# ResBaz workshop - The 21st Century Researcher

# Session: Persistent Identifiers for Research

This 40 minute session will cover:

* What are persistent identifiers?
* How do they apply to research?
* Which PIDs apply to research?
* Which PIDs do researchers need to know more about?
* Intro to DOIs for publications and data & how to get them
* Intro to ORCIDs & how to set one up
* Short intro to IGSNs

## What are persistent identifiers?

[ask audience] What is an identifier? What are examples of identifiers?

Answer: a label that is applied to a person, an object or a class of objects. Examples of identifiers are personal names, tax file numbers, credit card numbers etc.

An identifier can be made up of numbers, letters, symbols or a combination of these. They provide a means of naming a person, object or a class of objects so that someone who reads or sees the identifier can work out what it refers to.

[ask audience] Is it possible to have two identifiers that are exactly the same?

Answer: yes, for example personal names - refer to the Dr Seuss poem “too many Dave’s” (with apologies to any Dave in the room!)

Because of this problem, identifiers should be unique. For example, the International Standard Book Number or ISBN is assigned to identify a class of objects called books. ISBNs are a unique numeric identifier assigned to books and book-like products and are used primarily to facilitate the international book trade and the library market.

[show example] No other book has the same ISBN. This uniqueness is assured by the identity management structure whereby an ISBN Registration Agency controls the issuing of identifiers to these books.

[ask audience] What other identifiers can you think of in the scholarly world?

Answer: Some examples include: ISSN for periodicals; ISMN for printed music; LCCN for Library of Congress cataloguing records; DOIs for publications and data; ISNI, ResearcherID, ScopusID and ORCID for researchers.

Identifiers can be applied to both offline and online objects. The problem with identifiers such as URLs in the online environment is this [show page 404 page not found error].

[ask audience] What has happened to cause this error message?

Answer: generally caused when the resources are not maintained at the referenced URLs. For example, the URL has changed but the old version of the URL is the one you clicked and there is no automatic forward in place to take you to the new URL.

In research, the 404 type errors are a real problem. The error means that a resource - like a journal article - that may have been cited in a scholarly publication using its URL may no longer be accessed. This is frustrating for those who want to read the citation and also for those who want to confirm the claims made in the paper based on the citation.

A persistent identifier addresses this problem because it is an identifier that is guaranteed to be maintained over time in an online environment. A persistent identifier - or PID - should always resolve to an online page that contains information about the object it represents. This is the case whether the object has been moved to another location or even removed completely. The result is that resources with persistent identifiers can be cited with far more confidence than those which do not.

[ask audience] Whose responsibility do you think it is to maintain a PID?

Answer: usually PIDs are requested or minted by an organisation and it is up to the organisation to maintain the identifier. Having said that, it is also up to the researcher whose resource has been issued the PID to let the organisation know if things have changed.

## What are Digital Object Identifiers (DOIs)?

Let’s now look closer at a specific identifier: Digital Object Identifiers or DOIs.

[ask audience] Who can tell me what a DOI is?

Answer: A DOI is a globally unique, alphanumeric string or name that identifies digital content on the web [show example].

It consists of a prefix and a suffix separated with a forward slash [show]. A DOI prefix always begins with the number 10 and this distinguishes it from other implementations of the Handle System, of which DOIs are one implementation. The rest of the numbers in the DOI string refer to the registrant code and the item identifier.

DOIs can be expressed as a number [show example] or in a web resolvable form [show example]. They are quoted in citations [show APA style example]

DOIs evolved from the publishing industry to persistently identify online content, specifically scholarly journal articles.

DOIs are an international standard (ISO 26324) and they are distinguished from other PIDs in two aspects:

* The system is governed by the International DOI Foundation and the minting ( which is another word for issuing) of DOIs can only be done by a DOI Registration Agency;
* A minimal amount of metadata is required to mint a DOI like author, title, publisher, date and URL

[ask audience] I mentioned that DOIs are routinely assigned to scholarly publications, what else can they be the assigned to?

Answer: formal, scholarly publications including journal articles and book chapters; research data; and grey literature like conference papers, theses and reports.

[ask audience] How do you think DOIs could help you as a researcher?

[show slide]

Answer:

* Improves visibility of your research: easier to link to; can confidently be cited
* Facilitates access to your research: people can click on the link to open your publication, report or dataset
* Enables better metrics for your research: DOIs are much easier to track by citation algorithms then a non-DOI citation.
* Enables nice linking between research objects especially publications and data (an increasing requirement of journal publishers [show example from Nature and PLOS]) and also research (DOIs) and the researchers who publish the research (ORCIDs).

[show example from Dryad] How a DOI redirects to a landing page + scholarly publication with DOI that links to data via data DOI

I mentioned that DOIs are only issued by DOI Registration Agencies. For publications, the DOI RA is CrossRef [show] and for datasets and grey literature the DOI RA is DataCite [show]. Universities and other research institutions become members of the RA or it’s agency and begin minting DOIs for research objects that are held in their institutions. Publishers also routinely issue DOIs for journal articles as a standard part of the manuscript publication process. So to get a DOI you need to either get one from your publisher or your institution (library or IT usually).

[exercise] query CrossRef: <http://search.crossref.org/> and DataCite: <https://search.datacite.org/> for…[examples to be added]

Let’s finish with metrics. I mentioned that DOIs are great for getting metrics on your publications and datasets, much better than other types of online identifiers. Looking at an example from the data world: the Make Data Count Project in Europe looked at how to count citations of data. They found a major problem: data is not cited in consistent ways because sharing of datasets and citing data is a relatively new concept for researchers. So some cite data in the main body of an article, some in the reference section, some in a special place in the publication like supplemental data. This makes it very hard to pull citation statistics. Using DOIs makes this easier because the algorithm can search for the DOI rather than a URL or worse, a reference without an identifier altogether. There are other problems in counting data of course, for example, researchers have to cite their own data in a publication and this needs to be excluded from the count. But you get the idea that DOIs are useful for metrics and for maintaining accessibility, visibility and citability of your publications.

## What are ORCIDs?

We’ve looked at persistent identifiers - what they are, how they are used in research, plus we’ve examined DOIs and why you might want DOIs to be assigned to your research. In this next section, we’re going to look at PIDs for people, specifically, for researchers.

[ask audience] What do you think is the most common surname in Sydney?

Answer: Smith, but strong trend toward Nguyen [show].

Common names are a real problem, especially in the scholarly world and particularly in China [show]

Other name problems include: names changes (divorce, marriage, depoll), name variations in scholarly literature (David Wright, D Wight, DA Wright) and misspellings (D Right). These problems can have a big impact on research: mainly, that the wrong person is attributed to the research.

How do we fix this problem? The answer is in assigning persistent identifier for people, specifically ORCID [show].

ORCID stands for Open Researcher and Contributor Identifier. It is a unique, persistent identifier assigned to researchers to disambiguate one researcher from another. It groups all of your works under one online, resolvable web identifier ensuring you get correct attribution and credit for research [show Peter Doherty example].

[ask audience] Who has heard of ORCID? Who has an ORCID?

Shortly we will be creating or updating ORCID records. But first, a little more background.

ORCID is run by an international not-for-profit organisation which is supported by a large number of organisations representing every component of the research sector - publishers, universities and research organisations, research funders etc.

ORCID is free to join for researchers who simply sign up online. It only takes a few moments. Institutions can join ORCID for a small fee which enables them to interact with ORCID at a technical level (using the API).

In Australia, we have an ORCID consortium of 40 institutional members including 36 universities, CSIRO, the Heart Foundation and our two major funders, ARC and NHMRC. The Consortium follows the release of a Joint Statement of Principle in support of ORCID that has been signed by various key players in Australian research. Our funders are looking to integrate ORCID into their systems [show] - in fact the ARC will ask you for one now - and in the future, this will save you time on funding applications. For example, NIH uses your ORCID iD to pull your works into your funding application and select the 5 most relevant ones for that application.

Funders will also - with your permission - be able to write to your ORCID records and asset that you have been awarded that grant.

Similarly, for publications, you’ll be asked for an ORCID at manuscript submission and the publisher - with your permission - will write to your ORCID record to show you have published that material. Same with datasets and other research outputs. This will save you time as your institution will be able to get this information by reading your ORCID record rather than asking you to manually input it into their systems. And if you change institutions, they can also simply read your ORCID record to get your publications history.

ORCID is about researcher control. With ORCID, you control what you put in your ORCID, who sees what information and who can update it.

[demo] Looking at an ORCID record and updating it including privacy settings

[exercise] create or update your ORCID record. QUT people go to… UQ people go to….

Now that we have created or enhanced our ORCID records, let’s look at a few cool things we can do with it.

[demo] using ORCID to sign up for Impact Story

## What are IGSNs?

Now that we’ve looked at DOIs and ORCIDs, we will briefly look at IGSNs.

IGSN stands for International Geo Sample Number [show example]. The IGSN is a persistent unique identifier for physical samples and specimens that eliminates the problems associated with the ambiguous naming of samples. IGSNs also help with the discovery of physical specimens and enable a persistent link between a research publication, dataset, researcher etc and the specimen used in research. Applications for the IGSN may include rock and soil samples, archeological specimens and more - the “Geo” should really be renamed “General” as IGSNs can be applied to “all physical specimens”.

IGSNs work in a similar fashion to DOIs. As an individual researcher, your IGSNs for samples will be registered on your behalf by an Allocating Agent. Your Allocating Agent will advise you on how to create an IGSN and how to register it. In Australia, the allocating agents are CSIRO, Geoscience Australia and Curtin University.

## Concluding remarks

In this session we have looked at persistent identifiers and how they benefit the discovery and linking of research and researchers. I hope you have learned a lot and please feel free to follow up with myself or Ginny on any questions you have following this session. Thank you.