A system for easy access to scientific information using DOIs

Jan Brase
German National Library of Science and Technology (TIB)

Irina Sens
German National Library of Science and Technology (TIB)

A system for easy access to scientific information using DOIs

Irina Sens, Jan Brase

German National Library of Science and Technology (TIB), Hannover, Germany
irina.sens@tib-hannover.de
jan.brase@tib.uni-hannover.de

Abstract

Since 2005 the TIB is an official DOI registration agency. Scientific content, mostly primary data sets, resulting from public funded research can be awarded with a DOI by TIB to become accessible in any web browser worldwide.

1 Registration of Scientific data

Many publications are based on scientific data sets that can not be accessed, therefore re-evaluation or re-analysis of data is almost impossible. To enable citations of data that encourage good scientific practice and acknowledgement of scientific work, the *German Research Foundation* (DFG) started the project *Publication and Citation of Scientific Primary Data*.

Starting with the field of earth science *the German National Library of Science and Technology* (TIB) is now established as a DOI-registration agency for scientific primary data as a member of the *International DOI Foundation* (IDF).

Primary data related to geoscientific, climate and environmental research is stored locally at those institutions which are responsible for its evaluation and maintenance. In addition to the local data provision, the TIB saves the URL where the data can be accessed including all bibliographic metadata. When data are registered, the TIB provides a *Digital Object Identifier* (DOI) as a unique identifier for content objects in the digital environment

2 The Digital Object Identifier

To register the data, the TIB awards it with a DOI as a unique identifier. In may 2005 the TIB has become an official DOI Registration Agency.

DOIs are names assigned to any entity for use on digital networks. They are used to provide current information, including where they (or information about them) can be found on the Internet. Information about a digital object may change over time, including where to find it, but its DOI will remain stable.

A DOI consist of two parts: A prefix and a suffix. For scientific data, a DOI looks like this

doi:10.1594/WDCC/IPCC_EH4_OPYC_SRES_B2_MM

10.1594 is the prefix and identifies that this DOI belongs to a scientific data set, registered at the TIB, WDCC stands for the respective research institute (World data center climate in this case), followed by the internal name of the data record at the WDCC.

A DOI can be resolved in every web browser worldwide, using the *Handle system* from the *Cooperation for National Research Initiatives (CNRI)*. A Handle server for example is installed at the webpage of the *International DOI Foundation(IDF)*. Resolving of this DOI is therefore possible for example, by using the URL

http://dx.doi.org/10.1594/WDCC/IPCC_EH4_OPYC_SRES_B2_MM

It is furthermore possible to install a free plug-in into the Internet Explorer to resolve this DOI by simple typing it into the address bar of the browser.

3   Scientific data in the library catalogue

With every dataset registered at the TIB, metadata is stored. This metadata for the content is based on a well defined application profile for scientific data. The profile includes all metadata identified in the ISO 690-2 obligatory for the citing of electronic media, together with Dublin Core based standard metadata attributes. A detailed analysis of the metadata used can be found in [1].

All registered data sets are accessible via the catalogues of the involved data centres. Some datasets, usually supplementary data sets that accompany a publication, are also include in the online library catalogue of the TIB.

4 Infrastructure

A special infrastructure is needed for flexible registration of DOIs for datasets and migration of meta information into related library catalogues. The key element is a webservice as part of the middleware at the TIB that offers automatic and manual upload of registration information.

Protocols and data formats are XML based, making it easy for developers to comprehend. By utilizing HTTP/HTTPS on the transport layer, web services can work through many common firewall security measures without requiring changes to the firewall filtering rules.

As the TIB webservice is SOAP (Simple Object Access Protocol) conformant, data providers can embed the client stub into their middleware by importing the WSDL (WebService Description Language) file into their application server.

5 Citability

Any scientist working with this data is now able to cite the data in his work by its DOI. By this, scientific primary data is not exclusively understood as part of a scientific publication, but has its own identity.

If a scientist reads a publication where registered data is used, he might be interested in analysing the data under different aspects. He can now cite the data in his own publications using its DOI, referring to the uniqueness and own identity of the original data.

Since academic regard is often measured in so-called "citation-indexes" counting the number of citations of a scientist's work, collecting data can therefore be accomplished as an important part of academic work.

Authors of articles already have started to cite datasets using the DOI in the bibliography. The article [2] for example uses and cites:

Stendel, M., T. Smith, E. Roeckner, U. Cubasch (2004): *ECHAM4_OPYC_SRES_A2: 110 years coupled A2 run 6H values*, WDCC, doi:10.1594/WDCC/EH4_OPYC_SRES_A2.

The above given a typical example for a citation of a dataset with a DOI.

6 Other content types

During the last year we have extended the registration to other disciplines like medicine and chemistry and other content types like crystal structures or grey literature. Examples are:

Medical case studies

In cooperation with the *European Congress for Radiology* (ECR)

S. Cakirer , K. Demir, M. Beser, GM. Galip (2001, May 23). *Epiploic appendagitis*, doi:10.1594/EURORAD/CASE.1113

Grey literature

In cooperation with the *European Association for Computer Graphics* (Eurographics)

K. Debattista, L. Paulo Santos, and A. Chalmers *Accelerating the Irradiance Cache through Parallel Component-Based Rendering* Eurographics Symposium on Parallel Graphics and Visualization (2006) pp. 27-34, doi:10.2312/EGPGV/EGPGV06/027-034

Crystal structures

In cooperation with the *UK Office for Library Networking* (UKOLN)

6,7,9,10,12,13,15,16-Octahydro-benzo-1,4,7,10,13-pentaoxacyclopentadecin doi:10.1594/ecrystals.chem.soton.ac.uk/145

7 Status

Until now (summer 2007) the TIB has registered 435,276 data sets, 6,302 radiological case studies, 342 technical reports and 112 learning objects.

In the future the registration of any scientific content that is a result of community funded research shall be a primary task for the TIB. This will include the registration of various different content types like crystal structures, earth samplings, 3-D models, etc.

References

[1] Brase, J. 'Using digital library techniques - Registration of scientific primary data' (LNCS 3232) *Research and advanced technology for digital libraries,* (2004)

[2] Lorenz, S.J., Kasang, D., Lohmann, G. 'Globaler Wasserkreislauf und Klimänderungen -- eine Wechselbeziehung' *Warnsignal Klima: Genug Wasser für alle?*}, pp. 153-158. Wissenschaftliche Auswertungen, Hamburg, Germany.