

Using DataCite DOIs for ISIS Neutron Source Data

Michael Wilson¹, Brian Matthews¹, Sri Nagella¹, Antony Wilson¹

¹*Scientific Computing Department
STFC Rutherford Appleton Laboratory
Chilton, Didcot, Oxon, UK*

"We must give taxpayers more bang for their buck. Open access to ... data is an important means of achieving this." demanded Máire Geoghegan-Quinn, European commissioner for research, innovation and science, in July 2012. The EC and other funding agencies are putting pressure on large facility providers to ensure the best return is achieved on the investment made in them. Open access to data is being promoted as one way for facilities to achieve this [1, 2, 3]. Initial economic evaluation of a data service [7] reveals that for every pound invested in data infrastructure, the service returns £5.40 in net economic value to users and other stakeholders.

When data is made available users need to know: what it is, where it is and what its quality is, in order to find it, and decide if they can trust it enough to use it for their own purposes. The international DataCite consortium offers a service to mint, sustain, and discover digital object identifiers (DOI) which can be associated with data as the foundation for providing open access to facility data. DOIs are already used to cite journal articles, so they are already trusted as a mechanism by the community of scientific users. The DOI mechanism is already supported by bibliometric services which can describe how frequently datasets are cited, and commercial providers have started to implement citation indexes [4] to provide a basis for the judgement of trust in data providers and data sets.

Raw data collected from instruments on the ISIS facility are indexed on the ICAT data catalogue, and have the minimum metadata mandated by DataCite registered in the DataCite metadata registry through the provided API [5] so that a DOI is allocated to each investigation. The DOI allows ISIS users to cite the raw data which they have analysed to produce results that they publish in articles, so that other users can validate their analyses. The metadata registry also supports open searching which increases the availability of the data to potential users who have not read publications that cite the data.

Despite the implementation and operation of the system supporting open access to ISIS data, several important questions remain unanswered. Investigations are allocated beam time on facilities so they are an easy unit for facilities to allocate DOI too also, but should sustainable identifiers also be allocated at a finer granularity - DOIs or others? Should the minimum metadata be registered in the registry when data are collected, even though this is before publication of results or the expiration of the open access embargo period? Should extra metadata be published in the registry to increase the visibility of the data? How open should records of data access be made?

References

1. European Commission, Towards better access to scientific information: Boosting the benefits of public investments in research, European Commission: Brussels, 17 July 2012.
2. OECD, Principles and Guidelines for Access to Research Data from Public Funding, OECD Publications: Paris, 2007.
3. Royal Society, Science as an Open Enterprise, Royal Society: London, June 2012.
4. Thomson Reuters, Repository evaluation, selection and coverage policies for the data citation index within Thomson Reuters Web of Knowledge, Thomson Reuters: Philadelphia, London, Singapore, Tokyo, July 2012.
5. DataCite API, available at <https://www.datacite.org.uk/api/index>
6. DataCite Online Data Search Engine, available at <http://search.datacite.org/>
7. ESRC, Economic evaluation of the Economic and Social Data Service (ESDS), July 2012, available at: <http://www.esds.ac.uk/news/newsdetail.asp?id=3291>

Email corresponding author: michael.wilson@stfc.ac.uk

Preference: Oral Presentation