

COMET sighted in the east

JISC RDTF PROJECT

The **COMET**⁶ (Cambridge Open METadata) project aims to release over 2.2 million catalogue records under an Open Data Commons License.

This collaboration between Cambridge University Library (CUL), and the Centre for Applied Research in Educational Technologies (CARET) at the University of Cambridge builds on the previous release of 132,130 records from CUL as part of the JISC funded Open Bibliography Project⁷.



Cambridge University Library

The data to be released by COMET includes metadata derived from external sources, such as OCLC WorldCat. Such a step requires careful consideration of what intellectual property, contractual and licensing issues arise when releasing metadata records under an open license. In addition to the guidance already available from the JISC Guide to Open Bibliographic Data⁸ and JISC Legal⁹, the project has posted detailed information about the 'ownership' of MARC21 records in their catalogue¹⁰.

COMET will publish the records as Linked Data¹¹ enabling the library records to be linked to other data sources on the Web. Specifically COMET is working with staff at OCLC to link to the FAST and VIAF services dealing with subject terms and name authorities respectively. Project Manager Ed Chamberlain says that while Linked Data may have a high entry-bar it is "a great tool for sharing library data beyond the library community".

Footnotes & further reading

1. http://ie-repository.jisc.ac.uk/475/1/JISC%26RLUK_VISION_FINAL.pdf
2. <http://thinkquarterly.co.uk>
3. <http://www.guardian.co.uk/news/datablog/2010/jan/21/timbernerslee-government-data>
4. <http://www.bl.uk/bibliographic/datafree.html>
5. <http://jerome.library.lincoln.ac.uk>
6. <http://cul-comet.blogspot.com>
7. <http://openbiblio.net/2010/10/05/jisc-openbibliography-cul-data-release>
8. <http://obd.jisc.ac.uk/rights-and-licensing>
9. <http://www.jisclegal.ac.uk/Projects/TransferandUseofBibliographicRecords.aspx>
10. <http://cul-comet.blogspot.com/p/ownership-of-marc-21-records.html>
11. <http://linkeddata.org>
12. <http://www.bl.uk/bibliographic/datafree.html>
13. <http://openbiblio.net/p/bibliographica>

Do you have news items you'd like us to highlight in the newsletter or through other means? Contact us at: rdtf@sero.co.uk.

Open Bibliography Workshop

On May 6th at the London Knowledge Lab, the JISC Open Bibliography project held a workshop on open bibliographic data. Discussion and debate during the day ranged over both technical and strategic issues, including:

- » *The need for stable identifiers to refer to data (such as the BNB data released by the British Library)*¹²
- » *Mechanisms for keeping data sets up to date (such as the incremental data releases from PubMed)*
- » *Approaches to exposing RDF data in a simpler JSON format*
- » *The cost of sustaining the current data sets hosted in Bibliographica*¹³.

In the afternoon the workshop looked at the latest data release obtained by the project (over 19million records from PubMed), as well as discussing the possibilities and practicalities of machine-based disambiguation and 'sameness' between records.

A full report on the workshop, as well as further information about Open Bibliography is available from <http://openbiblio.net>.

Global news flashes

The **Amsterdam Museum** collection is now available as Linked Open Data. The data was retrieved from the Museum's Adlib collection database and converted to be RDF compliant with the Europeana Data Model. The result is more than 5 million RDF triples describing all of the more than 70,000 cultural heritage objects.

Beginning March 2011, the **University of Florida's** Smathers Libraries implemented a policy to include a Creative Commons license in all of its original cataloguing records. The records are considered to be in the public domain with unrestricted downstream use for any purpose.

CrossRef, the publishers' consortium, has made the metadata for 46 million DOIs (Digital Object Identifiers - uniquely identifying such as scholarly journal articles) available as Linked Open Data. Consequently wherever DOIs are available in the scholarly publishing ecosystem, it will be possible to use the Web to retrieve metadata associated with that document.



RDTF takes on a new identity ...

In May 2011 the work of the JISC and RLUK Resource Discovery Taskforce (RDTF) implementation plan entered a new phase of activity with a new identity: '**Discovery: A metadata ecology for UK education and research.**'

The **Discovery** website <http://discovery.ac.uk> will be launched on May 26th to coincide with our next community advocacy event at the Wellcome Trust, London.

open data open doors

Issue 2 - May 2011



Our aim is that **Discovery** will help to mobilise and energise the community, building on the work of RDTF¹ to further engage stakeholders in creating a critical mass of open and reusable data, and more importantly in exploring and demonstrating what open data makes possible through real-world exemplars and case studies.

It is only through demonstrating these benefits, and working collectively across the higher education and cultural heritage sectors, that we will create enough momentum to realise the Vision - **where our richest collections are discoverable thanks to our collective fostering of an ecosystem of reusable metadata.**

Open data - the missing link?

Ken Chad positions *Discovery* in the context of global and national thinking

In March 2011 the first issue of Google's *Think Quarterly*² online magazine was dedicated to data. Nigel Shadbolt of the University of Southampton writes that one of the key responses to the 21st century demand for information is open data. The data.gov.uk website and the influence of Shadbolt alongside Sir Tim Berners-Lee has positioned UK government as one of the leaders in open data³.



Discovery Chair, Prof. David Baker, emphasises that by opening up more data for reuse “we can better serve UK educators and researchers to excel in their work by increasing access to, and visibility of, relevant content”.

If we are to achieve the ambition of the *Discovery* initiative for a sustainable ‘metadata ecology’, two broad issues need to be addressed. The first is around making a clear business case. Key figures like Shadbolt and Berners-Lee have done much to clarify and advocate the broader business case especially for government data. However more remains to be done to help heads of libraries, museums and archives articulate the particular business case for their organisations - as *Discovery* is undertaking to do.

Secondly, a commitment to licensing open metadata will be vital. It is encouraging that this is central to a number of current projects in libraries, museums and archives with the British Library⁴ amongst those leading the way. At the same time *Discovery* is providing case studies and tools such as the Open Bibliographic Data Guide to support managers, practitioners and developers.

Read the full article by Ken Chad on: <http://discovery.ac.uk>.

A recommendation for open metadata release and licensing

As part of their commitment to *Discovery: A metadata ecology for UK education and research*, the Resource Discovery Task Force partners are inviting stakeholders across the UK libraries, archives and museums (LAM) community to join them in adopting a set of principles to enhance the impact of our knowledge resources for the furtherance of scholarship and innovation.

The RDTF will launch a ‘call to arms’ in London on 26th May based on three core statements:

- » **We recognise** the importance of promoting a clear endorsement of open metadata with practical guidance about licensing.
- » **We recommend** that institutions and agencies should proceed on the presumption that their metadata is by default made freely available for use and reuse, unless explicitly precluded by third party rights or licences.
- » **We strongly advocate** that all metadata releases require licensing, for which institutions and agencies should adopt a standard open licensing framework that is suited to their purposes.

The ‘call to arms’ is intended to be of value to institutions and agencies in the LAM sectors, increasing their confidence and motivation to enact these principles, to enquire further or to put them on their agenda.



Over the coming months, it is hoped that other institutions, large and small, will sign up to the statement, indicating that they support the principles and that they are committed to enacting them in appropriate areas of their work. Others will have their interest triggered and will be encouraged to seek clarification or pass it on to the right person in their organization.

To enable that process, supporting material will be available at: <http://discovery.ac.uk>.

“Anything that helps people to make more meaningful use of resources is a good thing ...”

Feedback from the community

Veronica Adamson and Jane Plenderleith report on recent interviews

Since March, we’ve carried out a series of interviews with leaders and managers in the library, archive and museum (LAM) community about what open data means for their users and communities. Discussions focused on benefits, issues and challenges for institutions, collections and users in this space. Some interesting and thought-provoking views have emerged, providing much food for thought on the development of the *Discovery* vision. Here are some key points:

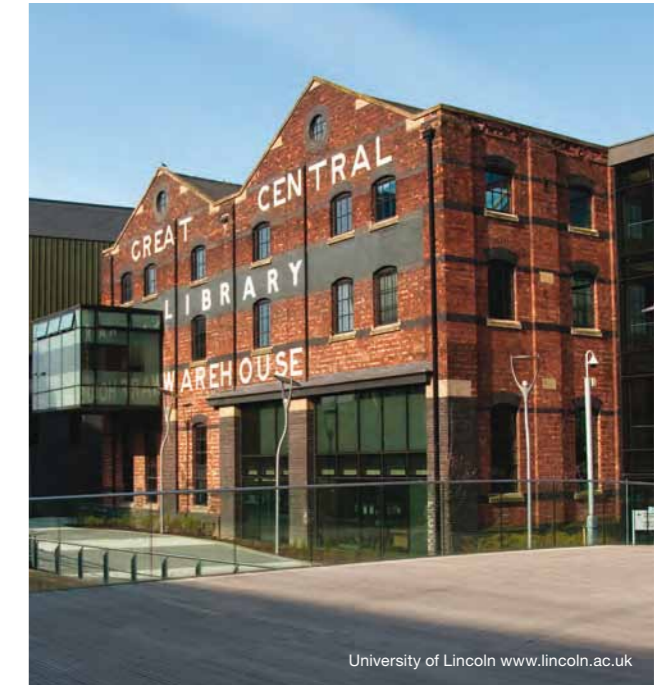
- » **Supporting open data** - the LAM community is keen that resources are available to a wide community of users and contribute as much as possible to the furthering of knowledge.
- » **Simplifying access** - there is strong support for systems which help users easily to discover resources and avoid the confusion caused by a multiplicity of disparate datasets.
- » **Communication** - to these ends, LAM professionals need accessible language, and clear evidence of the benefit of open data aggregation, aligned with institutional priorities.
- » **Local examples** - networks of libraries, museums and archives are already sharing data and developing local solutions to metadata challenges relating to standards, purpose and nomenclature.
- » **High quality aggregation** - we need to move beyond small-scale initiatives providing partial answers, which then sit on websites gathering digital dust.
- » **Special Collections as the archives of the future** - as more and more published material is available digitally, the role of the library is as custodian of unique collections, so data relating to these collections is an invaluable national resource.

Our thanks to all those who have been involved in this process so far. If you’d like to share your thoughts, aspirations, plans or reservations about these matters in our forthcoming round of interviews, please get in touch via rdtf@sero.co.uk.

Jerome at the University of Lincoln

JISC RDTF PROJECT

The stated purpose of the Jerome project is an ambitious one: “to develop a sustainable, institutional service for open bibliographic metadata, complemented with well documented APIs and an intelligent personalised interface for library users”. Not much there!



The project started life as an internal ‘un-project’ which aimed to deliver “an amazing way to interact” with the University of Lincoln’s library services in the wider context of the University’s user services and in the face of limited resources.

The funding as a JISC RDTF project has enabled the team to make much swifter progress with their aspirations and to document achievements so that they can share their expertise and developments with the wider community.

The key outputs for this current, JISC-funded phase of Jerome are:

- » A developers’ toolkit which will include APIs, web services, a technical ‘cook book’, user journeys and other documentation which will allow other developers to build and implement their own search tools.
- » Bibliographic records of books, journals and e-prints released as open data.
- » A user-controlled, personalised search interface.

The project has already gone live with the first implementation of a Jerome search interface⁵ at the end of March.