

The initial assumptions for the project’s sustainability were:

- THOR will demonstrate a portfolio of useful PID-based services within the project timeline
- These services will be scalable, be applicable across domains, and have good cost–benefit
- The cost of delivery and maintenance of THOR services will be marginal for both service providers and for service users
- Services developed by the project will be operationally robust, not prototypes
- Project partners and third parties will be able to enhance the service portfolio through the development of self-financing, value-added services on top of the core infrastructure

As far as service development is concerned, the assumptions are certainly still valid, as only halfway into the project the various partners have already demonstrated an expanded portfolio of operational PID services, a trend that is likely to continue over the project lifetime. ORCID and DataCite are themselves cross-domain platforms, so the risk of domain incompatibility is still low. Furthermore, it is in the best interest of the various project partners to take on only those services that will be scalable, with good cost–benefit, and will support additional value-added services. These tendencies act as a vetting mechanism for THOR outputs, thus reinforcing those assumptions and making the associated risk minimal.



3 Factors that Influence Sustainability

As previously emphasised, the THOR approach to sustainability relies on THOR outputs being absorbed by our partner organisations as part of their regular operations. This, in turn, means that the sustainability of THOR's outputs relies on the continued sustainability of those partner organisations.

There are several factors that influence sustainability, as demonstrated in Table 1, such as financial stability and open governance on the side of the PID *issuing* agencies, and the level of integration of PIDs on the side of PID service *adopting* institutions. However, the foremost factor is trust.

In keeping with Bilder, Lin and Neylon's *Principles for Open Scholarly Infrastructures* (2015), this trust spans organisational governance and oversight, financial stability, and the guarantee that the community retains control. Ultimately, establishing trust also encourages adoption and use of PIDs, and for the user to remain engaged with such practices. These considerations are currently being addressed in the PID service provider space, both by THOR partners and other service providers, and they will be examined further as this project moves forward.

3.1 THOR Partners

While the continued sustainability of all THOR partners is relevant to the sustainability of the work of THOR, this document will focus on the sustainability activities of DataCite and ORCID: as the primary fundamental infrastructure providers on the project, their success is the linchpin of the disciplinary partners. Their approaches to sustainability, given their more broadly applicable services, can serve as models for other organisations.

These THOR partners have demonstrated leadership in the exploration of sustainability issues relevant to long-lasting PID service models by considering openness and trust in the development of their service models. These are crucial elements of the organisations' overall business strategies, and not only specific THOR project tasks. Nevertheless, this work can help inform and validate THOR's approach to sustainability, while the THOR sustainability work can in turn put these efforts into a wider context.

3.1.1 DataCite

DataCite was founded in 2009, and is a leading global non-profit organisation providing persistent identifiers in the form of Digital Object Identifiers (DOIs) for research data. DataCite's strength is rooted in its active membership. Its global community of members, from more than 20 countries, includes data centres, libraries, government agencies, research universities, and more. DataCite members are the voting body of the organisation; membership is open to all organisations that support its data sharing mission. DataCite's members work with data centres, stewards, libraries, archives, universities, publishers and research institutes who have responsibility for managing, holding, curating and archiving data.

DataCite's goal is to help the research community locate, identify and cite research data with confidence. Work is carried out on several fronts to achieve this goal: primarily, the organisation supports the creation and allocation of DOIs and accompanying metadata; it provides services that support the enhanced search and discovery of research content; and at the same time, it promotes data citation and advocacy through community-building efforts, responsive communication and outreach materials.



DataCite’s operational services are funded by annual membership fees. It is currently growing its membership and hardening its services to meet the growing demand for DOIs for data. Some new initiatives are primarily funded by grant funding.

DataCite’s business model is illustrated in Figure 1, in business canvas format (Osterwalder et al., 2010). The business model shows on a relatively high level the unique position of DataCite among its partner organisations and the community it serves, and how the value propositions links the two sides.

DATA CITE










<p>Key Partners</p>  <ul style="list-style-type: none"> - DataCite members - Data centres - Identifier service providers (Crossref, ORCID, etc) - Metadata aggregators (COS, Plum Analytics, Thomson Reuters, etc) - The members of the THOR consortium 	<p>Key Activities</p>  <ul style="list-style-type: none"> - DOI creation, allocation, and management of DOIs and accompanying metadata. - Maintaining a robust and responsive service infrastructure. - Providing outreach, communication and support. - Engaging in research activities to improve quality of service. <p>Key Resources</p>  <p>Internal:</p> <ul style="list-style-type: none"> - Executive Board - Staff <p>External:</p> <ul style="list-style-type: none"> - DataCite DOI allocating agents (Member organisations) - Partner organisations 	<p>Value Propositions</p>  <ul style="list-style-type: none"> - Provide DOI allocation resource for data centres - Provide metadata creation support for data hosting organisations. - Connect data to publications, software and other materials. - Provide auxiliary data search and discovery functions for data management service providers. - Provide interoperability across platforms. - Contribute to research and development effort for the data service and scholarly communication service providers community. - Contribute to advocacy and community engagement effort for the wider researchers community. 	<p>Beneficiary Relationships</p>  <ul style="list-style-type: none"> - Co-creation of persistently identified datasets. - Automated public services through Find Data and Event Data. - Full voting rights on the decisions of the organisation for members. - Dedicated personal assistance on technical issues. - Maintain community relationship through social media and public events. <p>Channels</p>  <ul style="list-style-type: none"> - Membership/ PID allocation privilege - Partnership - Public services - Technical and general resources 	<p>Beneficiary Segments</p>  <ul style="list-style-type: none"> - Data centres - Data service providers - Data publishers - Conventional publishers - Scholarly communication service providers - Data related research groups - Government agencies - Higher Education sector - Institutional repositories - University libraries - Research institutes - Research libraries - Data service interest groups
<p>Cost Structure</p>  <p>Operation: Daily operations including running the service, providing technical support, finance and human resources tasks.</p> <p>Outreach: Marketing campaigns, outreach events, community presence.</p>		<p>Impact metrics</p> <ul style="list-style-type: none"> - Individual identifier adoption rate - Membership subscription rate - Service adoption rate - Community engagement level 	<p>Revenue Streams</p>  <ul style="list-style-type: none"> - Membership fees - Grants - Service fees 	

Figure 1: DataCite's Business Model

Sustainability Considerations and Change

In summer 2015, DataCite’s Executive Board hired an entirely new team to run the DataCite organisation. This triggered a number of organisational changes including growth in the number of DataCite members, changes in the service portfolio, and engagement with new communities. There was also a recognition that the administrative processes, bylaws, communication and engagement activities, and service portfolio needed to change to keep pace with the diverse and changing requirements of the research data community.

The bylaws and administrative processes changed in early 2016 to expand the type of organisations that could become DataCite members, and to streamline the process for membership. Previously, only non-profit organisations could become DataCite members. This limited DataCite’s growth and did not recognise the importance of the for-profit contributions to data sharing activities. Applications for membership were also only considered once a year, which severely limited DataCite’s growth. Both of these barriers to membership growth were changed. This had an immediate impact on DataCite’s growth, with many new member applications coming from the for-profit sector.



DataCite has traditionally worked through its immediate membership, which act as DOI allocating agents, to develop policies, sustainability practices, and partnerships. A large number of DataCite members are at national or large academic libraries. This membership practice has limited DataCite's direct interaction with the organisations (or better, data centres) that are actively involved in the creation, management, and sharing of research data. Without these direct connections, it is difficult to develop policies, services and communication strategies that meet the data community's needs. To mitigate against this barrier, DataCite has recently formed steering groups composed of representatives from groups such as data centres and funders to help guide its work.

With funding from the THOR project, DataCite is actively expanding its services to make data sharing activities more straightforward, and to provide direct incentives for the data community. These new services include direct interoperability with ORCID, and connection between DataCite's datasets and Crossref's journal articles. Together with quality communication, its services provide evidence of the importance of data sharing and are helping to grow the organisation to meet sustainability needs.

3.1.2 ORCID

ORCID is an independent non-profit organisation with an international reach and interdisciplinary scope. It partners with European Commission projects and organisations across the scholarly community to deliver value directly to institutions and individual users (Figure 2). Individuals can create and maintain their ORCID iDs free of charge, while institutions can become paid members in order to have access to the ORCID Member API and to support the ORCID mission. ORCID has expanded its membership base rapidly during the last year, and it is expecting to change its currently mixed funding model to 100% supported by membership dues by 2017 (Haak, 2016).

Sustainability Considerations and Trust

As ORCID has evolved into a core infrastructure provider, it has used the concept of trust as a lens for its considerations of sustainability. In spring 2016, ORCID launched the ORCID Trust programme (Paglione, 2016), a comprehensive policy programme outlining ORCID's approach to security, privacy, compliance, governance and persistence, and status (for example, incident reports and information about scheduled maintenance).

ORCID's sustainability is ensured through the development of the aforementioned membership fee-based business model, and through commitment to open source principles, including CC0 licenses, open software source code, documentation and technical specifications. The latter has also played a key role in shaping the organisation's succession planning process. Trustworthiness is an integral part of sustainability, as it encourages service integration significantly and secures the users' commitment to using and contributing to PID services.

The main goal of the Trust programme is to enhance transparency in those areas. This remains true to ORCID's ten founding principles about openness and availability, governance transparency, reliability and user control (ORCID, 2011), which are values that are also supported by ORCID's bylaws (ORCID, 2015). ORCID is an organisation built and managed by the ORCID community, and governed by a (mainly) non-profit board. As a non-profit organisation, it can only be purchased or managed by another non-profit organisation, if such a situation should arise. In order to further ensure transparency, summaries of board meetings and annual financial reports are made public, as are its US Internal Revenue Service Form 990 annual tax filings.



ORCID










<p>Key Partners</p>  <ul style="list-style-type: none"> - Members - Partner organisations - Consortial partners 	<p>Key Activities</p>  <ul style="list-style-type: none"> - ORCID creation, allocation and metadata management. - Research and development - Technical maintenance - Community outreach - Service support - Operation administration - Contribute to advocacy and community engagement effort for the wider researcher community. - Contribute to the R&D effort for the PID service providers. <p>Key Resources</p>  <p>Internal:</p> <ul style="list-style-type: none"> - Governance board - Executives - Technical team - Community support team - Technical capacity <p>External:</p> <ul style="list-style-type: none"> - Members - Partner organisations 	<p>Value Propositions</p>  <ul style="list-style-type: none"> - Provide researcher with free persistent identifiers. - Provide scholarly communication service providers with a solution for interoperability and new metrics possibility. - Provide higher education institutes with solution for data and researcher identity management procedure, metrics for assessing institutional impact, etc. - Provide research funders, facilities and platforms with streamlined information exchange possibilities. - Enable cross-sectoral efficiency gains by improving the quality and trustworthiness of individual attribution. - Provide open, community-governed infrastructure components and APIs. 	<p>Beneficiary Relationships</p>  <ul style="list-style-type: none"> - Free PID service for the researcher community. - Self-service through documentation resources. - Personalised technical support for members. - Community support through social media. - Community feedback through ambassador program and public events. <p>Channels</p>  <ul style="list-style-type: none"> - General ORCID registration service - Membership - ORCID system integration service - Technical and general resources 	<p>Beneficiary Segments</p>  <ul style="list-style-type: none"> - Global researcher community - Global scholarly communication sector - Research funders - Higher Education sector <ul style="list-style-type: none"> - Institutional repositories - University libraries - Research institutes - Research libraries
<p>Cost Structure</p>  <ul style="list-style-type: none"> - Operation: Daily operations including running the service, providing technical support, finance and human resources tasks. - Outreach: Marketing campaigns, outreach events, community presence. 		<p>Impact metrics</p> <ul style="list-style-type: none"> - Individual identifier allocation rate - Membership subscription rate - Service adoption rate - Community engagement level 	<p>Revenue Streams</p>  <ul style="list-style-type: none"> - Membership fees - Grants 	

Figure 2: ORCID's business model

As part of its activities within the THOR project, ORCID has taken significant steps to facilitate bi-directional communication links between our services through PIDs. Examples of this include the advances achieved with the integration of ORCID IDs into the systems of several disciplinary data repositories, namely PANGAEA, EMBL-EBI and CERN, thus creating automatic links between contributors and their datasets (de Mello et al., 2016).

3.2 Other PID Service Providers

In addition to DataCite and ORCID, there are a range of PID registration agencies for authors, contributors and digital objects with which THOR (or at least one of THOR's partner organisations) collaborates in a formal or informal way. It is especially relevant to explore how these agencies or providers sustain their business within the context of making PID services sustainable. It should be noted that for smaller or more 'localised' or 'specialised' identifier systems, it appears to be almost impossible to find reliable information about business strategies or models. Hence, this work will be continued and refined in years two and three.

For the purpose of this document, some highlights have been included pertaining to the sustainability practices of the agencies and systems most pertinent to this study. Of particular relevance is, of course, Crossref. As the core registration agency in scholarly publishing, Crossref has proven its viability for over a decade. Through the British Library, ISNI is a THOR collaborator and will be studied in more depth through year two. A range of database-specific IDs are briefly discussed, but will be taken into account for deeper analysis as the THOR project progresses.



3.2.1 Crossref

The goal of Crossref is to improve research communications via research and development of innovative technologies, operating PID infrastructure, providing metadata services, and community outreach (Crossref, 2015).

Crossref is operated by Publishers International Linking Association, Inc. (PILA), a non-profit membership association of publishers. Membership is open to any publisher of original scholarly content, and includes many organisations that do not consider themselves to be publishers primarily, such as university libraries and government agencies. Crossref’s services are still primarily focused on the publisher members, who are charged content registration fees on top of an annual membership subscription (see Figure 3 and Appendix A). Organisations that do not register content but want access to Crossref’s metadata can use the openly available REST API or sign up as an Affiliate and pay a fee for particular interfaces.

CROSSREF









<p>Key Partners</p>  <ul style="list-style-type: none"> - Member publishers - Affiliate organisations - Partner organisations - Funders - EC projects 	<p>Key Activities</p>  <ul style="list-style-type: none"> - Operation administration - Service running - Research and Development - Technical maintenance - Service support - Community outreach 	<p>Value Propositions</p>  <ul style="list-style-type: none"> - Provide metadata deposit service for publishers. - Provide value-adding services (e.g. Crossmark, similarity check, Conflict check) for publishers. - Provide DOI allocation resource for scholarly content hosting organisations. - Contribute to the R&D effort of the scholarly communication sector. - Provide metadata support for all member organisations. - Provide free DOI query service for general user 	<p>Beneficiary Relationships</p>  <ul style="list-style-type: none"> - Free metadata search, REST API and Funder Registry self-services for the researcher community. - Automated DOI allocation and management services for member agencies. - Community support through social media and public events. <p>Channels</p>  <p>Online:</p> <ul style="list-style-type: none"> - CrossRef website/ member service platform - Social media - Webinar - Ambassador meeting <p>Offline:</p> <ul style="list-style-type: none"> - CrossRef annual meeting - Conferences - Workshops 	<p>Beneficiary Segments</p>  <ul style="list-style-type: none"> - Global researcher community - Global scholarly communication sector - Research funders - Higher Education sector - Institutional repositories - University libraries - Research institutes - Research libraries - Public libraries
<p>Cost Structure</p> <p>Operation: Daily operations including running the service, providing technical support, finance and human resources tasks.</p> <p>Resource: DOI membership fee</p> <p>Outreach: Campaigns, events, community presence.</p> 		<p>Impact metrics</p> <ul style="list-style-type: none"> - Individual identifier allocation rate - Membership subscription/ renewal rate - Service subscription rate - Community engagement level 	<p>Revenue Streams</p> <ul style="list-style-type: none"> - Membership fees - Service fees 	

Figure 3: Crossref’s business model

3.2.2 ISNI

The International Standard Name Identifier (ISNI) is an identifier system defined by international standard ISO27729, which is managed by a non-profit governing organisation. Its registration agencies are predominantly national libraries from around the world. ISNI is contributing to the THOR project through the British Library, and ORCID incorporates ISNI into institutional identification.

For their services, ISNI charges an initial fee, an annual fee and a tiered fee per batch of identifiers (see Appendix A). Organisationally, ISNI is a mix of the models of Crossref and DataCite, consisting of founding members’ registration agencies as well as regular members.



3.2.3 Database-Specific IDs

In addition to PID systems intended as global unique identifiers (like DOIs and ORCID iDs), there are PID systems that are internal to specific services or databases. These PIDs can be either for people or objects. While many such systems exist, those summarised below are the most useful for our comparison, given their prevalence and efforts at compatibility.

ResearcherID and the Scopus Author ID are examples of database-specific person IDs. ResearcherID is a person identifier offered by Thomson Reuters. The identifier is complementary to ORCID. The two identifiers can be linked so that data can be exchanged between them (Thomson Reuters, n.d.). Registration is free for the individual and users retain rights to all their data (Thomson Reuters, 2013). The Scopus Author ID similarly grew locally and was then connected to ORCID, so that ORCID iDs permeate the Scopus system and are widely used to enhance Scopus data.

Database-specific IDs for objects appear more often than database-specific IDs for people. In comparison with other disciplines, there is a long tradition of use in the life sciences. A well-known case in point is the PMCID of PubMed Central (PMC), the openly accessible full-text archive of bio-chemical and life sciences literature maintained by the US National Institutes of Health National Library of Medicine (NIH/NLM). The functions of the PMCID exceed those of conventional database-specific object IDs, extending beyond system operation necessity. The PMCID acts as a 'badge' for publication that complies with the public access policy of the NIH/NLM.

The arXiv identifier from arXiv.org, the preprint server primarily for physics, mathematics and computer science, is another widely recognised database-specific object ID. ArXiv.org has a long history, having first appeared in 1991 at Los Alamos National Laboratory (arxiv.org, n.d.). Its compliance with the OAI-PMH protocol enables a high level of interoperability with other services.

Database-specific identifiers like those mentioned serve a useful purpose for automatically unifying works across a platform, allowing for ease of effort for the researcher. Though there is some cross-platform compatibility between these services and others outside the databases in question, these types of identifiers are ultimately insufficient for amassing a researcher's entire cross-platform body of work. Neutral third-party solutions are therefore a necessary and beneficial addition to the ecosystem. Developments in the area of database-specific identifiers are important to note, and it should be investigated further how local identifiers interoperate with global PIDs in a sustainable manner.

4 Open Questions

After the first year of THOR and several years of experience in PID service operations, our partners have demonstrated leadership in considering long-term sustainability issues. Tangible lessons have been learned to further improve the robustness of PID service infrastructure for the future. However, PID service sustainability is still a developing area. Based on these lessons learned, it became apparent that there remain many open questions that should be addressed through the second and third years of THOR.

4.1 Openness

THOR's sustainability approach and our initial business plan assumptions call for open development and transparency of service delivery. This approach is supported by the current policies (ORCID, 2013;



DataCite, 2016) and collective outlook of our various partners as well as others in the PID services space (Bilder et al., 2015). Given the ever-shifting landscape of PID service providers, it will be useful for THOR to characterise the benefits of our open approach. Specifically:

What is the role of non-profit and/or open resources in the sustainability of PID services?

What is the role of commercial services in the sustainability of PID services? How can they maintain openness?

How can we reassure the community not only that organisations and services will continue to exist for the long-term but that they will continue to maintain this spirit of openness?

All of these questions relate to the overall trust that the research community can expect to have in the long-term sustainability of PID services and scholarly infrastructure in general.

4.2 Dependencies

Services developed and maintained by the disciplinary partners within THOR rely on the services developed and maintained by the fundamental service partners, which in turn rely (at least in part) on services developed and maintained by other organisations.

How do dependencies between infrastructure components impact the sustainability of PID services?

This question of dependencies ties in with broader issues of preservation, persistence, and trustworthiness. How reliable is each link in the chain? How robust against breakage is the overall chain? This brings us to a related question:

What impact does an organisation's archival trustworthiness have on the sustainability of that organisation's services?

This type of question is of interest to all data-holding organisations, for example, data centres, libraries and long-term cultural heritage institutions.

4.3 Centralisation

Related to the question of reliability, dependency and robustness is the question of centralisation. This applies to centralisation of service operation, centralisation of data holding and centralisation of community practice. In the short term, it is convenient and beneficial to end users to have services maintained and operated in a centralised manner. Moreover, centralisation can assist the promotion and implementation of community-wide standardised practice. However, in the case of failure (either technical or organisational, or the case of loss of community accountability) centralisation can swiftly become a problem for sustainability. THOR should therefore explore:

What is the best balance between centralisation and decentralisation in terms of making PID services sustainable?

4.4 Persistence

When we refer to identifiers as being persistent, this comes with many possible assumptions. With all the different factors that feed in to an operational PID infrastructure, which components are we actually



saying are persistent? This is both a technical question, if databases or underlying data are considered, as well as a human question, if the persistence of the organisation as a functional entity is considered.

What does it really mean for an organisation or service to be persistent? What level of persistence is required for PID services?

Similarly, if the primary function of a PID infrastructure provider is to offer a service, to what extent, if any, can it guarantee the long-term preservation of the information that powers that service?

When talking about the sustainability and persistence of a service, where does long-term preservation of metadata fit? If long-term preservation of the scholarly record is not the primary task of PID providers, is that a role to be filled by institutions with a long-term mission, for example, a national library?

Exploring these questions is beneficial not only for the immediate needs of THOR but will also serve to further the conversation around these issues in the broader context of interoperable PID services.

4.5 Service Adopter Sustainability

Up to this point, our considerations of sustainability have been largely viewed from the perspective of PID issuing agencies, as these are the cornerstones of the sustainability of the PID system. However, to complete the circle of trust that is essential to sustainability across the board, it will be necessary to expand our view to include the perspectives of PID service adopters. Ultimately, PID users must be able to put their trust in PID service adopters (data centres, publishers, and so on), and the connecting services they develop and maintain, while those PID service adopters will in turn put their trust in the fundamental infrastructure of the PID-issuing agencies.

Furthermore, impact plays a direct role in determining which services should be sustained. Sustaining all services always is not scalable. Organisations will need to determine which services have the greatest potential for longevity, and sustain those services accordingly. These impact assessments should not be taken lightly. As the various PID infrastructure and service providers continue to become further entwined, determinations made by one entity will cascade throughout the system.

We recognise that a large-scale assessment of the sustainability of individual PID service adopters is untenable. Three open logistical questions that frame our continued efforts in years 2 and 3 are therefore:

How do we encourage and evaluate the sustainability of the tools and secondary infrastructure developed by PID service adopters in a way that is widely applicable and scalable?

Can the presence of these service adopters functioning as a bridge in the PID ecosystem, connecting PID issuing agencies and PID users, further secure the sustainability of the system as a whole?

Which indicators and/or metrics for determining impact should be considered by organisations in deciding which services to sustain?

These questions have been compiled in consultation with representatives from the leading PID services. Addressing these will ultimately allow THOR to put a more comprehensive and informed business plan together at the end of the project, which can be used as a roadmap to secure the overall sustainability of the PID ecosystem.



5 References

- arXiv.org (n.d.). *arXiv Primer*. Retrieved from: <https://arxiv.org/help/primer>
- Bilder G., Lin J., & Neylon C. (2015). *Principles for Open Scholarly Infrastructure-v1*. Retrieved from <http://dx.doi.org/10.6084/m9.figshare.1314859>
- Crossref (2015). *2014–2015 Annual Report*. Retrieved from http://www.crossref.org/crossref_annual_report_2014/index.html
- Dallmeier-Tiessen, Sünje, & Dasler, Robin. (2016). Analysis and Comparison of Persistent Identifier Use and Integration across Disciplines and Sectors. Zenodo. doi: [10.5281/zenodo.154592](https://doi.org/10.5281/zenodo.154592)
- Dallmeier-Tiessen, Sünje, Dasler, Robin, and Artemis Lavasa (2017, forthcoming). Interim Evaluation of Project Outputs and their Alignment with THOR Objectives.
- DataCite Mission* (2016). Retrieved from <https://www.datacite.org/mission.html>
- Geoffrey Bilder, Jennifer Lin, & Cameron Neylon. (2015). *Principles for Open Scholarly Infrastructures-v1*. Figshare. Retrieved from <https://doi.org/10.6084/m9.figshare.1314859>
- Haak, L. (2016, August 16). *We won't be Sold*. Retrieved from <https://orcid.org/blog/2016/08/16/orcid-we-wont-be-sold>
- de Mello, G., Graef, F., Stocker, M., Schindler, U., Dasler, R., McEntyre, J., & Dallmeier-Tiessen, S. (2016). *Demonstration of Services to Integrate ORCID's into Data Records and Database Systems*. <http://doi.org/10.5281/zenodo.58971>
- Osterwalder, A., Pigneur, Y., & Clark, T. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. Hoboken, NJ: Wiley.
- ORCID (2011). *Our Principles*. Retrieved from <http://orcid.org/about/what-is-orcid/principles>
- ORCID (2013). *ORCID Public data file use policy*. Retrieved from <http://orcid.org/content/orcid-public-data-file>
- ORCID (2015). *Bylaws of ORCID, Inc*. Retrieved from <https://orcid.org/bylaws>
- Paglione, L. (2016, March 3). *ORCID Trust*. Retrieved from <https://orcid.org/blog/2016/03/03/orcid-trust>
- Space data and information transfer systems: audit and certification of trustworthy digital repositories*. (2012). Geneva: ISO-16363:2012(E). Retrieved from <http://cds.cern.ch/record/2019804>
- Thomson Reuters (2013). *Your gateway to researchers and their published works: Researcherid.com - A free, global community where researchers connect*. Retrieved from: http://wokinfo.com/media/pdf/ResearcherIDFS_web.pdf
- Thomson Reuters (n.d.). *ResearcherID and ORCID Integration*. Retrieved from: <http://wokinfo.com/researcherid/integration/>
- THOR Consortium (2014). *THOR Project Proposal*.



Appendix A: Pricing Models

For easy reference, the pricing model of Crossref and ISNI services are summarised in comparable format in the tables below, as both organisations operate under membership model, as with THOR PID service providers DataCite and ORCID.

Crossref

While Crossref provides its service completely free of charge for individual users (researchers) and most public libraries, it charges membership fees for organisational and institutional users like publishers, archives, service providers, research libraries, and so on. Crossref provide three main types of membership: Publishers, Affiliates and Libraries. Affiliates are further divided into subtypes depending on the services demanded by the user organisation (Table 2).

Table 2: Crossref fee model

Membership type	Charging model	Service
Publishers	Mandatory annual fee based on member org. revenue	Membership subscription (m)
	Service fee	Locally host CrossRef metadata (Basic CMS)
	Per content type/ DOI/ volume	Deposit
	Per volume or annual fee	Similarity check
Affiliates	Per volume and item type	Crossmark deposit
	Per conflict	Conflict penalty
	Mandatory annual fee based on member org. revenue	Indirect deposit and query
		Direct query
		Locally host CrossRef metadata
Locally host CrossRef metadata		
End-user lookup	Free of charge	One at a time query
Libraries	Free of charge	DOI retrieval
		Metadata retrieval

These specifications generated a relatively complex charging matrix for Crossref to tailor their service according to each organisation profile, also allows multiple revenue streams to exist in parallel.



ISNI

ISNI provides two types of membership for its organisational users. To subscribe as a Registration Agency (RAG) allows a member to allocate ISNI identifiers for its users on behalf of ISNI; the RAG has access to both the API and the web interface that processes online ISNI requests. The remaining services provided by ISNI are shared by both RAG members and regular members, including access to whole non-confidential database (via API), access via additional indexes, modification of records, and so on. ISNI charge larger mandatory fees to its RAG members, while the optional service fees are the same for both types of memberships (Table 3).

Table 3: ISNI membership price model

Membership type	Charging model	Service
<p>Registration Agency (Org. based on geographic region and business sector)</p>	<p>Mandatory fee based on member org. revenue</p> <p>Initial fee</p> <p>Annual fee</p> <p>Service fee based on volume of record</p>	<ul style="list-style-type: none"> • ISNI allocation • Database access • API • One at a time query • Record editing • High quality linked data • Networking opp. <p>Batch upload</p>
<p>ISNI member</p>	<p>Mandatory fee based on member org. revenue</p> <p>Initial fee</p> <p>Annual fee</p> <p>Service fee based on volume of record</p>	<ul style="list-style-type: none"> • Database access • API • One at a time query • Record editing • High quality linked data • Networking opp. <p>Batch upload</p>



Appendix B: Terminology

Additional terms are defined below:

Term	Definition
API	Application programming interface
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
ID	Identifier
ISNI	International Standard Name Identifier
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
PILA	Publishers International Linking Association, Inc.
PMC	PubMed Central
PU	Deliverables are classified in the Grant Agreement as either Public (PU) or Confidential (CO)
RAG	Registration Agency
THOR	Technical and Human Infrastructure for Open Research
WP	Work Package