RatSWD
Working Paper Series

Working Paper   No. 163

DataCite:
The International Data Citation Initiative

Datasets Programme

Max Wilkinson

November 2010

Working Paper Series of the German Data Forum (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/2008 Heike Solga; since 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

DataCite: The International Data Citation Initiative

Datasets Programme*

Max Wilkinson

Datasets Programme Manager (max.wilkinson@bl.uk)

* This paper documents the presentation slides of a lecture given at a meeting of the UK Data Forum in November 2010

Abstract

DataCite is an international consortium which aims to increase acceptance of research data as legitimate, citable contributions to scholarly communication. To enable this DataCite assigns persistent identifiers for research datasets and manages the infrastructu res that support simple and effective methods of data citation, discovery and access. DataCite leverages the DOI infrastructure, which is already well-established. DOI names are the mostly widely used identifier for scientific journal articles, so researchers, authors, and publishers are familiar with their use. DataCite takes an open approach, however, and considers identifier systems and services that help forward its objectives.

DataCite is represented in the UK by the British Library. This summary of the British Library's involvement in DataCite was presented to the UK data Forum on Monday the 15th November 2010. Data publishers that wish to know more about DataCite or to use DataCite services are encouraged to contact the Library or their local DataCite members.

Further information and useful websites:

www.datacite.org.uk
www.datacite.org
www.doi.org

Keywords: Open Access, Research Data, Scientific Use, Public Use
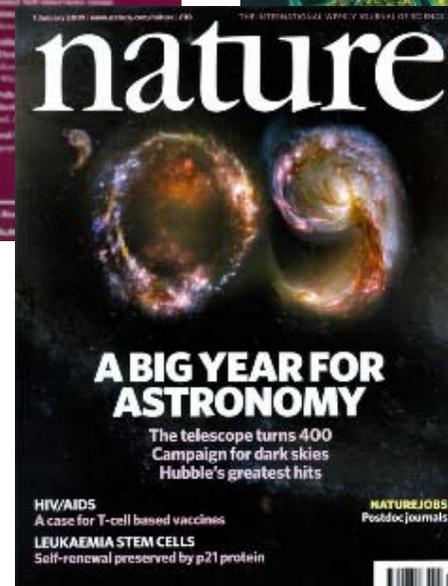JEL: C89, L86

Datasets Programme
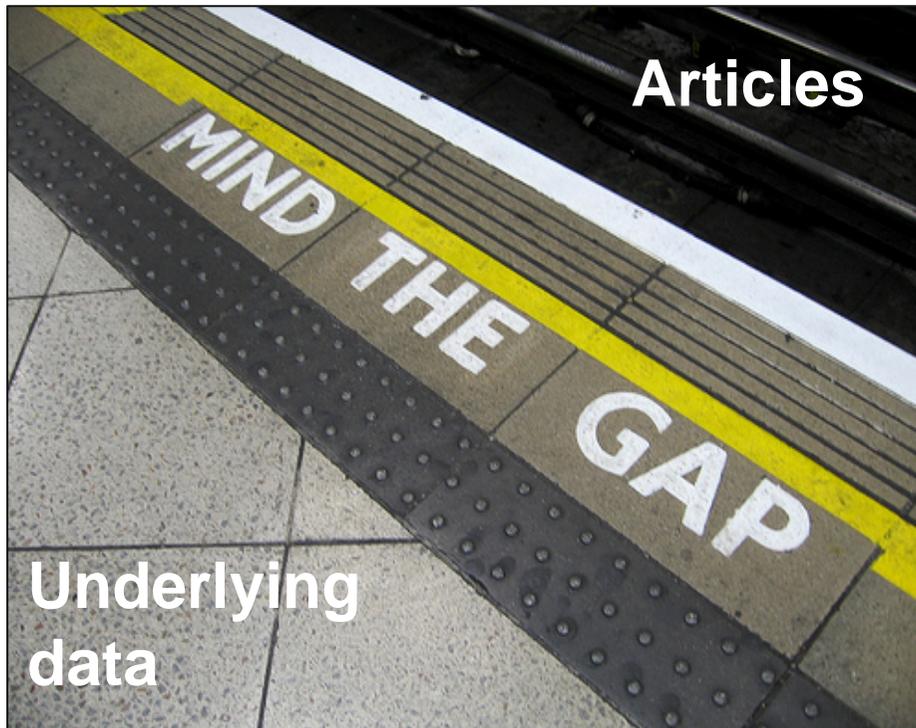
Nov 2010

Max Wilkinson

Datasets Programme Manager

Modern science relies on good data

When we talk about data, we are really talking about the evidence that underpins research

Currently…

Articles

Underlying data

- No effective way to link between articles and datasets

- No widely used method to identify datasets

- No widely used method to cite datasets

Project portfolio

A global consortium

DataCite is a global consortium founded in London on 1 Dec 2009, which:

- Aims to improve the scholarly infrastructure around datasets and other non-textual information

- Focuses on working with data centres and other data-rich organisations

- Provides standards, workflows, and best-practice

- Supports persistent identification of data using the DOI system

DataCite Membership

- German National Library of Science and Technology (TIB)

- British Library (BL), UK

- ETH Zurich Library, Switzerland

- Institute for Scientific and Technical Information (INIST-CNRS), France

- National Technical Information Center (DTIC), Denmark

- TU Delft Library, Netherlands

- Canada Institute for Scientific and Technical Information (CISTI)

- Australian National Data Service (ANDS)

- California Digital Library (CDL), USA

- Purdue University Libraries (PUL), USA

- German National Library of Medicine (ZB MED)

- GESIS - Leibniz Institute of Social Sciences, Germany

12 members from 9 different countries

Over 700,000 records registered with DOI names so far

Roles and responsibilities

The DataCite registration agency:

- Maintains the resolution infrastructure
- Maintains a searchable database of metadata
- Manages the identifiers over the long term
- Establishes and shares best practice

Publishing agents (data centres, research institutes, data publishers) are responsible for:
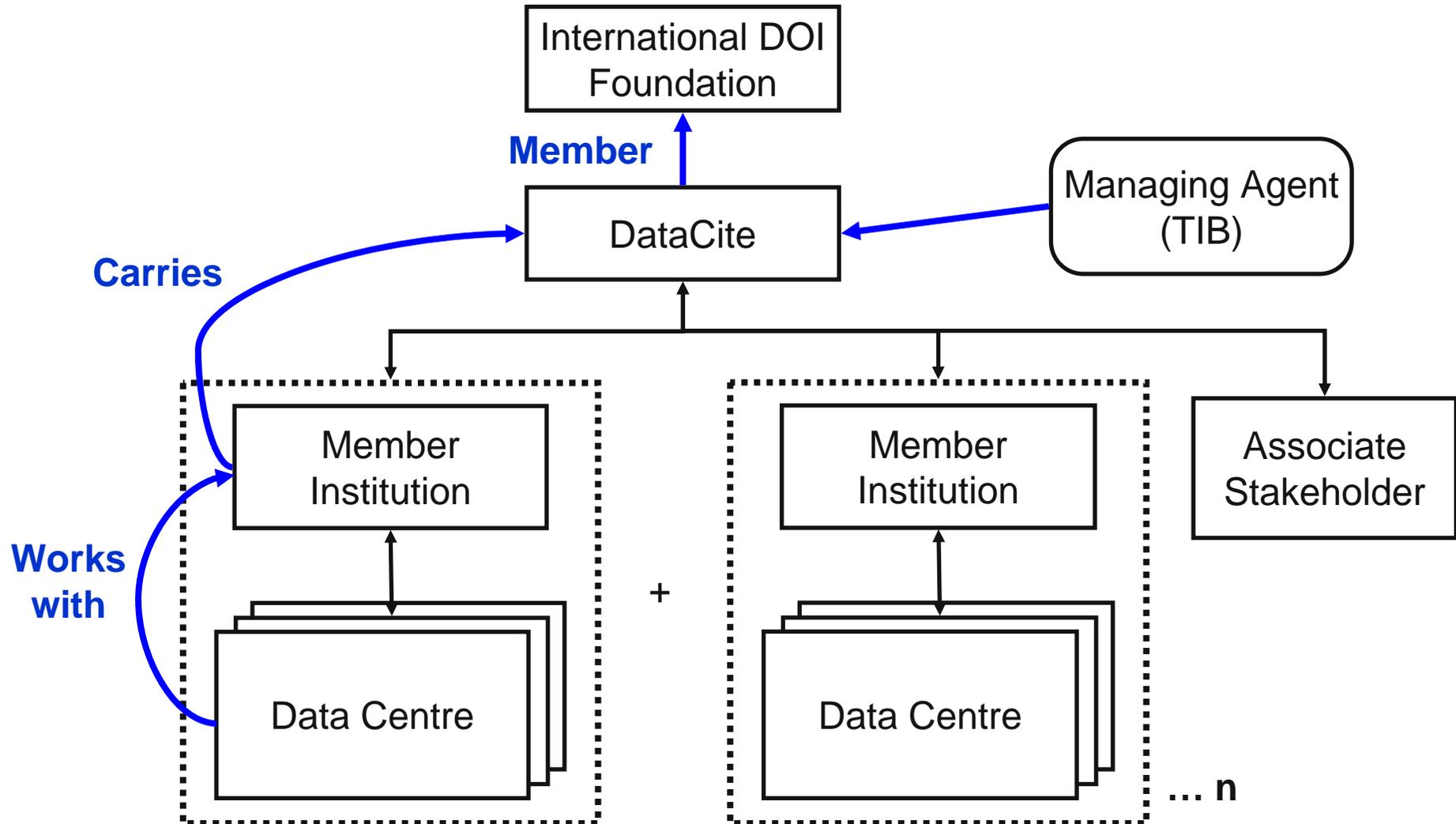
- Quality assurance
- Content storage and access
- Creating the identifier
- Creating and updating metadata

Our long term vision is to support researchers by providing methods for them to locate, identify, and cite research datasets with confidence

Memorandum of Understanding

Paris 2 March 2009

International DOI Foundation

Member

DataCite

Managing Agent (TIB)

Carries

Member Institution

Member Institution

Associate Stakeholder

Works with

Data Centre

Data Centre

+

… n

DataCite

- Support researchers by enabling them to locate, identify, and cite research datasets with confidence

- Support data centres by providing persistent identifiers for datasets, workflows and standards for data publication

- Support publishers by enabling research articles to be linked to the underlying data

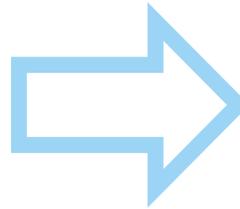**DataCite : Data Centres :: CrossRef : Publishers**

DataCite

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 (DOIs) offer a solution

A DOI is a unique identifier, similar in concept to an ISBN.

- Mostly widely used identifier for scientific articles
- Researchers, authors, publishers know how to use them
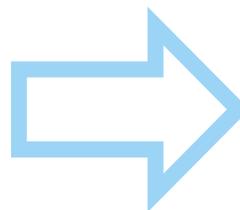- Put datasets on the same playing field as articles

The page cannot be found

The page you are looking for might have been removed, had its name changed, or is temporarily unavailable.

Please try the following:

- If you typed the page address in the Address bar, make sure that it is spelled correctly.
- Open the httpd.apache.org home page, and then look for links to the information you want.
- Click the ⇐ Back button to try another link.
- Click ◎ Search to look for information on the Internet.

HTTP 404 - File not found
Internet Explorer

Dataset
Yancheva et al (2007). Analyses on sediment of Lake Maar. PANGAEA.
doi:10.1594/PANGAEA.587840

In summary…

- Data are a crucial component of scholarly communication

- The British Library has a responsibility to address the challenges presented by data

- Datasets Programme works collaboratively to integrate datasets into its research services and to ensure the integrity of the scholarly record
- One such project is DataCite: The International Data Citation Initiative

Enquiries

Max Wilkinson

Programme Manager: Datasets

The British Library

96 Euston Road

LONDON. NW1 2DB

United Kingdom

max.wilkinson@bl.uk

Tom Pollard

Communication Officer: Datasets

The British Library

96 Euston Road

LONDON. NW1 2DB

United Kingdom

tom.pollard@bl.uk

Projects