DOIs for data: Using DataCite in the UK

Rachael Kotarski, British Library

Research Data Management Readiness "meeting the EPSRC policy framework"
February 2015
Addressing the challenges: DOIs

The Digital Object Identifier is a persistent identifier that directs users to an object, even if it changes location.

If an object moves, the new location is given to the existing DOI, so that it directs users to the same object in the new location.

EPSRC Expectation 5:

*Where the research data referred to in the metadata is a digital object it is expected that the metadata will include use of a robust digital object identifier (For example as available through the DataCite organisation)*
Other persistent identifiers are available!

Why DOIs?

• Global

• Widely adopted - most widely used identifier for research articles

• Governance and infrastructure

• Standardised - ISO 26324:2012
British Library's role in DataCite

- The British Library is one of 22 international members of DataCite
- We are an allocating agent
- We provide the DataCite infrastructure, enabling UK organisations to ‘mint’ DOIs for data
- Our aim is to support researchers, but we do not work with individuals - they must deposit to a repository
Are DataCite DOIs right for my institution?

• Do you have authority to assign DOIs to the data?
  – If not the data owner, you may need permission from the owner/creator

• Can you guarantee data persistence?
  – Data should be maintained and usable for the long term*

• Is the data accessible to external users?
  – Access restrictions may be required, but if no one can ever see the data again, why would it be cited?

• Does the data have ‘citation potential’?
  – Will it cited in future?
What is expected of data centres?

We supply DOIs for datasets as long as a data centre:

- Provides **mandatory CC0 metadata for its datasets**
- Provides a publicly accessible **landing page for each dataset**
- Can maintain **working URLs** for its registered datasets
- Has **curation and preservation policies in place**
Using DataCite DOIs

• No restrictions on DOIs providing criteria are met

• Access to User Interface and API to send metadata and mint DOIs

• Institution-specific prefix(es)

• *Institution specifies the DOI suffix*

10.5072/chosen_by.Institution-1

Prefix  Suffix
We are currently working with

<table>
<thead>
<tr>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Archaeology Data Service</td>
</tr>
<tr>
<td>2 Cambridge Crystallographic Data Centre</td>
</tr>
<tr>
<td>3 Cefas</td>
</tr>
<tr>
<td>4 Digital Repository of Ireland</td>
</tr>
<tr>
<td>5 Durham University</td>
</tr>
<tr>
<td>6 F1000Research</td>
</tr>
<tr>
<td>7 GigaScience</td>
</tr>
<tr>
<td>8 Imperial College London</td>
</tr>
<tr>
<td>9 Marine Science Scotland</td>
</tr>
<tr>
<td>10 Natural History Museum</td>
</tr>
<tr>
<td>11 NERC Data centres</td>
</tr>
<tr>
<td>12 Sir Alister Hardy Foundation for Ocean Science</td>
</tr>
<tr>
<td>13 STFC</td>
</tr>
<tr>
<td>14 UK Atomic Energy Authority</td>
</tr>
<tr>
<td>15 UK Data Archive</td>
</tr>
<tr>
<td>16 University of Bath</td>
</tr>
<tr>
<td>17 University of Bristol</td>
</tr>
<tr>
<td>18 University of Dundee</td>
</tr>
<tr>
<td>19 University of East London</td>
</tr>
<tr>
<td>20 University of Edinburgh</td>
</tr>
<tr>
<td>21 University of Glasgow</td>
</tr>
<tr>
<td>22 University of Leeds</td>
</tr>
<tr>
<td>23 University of Manchester</td>
</tr>
<tr>
<td>24 University of Oxford</td>
</tr>
<tr>
<td>25 University of Sheffield</td>
</tr>
<tr>
<td>26 University of Southampton</td>
</tr>
<tr>
<td>27 University of Strathclyde</td>
</tr>
<tr>
<td>28 University of Surrey</td>
</tr>
<tr>
<td>29 University of York</td>
</tr>
</tbody>
</table>

View online: http://bit.ly/1f9X0vV
How has it been done so far?
What are they being applied to? Not just ‘datasets’

- Grey literature
- ‘Collections’
- Crystal structure
- Poetry
- Etc…

Stoev, P; Komerički, A; Akkari, N; Liu, S; Zhou, X; Weigand, AM; Hostens, J; Porco, D; Penev, L (2013): Transcriptomic, DNA barcoding, and micro-CT imaging data from an advanced taxonomic description of a novel centipede species (Eupolybothrus cavernicolus Komerički & Stoev, sp. n.). GigaScience. http://doi.org/10.5524/100063
First World War Poetry Digital Archive, University of Oxford; (2011): Army Form B.218M, Pay and Terms of Service for His Majesty's Army (1); Bodleian Libraries, University of Oxford.
http://doi.org/10.5287/BODELIAN6XPC.2
Integration

ADS Collections Management System

Wroxeter Conservation Plan (Collection Id: 1001166)
See this archive on-line

Go to Tracking Project (1003607)  |  Go to Collection (1001166)

<table>
<thead>
<tr>
<th>DOI</th>
<th>Metadata</th>
<th>DataCite</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.5284/1017083</td>
<td><a href="http://archaeologydataservice.ac.uk/archives/view/wroxcons_eh_2012">http://archaeologydataservice.ac.uk/archives/view/wroxcons_eh_2012</a></td>
<td>[UPDATE]</td>
</tr>
</tbody>
</table>

[MOINT]
Providing mandatory metadata

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>This page represents DataCite's metadata for doi:10.14469/CH/123571. For a landing page of this dataset please follow <a href="http://dx.doi.org/10.14469/CH/123571">http://dx.doi.org/10.14469/CH/123571</a>.</td>
<td>This page represents DataCite's metadata for doi:10.5258/ECRUSTALS/913. For a landing page of this dataset please follow <a href="http://dx.doi.org/10.5258/ECRUSTALS/913">http://dx.doi.org/10.5258/ECRUSTALS/913</a>.</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://dx.doi.org/10.14469/CH/123571">http://dx.doi.org/10.14469/CH/123571</a></td>
</tr>
<tr>
<td><strong>Resource type</strong></td>
<td>RIS</td>
</tr>
<tr>
<td><strong>Collection</strong></td>
<td>NSC30323</td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>Create Commons Attribution (CC BY 3.0)</td>
</tr>
<tr>
<td><strong>Dates</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Issued</strong></td>
<td>2014-11-28</td>
</tr>
<tr>
<td><strong>Updated</strong></td>
<td>2014-11-28</td>
</tr>
<tr>
<td><strong>Available</strong></td>
<td>2014-11-28</td>
</tr>
<tr>
<td><strong>Alternate identifiers</strong></td>
<td></td>
</tr>
<tr>
<td>NCI</td>
<td>NSC30323</td>
</tr>
<tr>
<td>SMILES</td>
<td>N(CCC)(CCC)c1ccc(SC</td>
</tr>
<tr>
<td>CAS</td>
<td>6952-14-3</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Related identifiers</td>
<td></td>
</tr>
<tr>
<td>HasMetadata</td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="https://spectrum.pace.lib.imperial.ac.uk:8443/metadata/handle/10042/130224/mets.xml">https://spectrum.pace.lib.imperial.ac.uk:8443/metadata/handle/10042/130224/mets.xml</a></td>
</tr>
<tr>
<td>IsPartOf</td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="https://spectrum.pace.lib.imperial.ac.uk:8443/metadata/handle/10042/130224/core.xml">https://spectrum.pace.lib.imperial.ac.uk:8443/metadata/handle/10042/130224/core.xml</a></td>
</tr>
</tbody>
</table>
Providing mandatory metadata

<table>
<thead>
<tr>
<th>doi:10.7488/DS/72</th>
</tr>
</thead>
<tbody>
<tr>
<td>This page represents DataCite's metadata for doi:10.7488/DS/72. For a landing page of this dataset please follow <a href="http://dx.doi.org/10.7488/DS/72">http://dx.doi.org/10.7488/DS/72</a>.</td>
</tr>
</tbody>
</table>

**Citation**

**Resource type**
Dataset

**Subjects**
- conserved eukaryote signalling
- protein kinase phylogeny
- Ostreococcus tauri
- TOR signalling
- MAPK cascade
- circadian clock
- phosphorylation
- model kinome

**Dates**
- Issued: 2014-06-24
- Issued: 2014-06-24

**Alternate identifiers**
- [citation](http://dx.doi.org/10.7488/DS/72)
- [uri](http://hdl.handle.net/10283/663)

**Contributors**
- **DataManager**: University of Edinburgh
- **HostingInstitution**: University of Edinburgh
Open landing pages
Open landing pages
Open landing pages
Curation and preservation policies

Introduction and Purpose

The purpose of this policy is to: (i) create model research data management practices for all Researchers (defined below) at the University of Southampton (the University); (ii) foster responsibility for research data management through the development of research data management plans; and (iii) ensure that research data which is significant (as determined by the Researcher with guidance from the Faculty) is stored, retained, accessible and disposed of securely in accordance with all legal, statutory, ethical, contractual and funding requirements.

A robust Research Data Management Policy is required to demonstrate and ensure:

- good research practice and procedures;
- protection of intellectual property rights (IPR);
- proper recording, maintenance, storage and security of Research Data;
- compliance with relevant legislation and regulations; and
- appropriate access to Research Data is maintained.

This policy aims to address the dual objectives of making research data accessible to the public or wider academic community where possible whilst recognising that there are circumstances when access to research data must be restricted, either to achieve the highest standards for secure research data management or to fulfil commercial objectives.

1. Definitions

In this Policy, unless otherwise indicated:

1.1 "Principal Investigator (PI)" means the University employee primarily tasked with delivering a programme of research on behalf of the University, whether or not he/she is referred to as such in a research grant.

1.2 "Researcher" means any person undertaking research or involved in collecting, generating or creating Research Data, for or on behalf of the University which shall include but not be limited to employees, workers, visiting researchers, postgraduate research students and the PI.

1.3 "Research Data" means information in digital, computer-readable format or paper-based that:

1.3.1 is contained or presented in various ways including notes, facts, figures, tables, images (still and moving), audio or visual recordings; and

1.3.2 which is collected, generated or obtained during the course of or as a result of undertaking research (which includes both qualitative and quantitative data).
Curation and preservation policies

Preservation policy

Here is where the preservation policies are defined.

1. Items will be retained indefinitely.

2. The repository will try to ensure continued readability and accessibility:
   - Items will be migrated to new file formats where necessary.
   - It may not be possible to guarantee the readability of some unusual file formats.

3. The repository regularly backs up its files according to current best practice.

4. The original bit stream is retained for all items, in addition to any upgraded formats.

5. Items may be removed at the request of the copyright holder.

6. Acceptable reasons for withdrawal include:
   - Proven copyright violation or plagiarism
Curation and preservation policies

Research Data Management Policy for UEL

Preamble
The University of East London recognises that good research demands good data management in the support of academic integrity, openness and good stewardship. It will ensure that research data is managed to high standards throughout the research data lifecycle as part of its commitment to academic excellence. This policy will ensure UEL is in accordance with Research Councils UK’s Common Principles on Data Policy as well as the specific requirements of the Engineering and Physical Sciences Research Council. It is mandatory for data subject to contractual obligations from funders (such as a research council), and recommended for others. It was adopted by the Research & Knowledge Exchange Committee on 15 March 2012.

Research Data Management Policy

1. Research data created by UEL staff will be managed in compliance with RCUK’s Common Principles on Data Policy and community standards throughout the research data lifecycle as part of UEL’s commitment to academic excellence.

2. Primary responsibility for compliance with this policy lies with Principal Investigators (or UEL lead investigators) as Data Owners.

3. PIs will be offered a central source of information, guidance and contacts on research data management through the UEL website. This will encompass legal, IT, costing, data tools, storage, access restrictions and so on, and will be developed in consultation with research-active staff.

4. There is a presumption of releasing data created at UEL in order to meet regulatory and contractual requirements, with appropriate safeguards. In particular, the legitimate interests of participants will be protected in line with research ethics approval, and personal data will typically be anonymised if appropriate for release.

5. Deans will support compliance with this policy through their Research and Knowledge Exchange Leaders or equivalents, and make arrangements for stewardship of datasets in the absence of PIs.

6. Data management plans (DMP) are mandated at UEL for all externally-funded datasets created by UEL staff.
DOI suffixes

• ‘Must be copy-paste, so no mistakes’: Bristol
  – 10.5523/BRIS.14S2RMSTMX6AM1K8JKCD3G9UZ3

• ‘Based on an internal ID’: UKDS
  – 10.5255/UKDA-SN-7471-1

• ‘Sequential’: ADS
  – 10.5284/1024569

• Many institutions using a combination of these
Testing

• Test accounts are available to see how DataCite APIs can be integrated into your systems

• Just contact me: datasets@bl.uk
Men heard him now of pining stirrings, saw,
A thousand years ago even as now,
Black rooks with white gulls following the plough
So that the first are lost until a hero
Commits that last we first again — a law
Which men hold when one, like me, dreams how
A thousand years might dust lie in his brow
Yet some wise birds do between the sky it know.

Time swins before me, making as a day
A thousand years, while the sword ploughs on
Rows will lie: a man strike or turn the stroke
Of war never, audacions or resigned.
And God still sits aloft in the sky
That we have brought him, sole dry, stone blind.