DataCite - International Data Citation

Business Models Principles

Version 1

2012-10-01

doi:10.5438/0007

Members of the Business Practices Working Group

Karen Morgenroth, CISTI (Chair)

Paul Bracke, Purdue University

Lee-Ann Coleman, British Library

Cyndie Found, CISTI

Angela Gastl, ETH Zurich

Herbert Gruttemeier, INIST

Brigitte Hausstein, GESIS

Joan Starr, CDL

Frauke Ziedorn, TIB

Table of Contents

1 Introduction

DataCite is an international association dedicated to making it easier for everyone to identify, cite, discover, and use research data.

DataCite members enable data owners, stewards, or archives to assign persistent identifiers to research data. Currently, DataCite provides extensive support for Digital Object Identifiers (DOIs), although some members assign other sorts of identifiers as well. DOIs are used extensively for identifying research articles. They are particularly suited for resources that are intended to be cited.

This document describes some of the current best business practices and responsibilities for DataCite members and their clients. It will be reviewed at least yearly and may be updated more frequently if substantive changes are required.

2 Research Data

DataCite provides identifiers for research data.

Research data for which identifiers will be assigned must be located in data centres or repositories committed to persistence and maintenance.

Identifiers can be assigned to all data formats. For information about required and optional metadata associated with data formats, see the DataCite Metadata Schema.

Definitions:

For the purposes of this activity, research data is defined as:

- datasets
- data papers
- grey literature

Dataset: "Recorded information, regardless of the form or medium on which it may be recorded including writings, films, sound recordings, pictorial reproductions, drawings, designs, or other graphic representations, procedural manuals, forms, diagrams, work flow, charts, equipment descriptions, data files, data processing or computer programs (software), statistical records, and other research data." (from the U.S. National Institutes of Health (NIH) Grants Policy Statement via *DataCite's Best Practice Guide for Data Citation*).

Data paper: "A data paper minimally consists of a cover sheet and a set of links to archived artifacts. The cover sheet contains familiar elements such as title, authors, date, abstract, and persistent identifier (e.g., a DOI or ARK) — just enough to permit basic exposure to and discovery of data by internet search engines; also just enough to build a basic data citation, to instill confidence in the identifier's stability, and to be picked up by indexing services such as Google Scholar. This simple format represents only the first stage of the evolution of the data paper. There is room for the format to increase in complexity with the incorporation of other valuable elements, both general purpose and discipline specific, to enrich discovery, re-use, and archiving." (from *Practices, Trends, and Recommendations in Technical Appendix Usage for Selected Data-Intensive Disciplines*, http://www.cdlib.org/services/uc3/docs/dax.pdf).

Grey literature: "Documentary material in print and electronic formats, such as reports, preprints, internal documents (memoranda, newsletters, market surveys, etc.), theses and dissertations, conference proceedings, technical specifications and standards, trade literature, etc., not readily available through regular market channels because it was never commercially published/listed or was not widely distributed." (source: ODLIS - http://www.abc-clio.com/ODLIS/odlis_A.aspx).

Note: All references to "data" within this document are intended to reflect the above definitions of "research data"

3 Allocation Agency Responsibilities

DataCite members that provide identifiers for clients are allocation agencies. Allocation agencies work directly with clients who are responsible for providing access to data. In most cases, the allocation agency and client will be located in the same country. Some exceptions exist:

Countries with more than one DataCite member:
As each country's situation will be unique, members in countries with multiple DataCite allocation agencies are to work out the best arrangements amongst themselves, including cost and fee schedules.

Multinational consortia, international research organizations, or organizations in countries that do not currently have a DataCite member:
Deciding which allocation agency should provide the service will be done on a case by case basis. Determining factors can include, but are not confined to:

- The allocation agency with whom the client has established a previous relationship
- Polling of DataCite members (e.g., the DataCite member who receives the initial request for service forwards it to the DataCite allmembers list to allow each member the opportunity to provide the service).
- Geographic proximity of allocation agency to client's head office or the location of the client's key staff
- The allocation agency's ability to enter into an appropriate legal agreement with the client
- The primary language used by the client
- The domain in which the client specializes e.g. social sciences

See Costs and fee schedules for other exceptions.

Below are the allocation agencies' responsibilities for working with clients.

3.1 Contracts

Allocation agencies shall have contracts or appropriate legal agreements with each client for whom they register identifiers. Contracts should include, at minimum, all required client and allocation agency responsibilities.

Contracts must include a statement about the mandatory requirement for metadata. It is the intention and mission of DataCite to make all metadata available for sharing. DataCite recognizes that there are some circumstances under which it is necessary to temporarily suppress access to metadata. Examples include data under embargo or desire to synchronize visibility of data with article publication.

3.2 Services offered

DataCite offers a number of services and resources to its clients.

Individual allocation agencies are free to develop their own services for clients above and beyond those offered by DataCite.

3.3 Costs and fee schedules

Whether an allocation agency's business model will be cost neutral, cost recovery, revenue generation, or other model will depend on the organization's requirements

It will be up to individual allocation agencies to determine whether to accept private sector clients. Agencies are asked to share strategies for dealing with private sector clients.

Fee schedules will be different for every member.

It is intended that clients will work with the allocation agency in their home country. The following are guidelines for cases where clients wish to work with a member other than the one in their home country.

- An agreement should be made between the two agencies and the client regarding this arrangement, including with whom the client will have the contract.
- The fee for service will be paid to the allocation agency doing the work unless otherwise negotiated.
- This fee will be the greater of the fees charged by the home country's allocation agency and the allocation agency providing the service.

3.4 Prefixes

Allocation agencies provide DOI prefixes to their clients.

- Clients should not share prefixes. Shared prefixes hamper transfer of DOIs between organizations or allocation agencts.
- Each client may request multiple prefixes.

4 DataCite client responsibilities

DataCite clients are those who have signed contracts / legal agreements with a DataCite allocation agency.

DataCite recognizes that because of the great diversity in research data, data publishers and the hosting data centres, there can be no single or overriding practice with respect to data stewardship and data registration. Not all data publishers have the full resources required for registering data to the extent that they might wish. As well, depending on the nature of the data, thorough registration of their data might be either an easy or complex task. Therefore, these guidelines should in general be taken as just that: guidelines for best practices. Requirements (as opposed to guidelines) are clearly indicated.

Note: As DataCite allocation agencies have only limited enforcement capabilities, only client activities that allocation agencies can conceivably enforce are indicated as requirements.

The following are the clients' responsibilities:

4.1 Commitment to data persistence

By requesting a DataCite account, clients are expected to:

- ensure that objects assigned DOIs are stored and managed such that persistent access to them can be provided as appropriate
- maintain all URLs associated with the DOI

4.2 Metadata

Clients:

1. are responsible for ensuring that they have the authority to make available the metadata for the object to which they are assigning an identifier
2. will provide, at minimum, the mandatory metadata as defined in the DataCite Metadata Schema (REQUIREMENT)
3. agree to make their metadata freely available for discovery purposes baring a business reason prohibiting this, in which case the client agrees so to mark the metadata (see Contracts) (REQUIREMENT)
4. will ensure that the URL assigned to the identifier provides users with the necessary information for making meaningful use of the data. Often this will be in the form of a landing page as indicated below.

4.3 Landing Pages

It is best practice to have a landing page for all registered data. The landing page typically contains one or more of the following:

1. full citation of the data
2. statement on access to data (such as a link to the data or usage restriction information)
3. associated metadata

4. information, software, or context required for unpackaging, reading and interpreting the data

If it is not possible to have a landing page for the data then it should be considered whether the data is at too granular a level for assigning a DOI.

- A landing page is mandatory for any data that cannot be viewed using standard desktop software (eg. .xls, .pdf, .txt, etc.). (REQUIREMENT)
- A landing page is mandatory when the data has restricted access. (REQUIREMENT)
- All landing pages must be publicly accessible. (REQUIREMENT)
- All links on the landing page must be up-to-date and functional. (REQUIREMENT)

4.4 Tombstone Pages

In the case that data become unavailable, it is important to provide information to this effect. This is normally done by updating the URL associated with a DOI to point to a tombstone page. It is essential for the link to the tombstone page to be persistent. There should be a unique tombstone page for each such identifier and the page should explain why the data is no longer available. Normally, the client should provide the tombstone page, but the allocation agent may do so when required.

4.5 Granularity of Data

DataCite clients are the custodians of their data and therefore define the granularity of the data to which they are assigning DOIs. Because the aim of DataCite is to make data citable, however, all data registered should be easily and clearly citable.

4.6 DOI Syntax

Unless the allocation agency provides the means for generating a DOI automatically, DataCite clients determine the unique suffix for the prefix used. Allocation agencies may make recommendations or provide guidelines for DOI syntax, but clients are free to create their own DOI using whatever syntax they choose, provided the resultant DOI is unique.