COMET sighted in the east

The COMET (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.

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 be a great tool for sharing library data beyond the library community, the cost of sustaining the current data sets hosted by OCLC may prove prohibitive. Project Manager Ed Chamberlain says that while Linked Data may prove prohibitive. Project Manager Ed Chamberlain says that while Linked Data may prove prohibitive.

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

In the afternoon the workshop looked at the latest data release obtained by the project (over 19 million 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.

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 Library9)
» 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 Bibliographica11.

In the afternoon the workshop looked at the latest data release obtained by the project (over 19 million 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 - 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 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 data.

Discovery Chair, Prof. David Baker, emphasises that by opening up more data for reuse "we can better serve the British Library amongst those leading the way. At the heart of current projects in libraries, museums and archives with more remains to be done to help heads of libraries, business case especially for government data. However Lee have done much to clarify and advocate the broader business case for 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 data.

We recommend 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 RDFT 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 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.

A recommendation for open metadata release and licensing

Discovery: A metadata ecology for UK education and research

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 interface at the end of March.