Contents

1. Purpose ........................................................................................................... 3
2. Introduction .................................................................................................... 3
   2.1 Data citation ............................................................................................... 3
   2.2 The DOI system ......................................................................................... 4
   2.3 ANDS “Cite My Data” service ................................................................. 5
3. DOI management guide .................................................................................. 5
   3.1 Business rules for assigning a DOI to data ............................................ 5
   3.2 DOI landing page ...................................................................................... 6
   3.3 Level of granularity ................................................................................. 6
   3.4 Citation ....................................................................................................... 6
   3.5 Versioning .................................................................................................. 7
   3.6 Collaborative data collections ................................................................. 7
   3.7 Withdrawn collections ............................................................................ 8
   3.8 DOI maintenance ..................................................................................... 8
4. The DataCite Metadata Schema .................................................................... 8/9
5. DOI minting .................................................................................................... 9
6. References ...................................................................................................... 10
   Acknowledgements .......................................................................................... 10
1. Purpose

The purpose of this document is to provide an introduction and management guidelines to minting and maintaining Digital Object Identifiers (DOIs) at Macquarie University using the Australian National Data Service (ANDS) Cite My Data service.

2. Introduction

Huge volumes of research data, largely born digital and enabled by vast advances in computing power, are being generated worldwide. Research institutions are faced with the immensely difficult task of finding ways to store and manage data in a format that facilitates discoverability, accessibility, and re-use. Data that is richly described, organised, integrated and connected can be more easily discovered by other researchers who may pose new questions to be investigated, raise larger issues to be investigated, and identify data landscapes to be explored. (Australian National Data Service (ANDS), About ANDS)

2.1 Data citation

Traditionally, knowledge derived from research is shared in the form of a publication such as a journal article. However the data used to produce the research publication is effectively lost to poor archival practice and only a very small proportion of the original data is made available in conventional journals. As part of a global effort to improve access to research data, there is growing impetus for an international culture of data citation using the Digital Object Identifier (DOI) system.

“Data citation refers to the practice of providing a reference to data in the same way as researchers routinely provide a bibliographic reference to printed resources.” (Australian National Data Service (ANDS), Data Citation Resources.)

DataCite is a global not-for-profit organisation formed in London in December 2009 that is facilitating the growing culture of data citation for scientific content. A key aim of DataCite is to “increase acceptance of research data as legitimate, citable contributions to the scholarly record” and they work with organisations that hold data to provide persistent identifiers in the form of DOIs. (DataCite, What is DataCite.)
2.2 The DOI system

The DOI system provides a framework for persistent identification, managing intellectual content, managing metadata, facilitating commerce and linking customers with content. DOIs are an implementation of the Handle System for persistent identifiers and seamlessly transport the user from one interface to another without requiring specific software. Information about a digital object may change over time, including where to find it and who owns it, but its DOI will not change.

A DOI is made up of alphanumeric characters and must be unique. It consists of a prefix and a suffix separated with a forward slash. The prefix always begins with ‘10’ as this distinguishes it from other implementations of the Handle System and then states the registrant code designating the creating organisation or publisher that is registering the DOI.

The suffix identifies the individual work and is also known as the ‘item id’. It is assigned by the publisher/owner of the DOI. The system evolved from the publishing industry and solidified in 1998 with the founding of the International DOI Foundation (IDF), an open membership consortium.

A DOI Registration Agency infrastructure provides ongoing support and maintains quality and accuracy of DOI names. Agencies are appointed by the IDF to provide service, quality assurance and overall integrity of the DOI system. (1)

CrossRef, a consortium of around 3000 publishers, is one of many IDF Registration Agencies. (CrossRef)

In addition to the International DOI Foundation (IDF) and Registration Agencies structure, DOIs differ from other persistent identifiers in that they require a minimal amount of metadata to be provided at the point of obtaining each DOI.

They also require a commitment from the provider to maintain the URL associated with the DOI. In the context of research data, this requires the provider of the data to maintain access to, and preservation of, the data persistently over time. The “DOI has emerged as the most widely used standard for digital resources in the publication world”. (2)

DOI is an ISO International Standard and over 55 million DOI names have been assigned by DOI System Registration Agencies in the US, Australasia, and Europe. In December 2010, only one year after its formation, DataCite had registered over one million DOI names for research data material.
2.3 ANDS “Cite My Data” service

The Australian National Data Service (ANDS) joined DataCite and in 2011 launched the Cite My Data service to offer minting of DOIs to Australian research institutions. (ANDS, Cite My Data Service.)

The Cite My Data machine-to-machine service is offered free of charge to ANDS-partners and facilitates minting of DOIs for research data collections and datasets using DataCite as the DOI Registration Agency. From late 2014 ANDS has also made available manual DOI minting service.

As an ANDS-partner institution, Macquarie University is able to use the Cite My Data service to mint and manage DOIs for datasets.

3. DOI management guide

This section of the document provides guidance on minting and managing DOIs for research data. It is intended to inform best practice at the institution. Refer to the DataCite business models principles for more information. (DataCite (2012), Business Models Principles)

3.1 Business rules for assigning a DOI to data

A DataCite DOI should be minted (via the ANDS Cite My Data) service where the data collection meets the following criteria:

- Access to the materials that comprise the collection will be open, mediated or embargoed. Data collections which are closed (e.g. due to ethical or legal constraints) should not have a DOI assigned to them.
- The material is a citable contribution to the scholarly record, analogous to a journal article.
- The collection metadata supports the provision of the five mandatory metadata elements required for compliance with the DataCite Metadata Schema.
- Macquarie University will support management of the collection in the long-term. This includes access and storage.
- The material does not already have a DOI assigned to it.
- The material is held in databases or systems managed by Macquarie University and not by a third party.
3.2 DOI landing page

A DOI should point to a landing page that contains metadata describing the collection.

The URL (Handle) for the landing page should be used as the ‘Identifier’ element of the DataCite Metadata Schema.

The landing page should be part of a website or web portal managed by Macquarie University e.g. ResearchOnline.

3.3 Level of granularity

A DOI can be assigned to any level of granularity. In the Macquarie University context, the granularity is likely to be collection level and item level within each collection.

If material within a data collection is made up of a series of files with different landing pages, and it is considered useful for citation purposes to have a specific link to each file, then a DOI may be minted for the landing page of metadata describing each file in addition to the collection level DOI.

3.4 Citation

A citation should be automatically generated following the minting of a DOI using the 5 mandatory metadata elements required to mint the DOI.

The citation element should follow a style applicable to data, such as the DataCite format that recommends:

    Creator (PublicationYear): Title. Publisher. Identifier.

An example of the DataCite citation style is:

    Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. Geological Institute, University of Tokyo. http://dx.doi.org/10.1594/PANGAEA.726855

The DOI and the citation element (which includes the DOI) should be displayed on the metadata landing page for the collection (or item within the collection, as applicable).
3.5 Versioning

Some collections are subject to change. A high level change would be a

- data deletion or addition, e.g. a new wave of data from a longitudinal study
- a metadata change that would lead to a change in the citation, e.g. the addition or deletion of a creator.

A low level change would be correction of a spelling error.

If a DOI is minted and assigned to a collection or item within a collection that is later subject to a high level change then a new DOI should be minted for the later version of the material. In this case, Macquarie University should maintain both versions, each with a separate DOI.

If a lower level change is required, then a new DOI does not need to be minted. The ‘Related Identifier’ element of the DataCite Metadata Schema can be used to refer to previous versions. An exception to this rule may be where a data collection is maintained for a single research group – such as the ‘Australian Institute of Health Innovation Collection’ – and they keep adding items to the collection. In this case, only one DOI should be issued at the collection level. Item level DOIs may prove useful in this case, providing the items are open access and have a metadata landing page.

3.6 Collaborative data collections

If the material is a collaborative effort with those external to Macquarie University then those involved in producing the material need to decide who mints the DOI and whose landing page the DOI will be linked to.

Considerations might include:

- Who is the primary researcher or author?
- Who is the lead institution?
- Who provides access to the material?
- Who is going to maintain access to the material in the long-term?
- Which institution has the capacity to mint a DOI?

Where another institution has already minted a DOI for the jointly-produced material, then that DOI should be stored and displayed in the Macquarie University record rather than minting a new DOI. This is applicable even where the DOI landing page points at the other institution.
3.7 Withdrawn collections

Where a collection has been withdrawn (for whatever reason) and that collection had been issued with a DOI, Macquarie University needs to maintain the landing page that the DOI links to. The landing page should include reference to the withdrawn material so that users who click on the DOI can still find information about the collection. This is part of the commitment to maintain data persistence through use of the DOI infrastructure.

3.8 DOI maintenance

In minting DOIs, Macquarie University supports management of data collections for the long-term. This includes an implicit requirement to maintain DOIs. Where a landing page changes or breaks, the DOI identifier element needs to be updated by Macquarie University.

4. The DataCite Metadata Schema

ANDS currently supports version 2.1 of the schema. The table below describes the 5 mandatory elements required by the DataCite Metadata Schema to mint a DOI. These are therefore also the requirements of the ANDS CiteMyData service.

<table>
<thead>
<tr>
<th>Schema element</th>
<th>Description [Source: DataCite Metadata Schema]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifier</td>
<td>“The identifier is a unique string that identifies the resource”. Best practice would be to use a URL that is persistent e.g. a handle.</td>
</tr>
<tr>
<td>Title</td>
<td>“A name or title by which a resource is known”.</td>
</tr>
<tr>
<td>Creator</td>
<td>“The main researchers involved in producing the data, or the authors of the publication, in priority order”. “May be a corporate/institutional or personal name. The personal name format should be: family, given”.</td>
</tr>
<tr>
<td>Publisher</td>
<td>“A holder of the data (including archives as appropriate) or institution which submitted the work.” “In the case of datasets, “publish” is understood to mean making the data available to the community of researchers”.</td>
</tr>
<tr>
<td>Publication year</td>
<td>“The year when the data was or will be made publicly available”. In the format YYYY “If an embargo period has been in effect, use the date when the embargo period ends”.</td>
</tr>
</tbody>
</table>
There are a number of optional metadata elements in addition to the mandatory five. These are detailed in the DataCite document at: [http://schema.datacite.org/](http://schema.datacite.org/).

While only minimal metadata is required, Macquarie University should support as much of the DataCite Schema as possible, bearing in mind that:

- This will enable support for richer metadata. This is important in versioning and linking DOIs etc.
- This scheme is subject to updates due to the activities of the DataCite Metadata Working Group.

ANDS is a member of the working group and we can potentially provide feedback and suggestions to DataCite regarding the scheme via an ANDS representative.

5. DOI minting

The DOIs will be minted by the Macquarie University Library via ANDS provided manual DOI minting service.
6. References:


Acknowledgements:

This document is largely based on Griffiths University “Digital Object Identifiers (DOIs): Introduction and Management guide”, prepared by Natasha Simons, Senior Project Manager, eResearch Services, Samantha Searle, eResearch Senior Specialist, eResearch Services and Stacey Lee, Research Content Officer, Information Services