Data users, data producers….
….how can we help?

Elizabeth Newbold

UK-EOF Data Solutions Workshop, 25 June 2009

The British Library:
'This is the life blood of research and innovation'

Science and Innovation Investment Framework 2004-2014, H.M. Treasury (2004)
Information infrastructure
2.23 *The growing UK research base must have ready and efficient access to information of all kinds – such as experimental data sets, journals, theses, conference proceedings and patents.* ***This is the life blood of research and innovation.***

National library of the UK.
Serves researchers, business, libraries, education & the general public

Collection includes over 2m sound recordings, 5m reports, theses and conference papers, the world's largest patents collection (c.50m)

Collection fills over 600km of shelving and grows at 11km per year
16 Tb of digital material through voluntary deposit

Business and IP Centre:
Providing inspiration, and enabling protection of creative capital and business development

Helping people advance knowledge to enrich lives

One of the largest document supply service in the world. Secure e-delivery and 'just in time' digitisation enables desktop delivery within 2 hours

Generates value to the UK economy each year of 4.4 times public funding

GIA Funding 08/09:
£94.8m operational,
£12m capital

Other funding secured 07/08:
c.£33m

3 main sites in London and Yorkshire. Circa 2,000 staff

Role of the British Library in Science, Technology and Medicine……..

- Long history of collecting scientific and technical literature

- Serves business & industry, researchers, academics and students through dedicated reading rooms in London and our document supply services

- Indexing the UK input into Medline/PubMed

- Lead Partner in UK PubMed Central (www.ukpmc.ac.uk)

- Contributor to WorldWideScience.org

But times are changing…..what about data?

Area of increasing importance across all scientific disciplines is access to data and datasets.

- Increasing multi-disciplinary nature of research alongside technical advances
- Growing emphasis from funders and policy makers on maximising the value of data collected using public money
- Disciplinary disparities between collection, storage, preservation, management, access, confidentiality, use, reuse of data
- A widening gap in the scientific record between published research and the data that underlies it

Exploring ways to work with data centres focusing on curated datasets:

- Discoverability and linking
- Harvesting metadata
- Catalogues-access/ registries
- Ingest and preservation

Persistent identifiers: A key component of the solution

In order to provide global access to data sets and their metadata through existing catalogues we must have a method to persistently identify them!

This enables:

- Citation
- Increased visibility
- Easier re-use and verification
- Enhanced impact tracking for data producers (Citation Index)
- Support for the Brussels declaration on STM publishing
- Reduced duplication of research data
- Stimulation of new research

How can we do this?

A role for Digital Object Identifiers (DOIs)

The DOI system offers an easy way to connect the article with the underlying data

Several organisations have started to assign DOIs to datasets

- IUCR, ICPSR, OECD through CrossRef
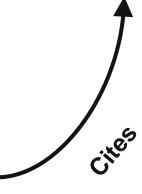- Pangea, Mare, and others through TIB

Dataset
G.Yancheva, N. R. Nowaczyk et al (2007) Rock magnetism and X-ray flourescence spectrometry analyses on sediment cores of the Lake Huguang Maar, Southeast China, PANGAEA
doi:10.1594/PANGAEA.587840

Article
G. Ycheva, N. R. Nowaczyk et al (2007) Influence of the intertropical convergence zone on the East Asian monsoon
Nature 445, 74-77
doi:10.1038/nature05431

Cites

What next?

Organisations with the role of a national science library are working together to establish a European and global infrastructure to

- Assign digital identifiers (DOIs) to datasets
- Manage the identifiers over the long term
- Establish and share best practice

Memorandum of Understanding: Paris, March 2, 2009

Recognizing the importance of research datasets as the foundation of knowledge and sharing a common commitment to promote and establish persistent access to such datasets, we, the signed parties, hereby express our interest to work together to promote global access to research data.

Our long term vision is to <u>support researchers</u> by providing methods for them to locate, identify, and cite research datasets with confidence.

[*MoU signatories: Technische Informationsbibliothek (TIB), Germany; Library or the ETH Zürich, Switzerland; L'Institut de l'Information Scientifique et Technique (INIST), France; Library of TU Delft, The Netherlands; Technical Information Center of Denmark; The British Library; Canadian Institute for Scientific Information (CISTI)*]

Thank you.

This is the start of a long and open dialogue and we welcome the opportunity to work with you to bring together the different parts of the picture.

British Library: www.bl.uk/science

Email: science@bl.uk

Research Information Network: www.rin.ac.uk

Twitter: research_inform

Some further information

- Press release Memorandum of Understanding http://www.icsti.org/documents/PressReleaseMarch2009-JointDOIforData.pdf
- Toby Green, We need publishing standards for datasets and data tables, OECD Publishing White Paper, http://dx.doi.org/10.1787/603233448430
- The DOI System, International DOI Foundation: www.doi.org
- Publication and Citation of Scientific Primary Data at WDC Climate. Lautenschlage, M et al preesented at MPG eScience Seminar 27- 28 March 2008: http://colab.mpdl.mpg.de/mediawiki/images/3/30/ESci08_Sem_1_Primary_data_registration_Lautenschlager.pdf