

Using DataCite DOIs for ISIS Neutron Source Data

Michael Wilson, Brian Matthews,
Sri Nagella, Antony Wilson

*STFC Rutherford Appleton
Laboratory*



“We must give taxpayers more bang for their buck. Open access to ... data is an important means of achieving this.”

Máire Geoghegan-Quinn,
European commissioner for research, innovation and science
July 2012.



Why Data Persistent Identifiers?

- *Citable* datasets become legitimate contributions to scholarly communication, paving the way for new metrics and publication models that recognise and reward data sharing.



DataCite DOI

- Use DataCite service to:
 - mint, sustain, search and discover digital object identifiers (DOI)
- Associated with ISIS raw data
- Data is citable like a journal article
- Bibliometric services count data citation frequency for impact, e.g. REF submission



What is DataCite ?

- Datacite is
 - Not-for-profit organisation
 - Formed in London on 1 December 2009
 - Managing Agent: German National Library of Science and Technology
 - To the wider DOI foundation
 - E.g. Crossref – widely supported by publishers
 - Members and Associate Members around the world – not limited to 1 per country
- Provide services for data citation
- www.datacite.org

Datacite members

- **Australia**
- [Australian National Data Service](#) (Member)
- **Canada**
- [Canada Institute for Scientific and Technical Information](#) (Member)
- **Denmark**
- Technical Information Center of Denmark (Member)
- **France**
- Institute for Scientific and Technical Information - INIST (Member)
- **Germany**
- [German National Library of Science and Technology - TIB](#) (Member)
- German National Library of Medicine - ZB MED (Member)
- [GESIS - Leibniz Institute for Social Science](#) (Member)
- [German National Library of Economics - ZBW](#) (Member)
- **Netherlands**
- [TU Delft Library](#) (Member)
- **Republic of Korea**
- [Korea Institute of Science and Technology Information - KISTI](#) (Associate Member)
- **Sweden**
- The Swedish National Data Service - SNDS (Member)
- **Switzerland**
- [ETH Zurich](#) (Member)
- **United Kingdom**
- [The British Library](#) (Member)
- Digital Curation Centre (Associate Member)
- **United States**
- California Digital Library (Member)
- [Office of Scientific and Technical Information, US Department of Energy](#) (Member)
- [Purdue University Libraries](#) (Member)
- Interuniversity Consortium for Political and Social Research - ICPSR (Associate Member)
- Microsoft Research (Associate Member)



Using the DataCite Metadata Store

1. Register for an account with a DataCite member
 - In the UK - the British Library
2. Receive: user name, password to store service
 - For STFC - DOI:10.5286/
3. Write software to call API
4. Create landing page for each DOI
5. Mint DOI for each dataset through API
 - e.g. DOI:10.5286/ISIS.E.24079627
6. Assign some suitable metadata to the DOI



Metadata Store API 1 / 2

- **Doi API - RESTful**

GET

- *URI: <https://mds.datacite.org/doi/{doi}> where {doi} is a specific DOI.*
 - This request returns an URL associated with a given DOI.

POST

- *URI: <https://mds.datacite.org/doi>*
 - POST will mint new DOI if specified DOI doesn't exist.
 - This method will attempt to update URL if you specify existing DOI.
 - A Data centre's doiQuotaUsed will be increased by 1.
 - A new record in Datasets will be created.

Metadata Store API 2 / 2

- **Metadata API - RESTful**

GET

- *URI: <https://mds.datacite.org/metadata/{doi}> where {doi} is a specific DOI.*
- This request returns the most recent version of metadata associated with a given DOI.

POST

- *URI: <https://mds.datacite.org/metadata>*
- This request stores new version of metadata. The request body must contain valid XML.

DELETE

- *URI: <https://mds.datacite.org/metadata/{doi}> where {doi} is a specific DOI.*
- This request marks a dataset as 'inactive'.
- To activate it again, POST new metadata.

- See <https://mds.datacite.org/static/apidoc>

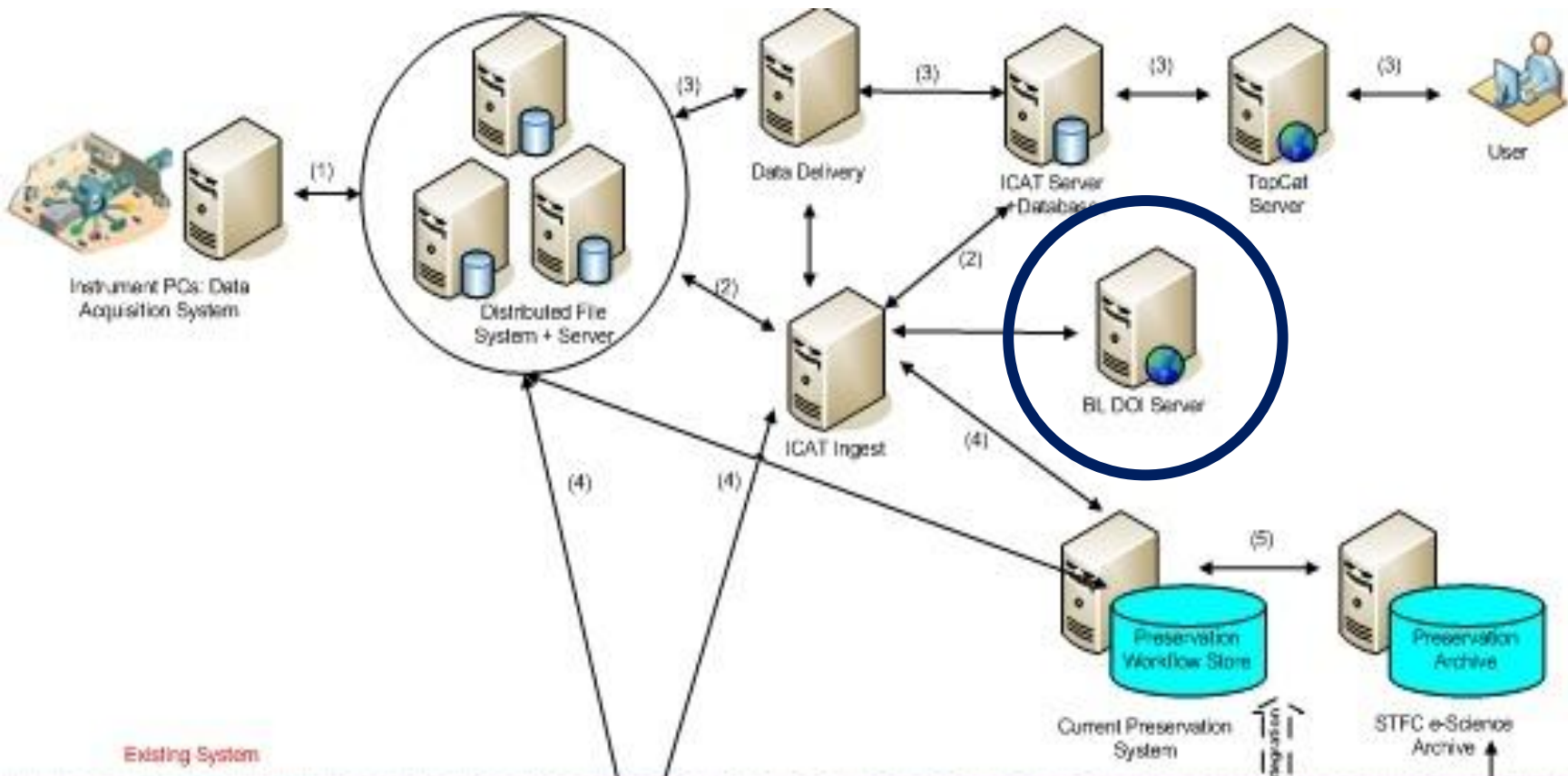
DataCite Metadata Schema

- Current Version 2.2, July 2011
- Available at: doi:10.5438/0005

<i>ID</i>	<i>Mandatory Property</i>	<i>ID</i>	<i>Optional Property</i>
1	Identifier	6	Subject
2	Creator	7	Contributor
3	Title	8	Date
4	Publisher	9	Language
5	PublicationYear	10	ResourceType
		11	AlternateIdentifier
		12	RelatedIdentifier
		13	Size
		14	Format
		15	Version
		16	Rights
		17	Description



ISIS Data Preservation Infrastructure



DOI Data Access Process

PHYSICAL REVIEW B 84, 075219 (2011)

http://search.datacite.org/ui - Microsoft Internet Explorer provided by STFC

http://search.datacite.org/ui#ui?&q=STFC

stereo ultimate 3d model viewer

Add-ons Gallery - Web Slice Suggested Sites Toshiba Places Web Slice Gallery

http://search.datacite.o...

Metadata


DataCite

Filter

- allocator
- datacentre
- prefix
- resourceType
- contributor
- creator
- publicationYear
- publisher
- language
- refQuality
- has_metadata

About STFC **Business & Innovation**

How we operate Collaborate with STFC



Data collected on the CRISP instrument at the ISIS facility

ISIS Data

Investigation
DOI: 10.5286/


Date of Expe
Publisher: S

Data format:
Select the data for

Data Citatic

The recomm
[author], [da

For Example
Griffin. et al;



RESEARCH COUNCILS UK

GLOSSARY : SITE-MAP : ACC

Science & Technology Facilities Council

Browse All Data

Download

- ISIS
 - ALF
 - ARGUS
 - CRISP
 - EMU
 - ENGINX
 - EVS
 - GEM
 - cycle_11_4
 - cycle_11_3
 - BaRuO3 8mm pos 8(id:CAL_GEM_2011-10-31T09:01:37)
 - GEM56174.raw
 - GEM56174.log
 - GEM56174_ICPdebug.txt
 - GEM56174_ICPevent.txt
 - GEM56174_ICPstatus.txt
 - GEM56174_Status.txt
 - GEM56175.raw
 - GEM56176.raw
 - GEM56174.nxs
 - GEM56175.nxs
 - GEM56176.nxs
 - Empty 6mm can 620(id:CAL_GEM_2011-10-10T17:23:28)

Recommended Data Citation

- **Creator (PublicationYear): Title. Publisher. Identifier**
- **Example:**
- **Easton,S; Barnes,C H W; Ionescu,A; (2011):
RB820232: Magnetic moment of EuO in spin
filtering magnetic tunnel structures.; STFC ISIS
Facility. doi:10.5286/ISIS.E.24066298**

Issues of using DOIs: 1/5

- DOIs are issued per experiment
- Easy for facility
- Should sustainable identifiers also be allocated at a finer granularity
 - Datasets, files
 - DOIs or others?



Issues of using DOIs: 2/5

- The minimum metadata are available from the registry when data are collected:
 - before publication of results
 - before expiration of embargo period.
- Is this too soon, should DOI be issued later ?
 - When data is collected ?
 - When data is released ?
- What does a persistent identifier “really” mean

Issues of using DOIs: 3/5

- Should extra metadata be published in the registry to increase the visibility of the data?
- If so when and what ?
 - Abstract, ?
 - Context ?
 - Published papers?



Issues of using DOIs : 4/5

- Access to the records of data access allow others to see where you are looking for data.
 - How open should records of data access be made?
 - Confidentiality vs. openness



Issues of using DOIs: 5/5

- Should users upload derived data to the ISIS archive for DOI access too?

