Working with the British Library and DataCite

A guide for Higher Education Institutions in the UK

BRITISH LIBRARY

DataCite
International Data Citation

About this guide

This booklet is intended as an introduction to the DataCite service that UK organisations can access through the British Library. It provides an overview of the DataCite universe and in-depth information about the application of Digital Object Identifiers (DOIs) to research data. Organisations that are considering working with DataCite will find best-practice guidance on how to register and manage DOIs for their data, alongside information on how to access the service through the British Library.

The information presented here has been informed by feedback received from attendees at a series of workshops on data citation, which took place at the British Library throughout 2012-13. Our work with existing DataCite clients has also influenced the guide.

Acknowledgements

Funding for the creation of this guide, and the workshops on which it was based, was provided by the Jisc Managing Research Data Programme 2011–13.

We would particularly like to thank Simon Hodson, Manager of the MRD Programme, and the contributors to the workshop series for their support.

Contents

1 The British Library and DataCite 02

2 How does DataCite work? 03

Administrative Infrastructure 03

Technical Infrastructure 03

3 Getting Started 05

DataCite Checklist 06

Client Responsibilities 06

British Library Responsibilities 06

4 Using the DataCite Metadata Store 07

5 Landing Pages 09

Basic elements of a landing page 09

6 DOI syntax 10

How to construct your suffix 10

7 Guide to DataCite Metadata 12

About the Schema 12

Mandatory metadata properties 12

Recommended metadata properties 12

Optional metadata properties 13

Citation Format 13

Metadata openness 13

8 Working with DOIs 14

Granularity: at what level should you assign DOIs? 14

Dealing with multiple versions 16

Embargoed data: when to apply the DOI 17

9 Glossary 18

10 Resources and References 19

1 The British Library and DataCite

A global platform for UK research data

About DataCite

DataCite is a global network of national libraries, data centres and other research organisations that works to increase the recognition of data as legitimate, citable contributions to the scholarly record.

Data sharing and re-use are becoming increasingly central to the research process, and so it is vital that we have effective tools to find, access and use that data. DataCite contributes to the emerging research data infrastructure by providing Digital Object Identifiers (DOIs) for datasets and other non-traditional research outputs. DOI assignment helps to make data persistently identifiable and citable.

In addition, DataCite promotes standards for data citation and description, and through its network of member organisations, creates an international community of users which helps establish best-practices for the citation and management of research data.

The British Library, DataCite and You

As one of the largest research libraries in the world, the British Library plays a key role in supporting research excellence in the UK.

The Library is one of the founding members of DataCite and helps make UK research data accessible and visible to an international audience by providing UK-based organisations with the ability to mint[1] DOIs for their data. Once signed up to the service, clients will have access to the full DataCite system, including:

- A simple web interface for DOI creation
- The Metadata Store: a centralised, searchable database which makes data easy to find and access
- A global infrastructure for data dissemination and citation
- Support and guidance on the best practices for data citation.

Digital Object Identifiers (DOIs)

DOIs are persistent identifiers that are used to uniquely identify an object (which itself may be physical or digital) in the digital environment.

Although other persistent identifiers are available, DataCite chose to work with DOIs as they have a number of features that are particularly compatible with the aims of the organisation. DOIs are:

- Already well-established as identifiers for research publications
- Recognised as an ISO standard[2]
- Centrally managed and governed by the International DOI Foundation (IDF).

[1] 'Minting' is the term used to describe the registration of a DOI for an item of data.
[2] ISO 26324:2012 **http://www.iso.org/iso/catalogue_detail?csnumber=43506**

2 How does DataCite work?

Administrative Infrastructure

As a UK-based DataCite client, your relationship is with the British Library. The Library acts as an intermediary between your organisation and the wider DataCite infrastructure: a hierarchy of organisations which provide the structure and governance that make DataCite work.

At the top of this hierarchy is the International DOI Foundation (IDF), the governing body responsible for the management of the DOI system worldwide. The IDF provides oversight of its member bodies, and policy and guidance in the form of the DOI Handbook[3].

DataCite is a member of the IDF and one of a handful of "Registration Agencies" (CrossRef is another) which have the authority to allocate DOIs. It is the only agency that focuses on identifiers for data.

DataCite operates the technical infrastructure (see below) that enables minting of DOIs for data and also provides policy and operating guidance for members.

Under DataCite's distributed model, member organisations serve regional communities around the world. The British Library is a member and "Allocation Agency" for DataCite, meaning that we have the authority to allocate DataCite DOI prefixes to our clients. We are the only Allocation Agency in the UK.

Technical Infrastructure

The DataCite system has two key components: the Metadata Store, which holds metadata for all DOIs registered by DataCite clients, and a Handle service, through which DOIs can be resolved by external users.

The DataCite Metadata Store (MDS)

The Metadata Store is at the heart of what DataCite does. It is the database which holds the metadata for all DOIs registered by DataCite clients worldwide.

The Metadata Store is centrally maintained by DataCite, ensuring secure, stable storage for your metadata.

Handle Service

The Handle System[4] is an established mechanism for managing identifiers ("handles") for resources in the digital environment.

The system is made up of numerous local handle services, each comprised of one or more servers on which handles are stored. DataCite's local handle service is hosted by CNRI[5]. An additional handle service called the Global Handle Registry holds information on the unique handle prefixes from each of the local services, meaning that a handle stored on any one of these services can be located through the Global Handle Registry.

The DataCite handle service maps handles to the URLs of landing pages, so when a user resolves a given DOI they are redirected to the landing page.

[3] DOI Handbook **http://www.doi.org/hb.html**
[4] Handle System website **http://www.handle.net/**
[5] The Corporation for National Research Initiatives **http://www.cnri.reston.va.us/**

How does it work?

The key processes carried out by the DataCite infrastructure are outlined below.

1 When the client repository registers or updates a DOI, metadata and the landing page URL are uploaded to the DataCite Metadata Store.

2 The Metadata Store sends information about the newly registered DOI and related URL to the DataCite Handle service, mapping each URL to a handle.

3 When users click on a DOI link or enter a DOI into a resolver service[6], the Global Handle Registry automatically redirects them to the landing page of the requested DOI.

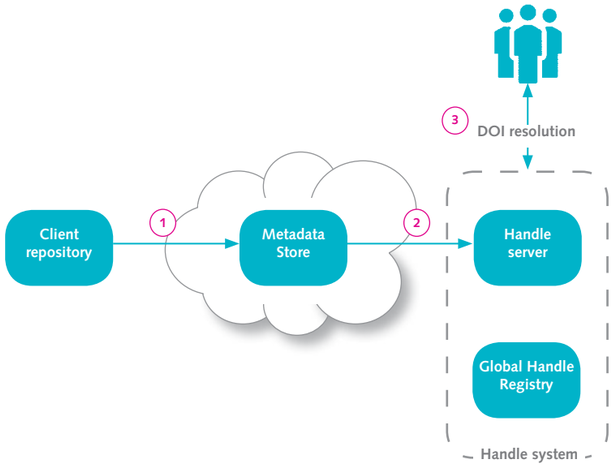

Figure 1 the DataCite technical infrastructure, showing the relationship between the Client, the Metadata Store, Handle Service and End User

DataCite DOIs and the Handle system

The DOI system operates on Handle technology, but this is only one element of the service that DataCite offers.

While the Handle system provides a secure method of resolution for identifiers, DataCite adds further infrastructure in the form of the Metadata Store, which securely links a data object to its description and ensures that your data is not only persistently identifiable, it also remains usable and searchable. The DOI system also offers a simple global resolution service through dx.doi.org.

In addition, the governance provided by the IDF and DataCite also means that there are standards, policies and guidelines in place to ensure that DOIs function as persistent identifiers for data, allowing them to be cited with confidence.

[6] **http://dx.doi.org/** is the global resolver maintained by the IDF.

3 Getting Started

If you are considering signing up for DataCite DOIs with the British Library, this flowchart sets out the basic pathway to getting started with the service.

1 Are DataCite DOIs right for my organisation?

To decide whether DOIs are the right option for your organisation and the data you hold, you should consider whether you meet the basic requirements set out in the British Library's DataCite Checklist.

There are also some ongoing Responsibilities that all DataCite clients must commit to, so please check that you will be able to fulfil these.

Refer to page 6 for further details of the Checklist and Responsibilities

2 Test account (optional)

Once you have decided that DataCite DOIs are an appropriate solution for your data you have the option to find out more before making a final decision by requesting access to a test account. A test account allows you to register temporary DOIs and to explore how the service fits within your workflows. The Metadata Store test environment replicates the live service almost exactly and is a good way to experiment with the system without any long term commitment.

3 Contract and Costs

The next step is to enter into a contract with the British Library. Contracts are for an initial term of three years; a period that indicates a level of commitment from both parties.

Institutions are charged a fixed annual fee which allows them to register an unlimited number of DOIs.

4 Implementing DataCite DOIs

Once the contract is agreed we will provide you with a Metadata Store account and unique DOI prefix so that you can begin minting DOIs for your data.

The rest of this guide provides detailed information on how to work with DataCite DOIs.

DataCite Checklist

Any organisation wishing to register DataCite DOIs through the British Library must meet the following conditions:

- **You have the authority to assign DOIs to your data**
  If you do not own the data you must have permission from the data owner/creator to assign DOIs.
- **You can guarantee data persistence**
  Before signing up with the British Library, you will be asked to demonstrate a commitment to maintaining your data in an accessible and usable form for the long term.
- **The data is accessible to external users**
  The central aim of DataCite is to make research data citable, therefore DOIs are not suitable for data that is completely unavailable to users outside your institution.
  We understand that, for certain types of data, it may be necessary to impose access restrictions or embargoes. In such cases, information about how to obtain and use the data should be made clear at the point of access (i.e. the landing page).
- **The data has citation potential**
  In other words, it is likely to be of interest to other researcher/users and may be cited in future works.

Client Responsibilities

As a British Library DataCite client, you agree to undertake the following tasks on an on-going basis:

- Provide and maintain at least DataCite mandatory metadata for each data item with a DOI
- Make metadata openly available without restriction (under Creative Commons Zero CC0[7] waiver)
- Maintain publicly accessible landing pages for each item with a DOI.

DataCite client requirements and responsibilities are fully detailed in the Business Models Principles[8] document.

British Library Responsibilities

When you sign a contract with the British Library to use the DataCite service, you can expect us to fulfil a range of responsibilities which help safeguard the persistence of your DOIs. We guarantee to:

- Maintain our membership of DataCite through payment of an annual subscription fee
- Issue a unique DOI prefix for your organisation
- Provide you with a Metadata Store account, allowing you to mint and manage your DOIs through both the user interface and API
- Provide support and advice on using the service, as required
- Work collaboratively with you to improve the service and develop new functionality.

[7] Creative Commons Zero (CC0) **http://creativecommons.org/about/cc0**
[8] DataCite Business Models Principles
http://www.datacite.org/sites/default/files/Business_Models_Principles

4 Using the DataCite Metadata Store

Once you have signed up with the British Library, we will provide you with a Metadata Store account which allows you to mint and manage DOIs for your data.

There are three primary tasks that you can perform through the Metadata Store user interface (Fig. 2):

Figure 2 The Metadata Store user interface

1 Mint a DOI

The primary task, minting a DOI for an individual dataset, can be done through the Metadata Store user interface.

To register a DOI for an individual item you will need:

- Metadata (in xml format)
- URL of its landing page
- DOI name that you wish to register (see DOI syntax).

When logged into your MDS account, you will find a "Register new dataset" option which allows you to enter the DOI and URL and to upload your metadata file. By clicking 'save' you have successfully registered a DOI.

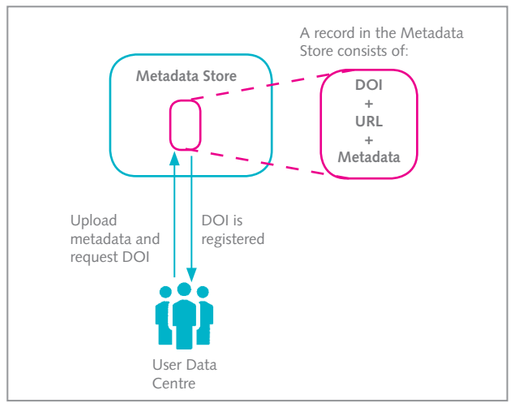

Figure 3 Minting of a DOI in the Metadata Store

2 Update metadata

Although the five mandatory properties cannot be changed, all other properties may be amended by uploading a new metadata record to the original DOI.

The MDS retains all versions of the metadata and records the version number, making it easy to track the history of your data.

3 Update URL associated with an existing DOI

If the location of the landing page changes, the new URL must be registered in the MDS to ensure that users resolving the DOI are directed to the correct target.
More information: You can view screencasts of the basic Metadata Store processes on the British Library's webpages[9].

Using the API

If you have a large number of datasets, it is possible to semi-automate the minting process by using the application programming interface (API). This means that you can integrate the minting process into your own systems. For details of how to use the API, please visit our online resource[10].

Making your data visible

Registering your data in the Metadata Store makes it easier to locate and access in a number of different ways.

The Metadata Store can be searched directly through the DataCite search page[11] which returns links to data landing pages. It can also be queried through the API, which allows for more tailored search and the option to return the results in a variety of formats, including xml, csv and json.

Furthermore, metadata is made available in OAI-PMH[12] format, which is the standard mechanism for repository interoperability. Both the API and OAI-PMH provider mean that the contents of the metadata store are available for search and harvesting by external services.

[9] Metadata Store screencasts
http://www.bl.uk/aboutus/stratpolprog/digi/datasets/datacitefaq/faqhome.html
[10] API documentation **https://mds.datacite.org/static/apidoc**
[11] Metadata Store search **http://search.datacite.org/ui**
[12] Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
http://www.openarchives.org/pmh/

5 Landing Pages

The web page that a DOI resolves to is called the landing page. The role of the landing page is to provide the user with information about the data and how it can be accessed. It is therefore essential that landing pages are publicly available with no barriers to access (such as registration). Each item with a DOI must have a landing page[13].

The URL of the landing page is uploaded to the Metadata Store at the point of DOI registration and should be updated by the client if it changes.

The landing page should display up-to-date information about the data and access/use conditions.

Basic elements of a landing page

The landing page must contain *at least* the following information:

- **The DOI**
  Prominently displayed to encourage users to cite the resource correctly.
- **Mandatory metadata**
  This provides basic information about the resource and its provenance. We suggest that you also include this information in the form of the full citation for the data.
  For information on DataCite's recommended citation format, please see Section 7 of this guide.
- **Information about how to access the data**
  This should include details of any access restrictions or registration requirements.

Users should be able to ascertain basic information about the nature of the data and its purpose from the landing page. Therefore, we recommend that you should include as much relevant descriptive metadata as possible; for example related resources, abstract or dates of creation.

You should also include any information necessary for the user to interpret or use the data. For example, specific software required to render the data accurately.

Landing Pages key points:

- Each DOI should resolve to a landing page
- Landing pages are open
- Landing pages provide information about how to access and use the data
- Even data that is no longer accessible must have a landing page.

More information: you can view a selection of best-practice landing pages on the British Library website at **http://www.bl.uk/datasets**.

[13] In some cases, it may be appropriate for multiple DOIs to resolve to the same landing page, for example collections of data or data with multiple versions.

6 DOI Syntax

A DOI name consists of a prefix and suffix separated by a forward slash.

10.1234/5678abcd
prefix **suffix**

Upon signing up to the DataCite service, each organisation is assigned a unique prefix. The format of the suffix is chosen by each organisation with only a few restrictions on syntax.

How to construct your suffix

The most important property of a DOI suffix is that it should be **unique** within the given prefix space. The DataCite system will reject any attempt to mint duplicate DOIs.

Please consider the following points when creating your suffixes:

- We suggest that the optimum suffix length is **8-10 characters**
  This is long enough to ensure uniqueness but short enough to avoid typing or text-wrapping errors. Please note that the DataCite system will not accept DOI strings of more than 255 characters.
- DOIs are **not case-sensitive**
  In other words 10.1234/**ABCD** and 10.1234/**abcd** are the same DOI.
- We strongly recommend that only the following characters are used within a DOI name: "**0-9**", "**A-Z**" and "**- . _ /**"
  Although the DOI name itself can accommodate a wider range of characters, some require encoding so that the DOI works correctly when used in URL form. If you need to use additional characters, please take care to follow the recommendations on encoding provided by the IDF[14].

Opaque vs. informational suffixes

Some data centres choose to use the suffix to convey information about the data that it represents. For example, the suffix could include a version number or local identifier to aid human recognition or to reflect some kind of hierarchy within the repository.

Please note, however, that the DOI system treats DOI names as opaque (i.e. "dumb numbers" which do not convey any information) and cannot interpret any meaning in your choice of suffix. Any information you choose to embed in the suffix should be for your internal use only.

You should also bear in mind that DOIs are designed to remain persistent over the long term and any meaning you embed in the DOI may not make sense to future users.

[14] DOI Handbook Section 2.5: Characters and Encoding
http://www.doi.org/doi_handbook/2_Numbering.html#2.5

Examples of DataCite suffixes

Below are some DOIs from existing DataCite clients which demonstrate the diversity of suffix structures currently in use.

Archaeology Data Service: 10.5284/**1000141**
Suffix chosen from a numerical sequence

NERC (BODC): 10.5285/**AD314FA1-7118-11E2-938C-1803734A77FB**
Randomly generated string

UK Data Service: 10.5255/**UKDA-SN-6918-2**
10.5255/**UKDA-SN-6918-1**
UKDS (formerly the UK Data Archive, UKDA) suffixes contain the organisation name, the study number (6918) and the version number (1, 2)

7 Guide to DataCite Metadata

Metadata is an integral part of the DataCite system. By associating good quality metadata with a DOI in the Metadata Store, your data is made more visible, accessible and usable.

DataCite provides a metadata schema that was designed for use with datasets and other non-traditional research outputs.

About the Schema

The DataCite Metadata Schema[15] is a list of core elements which were chosen to support the citation and retrieval of data. Some key benefits of the DataCite Schema are:

- It forms the basis for a recommended standard citation format for datasets (see page 13)
- It is interoperable with other data management schemas, such as Dublin Core
- It can accommodate numerous types of data
- Its development is overseen by a Working Group composed of DataCite members, allowing the schema to evolve in response to feedback from DataCite clients.

The schema consists of five mandatory properties plus a range of recommended/optional properties that may be used to describe and identify your data more fully. These can be further refined through the use of sub-properties.

Mandatory metadata properties

The following properties *must* be provided when you register a DOI.

Identifier sub-property: **identifierType**	The only allowed value for identiferType is "DOI"
Title	A descriptive name for the resource
Creator sub-property: **creatorName**	The author or producer of the data. There may be multiple Creators, in which case they should be listed in order of priority
Publisher	The data holder. This is usually the repository or data centre in which the data is stored
PublicationYear	The year when the data was (or will be) made publicly available.

Once registered, the mandatory metadata cannot be changed, so care must be taken to ensure that it is free of errors.

Recommended metadata properties

(sub-properties are specified only where they are required)

There is a further tier of properties that, while not compulsory, are recommended to increase the visibility and usability of your content. These are:

Subject
Contributor (sub-properties: **contributorType, contributorName**)
Date (sub-property: **dateType**)
ResourceType (sub-property: **resourceTypeGeneral**)
RelatedIdentifier (sub-properties: **relatedIdentifierType, relationType**)
Description (sub-property: **descriptionType**)
GeoLocation

[15] All information relates to version 3.0 of the Schema, which was released in August 2013 **http://schema.datacite.org/**

Optional metadata properties

A final series of properties can be applied to further refine the description of your data.

Language
AlternateIdentifier (sub-property: **alternateIdentifierType**)
Size
Format
Version
Rights

Not all of these properties will be relevant to all data (for example you may not have GeoLocation or RelatedIdentifier information for all data) but, if you hold this information about your data, we strongly recommend that you include it in the metadata record.

Citation Format

The five mandatory properties form the basis of the recommended standard citation format:

Creator (PublicationYear): Title. Publisher. Identifier

Where possible, we recommend that you should also include the "Version" and "ResourceType" properties in your metadata to ensure that your resources are easily identifiable. The expanded recommended citation format then becomes:

Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier

Please note that this is a recommended format only. You may choose to use any format that meets your needs or is required by the publisher.

Metadata openness

The British Library requires that all metadata held in the Metadata Store is published under a Creative Commons Zero (CC0) waiver, meaning that it may be freely re-used for any purpose. This ensures that the metadata is fully searchable and removes barriers to interoperability with other systems.

If you are unable to share even the mandatory metadata for your data under CC0, then the data is unlikely to be suitable for sharing and citation, and so a DataCite DOI is unnecessary.

DataCite Metadata: key points

- Metadata (at least the mandatory fields) *must* be provided at the time of DOI registration
- The Metadata Store will only accept metadata in xml format
- Mandatory metadata fields are fixed at the point of registration and cannot be updated
- Optional/recommended metadata fields may be updated or added to as required
- Metadata is open under CC0.

8 Working with DOIs

Deciding how to allocate DOIs to content is one of the first issues to be addressed by new DataCite users. Research data takes many forms and is not always presented in clearly defined 'units'. This means that assigning identifiers to data so that it is usable, meaningful and manageable can be challenging.

The key questions when considering how to implement DOIs for your data can be summarised as:

WHAT should you assign the DOI to?	**WHEN** should you mint the DOI?
• What can be considered a 'meaningful' unit of data? (see **Granularity**) • What level of data change triggers a new DOI (see **Dealing with multiple versions**)	• Upon deposit or publication? • Dealing with data embargoes

DataCite does not impose strict rules with respect to these issues and recognises that practices vary by discipline. However, there are a few general principles that should guide you when establishing institutional policies for DOI allocation.

Granularity: at what level should you assign DOIs?

The level of granularity at which DOIs are applied is determined by the individual data centre. When deciding upon the most appropriate level you should consider the following:

- The current citation and research practices among your user community: what is likely to be cited?
- The needs of various stakeholders: how will funders/publishers/administrators/etc. use the data?
- The type of resource: for example, a complex dataset may require a more granular identifier structure than a document or image file
- Sustainability: you must be able to maintain each item with a DOI in accordance with DataCite client Responsibilities.

With respect to this last point, the only DataCite and British Library requirements relevant to your choice of data granularity are that you must be able to provide meaningful metadata and a landing page for each item with a DOI.

If you cannot meet these basic requirements for each item that you want to assign a DOI to, you might consider using internal identifiers for lower-level data identification.

Expressing relationships between content items

Relationships between items can be encapsulated in the metadata using the **RelatedIdentifer** field. The **relationType** attribute of this property allows the precise nature of the relationship to be specified. Relevant allowed values include **isPartOf, hasPart, isSupplementTo** and **isSupplementedBy**[16].

Aggregate data

Objects stored separately can be placed in logical collections by applying an 'umbrella DOI' over individual components.

[16] Refer to the schema for the full list of allowed values **http://schema.datacite.org/**

Hierarchical data

Just as publications have identifiers at different levels of hierarchy (journal > article > section), some types of data may have similar relationships that it is useful to record. For example, a group of related items may exist at collection > study > file levels.

Relationships between items in a hierarchy may be expressed in DataCite metadata using the **RelatedIdentifier** property:

```
<relatedIdentifiers>
  <relatedIdentifier relatedIdentifierType="DOI"
    relationType="HasPart">10.1234/subDOI1</relatedIdentifier>
  <relatedIdentifier relatedIdentifierType="DOI"
    relationType="HasPart">10.1234/subDOI2</relatedIdentifier>
 </relatedIdentifiers>
</resource>
```

Figure 4 Excerpt of metadata of the lower level data object in a hierarchy

```
 <relatedIdentifiers>
   <relatedIdentifier relatedIdentifierType="DOI"
     relationType="IsPartOf">10.1234/umbrellaDOI</relatedIdentifier>
  </relatedIdentifiers>
</resource>
```

Figure 5 Excerpt of metadata of the lower level data object in a hierarchy

Linking to external items

You can also use the **RelatedIdentifier** property to link a dataset to a related publication, or any other item, both within the Metadata Store or elsewhere.

Again, the nature of the relationship between the items can be described using the relationType property. Relevant allowed values include **IsCitedBy**, **Cites**, **IsReferencedBy**, **References**, **IsDocumentedBy** and **Documents**.

For example, a dataset that forms the basis of a published article may contain a link to the publication via its DOI:

```
<relatedIdentifiers>
  <relatedIdentifier relatedIdentifierType="DOI"
    relationType="IsCitedBy">10.1234/articleDOI</relatedIdentifier>
 </relatedIdentifiers>
</resource>
```

Figure 6 Excerpt of metadata of a data object which has been cited in a publication

Note that the **RelatedIdentifier** field can accommodate many identifier types (not just DOIs), so you may link to content that is identified by, for example, URN, ARK or ISSN. Refer to the schema for the full controlled list.

Data with multiple access or use conditions

Having a greater level of granularity allows for greater flexibility and control over your data. For example, if a subset of a dataset contains confidential information, you can allocate a separate DOI to this subset so that access restrictions can be imposed while leaving the remainder of the dataset open.

Dealing with multiple versions

It is best practice to issue a new DOI when an item changes. Bear in mind that the original DOI may have been cited and the item reused, with published work based on it. Changing the data may undermine that work. So it is important to ensure that, when the DOI is accessed in the future, new users are able to exactly replicate work based on it. If you assign a new DOI to updated items, this issue can be avoided.

If a dataset undergoes a change requiring a new DOI to be issued, the **Version** property allows the update to be recorded. You can also link to previous/subsequent versions using the **RelatedIdentifier** property.

Very minor changes (such as the correction of a typing error that does not affect the meaning or authenticity of the resource) may, at the discretion of the data holder, be made without issuing a new DOI. Please contact us if you are unsure about how to deal with changes to your data.

```
<relatedIdentifiers>
   <relatedIdentifier relatedIdentifierType="DOI"
    relationType="IsNewVersionOf">10.1234/earlierversionDOI
    </relatedIdentifier>
  </relatedIdentifiers>
  <version>2</version>
</resource>
```

Figure 7 Metadata excerpt from a data object with an earlier version

Dynamic resources

Resources that undergo regular or ongoing changes (such as longitudinal data) require a strict versioning policy to ensure that users are served with the most recent or relevant dataset.

Depending on the nature of the resource, you may take a snapshot or slices approach[17] to versioning, with a new DOI issued for each snapshot or slice and clear information about what information is being presented and how it relates to other versions.

The schema also permits the citation of a continuously evolving dataset. In this case, you should add an access date and time to the citation to give users as much information as possible about the resource at the time it was used. We urge caution with this approach as, inevitably, when someone tries to access the data with this DOI in future, they will no longer be able to use the data as cited.

[17] Ball, A. & Duke, M. (2012). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: **http://www.dcc.ac.uk/resources/how-guides**

Embargoed data: when to apply the DOI

In cases when data can't be made accessible immediately upon deposit in the repository, it may be unclear when the most appropriate time to register the DOI is. While this depends largely on the policies and workflows of your repository, the following pros and cons may help you make your decision:

	Pros	Cons
DOI on embargo end date	No risk of revealing sensitive information	More difficult to cite data while under embargo Lack of publicly available information about the data
DOI on deposit	Record of the data exists in the Metadata Store Data creator can cite their own data The landing page and metadata can provide information on release date for external users	Metadata could reveal too much information Data may change significantly between creation and embargo end date

If you choose to mint a DOI for data that is not yet available for use, please remember that you must provide a landing page with information about the embargo and, preferably, the date when the data is likely to become available.

9 Glossary

Allocating Agent
A DataCite member organisation that provides access to the DOI system for clients.

API (Application Programming Interface)
A set of routines, protocols, and tools for building software applications. An API acts as an interface between software components, allowing them to communicate directly.

Content resolver
A service that exposes the metadata stored in the DataCite Metadata Store in multiple formats. It can also redirect to content hosted by DataCite participating data centres, making it possible to access data directly using a DOI.

Data Centre
Any organisation that takes responsibility for the storage, management and curation of data.

Data object
A resource to which a DOI can be assigned.

DOI (Digital Object Identifier)
A type of persistent identifier used to uniquely identify an object (which itself may be physical or digital) in the digital environment.

Domain name
An identification string that defines a realm of administrative autonomy, authority, or control on the Internet. E.g. "mywebsite.ac.uk"

Handle System
An infrastructure for the assignment, management and resolution of persistent identifiers ("handles") for digital resources in a network environment (such as the Internet). The DOI system is an implementation of the Handle System.

International DOI Foundation (IDF)
The international not-for-profit organisation that provides governance and management of the DOI system.

Landing page
The publicly accessible web page to which a DOI resolves. The landing page should contain metadata and access information about the object represented by the DOI.

Metadata Store
The database established and maintained by DataCite to act as a repository for metadata associated with DataCite DOI minting.

Mint/minting
The term used to describe the registration of a DOI for a data object by uploading metadata and landing page information to the Metadata Store. i.e. to "mint a DOI".

Persistent identifier
Generic term for a type of identifier that will continue to point to a resource even if the location of that resource changes.

Registration Agency
A member of the IDF that offers services for the registration of prefixes and DOI names using the DOI system. DataCite is a Registration Agency.

Resolution
The process by which an identifier is input into a network service and information about that identifier is returned. In the case of DataCite DOIs, the usual resolution target is the landing page of the object represented by the DOI.

10 Resources and References

Using DataCite DOIs

DataCite
http://www.datacite.org/

DataCite Schemas Repository
http://schema.datacite.org/

Metadata Search
http://search.datacite.org/ui

API documentation
https://mds.datacite.org/static/apidoc

Business Models Principles
http://www.datacite.org/sites/default/files/Business_Models_Principles_v1.0.pdf

British Library: Information for potential DataCite clients
http://www.bl.uk/datasets

The International DOI Foundation (IDF)
http://www.doi.org/

DOI Handbook
http://dx.doi.org/10.1000/182

Global DOI resolver
http://dx.doi.org

Other useful references

Information about the projects funded by the JISC Managing Research Data programme 2011-13
http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata

Handle System
http://www.handle.net/

ISO 26324:2012, Information and documentation -- Digital object identifier system
http://www.iso.org/iso/catalogue_detail?csnumber=43506

Creative Commons Zero (CC0)
http://creativecommons.org/about/cc0

Open Archives Initiative Protocol for metadata Harvesting (OAI-PMH)
http://www.openarchives.org/pmh/

Ball, A. & Duke, M. (2012). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online
http://www.dcc.ac.uk/resources/how-guides

Contact

If you'd like to talk to us about using DataCite DOIs in your organisation, please email us at **datasets@bl.uk**