discovery

A metadata ecology for UK education & research

open data open doors

Issue 4 - October 2011

Discodev - Dancing with data
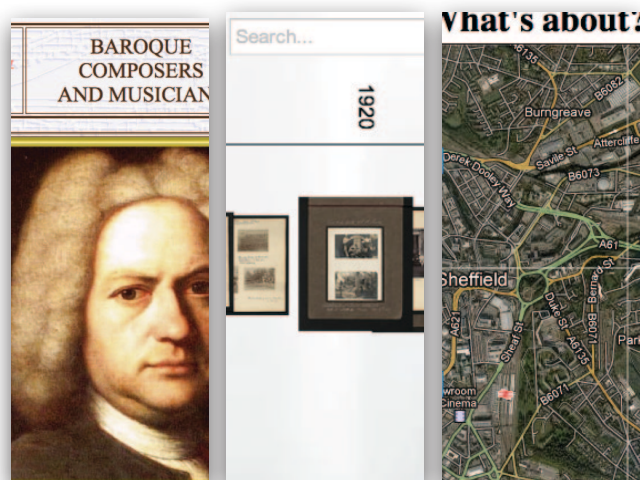
When the Discovery Developer Competition was launched in July, the Twitter tag #discodev encapsulated the challenge – not only to have fun, but also to enable the data to 'dance', to express itself in a variety of ways, even to find new partners (there was a special prize for combining data sources).

Developers were asked to create useful applications from one or more of 10 openly licensed UK datasets from libraries, archives and museums, selected from releases made during the life of the Discovery initiative. The 11 entries were impressive in terms of variety, imagination, technical execution and the user experiences that they captured.

Andy McGregor of JISC commented, 'We were enthused by the range, the imagination and the practical potential of every one of the entries. These exemplars offer tangible encouragement to institutions considering the business case for opening up and licensing their library, archival and museum metadata. Furthermore, whilst the judging criteria focused on the user experience, we also recognised the potential value for institutions of entries falling outside the principal focus on data processing and associated visualisation.'

The judges were unable to separate three joint winners:

» *Discobro* – based on the Archives Hub and MusicNet datasets, by Mathieu d'Aquin from the Open University KMI team

» *Timeline* – based on the National Archives dataset, by Alex Parker, a Southampton computer science undergraduate

» *What's About* – based on the British National Bibliography and English Heritage Places datasets, by Jason Cooper from the Loughborough library systems team

You can explore all the entries at:
http://discovery.ac.uk/developers/competition

RDTF Project outcomes

The eight JISC Resource Discovery projects have now completed. These projects were awarded to institutions and consortia for short, sharp explorations of how the Discovery technical and licensing principles might open up metadata and the underlying collections. The practical lessons learned will shortly be published on the Discovery website in a series of themed Guides (including Licensing and Business Case) and Project Reviews.

The previous newsletter reported on the two library projects, Jerome from Lincoln and Comet from Cambridge. Now we hear from one of two archival projects with contrasting but complementary responses to the same challenge – how collections can economically and effectively enhance their indexing and therefore become more discoverable in the wider world of scholarly resources.

SALDA (Sussex Archive Linked Data Application) aimed to extract metadata records for the Mass Observation Archive (MOA) from the university's archival system (CALM) and convert them into Linked Data. This data is now publicly available under the ODC-PDDL license and will be fed into the unified discovery service required by The Keep, a cross-sector aggregation project.

Sussex had a problem. The source MOA descriptions have no separate fields for date or description and don't use search access points – not a good starting point for making the data more discoverable, and especially not for generating linked data. Therefore the project had to find an efficient and affordable way to augment the content with access points.

On the plus side, like most archives, there were implicit connections in the data, potentially a basis for generating the explicit associations needed for any indexing and for linked data. The team identified from the data 28 names in authorised form using National Register of Archives rules, and over 100 keywords covering subjects from air raids to sex, places and organisations, events and wider concepts like class, family, education and death.

Given these 'hooks', the solution was not a major cataloguing and indexing exercise, but rather involved handing the MOA data and the lists of terms to a skilled developer (Pete Johnson of Eduserv) who was able to build software to join them up and therefore create Linked Data. Pete based this on the Archives Hub LOCAH framework, so there is the possibility of reuse for other archives who also want to cut the journey time and the cost of going from no visibility to a highly discoverable well indexed collection.

The next issue will cover the cataloguing techniques developed by AIM25 consortium's Open Metadata Pathway project to the same end – creating well-indexed and discoverable descriptions without increasing the burden or cost to the archive.

Announcement puts Europeana and Discovery in step

In late September 2011, Europeana made a potentially game-changing announcement concerning the licensing applying to its aggregation service. The Foundation has officially adopted a Data Exchange Agreement using the Creative Commons Universal Public Domain Dedication Licence.

This explicitly drops previous 'Non-Commercial use only' (NC) and 'Share Alike' (SA) restrictions on the basis that they preclude too many channels for effective reuse and discovery – exactly as put forward in our Discovery principles. The agreement is supported by Usage Guidelines, which set out the responsibilities that people should be aware of in reusing the data.

Discovery particularly endorses two of the associated Europeana metadata principles:

» Data providers are not required to provide an aggregator with complete metadata for digital objects
» Providing metadata relating to some works in a collection does not create the obligation to provide metadata about complete collections

These principles are clearly relevant to specialist collections that are considering open licensing in any context.

Open and Linked Metadata Initiatives - what the rest of the world is doing

Ken Chad of the Discovery Communications team has recently undertaken a horizon scan of some of the key open and linked metadata initiatives that are taking place outside the UK.

Open data remains a novelty, and in many instances a luxury. In the LAM domain, developments have been supported through national libraries and related bodies which because of their long term vision and public mission can forego an immediate return on investment.

In an international context, initiatives are typically moving from open data to linked data, making the output more useful in the wider linked or semantic web world beyond the LAM domain. Whilst open data is not necessarily linked data (and visa versa) the two terms increasingly inhabit the same space. Their shared purpose is to encourage reuse and add value.

In general, the extent to which the LAM domain is realising the potential of reusing and adding value to data that is made available remains relatively low. However, initiatives such as UK Discovery are helping to generate interest in using open data and kickstart enterprise in adding value. Making sense of open data in academic and commercial contexts remains a key challenge across the globe.

Discovery, the story so far...

Supported by JISC and RLUK, Discovery has been working since January 2011 on realising the vision of open metadata for higher education in the UK. Achievements to date include:

» Well-attended cross-sector community events to share ideas
» A communications infrastructure to engage and inform the community
» Working with partners and experts in the sector to establish technical requirements and understand issues in different domains
» Working with partners to identify opportunities, example activity and potential business benefits
» Launch of the Open Bibliographic Data Guide website featuring use cases to help libraries deal with practical issues such as how to license data
» Launch of the Discovery Open Metadata Principles advocating confident and appropriate usage of standard open licences
» A Discovery and DevCSI Developers competition illustrating potential applications of open data sets, with encouraging results
» Guidelines and open data sets from eight JISC-funded projects evenly spread across libraries, archives and museums working on standards and licensing to enable reuse
» Service development including open linked data from Suncat and Go Geo and a Copac collections management prototype

This is just the start - the journey of Discovery continues. Find out about recent developments and sign up to be part of the Discovery experience at http://discovery.ac.uk

Podcast: The British Museum's Collections Online

The British Museum is one of the world's great museums, with rich collections encompassing cultures from around the globe. The museum building in London is freely accessible to the public, and now the institution's online collections are also being opened up for use and reuse by a new type of visitor.

Paul Miller of Cloud of Data (and a member of the Discovery Communications team) talks to Dominic Oldman, Deputy Head of Information Systems at the British Museum about recent efforts to share almost two million objects from the Collections Online database as Linked Data, and to licence the data as permissively as possible. The conversation focuses on examples of how resource providers and researchers might make use of the data.

Discovery - where next?

The JISC-funded Discovery programme was launched in May 2011 to create 'a metadata ecology' that will support better access to vital collections data in libraries, archives and museums and facilitate new services for UK education and research.

Initial work has focused on engaging with the community, advocating open data and making headway with technical and licensing issues. You can see a full round-up of progress to date in the Discovery News Column.

Discovery will continue to the end of 2012 and is now entering its second phase, focusing on aggregation and building services with a particular emphasis on information, advice and training to institutions to help with practical implementation. Funded projects will support collaborative and cross-sector working with the aim of releasing more sustainable open metadata and developing innovative tools and services. Discovery will also work with national stakeholders to develop compelling exemplars on high-profile topics of significant research and learning interest to illustrate the potential benefits of an open metadata ecology.

The overarching Management Framework and Communications activity will support this work by:

» Providing case studies, principles, guidelines and training to help embed learning and build confidence, addressing technical and licensing concerns

» Resourcing developers to drive innovation and overcome technical barriers

» Entering dialogue with aggregators to identify efficiencies in dataflow and aggregation that can be achieved by key shared UK bibliographic data services

» Securing buy-in to the Discovery Open Metadata Principles and supporting implementation, with particular emphasis on making available hidden or strategically important collections under open licences

» Collaborating with JISC library and archives services and other key aggregations to identify opportunities for establishing open licences

» Establishing dialogue with commercial providers

» Engaging with related initiatives internationally as well as in the UK to ensure joined up thinking and solutions

For more information please visit the Discovery website or read The Discovery blog.

Your feedback and ideas for this next phase of the programme would be very welcome, as well as your support for the Discovery Open Metadata Principles. To get in touch, please contact Andy McGregor at JISC, Joy Palmer at Mimas or Liz Wallis at Sero (contact details are at http://discovery.ac.uk/contact).

discovery A metadata ecology for UK education & research

New signatories to the Discovery Open Metadata Principles

Since the Principles were launched in May 2011, over 30 people representing a range of significant stakeholders and collections have signed up, leading to a regular stream of enquiries.

Recent signatories include:

Helen Cooper,
Repository Manager & University Archivist,
University of Central Lancashire

Ian Piper,
Creator of the Open Vocabularies Service,
Tellura Information Services

Matthew Woollard
Director,
UK Data Archive and Economic and Social Data Service.