

DataCite – International Data Citation Initiative (DOIs for Datasets)

Presented at Links that Last, Digital Preservation
Coalition workshop, Cambridge
19 July 2012

Elizabeth Newbold, British Library



What is DataCite?

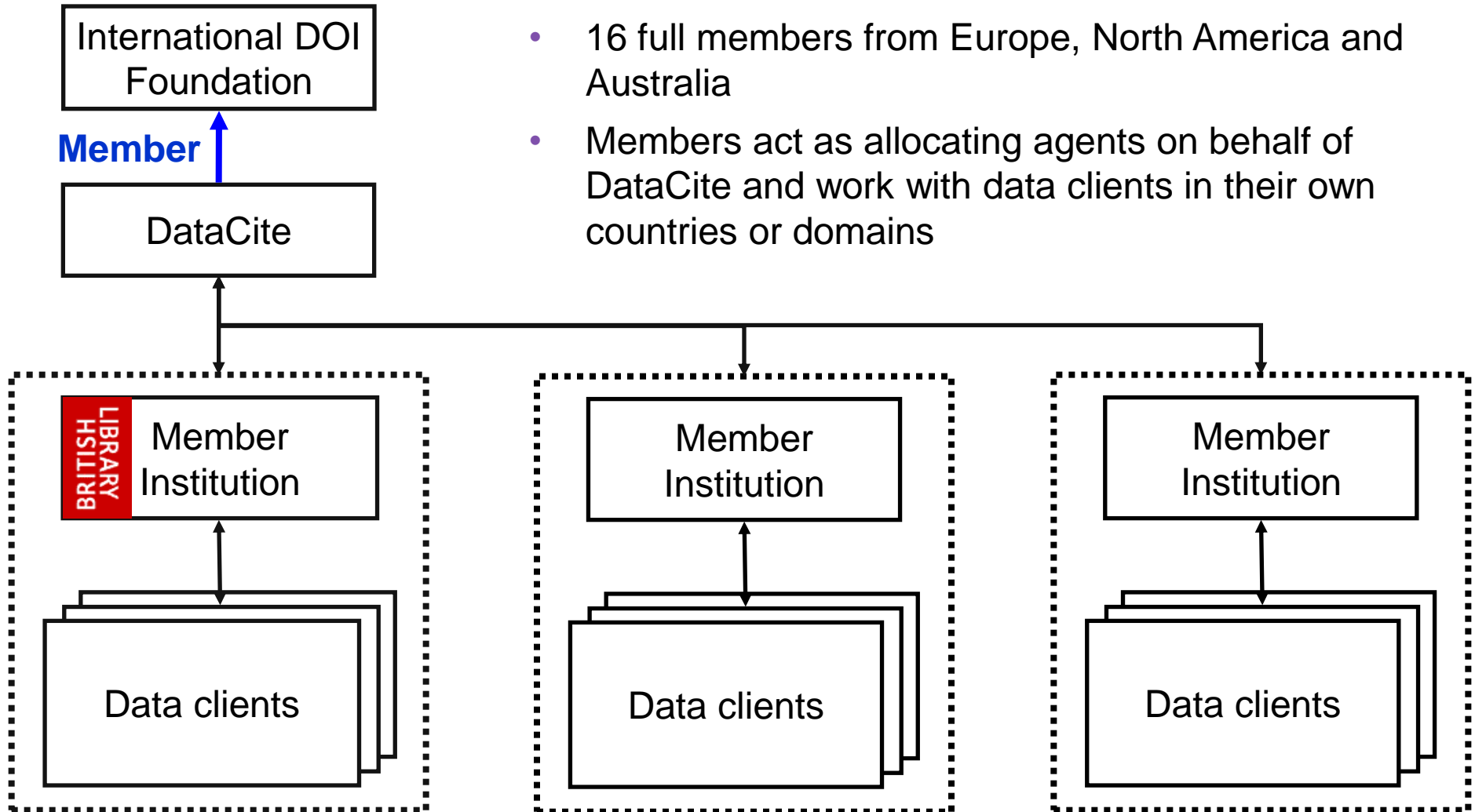
The image shows the DataCite logo and tagline on a blue background with a network diagram. The logo 'DataCite' is in a white box. The tagline 'Creating a global citation framework for data' is in white text below it. The background features a network of nodes and lines.

DataCite

Creating
a global
citation
framework
for data

- Established in 2009 (December)
- A Registration Agency for DOI names
- A member of the International DOI Foundation
- Not for profit organisation
- Provide a shared infrastructure for minting DOIs

DataCite structure



Stepping back - why the need?

Finding data it not easy!

- No widely used methods to **identify** and **locate** data
 - How can data be found and reused?
- No widely used method to **cite** data
 - How can researchers get credit for data sharing
 - How can data creators obtain credit of dataset creation efforts
- No effective way to **link** between articles and datasets

Why use DOIs?

What is the purpose of the DOI?

- provides the definition of the structure and syntax of the identifier itself and the provision of a technical infrastructure for resolving
 - Persistent **digital identifier** for an object
 - Level of indirection for URL

Example:

10.5524/100001 -> <http://gigadb.org/e-coli/>

Why use DOIs?

- Global
- Widely adopted - most **widely used** identifier for research articles
 - Permits the scientists and publishers to use the same syntax and technical infrastructure for datasets and articles
- Governance and infrastructure
 - Standardised - ISO 26324:2012

The **Digital Object Identifier** is a persistent identifier that directs users to an object, even if it changes location.

If an object moves, the new location is given to the existing DOI, so that it directs users to the **same object in the new location**.

Maintaining persistence

- The 'persistence' of identifiers can only be determined by the holders of the data
- So data must be stored somewhere that is **accessible**, and **well managed** for **the long term**
- This means that we need to make sure data centres we work with are appropriately managed for data curation and preservation

How do we do this – the role of data centres?

We supply DOIs for datasets as long as a data centre:

- Provides **mandatory metadata** for its datasets
- Provides a **landing page** for each dataset
- Can maintain **working URLs** for its registered datasets
- Has trusted **curation and preservation policies** in place

Data Persistence

Data clients must:

- have a clear and public indication to preserve the data

or

- have responsibility for providing access to the data over the long term

How can this be satisfied?

- By having a formal data preservation plan
- An operational service level agreement
- Clear intention in a mission statement

Open Resolution Target

- DOIs must resolve to a publicly accessible landing page
- The landing page must be open
- Describe the data objects and include at least the mandatory metadata and DOI

Economic and Social Data Service



Economic and Social Data Service

[About](#)[Data](#)[Access and support](#)[Resources](#)[Offer and deposit](#)[News](#)[Events](#)[Which service?](#)

National Child Development Study: Sweep 4 Feasibility Study and Tobacco Research Council Study, 1978

A new Digital Object Identifier (DOI) is assigned to the data collection each time there is a major change to data, documentation or metadata. The new DOI will resolve to an updated version of this page containing a log of changes to this data collection since the allocation of its first DOI. The DOI system supports resource discovery and simplifies citation for users of data collections. Data producers benefit directly through increased visibility of their work.

10.5255/UKDA-SN-2025-1

Citation:

National Children's Bureau, *National Child Development Study: Sweep 4 Feasibility Study and Tobacco Research Council Study, 1978* [computer file]. Colchester, Essex: UK Data Archive [distributor], 1984. SN: 2025, <http://dx.doi.org/10.5255/UKDA-SN-2025-1>

Change log:

7 October 2011: DOI created

Minor changes

5 April 2012: Minor changes to catalogue record

11 April 2012: Minor changes to catalogue record

[Full catalogue record >](#)

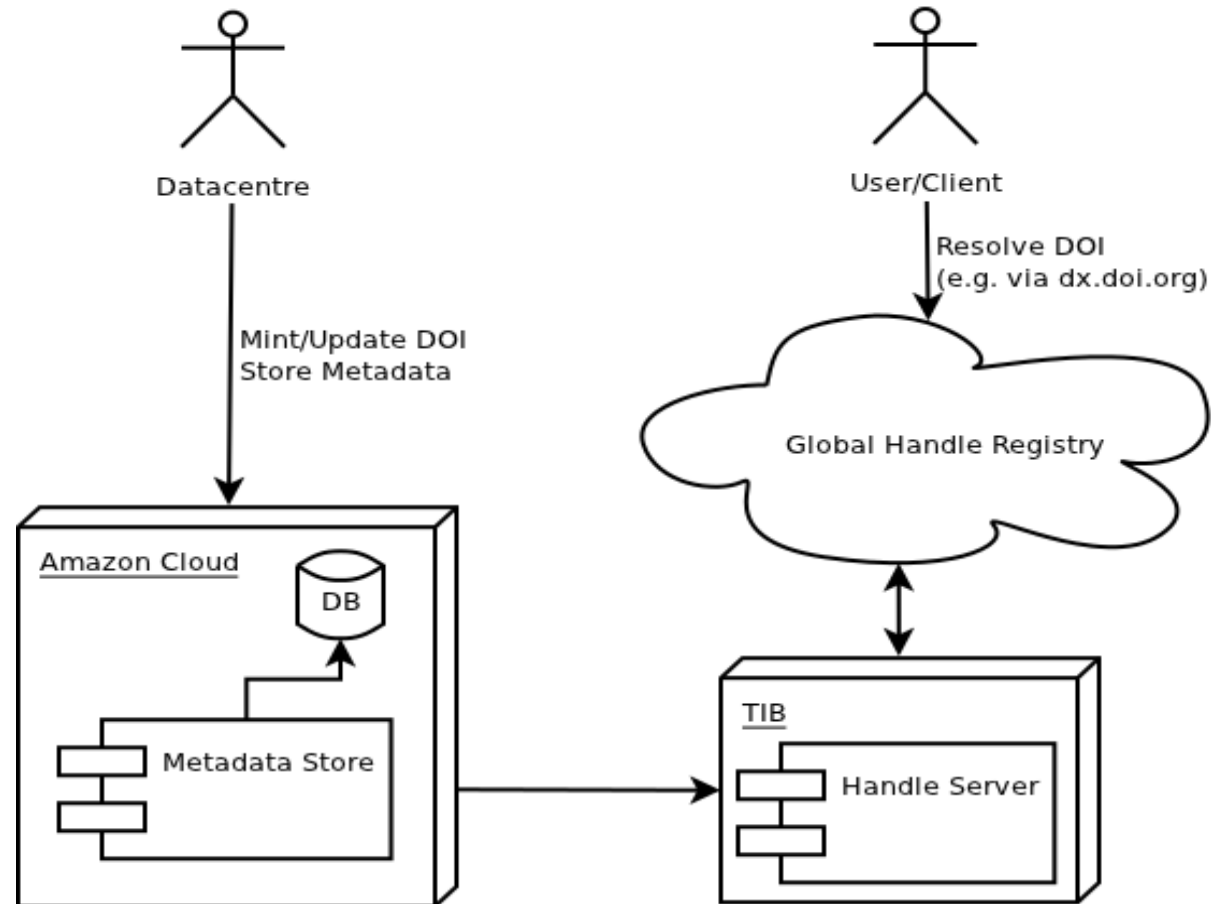
This is the latest release

Data Discovery

- Mandatory metadata must be provided according to the DataCite metadata scheme
- Metadata must be made freely available for discovery purposes
- Develop and implement suitable quality control measures to ensure compliance with the metadata scheme

DataCite infrastructure - Metadata Store (MDS)

- MDS is the DataCite central infrastructure for DOI management
 - Minting of DOIs
 - Storing of metadata
- Is only available to DataCite members and their clients



DataCite infrastructure - Metadata Schema mandatory properties

- Identifier
 - A unique string that identifies a resource – the allowed value is the DOI
- Creator
 - The main researchers involved in producing the dataset
- Title
 - A name by which the resource is known
- Publisher
- PublicationYear
 - Year in which the data was or will be made publicly available

The mandatory metadata elements are required for registration and minting DOIs but also provide the elements for the citation.

Creator (Publication Year): Title. Publisher. Identifier

In addition to the mandatory properties there are optional properties that help describe the data

In summary DataCite aims to provide a framework of support

- Supports **researchers** by enabling them to locate, identify and cite research datasets with confidence
- Supports **data centres** by providing workflows and infrastructure to identify and cite datasets
- Supports **publishers** by enabling research articles to be linked to underlying data

Further information

- DataCite www.datacite.org
 - [Approach for a Join Global Registration Agency for Research data](#). This white paper describes the initial principle of DataCite
- DataCite Metadata
 - The DataCite metadata schema is available at <http://schema.datacite.org>
 - [IsCitedBy: A Metadata Scheme for DataCite](#). Joan Starr from CDL explains the background to DataCite's metadata schema
- DataCite Metadata Store (MDS)
 - Documentation on the API and using the metadata store is available from <http://mds.datacite.org>
 - Metadata search <http://search.datacite.org>
- British Library www.bl.uk/datasets
- International DOI Foundation www.doi.org

Thank you

Email: datasets@bl.uk

Web: www.bl.uk/datasets
www.datacite.org