Unique Identifiers:

Current Landscape and Future Trends

Research Data Canada
IDs Working Group, Standards and Interoperability Committee

Developed by the RDC stakeholder community, including:
Mark Leggott, Kathleen Shearer, Chantel Ridsdale, David Baker, Eugene Barsky

Version 1.0
doi:10.5281/zenodo.557106

September 9, 2016

Table of Contents

1. Introduction
2. The Identifier Landscape
    2.1 Researcher Identifiers
    2.2 Object Identifiers
    2.3 Organizational Identifiers
    2.4 Equipment Identifiers
    2.5 Other Identifiers
3. Discussion
4. Recommendations
5. Acknowledgements
6. Bibliography

1. Introduction

The research enterprise generates a great deal of information, both in physical and digital form, representing a wide range of data, from descriptive information about researchers, to publications, to datasets resulting from a research project. This information is scattered across many systems and technologies including human resource systems, grant management systems, publication databases, repositories, web pages and so on.

Persistent Identifiers (PIDs) are the anchors that facilitate links between related information. Essentially, *IDs* are labels that refer to a specific entity in the information landscape, such as an object, organization, person, dataset, etc. For example, in the same way that a 'Person' record has the fields 'First Name' and 'Last Name' it should be a best practice for each such record to also have a unique field for 'ID' to make the record more useful as a research information source since there can be more than one author with the same first and last names. A *persistent* ID adds value to an ID by providing *a long-lasting reference to a digital object that gives information about that object regardless what happens to it.*

PIDs serve two major functions: using a label or ID to uniquely identify an object, person or organization so that a reference to the entity can be unambiguous; and to provide a mechanism to locate an entity over time even when they move (persistence). A PID System adds a third function: a framework (e.g. a software and associated repository) for discovering objects described by a PID and doing something with them (such as viewing the object) and doing so in a sustainable way. The lack of PIDs, or the poor use/maintenance of them, stifles the power of discoverable or actionable information gained from linking multiple items, or objects, together - a fundamental characteristic of information systems.

PIDs are needed in research to unambiguously locate, link and cite journal articles, data and other research products such as samples, software, formulas, as well as entities in the research process, such as authors, funders and institutions. PIDs are now available for literature, data, samples, authors and more.

This paper is the first in a series of best practice documents produced by Research Data Canada that will be combined with other outputs from RDC and other organizations to provide a foundation and roadmap leading towards a comprehensive national framework for research data management in Canada. It is intended to provide an overview of the current landscape of research PIDs, and will provide insights into their role in developing a more cohesive virtual research environment, supporting the preservation, discoverability, and reuse of research information.

The aims of this paper are to:

1. provide all readers, regardless of expertise, with a better understanding of the role of unique identifiers;

2. identify best practices in the international research data management community that facilitate the development of robust and sustainable systems for identifying and linking research outputs;
3. describe the current state of adoption of PIDs in Canada and internationally;
4. offer recommendations to the Canadian research community about the adoption of PIDs in various contexts.

2. The Identifier Landscape

There are many types and formats of identifiers currently used in the research and scholarly context. This report focusses on those IDs most relevant for the research data management landscape, and for which there is a considerable agreement in the broader community. Other PIDs will be listed below, but will not be discussed in detail in this document.

The ID landscape is large and complex, but some basic concepts are defined here to facilitate the reader's review of this document. The **ID** is *a name that identifies (that is, labels the identity of) either a unique object or a unique class of objects, where the "object" or class may be an idea, physical [countable] object (or class thereof), or physical [noncountable] substance (or class thereof)*.[1] A **PID** is *a long-lasting reference to a digital object that gives information about that object regardless of what happens to it*.[2] A **PID minting organization** (or **service**) is an organization that provides a service (typically a software service) that when requested creates a unique PID (typically according to a standard) that can used in a system: for example to create an HTTP URI that resolves to a web page with description of that object. A **URI** is *a string of characters used to identify or name a resource on the Internet.*[3] A **portal**, or **discovery system**, is an interface (typically a web page and associated software components) that provides access to a collection of IDs and associated URIs, generally via a search or browse option.

The ultimate goal of an effective and sustainable research data ecosystem is to provide the user (human or machine) with the ability to discover unique outputs from the research data lifecycle and link those outputs with the individuals and resources (e.g. equipment) that were involved in their creation. It is through this ecosystem, or network of linked data (metadata and binary data files), that all stakeholders in the research data community are able to answer the full range of questions they may have.

2.1 Researcher Identifiers

Researcher Identifiers, also known as Author Identifiers or Scholar Identifiers, are alphanumeric strings that establish a unique identity for a given author or creator. Researcher IDs (RIDs) are becoming more critical in the scholarly communication system. As the number of scholars

[1] "Identifier," *Wikipedia.org*, last modified July7, 2016, https://en.wikipedia.org/wiki/Identifier.
[2] "Persistent Identifier," *CASRAI.org,* last modified August 13, 2015, http://dictionary.casrai.org/Persistent_identifier.
[3] "Uniform Resource Identifier," *CASRAI.org,* last modified August 13, 2015, http://dictionary.casrai.org/Uniform_resource_identifier.

grows, and there is an increasing likelihood of identical names, the resulting confusion about which author one is referring to will also grow. Additionally, a researcher may publish using several variations of their name over time; or they may change their name or institutional affiliation, leading to their work being incorrectly associated with the wrong author. RIDs help to disambiguate researchers by connecting a researcher with a unique number that is associated with them throughout their career. These PIDs also greatly facilitate the tracking or linking of scholars and their outputs, whether by people or by software agents and systems.

Some typical use cases for RIDs are:

1. A funder would like to retrieve all the publications by the researchers associated with a specific grant.
2. A researcher would like to retrieve all full-text articles published by a colleague in a specific subject area. The colleague has changed institutions several times in her career.
3. A university administrator would like to assess the intellectual impact of a department or research group to determine their areas of strength.
4. A publisher or repository would like to validate the name of an author and distinguish them from other authors with similar names.

Over the years, different actors or systems have utilized a number of different types of RIDs, usually for their own internal purposes including publishers, archives, libraries, commercial aggregators, database providers and funders. Some examples of these are:

- AUID - used in the Elsevier Scopus database
- Google Scholar ID - used in the Google Scholar database (e.g. - https://scholar.google.com/citations?user=H-M7ztAAAAAJ)
- VIAF Number - The Virtual International Authority File is an aggregation of national authority data and provides a single unique ID for specific individuals. (e.g. - https://viaf.org/viaf/104750011/)

The problem with these IDs is that they are system-specific and often maintained by a single commercial organization for a specific internal purpose and therefore do not have the potential to be broadly adopted and cannot evolve into truly comprehensive services. In some cases (e.g. Ringgold IDs for institutions) commercially maintained IDs are an important part of otherwise open ID systems (e.g ORCID, which uses the Ringgold system), an issue which needs attention to ensure a sustainable and fully open ID ecosystem. For RIDs to be really effective and powerful, they must be broadly and publicly accessible, and be both human and machine-readable.

In the last several years, the ORCID organization has been working on a global solution for RIDs, which is gaining in adoption around the world. ORCID stands for *Open Researcher and Contributor ID*, and is often used interchangeably to describe the ORCID Organization (properly referred to as *ORCID*), the unique ORCID ID string (properly referred to as the *ORCID Identifier*), and the ORCID Registry (properly referred to as the *ORCID Registry*), which stores

and provides access to the full collection of ORCID Identifiers.[4] In the last couple of years ORCID Identifiers have gained ground as a de facto standard for RIDs. ORCID Identifiers are free for individual researchers to register and organisations can integrate ORCID identifiers into research systems and other local systems and workflows.

ORCID's focus is on connecting people and research across disciplines, borders, and time. They are using IDs to more easily connect these elements of research, which have traditionally been a challenge for the community. ORCID connects person identifiers with key workflows, from manuscripts to grant applications, encouraging the research community to interact with features such as tools, use cases, documentation, examples, and open-source code.

ORCID allows researchers to control their own profiles, and include information about themselves, their institutional affiliations, research funding and research outputs, as well as determine what is public/private. This last feature is important, because users can search the publicly available part of the registry for first names, last names, ORCID Identifiers, or keyword. There is no browse capability, but if the user has a clear idea of what or whom they are looking for, the platform is quite robust. The recent addition of institutional authorization using Shibboleth also provides an additional level of "institutional authority" to the maintenance of ORCID Identifiers. Examples of resolvable ORCID Identifiers:

- Preferred citation format - http://orcid.org/0000-0003-1392-7799
- QR Code version -

ORCID Identifiers have been gaining traction internationally and are being integrated into institutions and systems across the world. According to the ORCID landing page, at the time of writing this paper, there were just over 2,250,000 IDs assigned to researchers around the globe. A growing number of these are coming through institutional membership in ORCID, either individually or via ORCID consortia (generally national efforts). Membership in ORCID also provides sustainable funding and support for the ORCID organization, a critical aspect of maintaining an open and non-profit effort. Organisations can support ORCID through joining as a paying member in exchange for access to specific API features and technical support as well as other member benefits. There are currently 382 member organizations in ORCID. In addition, a growing number of research funders require the adoption of ORCID Identifiers by their researchers[5] and, as of May 31, 2016, ORCID has been integrated into about 250 systems, with

[4] "What is ORCID?" *ORCID.org,* last modified August 15, 2013, https://orcid.org/trademark-and-id-display-guidelines.

[5] For example: the Australian Research Council, CONCYTEC (Peru), FWF (Austria), MQ (UK), National Health and Medical Research Council (Australia), National Research Foundation (South Africa), and Research Councils UK.

more in the works[6]. Publishers are also increasingly requiring ORCIDs to represent the unique ID of an author in their systems.[7]

Beyond individual organizations, there have also been national approaches to the adoption of ORCID Identifiers. Australia has taken out a consortial membership which brings together 40 institutions (36 universities, plus the Heart Research Institute, CSIRO, the Australian Research Council and the National Health and Medical Research Council). In some countries, such as Italy, almost every researcher has an ORCID Identifier, facilitated by the efforts of national funders and other research agencies working together with ORCID.[8] The aim of these national initiatives is to achieve economies of scale, while still allowing for flexibility in how ORCID Identifiers are implemented across the organization.

In Germany, the ORCID DE project was launched in February 2016 to support German universities and research institutions that are considering implementing ORCID Identifiers in a coordinated and sustainable approach. The project will address organizational, technical, and legal issues in addition to providing a central contact point for universities and research institutions. ORCID DE will also focus on the cross-linkage and use of ORCID Identifiers in open access repositories and publication services. ORCID DE's project partners are the Helmholtz Open Science Coordination Office at the GFZ German Research Centre for Geosciences, the German National Library (DNB) and the Bielefeld University Library. The project was initiated by the German Initiative for Network Information (DINI).[9]

In Canada, uptake of the ORCID Identifier has been slower. There is currently only one institutional member of ORCID from Canada (Carleton University) as compared to 96 members in the US as of June 6, 2016; however, there are currently over 34,000 Canadian researchers with ORCID Identifiers[10].

In May 2016, two workshops were organized in Ottawa[11] and Toronto[12] to raise awareness of ORCID Identifiers and recruit institutional members. The aim was to discuss the importance of RIDs in enabling a strong research infrastructure and demonstrate value of cross-sector approach and application of ORCID Identifiers in research infrastructure with various

---

[6] Meadows, Alice, "Collect & Connect: Turning ORCID's Vision into Reality" *ORCID,* last modified May 31, 2016, http://orcid.org/blog/2016/05/31/collect-connect-turning-orcid%E2%80%99s-vision-reality.
[7] "Manage My Author Profile," *Elsevier.com,* last modified 2016, https://www.elsevier.com/solutions/scopus/support/authorprofile.
[8] Mennielli, Michele, Andrea Bollini, Josh Brown, and Susanna Mornati, "Identify to simplify: improving interoperability at a national level with the ORCID HUB," *EUNIS,* 2016, http://www.eunis.org/eunis2016/wp-content/uploads/sites/8/2016/02/EUNIS2016_paper_17.pdf.
[9] http://orcid.org/blog/2016/03/14/announcing-orcid-de-project-foster-orcid-adoption-germany.
[10] Personal correspondence with Matthew Buys, ORCID, June 28, 2016.
[11] "2016 ORCID Ottawa Workshop: Why identifiers matter to scholars and researchers," *ORCID.org,* accessed August 24, 2016, https://orcid.org/content/2016-orcid-ottawa-workshop.
[12] "2016 ORCID Toronto Workshop," *ORCID.org,* accessed August 24, 2016, https://orcid.org/content/2016-canada-orcid-workshop.

communities. RDC participated in the workshops, along with other stakeholders including CASRAI, CARL, CRKN, NRC, the Tri-agencies and others.[13]

Based on discussions at the two ORCID workshops in late May, three things were clear:

- there is an international trend towards ORCID Identifiers as default for researcher IDs;
- there is a high level of interest from Canadian research organizations in forming a Canadian-based ORCID Consortium: ORCID-CA;
- the stakeholder community has a vested interest in the outcome.

To that end, a number of participants agreed to form a working group that will:

- create a report detailing the advantages of a Canadian ORCID consortium, and the process and timelines to get there;
- communicate with the broader community of research organizations in Canada to determine the interest in membership on an ORCID-CA;
- propose an ORCID-CA governance framework to move the process forward.

These events clearly indicate significant interest in the use of ORCID Identifiers in Canada and the potential strong support for the formation a national consortium to support its adoption.

2.2 Object Identifiers

Object Identifiers (OIDs) are (typically alphanumeric) strings that establish a unique identity for a specific object. In *Section 2.4, Equipment Identifiers* and *2.5, Other Identifiers*, we highlight a number of ID systems for describing physical objects of interest to research, but in this case we are interested in specific types of *digital objects*, which can be defined as *a machine-independent data structure consisting of one or more elements in digital form that can be parsed by different information systems; the structure helps to enable interoperability among diverse information systems in the Internet.*[14] In the research data context the digital objects of interest can be any number of research outputs created during the data lifecycle, but especially a final output such as a journal article or conference presentation, the dataset associated with such an article or project, or a research data management plan. For our purposes we are interested in an OID system that: 1) uniquely identifies an object; 2) includes a description of that object; and 3) provides a context with which to find that object and its description (as ORCID Identifiers do for RIDs). This means that an OID will provide an actionable, interoperable, persistent link to a digital object. Once assigned to an item, a OID remains a constant locator, not changing even if an object moves location. While information about a digital object may change over time, including where to find it, the OID for that object will stay constant. When properly maintained, OIDs help solve the problem of dead links and link rot.

[13] "ABC ORCID in Canada/Outcomes of May 2016 Meetings," *CASRAI.org*, last modified June 21, 2016, http://docs.casrai.org/ABC_ORCID_in_Canada/Outcomes_of_May_2016_Meetings.
[14] "Digital Object," *CASRAI.org,* last modified August 13, 2015, http://dictionary.casrai.org/Digital_object.

Just as with RIDs, OIDs are critical to uniquely identifying research outputs in a landscape where such outputs are created at an ever-increasing pace. When an OID is associated with one or more RIDs the resulting network of associations can present a very powerful way of describing scholars and their outputs.

Some typical use cases for this type of identifier are:

1. A funder would like to have a list of links to all publications to ensure a funded project is adhering to their open access policy.
2. A researcher would like to cite the data in their paper using a link so that others can access the data and verify their conclusions.
3. A university administrator would like to track all research outputs for their annual reporting.
4. A repository would like to assign a PID to all related objects from a specific research output so that they can be referenced and linked to related material.

There are a number of IDs that are being used to uniquely label individual digital objects: the GUID and ARK systems are two examples. A GUIDs (globally unique identifier) is *a unique reference number used as an identifier in computer software.*[15] The GUID (and associated *UUID*, or universally unique identifier) can be used to provide the unique string or label for an object, but in and of itself does not meet our complete needs for an effective OID system. The ARK (Archival Resource Key) is *a multi-purpose persistent identifier for information objects of any type.*[16] The ARK system was developed by the California Digital Library to provide PIDs for digital objects being created as part of large digitization efforts, but is much more flexible and designed to provide IDs for any type of object, digital, physical, living or intangible.[17]

The ARK scheme does provide a resolvable OID for research outputs, but ARKs have not been widely adopted in the research landscape to describe the types of research outputs highlighted earlier. Having said that, the use of ARK is increasing for other use cases and it does provide a granular ID system that accommodates the full range of objects. In that broader context, ARKs are a system that should be considered as part of the broader landscape of digital objects, which includes research outputs as a subset. Take the example of a journal article; the article itself can be identified using a widely accepted identifier like a DOI (see below), but the individual *digital components* of the article should also be identified with a PID like an ARK. A standard scientific article has a number of components, including images, tables, and graphs, and in the case of online journals may also have multimedia elements such as videos or animations. In an ideal scenario each of these individual digital objects would have a unique PID that provides access that individual component outside the context of the article. In the case of a

[15] "Globally Unique Identifier," *Wikipedia.org,* last modified August 10, 2016, https://en.wikipedia.org/wiki/Globally_unique_identifier.
[16] "Archival Resource Key," *Wikipedia.org*, last modified June 14, 2016, https://en.wikipedia.org/wiki/Archival_Resource_Key.
[17] Perry Willet, "ARK (Archival Resource Key) Identifiers," *Confluence,* Last modified June 28, 2016, https://confluence.ucop.edu/display/Curation/ARK.

typical journal system, this level of granularity is not supported, as articles are often represented as a single digital object, such as a PDF file. As journal systems evolve, the opportunity to represent individual components will become more important and systems like ARK may well provide the best example of a robust PID for those components. An added advantage of ARKs is that the same URL is used for the metadata of the object and the digital object itself. Also, some organizations like ANDS (Australian National Data Service), recommend the use of ARKs with research outputs,[18] and we anticipate that this practice will increase.

The DOI (Digital Object Identifier) is another type of OID and the most commonly used type of PID in the scholarly communication context for the types of research outputs we have previously highlighted.

The DOI system is managed by the International DOI Foundation (IDF)[19], which provides oversight to DOI registration agencies and maintains the DOI resolver. Organizations that meet the contractual obligations of the DOI system and are willing to pay to become a member of the system can assign DOIs. According to the IDF, over 120 million DOI names have been assigned to date, with an annual growth rate 18%.[20] Like ARKs, DOIs can be assigned for other scholarly outputs beyond publications and datasets, such as software, images, samples, audio/video resources, learning outputs, and so on, enabling the citation and re-use of these resources over time. Journal articles represent the majority of DOIs assigned to date, with datasets rapidly becoming the second most common use of DOIs. You will encounter DOI links using both the standard DOI URI (doi:10.1038/nature19057) and the fully resolvable URL (http://www.nature.com/nature/journal/v536/n7616/full/nature19057.html). While either URI should be resolvable to the specific object, in practice not all web browsers detect a valid DOI URI, hence the common practice of representing the complete URL of the item accessed at a specific time. Like ARKs, DOIs can also represent components of a complex digital object like a journal article, although this practice is not widespread.

The DOI system is itself based on a framework called the Handle System, which, like ARK, is used to provide PIDs for digital objects. One of the primary differences between the Handle System and ARKs is that the ARK system relies on domain names in the URL of the PID, while Handles do not. Also, ARKs include a metadata record (defined by the INDECS Content Model)[21] as part of the framework, while the Handle System does not. The DOI system adds additional components to the Handle System to provide a more seamless and functional approach to the provision and resolution of PIDs, hence the close association between the two. As suggested above with ARKs, Handles can be used to provide a PID for any digital object in the research ecosystem, and both ARKs and Handles are used in this way. The decision about which to use is often an institutional or

[18] "Research Data Australia Content Providers Guide: Identifier," *Research Data Australia*, last modified July 22, 2016, http://guides.ands.org.au/rda-cpg/identifiers.
[19] "The DOI System," *DOI.org,* last modified May 3, 2016, http://www.doi.org/index.html.
[20] "Factsheet: Key facts on Digital Object Identifier Systems," *doi.org,* last modified June 6, 2016, http://www.doi.org/factsheets/DOIKeyFacts.html.
[21] "indecs Content Model," *Wikipedia.org,* last modified August 17, 2016, https://en.wikipedia.org/wiki/Indecs_Content_Model.

domain-specific decision, but the critical point is that they both provide OIDs at a very granular level and can be used together in a research data management context without loss of functionality.

DOIs are also increasingly being assigned to datasets via the DataCITE organization[22], an official DOI Registration Agency. DataCITE is a not-for-profit, international organization that works with data centres and researchers to assign persistent identifiers to datasets and other research objects. The aim is to develop an infrastructure that supports data citation, discovery, and access. Once data are citable, they can become legitimate contributions to scholarship. DataCite does not allocate persistent identifiers itself; this is done by its members, who act as allocating agents. In Canada, the National Research Council is a full member of DataCITE and allocates DOIs to participating organizations in Canada.

DataCite Canada's registration service enables Canadian research organizations to obtain and manage DOIs for their research outputs. As part of this role, “DataCite Canada provides local support for Canadian data centres and libraries; contributes to ongoing development of DataCite's services and resources on behalf of Canadian clients; and promotes the value of data archiving, citation and discoverability within Canada”[23]. One of the impediments with the Canadian DataCITE service is that it positions an organization to create its own DOIs, and doesn’t yet provide software interfaces that allows efficient local minting of DOIs. While some organizations have registered with the National Research Council (NRC) and built their own software services, such as University of PEI[24], others are using DataCITE services elsewhere that offer more support. The University of British Columbia is addressing the challenge of providing a generic Canadian DOI minting service and has a web-based interface for creating DOIs one at a time, or en mass. This is a good example of how current efforts at Canadian institutions can work in collaboration with an organization like NRC to build a more robust RDM ecosystem.

Software ID’s and citations are also becoming a focus of discussion, due to the need for reproducibility in research, and the reliance on software at several levels throughout the research lifecycle[25]. DOIs have been recognized as a best practice within the software community and Github has been working in conjunction with other stakeholders to provide a DOI minting service[26], however, the practice is still in its infancy. The conversation is ongoing and should be considered an area of development; in the interim, users should consider creating a DOI for specific software applications, linking to a landing page such as a Github site where they fully describe their methods sufficient for replication.

---

[22] “Welcome to DataCite,” *DataCite.org,* access August 24, 2016, https://www.datacite.org/.
[23] DataCITE Canada, “DOI registration service,” *nrc.gov.ca,* last modified January 26, 2016, http://www.nrc-cnrc.gc.ca/eng/publications/library_services/datacite/.
[24] “Manage Research Data at UPEI,” *UPEI.ca*, accessed August 24, 2016, https://data.upei.ca/.
[25] Victoria Stodden, and Sheila Miguez, “Best Practices for Computational Science: Software Infrastructure and Environments for Reproducibility and Extensible Research,” *Journal of Open Research Software* 2 no. 1 (2014): 1-6, DOI: http://dx.doi.org/10.5334/jors.ay.
[26] “Making Your Code Citable,” *GitHub Guides,* Last modified 2014, https://guides.github.com/activities/citable-code/.

2.3 Organizational Identifiers

An Organizational Identifier (OrgID) is a (typically alphanumeric) string that establishes a unique identity for a specific organizational entity. Organizational affiliations are another important identifier for a variety of stakeholders, including academic administrators, funders, publishers, repository managers, software developers, rights agencies and individual researchers. Currently, identifying and tracking organizational affiliations is very challenging, because an organization may be known by a variety of names and may have several affiliated departments or research centers that are also recognized entities in themselves. It provides the means to both find and identify an organization accurately and to define the relationships with its sub-units.

> *"The use of persistent identifiers for organisations lags behind the use of persistent identifiers for research outputs and people."*
>
> - Project THOR, Artefact, Contributor, and Organisation Relationship Data Schema

Some typical use cases for organizational identifiers are:

1. A funder would like to collate all publications related to their funding for a given year.
2. A researcher would like to compile a list of institutions participating in a proposal with multiple partners.
3. A university administrator would like to identify all publications by their affiliated researchers to assess impact.
4. A repository would like to download all articles published by institutional researchers to ensure that they are complying with an open access policy.

The scholarly community has not yet coalesced around a single solution for organizational identifiers, leading to large gaps in the adoption of these types of IDs. A recent study, undertaken by CASRAI (Consortia Advancing Standards in Research Administration Information) and Jisc in the UK in 2014, found a wide variety of organizational IDs, none of which fulfils the role of being an authoritative list of organisations involved in research.[27]

One of the most commonly used organizational IDs with international scope the ISNI (International Standard Naming Identifier), which is an ISO (International Standards Organization) standard that is used widely and can be assigned to individuals and organizations. The ISNI database currently holds public records of 565,000 organisations and over 8 million individuals. Ringgold, a for-profit company, also maintains a list of organizational identifiers, which is used by ORCID to authoritatively identify organizations. CrossRef's 'Open Funder Registry' provides a common taxonomy of over 11,000 funding body names and is licensed with a CC-0 Creative Commons license, removing any restrictions on use.

[27] Review of selected organisational IDs and development of use cases for the Jisc CASRAI-UK. Organisational Identifiers Working Group. http://repository.jisc.ac.uk/5853/.

A more recent development is the creation of the GRID database (created by Digital Science), which is a free and open database of over 63,000 research institution and their identifiers. GRID provides an important service in ensuring the interoperability of multiple OrgIDs for the same organization. The fact that it is open (with a CC BY Creative Commons license) is also important to ensuring a sustainable ID ecosystem. You can see an example of GRID in use via the OpenVIVO database.[28] A sample record from the GRID database illustrates the value provided.

- https://grid.ac/institutes/grid.139596.1
- https://grid.ac/institutes/grid.4991.5

In the scholarly communication context ORCID Identifiers are increasingly used over ISNIs as the de facto standard for individuals. In order to ensure interoperability in this ecosystem, ORCID and ISNI have signed an MOU to facilitate this effort.[29] For example, ORCID Identifiers are format compatible with the ISNI standard and ISNI has assigned a block of IDs for ORCID, preventing overlapping ID assignments.[30] The picture for OrgIDs is less clear and continues to evolve.

The CASRAI-Jisc study concluded that "[t]he most desirable vision for the future for UK research would be for ISNI to emerge as a strong, sustainable and internationally well supported baseline"[31]. The study also suggested the importance of an "ISNI+" approach, which would integrate the ISNI standard with a service (such as that represented by Ringgold or Digital Science) which would provide all the features necessary to create an OrdID framework similar to what we see for RIDs and OIDs. The Digital Science GRID approach is the most open and is growing at a rapid rate, suggesting that it may well be an important example of the ISNI+ approach for the global research community.

While not as far along as other ID frameworks, clearly the OrgID context is starting to receive more focused attention. For example ORCID is leading the conversation to provide a more complete integration or interoperability between the ISNI standard and associated services. This also serves to highlight the role a non-profit organization like ORCID can play in this landscape: the interests of the research community must take precedence over individual organizational interests.

2.4 Equipment Identifiers

[28] "Open Vivo Homepage," *OpenVivo.org,* last modified 2016, http://openvivo.org/.
[29] Andrew MacEwan and Laure Haak, "ISNI and ORCID sign Memo of Understanding - 2014-01," *ISNI.org*, last modified January 2014, http://isni.org/content/isni-and-orcid-sign-memo-understanding-2014-01.
[30] "What relationship between ISNI and ORCID," *ORCID.org,* accessed August 24, 2016, http://orcid.org/content/what-relationship-between-isni-and-orcid.
[31] Review of selected organisational IDs and development of use cases for the Jisc CASRAI-UK Organisational Identifiers Working Group. http://repository.jisc.ac.uk/5853/.

Like the previous examples of IDs, an equipment ID (EID) is a (typically alphanumeric) strings that establishes a unique identity for a specific piece of research equipment, or subcomponents of such equipment. Equipment and scientific instruments are increasingly a critical part of many research projects, often generating high volumes of data in a short time.

Some typical use cases for equipment identifiers are:

1. A funder would like to collate all publications related to the use of a piece of equipment that they helped fund.
2. A researcher would like to compile a list of all collaborators who have used data from a specific piece of equipment that they acquired and manage.
3. A collaborator would like to see the history of actions (e.g. installations, configuration changes, software upgrades) taken with a piece of equipment that generated data used in their study.
4. An equipment manufacturer would like to identify uses of their equipment, including specific examples of research outputs.

The use of EIDs has only recently started to receive significant attention, so recommendations regarding best practices are only now being developed. However, given the increasing importance of this component of the research ecosystem we felt it important to highlight the latest developments. One of the most useful efforts with EIDs is 'equipment.data', a project co-funded by Jisc and the Engineering and Physical Sciences Research Council. The goal of the project is to *to improve visibility and utilisation of UK research equipment*.[32] In addition to providing a searchable database of research equipment in the UK, the system has defined a metadata record (UNIQUIP[33]) that is used to ensure a common record format for the equipment. When institutions submit a record it includes a local ID (ie. an institutionally derived ID for a piece of equipment) and the equipment.data project then mints a unique ID based on an PD5 hash value derived from a specific algorithm.[34] A sample ID is reflected below.

- http://equipment.data.ac.uk/item/1719606459dda516db9917ff23b55620.html

While the equipment.data project is a good example of the development of an approach to EIDs, it is used in a specific context and does not offer a solution for an international standard to the generation of PIDs and associated URIs for equipment.

Another important initiative in this space is the MERIL project in the EU, which is *the only systematic inventory of significant research infrastructures that is comprehensive over all European countries and scientific domains. MERIL provides high-quality information about*

[32] "Equipment Data Homepage," *Equipment.data.ac.uk,* accessed August 24, 2016, http://equipment.data.ac.uk.
[33] "UNIQUIP Data Publishing Specification," *Euipment.data.ac.uk,* accessed august 24, 2016, http://equipment.data.ac.uk/uniquip.
[34] Personal correspondence, Adrian Cox, University of Southampton. Jul 29, 2016.

research infrastructures and data extraction capabilities on an open-access web portal.[35] The MERIL database provides access to information at a higher level than individual pieces of equipment (e.g. facilities or Centres), but the project is starting to consider the issue of individual EIDs.

2.5 Other Identifiers

In the previous sections, we have described IDs that together contribute to a more cohesive and linked scholarly landscape. However, identifiers can also be used in a wide variety of other contexts and practices, providing additional citable components in dissemination of research and further facilitating discoverability. Although not the focus of this report, it is important to note that the generalized use of other IDs in other contexts will create an environment that improves access, discovery, impact and recognition in the research domain. As with researchers, digital objects, institutions and equipment, using unique IDs for other entities will facilitate name-disambiguation, discovery of knowledge, and ultimately the reuse of valuable scholarly materials.

The National Information Standards Organization (NISO) recently published a draft document[36] listing persistent identifiers, including a dynamic table[37] with information on PIDs used in scholarly communication. The list below provides some examples of best practice IDs for specific areas of research, with a focus on IDs that are supported by non-profit organizations and in active use.

Research Domain	PID Name	More Information
GeoScience/General	IGSN	WikiPedia, SESAR
Life Sciences/Biodiversity	LSID	Wikipedia, ZooBank
Science	MIRIAM	Wikipedia, Registry
Mycology	MycoBank ID	Wikipedia, MycoBank

[35] "MERIL- Mapping of the European Infrastructure Landscape," *European Science Foundation,* last modified 2016, http://www.esf.org/serving-science/ec-contracts-coordination/meril-mapping-of-the-european-research-infrastructure-landscape.html.

[36]NISO Persistent Identifiers and Alternative Outputs Working Group, "Persistent Identifiers in Scholarly Communications," *NISO,* RP-25-201x-2B, May 2016. http://www.niso.org/apps/group_public/download.php/16554/NISO%20RP-25-201x-2B%2C%20Persistent%20Identifiers%20in%20Scholarly%20Communications.pdf.

[37] Google Document, Persistent Identifier Lansdcape Analysis, May 2016, https://docs.google.com/spreadsheets/d/1TnTF2r0fngfeSReoPX44kNIjH6zM3X2VKgFvGuGTPTc/edit#gid=650790253.

Science	RRID	WIkipedia, SciCrunch
Chemistry	InChI	Wikipedia, InChI Trust
Pharmaceuticals	DIN	Wikipedia, DPD
Enzymes	EC Number	Wikipedia
Proteins	PDB ID	Protein DataBank
General	UPC	Wikipedia, UPC Search

3. Discussion

Research is increasingly international and multidisciplinary, resulting in a very complex research environment. Unique identifiers facilitate discovery and reuse of content, global interoperability, and a better understanding of the impact and value of scholarship. They also reduce time and administrative burden by enabling the entry of data into one system that can be automatically reused in the context of other systems, supporting more streamlined and automated processes for researchers, funders, vendors and institutions. They also facilitate the transfer of information across organizations and establish links across systems institutionally, nationally and internationally. Used together RIDs, OIDs, OrgIDs, EIDs and other IDs represent a sophisticated infrastructure of interconnected people, organizations and resources that enable the community to innovate in new ways. Clearly there are many advantages to adopting unique identifiers for research entities and outputs.

However, to be truly useful, identifiers must be more than just distinctive strings of alphanumeric characters; they should be integrated and compliant with the various infrastructures and systems that use them. They should also be both human and machine readable and actionable, ideally with API access, resolvability and a recognisable, open licences such as CC-0 (Creative Commons).[38] When identifiers are standard and machine readable it facilitates the synchronization of disparate data events via “triggers”. Two systems could automatically update each other every time a new record is added, or an existing record is changed. For example, The US Funding application system, SciENcv, is able to automatically push an update from an application’s metadata to the author’s ORCID profile using this kind of actionable metadata infrastructure. The opportunity this presents to ease both researcher's and organization’s administrative burden is substantial, and critical to ensuring that research outputs are accessible.

Despite their obvious advantages, there have been some concerns expressed about the widespread use of PIDs. For example, there is a lack of privacy in a scholarly environment

[38] Chris Phillips, CANARIE Inc., Personal communication, July 14, 2016.

which relies heavily on PIDs. Once a PID is associated with a person or object related to a person, that information is in the public sphere forever. While the scholarly community has always relied on personal identifiers, such as student IDs or email addresses, in the past this information has not been made available to the public and organizations make significant efforts to ensure that this information remains confidential. It is worth recognizing that once PID services are widely adopted, there will be no way to remove this information from the internet.

Another issue is that PIDs may be vulnerable to failure, commercialization, or a change in mission. It is important, therefore, that the solutions we adopt adhere to principles and governance that ensure services are developed based on the needs of the community and universally accepted best practices for the governance of open standards. As such, PID services should be open, freely available, easily accessible, and maintained in a well-governed, independent, trusted and sustainable manner.

The ODIN project (ORCID and DataCite Interoperability Network) has defined the term "trusted identifier" as part of a conceptual model of PID interoperability. In this model, a trusted identifier is one that is *unique, persistent/resolvable, descriptive/discoverable, interoperable* and *governed*. These characteristics are further defined:

- ***Unique*** identifiers are unique on a global scale, allowing large numbers of unique identifiers
- ***Persistent*** identifiers resolve as HTTP URIs with support for content negotiation, and these HTTP URIs should be persistent.
- ***Descriptive*** identifiers come with metadata that describe their most relevant properties, including a minimum set of common metadata elements. A search of metadata elements across all trusted identifiers of that service should be possible [to aid discovery].
- ***Interoperable*** identifiers are interoperable with other identifiers through metadata elements that describe their relationship.
- ***Governed*** identifiers are issued and managed by an organization that focuses on that goal as its primary mission, has a sustainable business model and a critical mass of member organizations that have agreed to common procedures and policies, has a governing body, and is committed to using open technologies.

Additionally, not all persistent identifiers are made equal. A robust PID system should adhere to best practices including[39]:

Actionable: They should be resolvable on the web.

[39]"Persistent identification and PIDs," *The Hope Wiki,* last modified April 23, 2013, http://hopewiki.socialhistoryportal.org/index.php/Persistent_Identification_and_PIDs.

Syntax: The syntactic qualities of the PID should be flexible and scaleable.

Supporting Services, Interoperability, Community:

- The PID service should not come bundled together with other services nor create technical or administrative dependencies
- The PID service should be reliable, sustainable, secure, and cost effective and there should be no administrative or technical obstacles to using the services
- The PID services should use a formal and well-documented standard, flexible enough to interoperate with other schemes, and not dependent on protocols which may change over time or become obsolete
- The PID should be mature, well-supported, and widely-adopted system with a committed community of users
- The PID service should enable privacy controls so that individuals can set their own levels of access

4. Recommendations

Given the many potential benefits, but also some of the implications of using PIDs, RDC recommends the adoption of PID services that adhere to best practices and the principles of trust and good governance. The recommendations are designed to support the development of a rich national framework for research information in Canada including improved management, discovery, and the interoperability of research data.

Recommendations are organized according to the major stakeholder community. However, there are also some important recommendations that apply to all of these groups. RDC would like to work with all these communities to develop a strategy for awareness of the benefits of PIDs and their adoption in the Canadian context.

All stakeholders

General	1. Commit to the adoption of best practice PIDs and ensure that the solutions adopted are not-for-profit, with appropriate community governance.
Researcher IDs	2. Join the ORCID-CA consortium to develop a comprehensive national solution for the adoption of PIDs in the research ecosystem.
Terminology	3. Agree on a common glossary of metadata terms that facilitate adoption of the PID standards.

Research Funders

Researcher IDs	4. Require researchers to have an ORCID when applying for grants. 5. Work with Common CV Office to include the integration of ORCID Identifiers and the ORCID Repository into the CCV system.
DOIs	6. Require the adoption of DOIs for data sharing and as an outcome of approved Funder data management policies and principles.
Institutional IDs	7. Adopt ISNI+ as the standard for institutional identifiers. 7.1 Ensure that the information about all relevant research centres and affiliates are accurately documented in the ISNI database and the open GRID (Global Research Identifier Database) system to facilitate the integration with local, regional and national systems.

Universities and Research Centres

Researcher IDs	8. Develop institutional policies and strategies that encourage the adoption of ORCIDs, and best practices in describing and citing research outputs. 9. Work with CARL, regional library consortia, and the CARL Portage Network on the development and deployment of educational and awareness resources that highlight the value of ORCIDs.
DOIs	10. Require the adoption of DOIs for data sharing in the context of data management policies and principles.
Institutional IDs	11. Adopt ISNI+ as the standard for institutional identifiers 10.1 Ensure that the information about all relevant research centres and affiliates are accurately documented in the ISNI database and the open GRID (Global Research Identifier Database) system to facilitate the integration with local, regional and national systems.

Science-based Government Departments

Researcher IDs	12. Require researchers to have an ORCID when applying for grants.

DOIs	13. The National Research Council enhance the DataCITE services available for institutions to create DOIs. 11.1 Work with the University of British Columbia to integrate their DOI minting services into the DataCite Canada operations.
Institutional IDs	14. Adopt ISNI+ as the standard for institutional identifiers 14.1 Ensure that the information about all relevant research centres and affiliates are accurately documented in the ISNI database and the open GRID (Global Research Identifier Database) system to facilitate the integration with local, regional and national systems.

Repositories and Publishers

General	15. Support the adoption of Researcher IDs, DOIs, ISNIs and other IDs into metadata records in order to facilitate the development of a linked metadata framework.

Researchers

Researcher IDs	16. Register for an ORCID Identifier and complete the ORCID Profile by adding research outputs.
DOIs	17. Register datasets and other research outputs with DataCITE Canada and add the DOI to appropriate metadata records.

CASRAI

General	18. Promote the use of PIDs in the Canadian context and internationally.
Researcher IDs	19. Work with Common CV Office to include the integration of ORCID into the CCV system 20. Work with the ORCID and the ORCID-CA organizations to facilitate the adoption of appropriate standards, including Lists/Controlled Vocabularies.

Research Data Canada

General	21. Develop a strategy for the widespread adoption of PIDs, working with other stakeholders to facilitate those developments. 21.1 Create an RDC “PID Working Group” to develop and deploy a strategy for national adoption of PIDs. This Working Group should include: RDC, CASRAI, Scholar’s Portal, Portage, GOC Open Data Committee, Tri-Agencies, National Research Council, Open Science representatives. 21.2 Task the RDC PID WG with updating this document, and creating a WG Action Document, as advancement in PID systems warrant. 21.3 Develop a communications strategy to encourage the adoption of PIDs in Canada 22. Require the use of PIDs in all outputs from RDC Committees and Working Groups. 23. Engage with the international research data management community on the development and adoption of PIDs.
Researcher IDs	24. Facilitate the development of an ORCID-CA consortium. 25. Contact Canadian journal publishers with the goal of encouraging the adoption of a “PID Mandate” policy, requiring the use of ORCIDs, DOIs, ISNI+, EIDs, and other best practice PIDs in article/dataset submission.
DOIs	26. Promote the use of DataCITE DOIs for research data sets in Canada and engage with developments internationally to ensure best practice in Canada.

5. Acknowledgements

We are very grateful to the RDC community for their contribution to this document, but especially David Baker (CASRAI-CA) and Eugene Barsky (University of British Columbia) who have contributed to the quality of this white paper through peer review and providing valuable feedback.

6. Bibliography

--. “ABC ORCID in Canada/Outcomes of May 2016 Meetings.” *CASRAI.org*. Last modified June 21, 2016. http://docs.casrai.org/ABC_ORCID_in_Canada/Outcomes_of_May_2016_Meetings.

--. “Digital Object.” CASRAI.org. Last modified August 13, 2015. http://dictionary.casrai.org/Digital_object.

--. “Persistent Identifier.” *CASRAI.org.* Last modified August 13, 2015. http://dictionary.casrai.org/Persistent_identifier.

--. “Uniform Resource Identifier.” *CASRAI.org.* Last modified August 13, 2015. http://dictionary.casrai.org/Uniform_resource_identifier.

DataCITE Canada. “DOI registration service.” *National Research Council Canada*. Last modified January 26, 2016. http://www.nrc-cnrc.gc.ca/eng/publications/library_services/datacite/.

--. “Welcome to DataCite.” *DataCite.org.* Access August 24, 2016. https://www.datacite.org/.

--. “Factsheet: Key facts on Digital Object Identifier Systems.” *doi.org.* Last modified June 6, 2016, http://www.doi.org/factsheets/DOIKeyFacts.html.

--. “The DOI System.” *DOI.org*. Last modified May 3, 2016. http://www.doi.org/index.html.

--. “Manage My Author Profile.” *Elsevier.com.* Last modified 2016. https://www.elsevier.com/solutions/scopus/support/authorprofile.

--. “Equipment Data Homepage.” *Equipment.data.ac.uk.* Accessed August 24, 2016. http://equipment.data.ac.uk.

--. “UNIQUIP Data Publishing Specification.” *Equipment.data.ac.uk.* Accessed august 24, 2016.

http://equipment.data.ac.uk/uniquip.

--. “MERIL- Mapping of the European Infrastructure Landscape.” *European Science Foundation.* Last modified 2016. http://www.esf.org/serving-science/ec-contracts-coordination/meril-mapping-of-the-european-research-infrastructure-landscape.html.

Feeney, Patricia. “DOIs for Journals: Linking Data and Beyond.” *Information Standards Quarterly,* 22, no. 3 (2010): 27-31. doi: 10.3789/isqv22n3.2010.06.

--. “Making Your Code Citable.” *GitHub Guides.* Last modified 2014. https://guides.github.com/activities/citable-code/

Herbert Van de Sompel, Carl Lagoze, Jeroen Bekaert, Xiaoming Liu, Sandy Payette, and Simeon Warner. “An interoperable fabric for scholarly value chains.” *D-Lib Magazine, 12(10).* October 2006. http://www.dlib.org/dlib/october06/vandesompel/10vandesompel.html.

--. “Persistent identification and PIDs.” *The Hope Wiki*. Last modified April 23, 2013. http://hopewiki.socialhistoryportal.org/index.php/Persistent_Identification_and_PIDs.

Ian Lyne, Catherine Grout, and Balviar Notay. “Overview of Systems Interoperability Project: Report to the RCUK Research Funding Programme Board.” *Jisc Research Council UK.* June 2015. http://www.rcuk.ac.uk/RCUK-prod/assets/documents/publications/osipreport.pdf.

MacEwan, Andrew and Laure Haak. “ISNI and ORCID sign Memo of Understanding - 2014-01.” *ISNI.org*. Last modified January 2014. http://isni.org/content/isni-and-orcid-sign-memo-understanding-2014-01.

Meadows, Alice. “Collect & Connect: Turning ORCID’s Vision into Reality.” *ORCID.* Last modified May 31, 2016. http://orcid.org/blog/2016/05/31/collect-connect-turning-orcid%E2%80%99s-vision-reality.

Mennielli, Michele, Andrea Bollini, Josh Brown, and Susanna Mornati. “Identify to simplify: improving interoperability at a national level with the ORCID HUB.” *EUNIS.* Last modified 2016. http://www.eunis.org/eunis2016/wp-content/uploads/sites/8/2016/02/EUNIS2016_paper_17.pdf.

--. “Editorial: Credit where credit is due,” *Nature 462, no. 825* (2009), doi:10.1038/462825a.

NISO Persistent Identifiers and Alternative Outputs Working Group. “Persistent Identifiers in Scholarly Communications.” *NISO,* RP-25-201x-2B. May 2016. http://www.niso.org/apps/group_public/download.php/16554/NISO%20RP-25-201x-2B%2C%20Persistent%20Identifiers%20in%20Scholarly%20Communications.pdf.

--. “Open Vivo Homepage.” *OpenVivo.org.* Last modified 2016. http://openvivo.org/.

--. “2016 ORCID Toronto Workshop.” *ORCID.org.* Accessed August 24, 2016. https://orcid.org/content/2016-canada-orcid-workshop.

--. “2016 ORCID Ottawa Workshop: Why identifiers matter to scholars and researchers.” *ORCID.org.* Accessed August 24, 2016. https://orcid.org/content/2016-orcid-ottawa-workshop.

--. “What is ORCID?” *ORCID.org.* Last modified August 15, 2013. https://orcid.org/trademark-and-id-display-guidelines.

--. “What relationship between ISNI and ORCID.” *ORCID.org.* Accessed August 24, 2016. http://orcid.org/content/what-relationship-between-isni-and-orcid.

--. “Persistent Identifier Lansdcape Analysis”. Last modified May 2016, https://docs.google.com/spreadsheets/d/1TnTF2r0fngfeSReoPX44kNIjH6zM3X2VKgFvGuGTPTc/edit#gid=650790253.

Peter Kraker, Asura Enkhbayar, and Elisabeth Lex. “Exploring Coverage and Distribution of Identifiers on the Scholarly Web.” *International Symposium of Information Science*. March 2015. arXiv:1503.05096.

Powell, James, and Nigel Shadbolt. “Creating value with identifiers in an open data world.” *Open Data Institute.”* October 2014. http://innovation.thomsonreuters.com/content/dam/openweb/documents/pdf/corporate/Reports/creating-value-with-identifiers-in-an-open-data-world.pdf.

--. “Research Data Australia Content Providers Guide: Identifier.” *Research Data Australia.* Last modified July 22, 2016. http://guides.ands.org.au/rda-cpg/identifiers.

Smith-Yoshimura, Karen, Janifer Gatenby, Grace Agnew, Christopher Brown, Kate Byrne, Matt Carruthers, Peter Fletcher, Stephen Hearn, Xiaoli Li, Marina Muilwijk, Chew Chiat Naun, John Riemer, Roderick Sadler, Jing Wang, Glen Wiley, and Kayla Willey. “Addressing the Challenges with Organizational Identifiers and ISNI”. *OCLC Research.* 2016. Dublin, Ohio. http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-Organizational-identifiers-and-isni-2016.pdf.

Stodden, Victoria and Sheila Miguez. “Best Practices for Computational Science: Software Infrastructure and Environments for Reproducibility and Extensible Research.” *Journal of Open Research Software* 2, no. 1 (2014): 1-6, DOI: http://dx.doi.org/10.5334/jors.ay.

--. “Overview.” *THOR.* Accessed July 15, 2016. http://dashboard.project-thor.eu/dashboard/.

--. “Welcome to THOR.” *THOR.* Access July 15, 2016. https://project-thor.eu.

--. “Manage Research Data at UPEI.” *UPEI.ca*. Accessed August 24, 2016. https://data.upei.ca/.

--. “Archival Resource Key.” *Wikipedia.org*. Last modified June 14, 2016. https://en.wikipedia.org/wiki/Archival_Resource_Key.

--. “Globally Unique Identifier.” *Wikipedia.org*. Last modified August 10, 2016. https://en.wikipedia.org/wiki/Globally_unique_identifier.

--. “Identifier.” *Wikipedia.org*. Last modified July 7, 2016. https://en.wikipedia.org/wiki/Identifier.

--. “indecs Content Model.” *Wikipedia.org.* Last modified August 17, 2016. https://en.wikipedia.org/wiki/Indecs_Content_Model.

Willet, Perry. “ARK (Archival Resource Key) Identifiers.” *Confluence.* Last modified June 28, 2016. https://confluence.ucop.edu/display/Curation/ARK.