

# Making Data Citeable: DataCite

Jan Brase

**Abstract** In 2005 the German National Library of Science and Technology started assigning DOI names to datasets to allow stable linking between articles and data. In 2009 this work led to the funding of DataCite, a global consortium of libraries and information institutions with the aim to enable scientists to use datasets as independently published records that can be shared, referenced and cited.

Data integration with text is an important aspect of scientific collaboration. It allows verification of scientific results and joint research activities on various aspects of the same problem. Only a very small proportion of the original data is published in conventional scientific journals. Existing policies on data archiving notwithstanding, in today's practice data are primarily stored in private files, not in secure institutional repositories, and effectively are lost to the public. This lack of access to scientific data is an obstacle to international research. It causes unnecessary duplication of research efforts, and the verification of results becomes difficult, if not impossible. Large amounts of research funds are spent every year to recreate already existing data.

Handling datasets as persistently identified, independently published items is a key element for allowing citation and long term integration of datasets into text as well as supporting a variety of data management activities. It would be an incentive to the author if a data publication had the rank of a citeable publication, adding to their reputation and ranking among their peers.

The German National Library of Science and Technology (TIB) developed and promotes the use of Digital Object Identifiers (DOI) for datasets. A DOI name is used to cite and link to electronic resources (text as well as research data and other types of content). The DOI System differs from other reference systems commonly used on the Internet, such as the URL, since it is permanently linked to the object

---

J. Brase (✉)

National Library of Science and Technology (TIB), Hanover, Germany  
e-mail: Jan.Brase@tib.uni-hannover.de

itself, not just to the place in which the object is located. As a major advantage, the use of the DOI system for registration permits the scientists and the publishers to use the same syntax and technical infrastructure for the referencing of datasets that are already established for the referencing of articles. The DOI system offers persistent links as stable references to scientific content and an easy way to connect the article with the underlying data. For example:

The dataset:

G.Yancheva, N. R. Nowaczyk et al. (2007)

Rock magnetism and X-ray fluorescence spectrometry analyses on sediment cores of the Lake Huguang Maar, Southeast China, PANGAEA

doi:[10.1594/PANGAEA.587840](https://doi.org/10.1594/PANGAEA.587840) (<http://dx.doi.org/10.1594/PANGAEA.587840>)

Is a supplement to the article:

G. Ycheva, N. R. Nowaczyk et al. (2007)

Influence of the intertropical convergence zone on the East Asian monsoon

Nature 445, 74-77

doi:[10.1038/nature05431](https://doi.org/10.1038/nature05431) (<http://dx.doi.org/10.1038/nature05431>)

Since 2005, TIB has been an official DOI Registration Agency with a focus on the registration of research data. The role of TIB is that of the actual DOI registration and the storage of the relevant metadata of the dataset. The research data themselves are not stored at TIB. The registration always takes place in cooperation with data centers or other trustworthy institutions that are responsible for quality assurance, storage and accessibility of the research data and the creation of metadata.

Access to research data is nowadays defined as part of the national responsibilities and in recent years most national science organisations have addressed the need to increase the awareness of, and the accessibility to, research data.

Nevertheless science itself is international; scientists are involved in global unions and projects, they share their scientific information with colleagues all over the world, they use national as well as foreign information providers.

When facing the challenge of increasing access to research data, a possible approach should be global cooperation for data access via national representatives.

- a global cooperation, because scientist work globally, scientific data are created and accessed globally.
- with national representatives, because most scientists are embedded in their national funding structures and research organisations .

The key point of this approach is the establishment of a Global DOI Registration agency for scientific content that will offer to all researchers dataset registration and cataloguing services. DataCite was officially launched on December 1st 2009 in London to offer worldwide DOI-registration of scientific

data to actively offer scientists the possibility to publish their data as an independent citable object. Currently DataCite has 17 members from 12 countries:

The German National Library of Science and Technology (TIB), the German National Library of Medicine (ZB MED), the German National Library of Economics (ZBW) and the German GESIS—Leibniz Institute for the Social Sciences. Additional European members are: The Library of the ETH Zürich in Switzerland, the Library of TU Delft, from the Netherlands, the L'Institut de l'Information Scientifique et Technique (INIST) from France, The technical Information Center of Denmark, The British Library, the Sedish National Data Service (SND), the Conferenza dei Rettori delle Università Italiane (CRUI) from Italy. North America is represented through: the California Digital Library, the Office of Scientific and Technical Information (OSTI), the Purdue University and the Canada Institute for Scientific and Technical Information (CISTI). Furthermore the Australian National Data Service (ANDS) and the National Research Council of Thailand (NRCT) are members.

DataCite offers through its members DOI registration for data centers, currently over 1.7 million objects have been registered with a DOI name and are available through a central search portal at DataCite.<sup>1</sup>

Based on the DOI registration DataCite offers a variety of services such as a detailed statistic portal of the number of DOI names registered and resolved.<sup>2</sup> In cooperation with CrossRef, the major DOI registration agency for scholarly articles a content negotiation service has been established that allows persistent resolution of all DOI names directly to their metadata in XML or RDF format.<sup>3</sup>

In June 2012 DataCite and the STM association<sup>4</sup> signed a joint statement to encourage publishers and data centers to link articles and underlying data.<sup>5</sup>

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

---

<sup>1</sup> DataCite Metadata Search: <http://search.datacite.org/ui>

<sup>2</sup> DataCite Statistics: <http://stats.datacite.org>

<sup>3</sup> DOI Content Negotiation: <http://www.crosscite.org/cn>

<sup>4</sup> STM Association: <http://www.stm-assoc.org>

<sup>5</sup> Joint statement from STM and DataCite: <http://www.datacite.org/node/65>