PaN-Data ODI

Deliverable D7.1

Implementation of persistent identifiers for PaNdata datasets

Grant Agreement Number	RI-283556
Project Title	PaN-data Open Data Infrastructure
Title of Deliverable	Implementation of persistent identifiers for PaNdata datasets
Deliverable Number	D7.1
Lead Beneficiary	STFC
Deliverable Dissemination Level	Public
Deliverable Nature	Report
Contractual Delivery Date	01 January 2013 (Month 15)
Actual Delivery Date	15 February 2013

The PaN-data ODI project is partly funded by the European Commission under the 7th Framework Programme, Information Society Technologies, Research Infrastructures.

Abstract

This document describes the processes leading to the deployment of DOIs as persistent identifiers for datasets produced in the PaN-Data facilities.

Keyword list

Digital Object Identifiers, metadata, DataCite

Document approval

Approved for submission to EC by all partners on Feb. XX, 2013

Revision history

Issue	Author(s)	Date	Description
0.1	Jean-François Perrin	4 Jan 2013	First draft version
0.2	Jamie Hall	5 April 2013	Technical description of the implementation
0.3	Tom Griffin	25 April 2013	ISIS process and implementation. Minor modifications.
0.4	Brian Matthews	3 June 2013	Minor modification and corrections.
1.0	Jean-François Perrin	6 June 2013	Final version

Table of Contents

1 Introduction 4
2 Persistent Identifier technology 4
2.1 Choice of the technology 4
2.2 Short technical Introduction to DOI 5
3 DOI Provider 6
4 Contract 7
4.1 Non-disclosure period 7
4.2 Quality assessment and facility business model 7
5 Workflow 8
5.1 STFC/ISIS workflow 8
5.2 ILL workflow 8
6 Implementation 9
6.1 General process 9
6.2 STFC/ISIS 9
6.3 ILL 9
6.4 Registration of DOIs 10
6.5 Landing page 12
7 Citation 14
8 First result 15
9 Conclusion 16

1 INTRODUCTION

The PaN-Data consortium brings together thirteen major European research infrastructures to create an integrated information infrastructure supporting the scientific process where Neutrons or Photons facilities have been involved. PaN-Data-ODI will develop, deploy and operate an Open Data Infrastructure across the participating facilities with user and data services which support the tracing of provenance of data, preservation, and scalability through parallel access.

The project aims at the standardisation and integration of the consortium's research infrastructures in order to establish a common and traceable pipeline for the scientific process from scientists, through facilities to publications.

WP7 aims at providing tools for fostering long term preservation of the datasets, in our case long term preservation means at least 15 years:

- Up to 5 years could be necessary for scientists in order to publish their work based on the data produced in the facility;
- 10 years is often considered necessary for further exploitation of the publication and associated datasets. Keeping the data after publication is becoming a recommendation of publishers and funding agencies, the period could vary from three to 10 years.

The first task was to adopt or develop a mechanism for identifying persistently and uniquely (worldwide) the data produced on our instrument.

In this document we describe the choice of the technology, the contracts that have been signed with the technology provider, the workflow that have been set in place and the implementations that lead to set up the identification on the data sets.

2 PERSISTENT IDENTIFIER TECHNOLOGY

2.1 CHOICE OF THE TECHNOLOGY

In this work package dedicated to preservation, we are looking for technical solutions that will allow facilities to persistently identify experimental datasets. We are looking for solutions that could: identify a digital object, be persistent over the years, even subsist after the closure of a data repository and be easily citable. Many different technologies could perfectly fit the needs of the objective of this work package, including: Uniform Resource Name (URN), Handle System/Digital Object Identifier (DOI), Persistent Uniform Resource Locators (PURL/Z), Archival Resource Key (ARK) and others. A technical review of ARK, PURL(Z) and DOI, done during the CRISP project highlighted that based on the technical criteria of persistency and ability to reference experimental data, they were all valid choices. Therefore the choice of DOIs is mainly based on the level of maturity and the effort of the different organizations for building a complete environment and developing its usage in the scientific community.

- As of 2012, more than 65 Million DOIs have been assigned by 5000 naming authorities registered through 10 Registrations Agencies (RA);

- In May 2012, after years of effort, DOI has become an International standard under the reference ISO 26324:2012, this recognition is a major step forward and provides confidence in the openness and sustainability of the solution. **DOI is becoming the persistent linking standard**.
- Digital Object Identifiers has the advantage of being already widely used for identifying scientific publications, this is the technology chosen by almost all scientific publishers (including the publication office of the EU) in order to electronically identify publications. It is of key interest for this project as we want to provide links between publications and data sets, this commonality will certainly help to introduce data citation in the standard scientific publication workflow and tools;
- Some organizations have already created synergies with the different partners (publishers[1], data producers, archive centers …) they are paving the way, our best option is certainly to contribute to this momentum;
- DOI is based on the handle system infrastructure which ensure that even in a worst case scenario where DOI registries disappear, we could technically be able to set up our own Handle infrastructure and maintain a minimum level of compatibility (at least suffix and metadata) for the already registered data sets. This option is already investigated by the European Persistent Identifier Consortium[2] who has set up its own handle infrastructure and wants to mint DOI when a citable and irrevocable ID is needed[3].

2.2 Short technical Introduction to DOI

This short paragraph is intended for readers who are not familiar with the concepts of persistent identifiers and DOIs in order to facilitate the reading of this document. Please refer to the DOI HandBook for a deeper explanation (http://www.doi.org/hb.html).

A DOI name such as 10.5291/ILL-DATA.6-05-579 is made up of a prefix (i.e. 10.5291) and a suffix (i.e. ILL-DATA.6-05-579) separated by a slash.

The prefix comprises a directory indicator, always 10 in case of DOI and a registrant code (i.e 5291 in the case of the ILL). The DOI suffix consists of a character string chosen by the registrant that should ensure uniqueness of the ID.

The registrant when minting a DOI should provide the suffix, some metadata describing the object and the address for accessing the object. The latter plays a special role for getting access to the identified object, in our case this address is the URL of a web page, referenced hereafter as “Landing Page”. The registrant should ensure persistent and open access to the landing page.

The registration agency (RA) should ensure persistent resolution of the DOI, to make it simple, the registration agency should ensure that if you provide a DOI name you should get the URL of the landing page, in a first approximation this could be compared to the role of Domain Name Server for IP resolution. The RA should also provide tools for minting DOIs and updating the metadata, the latter is particularly important in the case we need to move the landing page, for maintaining the consistency of the system.

[1] http://www.stm-assoc.org/2012_06_14_STM_DataCite_Joint_Statement.pdf
[2] http://www.pidconsortium.eu
[3] http://www.pidconsortium.eu/activities/EPIC_User_Meeting_2012_CSC/EPIC-Helsinki-Vortrag_Weigel_2012.pdf

3 DOI PROVIDER

Once the choice of DOI technology was clear, the choice of the registration agency was fairly simple. Since 2012, amongst the 10 DOI registration agencies[4], only DataCite provides solution for scientific data sets, whereas the others focus their model on journals, articles, books, movies or other content types. Another important element is that DataCite is a worldwide organization and not restricted to a particular region (EU, China…), an organization limited to a particular region could be problematic for the future of PaNData as at some point we should link our work with the one of our colleagues from America, Asia and Australia.

Even if in theory DOI could digitally identify any electronic or physical object, as scientific facilities, data producers and archive centers, our data descriptions, workflows and models are different from the one of publishers and could not fit the solutions proposed by publication oriented RA.

As a matter of example, if we take a closer look at CrossRef metadata schema[5], the other RA which is already well developed in the scientific community, the 3 hierarchies of the schema are: book, journal and conference, and they allow the description of metadata fields including: pages, ISBN, code, etc This is well suited for describing a book or conference proceeding, but is not relevant in our context and does not fit our needs for the description of our datasets.

DataCite is also very active in fostering the identification and proper citation of scientific datasets. DataCite is also involved in many projects and collaboration like ODIN, OpenAire, STM etc, in order to try to set up a whole data continuum and proper references and citation.

DataCite also proposes a number of useful additional services such as:

- Web Service API (https://mds.datacite.org), in order to massively mint or update DOIs, which is absolutely necessary for organization, like us, generating an important number of DOIs;
- Metadata Export (http://oai.datacite.org), service for harvesting metadata stored in their catalogue (MDS) using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH);
- Citation Formatter (http://crosscite.org/citeproc), in order to generate proper citation in standardized format (e.g. BibTeX) or publisher specific format (e.g. Nature);
- Metadata Search (http://search.datacite.org/ui), in order to search through the metadata, even with complex query, for DOIs;
- Statistics Portal (http://stats.datacite.org): an analytics tool for DOIs.

As it will be discuss in the next chapters, the steps to the full deployment of DOIs with DataCite in our research infrastructures was not so simple, but the objective alignment, the services and the general momentum created by DataCite has made our choice fairly simple. The other alternative, which would have been to create our own DOI RA in order to have a fine-grained adaptation to our model, was quickly discarded.

[4] http://www.doi.org/registration_agencies.html
[5] http://www.crossref.org/08downloads/CrossRef_Schema_Documentation.pdf

4 CONTRACT

Now that the choice of DOIs and DataCite had been done, it was time to start contract negotiation with DataCite.

The cost of registering DOI have never been a difficulty in our case. This cost is currently kept modest, in the order of 10 Cents of Euro per DOI, and appears necessary in order to allow DataCite to sustainably maintain the infrastructure.

The first difficulty encountered by the PaNData consortium was the national or regional organizational nature of DataCite. Discussion couldn't take place at the consortium level but had to be organize, in the case of Europe, with the national representatives. In the case of STFC/ISIS discussion took place with the British Library in the UK, in the case of ILL they took place with the INIST-CNRS (Institut de l'Information Scientifique et Technique) in France. This discussion took several months for both institutes, due to two main difficulties related to the business model of our facility.

4.1 NON-DISCLOSURE PERIOD

ISIS and ILL have each published a Data Policy[6][7] based on the joint work done during the 2nd work package[8] of the PaNData Europe project. Those Data Policies restrict the access to the data and metadata to the experimental team during a non-disclosure period, in order to allow sufficient time for our users to publish their results.

Ideally DOIs should be generated at the end of the experiment, in order to have only one reference during the whole life cycle of the Data. The DataCite process required some mandatory metadata when minting DOIs, amongst them the experiment title and experimental team. For some hot scientific topics, publishing the title of the experiment and the team involved could be considered as sensitive. This is problematic but not really a showstopper, ISIS and ILL have adopted 2 different strategies that are explained in the next chapter.

4.2 QUALITY ASSESSMENT AND FACILITY BUSINESS MODEL

DataCite requests that registrants certified the quality of the data for which DOIs are minted. Even with the best efforts it is difficult to achieve this for an experimental facility:

- The data are preliminary results, they are experimental raw data. The quality and value of data could only be understood after analysis, the analysis process could be long and necessitate in some cases a second experiment.
- ILL and ISIS provide the beam, the instruments and experts, but in most cases the users prepare the samples, sometimes with very experimental techniques. It is not possible for either the user or for the facility to assess the quality of the data. Nevertheless we can ensure that the best scientific practices have been used.

Discussion took places over several months, with good will and mutual understanding of the partners we were able to modify the initial text in order to take into account the model of the scientific facilities.

For instance sentences such as:

[6] http://www.isis.stfc.ac.uk/user-office/data-policy11204.html
[7] http://www.ill.eu/users/ill-data-policy
[8] http://wiki.pan-data.eu/imagesGHD/0/08/PaN-data-D2-1.pdf

> *"*ILL has to ensure that the data is valid according to ILL's own standards and according to the general rules of scientific good practice in the relevant field of research.*"*

has been completed with statement:

> *"It must be emphasised when ILL is acting under mandate of its partners that the roles as to the storing, persistency, evaluation and quality assurance can be shared among such partners."*

Contacts have been signed with DataCite at the end of 2012 between STFC/ISIS and the British Library and beginning of 2013 between ILL and INIST-CNRS.

5 WORKFLOW

STFC/ISIS and ILL have implemented a slightly different workflow.

5.1 STFC/ISIS WORKFLOW

For experiment proposals submitted before September 2013, ISIS is required by the data policy to keep all metadata regarding the experiment private for 3 years. This is required in order to allow the data creator to gain full value from their data. In order to issue a DOI immediately following the end of the experiment, very limited metadata could be provided to the DOI server. The mandatory metadata at DOI creation is:

Metadata Item	**Typical Value**
DOI	10.5286/ISIS.E.24079765
Creator	Smith, T; Jones, F
Publisher	STFC ISIS Facility
Year of publication	2012
Title	RB1210123

The Title is set as the experiment reference number. This is sufficient to identify the experiment, but does not release any sensitive metadata. When the data reaches the end of the embargo period, the metadata is updated with the more meaningful and searchable text title. Should the data owner wish, they may request that the facility publishes the full metadata before the end of the embargo period.

For experiment proposals submitted after September 2013, the proposal submission system terms and conditions have been modified to ask the users to provide a title which will be published immediately if their proposal is successful. This will allow us to mint DOIs with full metadata immediately after the experiment finishes.

5.2 ILL WORKFLOW

Three main types of proposals are offered by the ILL. The standard proposal, where one proposal could lead to one experiment that will take place on one or two instruments usually during the same cycle/year. The Long Term Proposal (LTP), where different experiments will

take place over 2 or 3 years. The Block Allocation Groups (BAG), where different proposals will lead to a single experiment.

In order to identify data with a single ID for their whole life, the ILL has decided to mint DOI right after the end of the first experiment of a proposal. The DOI metadata generated will correspond to the minimal set of mandatory metadata requested by DataCite as defined by the schema version 2.2[9]. Amongst the five properties only the title and the creator property, which corresponds to the individual name of the proposal team, could be sensitive. The ILL choice is to publish them, others alternatives could have been foreseen but they either introduce multiple ID or request an action from the users which could never happen.

6 IMPLEMENTATION

6.1 GENERAL PROCESS

There are three different steps involved in the implementation process:

1) Collect the DOI metadata
2) Mint the DOI
3) Provide the landing page

DataCite provides a RESTful API to register datasets to their Metadata Store.

The API provides three endpoints: *https://mds.datacite.org/doi/{doi}, https://mds.datacite.org/metadata/{doi} and https://mds.datacite.org/media/{doi}.*

The API requires that we authenticate using the HTTPBasic authentication mechanism for every request.

6.2 STFC/ISIS

The STFC/ISIS DOI generation process is driven by the data catalogue (ICAT). When an experiment has been catalogued, the ISIS DOI server is able to query ICAT using the ICAT API, extract the relevant metadata and send this to DataCite. The DOI is then stored in the ICAT. The landing pages are dynamically generated from the ICAT, ensuring the most up-to-date metadata is provided. The landing pages also provide access to the underlying dataset.

The code will be available under and open source license as part of the ICAT release from ICAT version 4.5

6.3 ILL

In order to facilitate the process of registering a dataset to DataCite, ILL has opted for developing its own open-source solution licensed under the MIT license. Hereafter we briefly describe this solution, for a more technical documentation please refer to the description and code available on the public repository[10].

The open-source solution provided by ILL is completely agnostic to the ILL; it is highly configurable and therefore there are no specific constraints that are unique to ILL. If another facility that has a contract with DataCite would like to use the solution provided by the ILL then they are free to do so. The solution is highly configurable therefore there are no specific ILL

[9] http://schema.datacite.org/meta/kernel-2.2/index.html
[10] https://github.com/ILLGrenoble/ILLDataCiteDOIBundle

constraints that are enforced. By all means if a third-party wishes to build on the work done by ILL, then due to the MIT license, they are permitted to do so. It is also worth noting that the current development state of the code is still a work in progress.

Many of the web application solutions provided at ILL are written in the Symfony2 PHP framework and as a result, ILL has opted to develop the solution in accordance to the Symfony2 architecture and best practices.

The solution works by using the DataCite API as its persistent layer for minting DOIs and storing metadata for the datasets. It provides a high-level and decoupled object architecture to work with the API. It conforms to the latest metadata scheme version of 2.2. If the metadata scheme changes then the code would be updated accordingly.

There are three main objectives achieved by the solution to allow a developer to:

- Retrieve metadata, register metadata, update metadata and to mark metadata as inactive;
- Mint a DOI to an associated metadata identifier and update the URL of a DOI (if necessary);
- Serialise the Metadata XML to and from different formats and perform validation for any metadata which is being registered to DataCite

The solution also provides a rich web interface, straight out of the box,with the following functionalities :

- Manually register a dataset using a web form (including client side and server side validation);
- View a dataset;
- Update a dataset;
- Provide a global overview of all registered datasets

There is also a pluggable authentication system, which enables a developer to use any authentication mechanism of his choice (LDAP, local database, CAS etc.) to limit who can administrate datasets or indeed register new datasets.

6.4 Registration of DOIs

When minting a DOI, the registrant has to provide a set of XML structured information that describes the data for citation and retrieval purposes.

The Metadata properties as requested by the DataCite Metadata schema 2.2 are mostly stored in the proposal system of the facilities. Therefore this step consists of extracting the information from the current database, creating a metadata object using the solution provided, validating against it and then uploading it DataCite.

When uploading metadata to the DataCite metadata store, there are *5* required properties:

- Identifier (with type attribute)
- Creator (with name identifier attributes)
- Title (with optional type attribute)
- Publisher
- Year of publication

If these properties are not satisfied, the metadata will not be valid. The DataCite Schema also specifies 12 optional fields, which are currently not used by ISIS or ILL.

Here is an example of registering basic (only uses required properties) metadata and assigning a DOI using the solution provided by the ILL.

First we need to create a metadata object to represent the metadata (we always take the metadata details from our proposal database, however, for this example, static include static text for clarity):

```
$metadata = new Metadata();

$metadata->setIdentifier("10.5291/ILL-DATA.6-05-589")

          ->setPublisher("Institut Laue-Langevin")

          ->setPublicationYear("2004")

          ->addCreator((new Creator)->setName("Angell, Caustin"))

          ->addCreator((new Creator)->setName("Schober, Helmut"))

          ->addCreator((new Creator)->setName("Scopigno, Tullio"))

          ->addCreator((new Creator)->setName("Yue, Yuanzheng"))

          ->addTitle((new Title)->setTitle("Study of vibrational dynamics in
hyperquenched glasses"));
```

The XML representation of this object for use with the Metadata API is the following:

```
<?xml version="1.0" encoding="UTF-8"?>

<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://schema.datacite.org/meta/kernel-
2.2/metadata.xsd" lastMetadataUpdate="2006-05-04" metadataVersionNumber="1">

  <identifier identifierType="DOI">10.5291/ILL-DATA.6-05-589</identifier>

  <creators>

    <creator>

      <creatorName>Angell, Caustin</creatorName>

    </creator>

    <creator>

      <creatorName>Schober, Helmut</creatorName>

    </creator>

    <creator>

      <creatorName>Scopigno, Tullio</creatorName>

    </creator>

    <creator>

      <creatorName>Yue, Yuanzheng</creatorName>

    </creator>

  </creators>

  <titles>

    <title>Study of vibrational dynamics in hyperquenched glasses</title>
```

```
  </titles>
  <publisher>Institut Laue-Langevin</publisher>
  <publicationYear>2004</publicationYear>
</resource>
```

To upload the metadata and mint a DOI, we can write the following code. When the metadata is uploaded to DataCite it is automatically converted into a XML representation conforming to the DataCite metadata schema 2.2. For the sake of simplicity, the handling of exceptions won't be included in this code example.

```
// get the metadata manager from the service container
$mdm = $container->get("ill_data_cite_doi.metadata_manager");

// check our metadata object is valid
if($mdm->isValid($metadata)) {
    // metadata is valid, send to datacite
    if($mdm->create($metadata)) {
      // the metadata has been registered to DataCite. Next step is to mint the DOI
      $doi = new DOI();
      $doi->setMetadata($metadata);
      $doi->setPath("10.5291/ILL-DATA.6-05-589");
      $doi->setDomain("doi.ill.fr");
      // get the DOI manager from the service container
      $dm = $container->get("ill_data_cite_doi.manager");
      if($dm->create($doi)) {
        // the DOI has been registered to DataCite
      }
    }
  }
}
```

Because we already have a representation of our metadata in our proposal database there is no need to store any other metadata into another database.

6.5 LANDING PAGE

The landing page for DOIs at ISIS is integrated with the TopCat client to generate the appropriate web page; this page provides a link to access the raw data, given appropriate registration and permissions. The landing page for the DOIs is also provided by a PHP Symfony solution at the ILL.

Hereafter, screenshots of the landing pages of the ISIS and ILL.

Figure 1 ISIS's DOI Landing page.

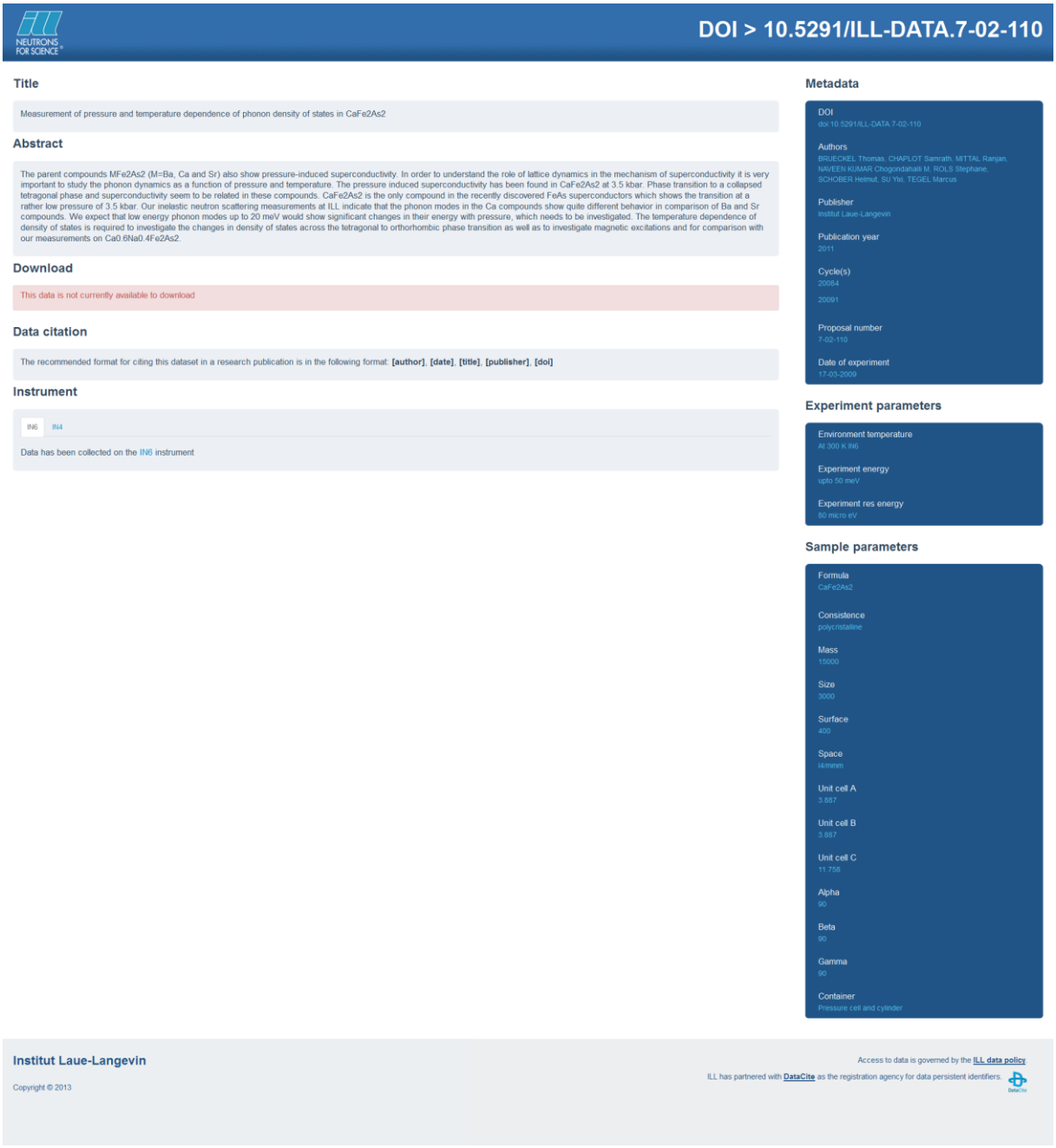


Figure 2 ILL's DOI landing page

7 CITATION

We recommend to the scientists to cite the DOI of the data when work based on the data is being published:

"Researchers must acknowledge the source of the data and cite its unique identifier and any publications linked to the same raw data." Extract form ISIS and ILL data policy.

The citation should reference:

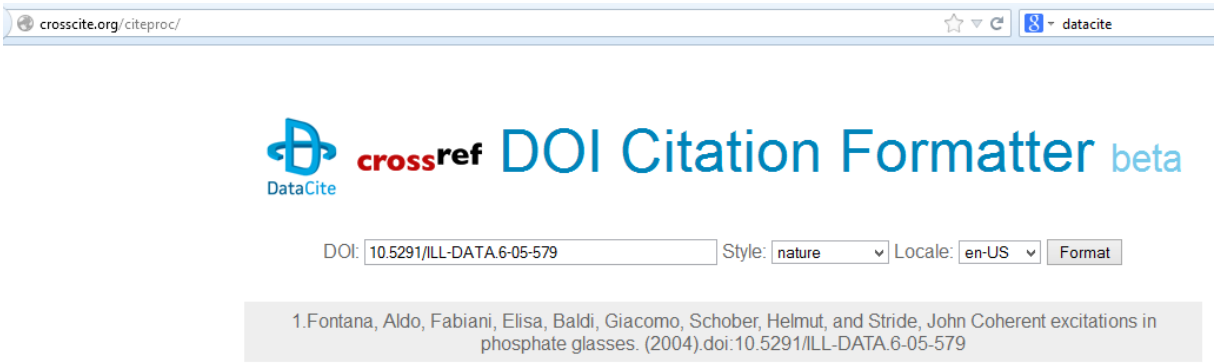
- The landing page;
- The proposal team;
- The title of the experiment;

- The publisher, in our case the facility;

Hereafter an example of a BibTeX citation:

```
@data{8e96ca6f-c3cc-494c-a32b-d3b4850336cc,
  doi = {10.5291/ILL-DATA.6-05-579},
  url = {http://dx.doi.org/10.5291/ILL-DATA.6-05-579},
  author = {Fontana, Aldo; Fabiani, Elisa; Baldi, Giacomo; Schober,
Helmut; Stride, John; },
  publisher = {Institut Laue-Langevin},
  title = {Coherent excitations in phosphate glasses},
  year = {2004}
}
```

The formatting of the citation is specific to each journal, we invite our users in order to properly cite the data to use the new citation formatting service[11] of DataCite and CrossRef.



8 FIRST RESULT

Despite the already 2000 registered DOIs only one has been cited in a publication, this is due to the necessary time needed for scientists to publish their results. Nevertheless this first result demonstrates that the whole system is functional. The following extract shows the use of the DOI in the paper. Note that this does not use the complete recommended citation format as the paper was prepared early in the process and the citation format had not been defined.

[11] http://crosscite.org/citeproc/

PHYSICAL REVIEW B **84**, 075219 (2011)

Thickness-dependent magnetic properties of oxygen-deficient EuO

M. Barbagallo,[1,*] T. Stollenwerk,[2] J. Kroha,[2] N.-J. Steinke,[1] N. D. M. Hine,[1,3] J. F. K. Cooper,[1] C. H. W. Barnes,[1,†] A. Ionescu,[1] P. M. D. S. Monteiro,[1] J.-Y. Kim,[1] K. R. A. Ziebeck,[1] C. J. Kinane,[4] R. M. Dalgliesh,[4] T. R. Charlton,[4] and S. Langridge[4]

[1]*Cavendish Laboratory, Physics Department, University of Cambridge, Cambridge CB3 0HE, United Kingdom*
[2]*Physikalisches Institut and Bethe Center for Theoretical Physics, Universität Bonn, D-53115 Bonn, Germany*
[3]*Thomas Young Centre, Department of Materials and Department of Physics, Imperial College London, Exhibition Road SW7 2AZ, United Kingdom*
[4]*ISIS, Harwell Science and Innovation Campus, STFC, Oxon OX11 0QX, United Kingdom*

(Received 24 May 2011; published 26 August 2011)

We have studied how the magnetic properties of oxygen-deficient EuO sputtered thin films vary as a function of thickness. The magnetic moment, measured by polarized neutron reflectometry, and the Curie temperature are found to decrease with reducing thickness. Our results indicate that these surface-induced effects are caused by the reduced number of nearest neighbors, band bending, and the partial depopulation of the 4*f* states of Eu.

DOI: 10.1103/PhysRevB.84.075219 PACS number(s): 75.50.Pp, 75.70.Ak, 75.30.Et, 75.60.Ej

I. INTRODUCTION

Electron-doped EuO is a semiconductor which undergoes a simultaneous ferromagnetic and insulating-conducting phase

II. EXPERIMENTAL METHODS

Thin films of EuO_{1-x} with $x = 4\%$ were deposited by cosputtering of Eu_2O_3 and Eu on Si substrates with a Pt

°°°

[18]doi: 10.5286/ISIS.E.24066298 (2009).
[19]T. S. Santos, J. S. Moodera, K. V. Raman, E. Negusse, J. Holroyd

DOI referring to the data collected at ISIS in 2009

[26]See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.84.075219 for a best-fit of the PNR data with

9 CONCLUSION

ISIS and ILL have now fully implemented the persistent Identification of the datasets using DOIs, contracts with DataCite have been signed and the tools for generating and minting the DOI have been deployed. ESRF has planned to proceed in September 2013.

Due to the time necessary for publication, it will now take 2 to 3 years to get important enough citation results to really assess the impact of this work in term of cultural shift. Nevertheless in term of datasets identification in view of long term preservation the DOI represent a step forward, they have already demonstrated with publication and article identification that they fulfil the requirements of long term identification.

The use of DOIs gives an important step in the process of encouraging data publication and citation, which is underpinned by persistence data archiving and preservation. A potential next step would be the development of a systematic publication route for faciltiies, a form of “data journal” providing the persistent record of experiments from the facility, available for citation (and thus credit) and reuse.