DataCite revisited – Citing data in the XXIst century, at long last

Jan Brase, DataCite

APE 2012
January 25th
Berlin
Problem with data: The research trajectory

Data → analysised → synthesised → interpreted → become Information → is published → becomes Knowledge → ... is lost!

... is traceable

... is accessible
**DOI names for data**

**URLs are not persistent**
- (e.g. Wren JD: **URL decay in MEDLINE- a 4-year follow-up study**. Bioinformatics. 2008, Jun 1;24(11):1381-5).

**Digital Object Identifiers (DOI names) offer a solution**
- Mostly widely used identifier for scientific articles
- Researchers, authors, publishers know how to use them
- Put datasets on the same playing field as articles

**Dataset**
Yancheva et al (2007). Analyses on sediment of Lake Maar. PANGAEA.
doi:10.1594/PANGAEA.587840
What if data would be citable?

High visibility of the data
Easy re-use and verification of the data sets.
Scientific reputation for the collection and documentation of data (Citation Index)
Encouraging the *Brussels declaration on STM publishing*
Avoiding duplications
Motivation for new research
How to achieve this?

**Science is global**
- it needs global standards
- Global workflows
- Cooperation of global players

**Science is carried out locally**
- By local scientist
- Being part of local infrastructures
- Having local funders
Global consortium carried by local institutions focused on improving the scholarly infrastructure around datasets and other non-textual information. Focus on working with data centres and organisations that hold data. Providing standards, workflows and best-practice. Initially, but not exclusively based on the DOI system. Founded December 1st 2009 in London.
• DFG funded project with German WDCs
• TIB begins to issue DOI names for datasets
• Paris Memorandum
• DataCite Association founded in London
• 7 members
• 12 members
• All members assigned DOIs
• Over 800,000 items registered
• Pilot projects with Data Centres
• 16 members
• Over 1,2 million DOI names
• Metadata store
DataCite members

Technische Informationsbibliothek (TIB)
Canada Institute for Scientific and Technical Information (CISTI),
California Digital Library, USA
Purdue University, USA
Office of Scientific and Technical Information (OSTI), USA
Library of TU Delft,
   The Netherlands
Technical Information Center of Denmark
The British Library
ZB Med, Germany
ZBW, Germany
Gesis, Germany
Library of ETH Zürich
L’Institut de l’Information Scientifique et Technique (INIST), France
Swedish National Data Service (SND)
Australian National Data Service (ANDS)
Conferenza dei Rettori delle Università Italiane (CRUI)

Affiliated members:
Digital Curation Center (UK)
Microsoft Research
Interuniversity Consortium for Political and Social Research (ICPSR)
Korea Institute of Science and Technology Information (KISTI)
DataCite structure

International DOI Foundation

Member

DataCite

Managing Agent (TIB)

Carries

Member Institution

Data Centre

Member Institution

Data Centre

Associate Stakeholder

Works with
What type of data are we talking about?

Anything that is the foundation of further research is research data.

Data is evidence.
DataCite‘s main goals

Act as DOI registration agency

Actively involved in developing standards and workflows
CODATA-TG, STM, ICSTI, Data citation index

Central portal allowing access to the metadata from all registered objects. (OAI)

Community for exchange of all relevant stakeholders in the area access to and linking of data (data centers, publishers, libraries, research organisation, science unions, funders)
DataCite in 2012

Over 1,300,000 DOI names registered so far

DataCite Metadata schema published (in cooperation with all members) [http://schema.datacite.org](http://schema.datacite.org)

DataCite MetadataStore
[http://search.datacite.org](http://search.datacite.org)

OAI Harvester
[http://oai.datacite.org](http://oai.datacite.org)
DataCite search

Searchterm: *

Searchterm: uploaded:[NOW-7DAY TO NOW]

Searchterm: relatedIdentifier:*

Searchterm: relatedIdentifier:issupplementto:\10.1029*

Searchterm:relatedIdentifier:*\:10.1055*
DataCite Content Service

Service for displaying DataCite metadata

Different formats (BibTeX, RIS, RDF, etc.)

Content Negotation (through MIME-Typ)

• Access through DOI proxy (http://dx.doi.org)
• First implemented by CNRI and CrossRef:

Alpha available:

http://data.datacite.org
Examples

curl -L -H "Accept: application/x-datacite+text" "http://dx.doi.org/10.5524/100005"

curl -L -H "Accept: application/rdf+xml" http://dx.doi.org/10.5524/100005
⇒ RDF-file

curl -L -H "Accept: application/raw" http://dx.doi.org/10.5524/100005
⇒ ?
The dataset:
Storz, D et al. (2009):
*Planktic foraminiferal flux and faunal composition of sediment trap L1_K276 in the northeastern Atlantic.*
http://dx.doi.org/10.1594/PANGAEA.724325

Is supplement to the article:
Storz, David; Schulz, Hartmut; Waniek, Joanna J; Schulz-Bull, Detlef; Kucera, Michal (2009): *Seasonal and interannual variability of the planktic foraminiferal flux in the vicinity of the Azores Current.*
Deep-Sea Research Part I-Oceanographic Research Papers, 56(1), 107-124,
http://dx.doi.org/10.1016/j.dsr.2008.08.009
Where does a data citation belong?

eColi outbreak: First genome sequencing:
http://dx.doi.org/10.5524/100001

Nature put the DOI name in Accession codes, not reference:

"Crossref strongly recommends using the DataCite DOI in the citation, and also putting the citation in the reference section of the article (not just inline in the text)"

http://www.crossref.org/10quarterly/quarterly.html#dois_in_use
We don’t use the Web.
Berners-Lee created the Web as a scholarly communication tool.
Today the Web has changed everything but scholarly communication.
Online journals are essentially paper journals, delivered by faster horses.

But journals and citation are technology of the 18th century
Future of data citation

Try to measure various kinds of use to describe the importance of a data set:

• Resolution
• Downloads
• Mentions
• Citations
• Other types of linking

(Example: http://total-impact.org/, H. Piwowar; J. Priem)
Meet us and discuss with us

• ICSTI workshop ’Delivering Data in Science’, March 5th, Paris

• DataCite summer meeting, June 14th, Copenhagen (in conjunction with Nordbib conference „Structural frameworks for open, digital research”, June 11.-13.)

• [http://www.datacite.org](http://www.datacite.org)
• [contact@datacite.org](mailto:contact@datacite.org)