



# discovery

A metadata ecology for  
UK education & research

## open data open doors

Issue 3 - July 2011

### In this edition:

*Discovery launches a competition for open source applications*

*Information about television, radio and film programming gains a unified search environment*

*RDFT profiles of two projects using linked data*

*Report from the 2011 Discovery Conference at the Wellcome Trust, London*

*Focus on Archives and Museums*

*Discovery and Special Collections: RLUK's Mike Mertens discusses Unique and Distinctive Collections*

*Our monthly news digest*

To endorse the Discovery Open Metadata Principles visit:

<http://discovery.ac.uk/businesscase/principles>

## Worldwide Developer Competition Underway

UK Discovery and the Developer Community Supporting Innovation (DevCSI) are running a global Developer Competition to build open source software applications, using at least one of our 10 open data sources collected from libraries, museums and archives.

The competition started on Monday 4 July – Independence Day, a good day for liberating data – and closes on Monday 1 August. It's open to anyone, anywhere and the rules are very simple:

1. Build a software application that uses at least one of our data sources
2. Optionally combine our data with any other data
3. The code should be Open Source - available for others to use
4. Ask any questions at <http://getthedata.org> or directly to [m.mahey@ukoln.ac.uk](mailto:m.mahey@ukoln.ac.uk) (*Mahendra Mahey, DevCSI Project Manager*)
5. Entries must be submitted by **23:59** on Monday 1st August 2011

We've gathered data from 10 sources in libraries and archives and museums, all licensed to reuse freely and to aggregate with any other data. The data describes things ranging from books and electronic journals to archival collections and museum artefacts. Some of the sources are from famous collections, some are quite technical, some are very descriptive. To help developers, we've created a directory to tell you more about each resource, how to access the data (APIs etc) and what format it's in.

There are 13 prizes including the best entry for each dataset and the top example of 'Data Munging' (that's consolidating, aggregating, matching or de-duplicating). The overall winner will get an EEE Pad Transformer, the runner up will get a £200 Amazon voucher.

Read the full rules the competition small print here:

<http://discovery.ac.uk/developers/competition/#enter>

Tweet [#discodv](#) and [#devcsi](#)

An astronaut in a white spacesuit with an American flag patch on the shoulder is standing on the moon's surface. The background is the dark, cratered lunar landscape under a black sky. A thin orange line curves from the top left corner towards the title.

# The Voyage of Discovery

## Lively conference says 'Just do it!'

Around 90 library, archive and museum managers, information specialists and other professionals attended the Discovery Conference at the Wellcome Trust in London on 26 May.

Chaired by Professor David Baker, the event was about making the case for opening collections, brokering collaboration opportunities, and identifying support requirements for moving the Discovery initiative forward. It also marked the launch of Discovery's Open Metadata Principles. Professor Baker's introduction outlined achievements so far, and challenged delegates to inspire further developments in open data.

The morning focused on the demand side of Discovery - user expectations in teaching, learning and research. In a video address, Stuart Lee emphasised the value of open data for researchers in the arts and humanities. Peter Murray-Rust sparked controversy with a critical account of what science and technology researchers need, and how far the UK open data policy lags behind other countries. Drew Whitworth focused on the opportunities for teaching and learning in formal and informal settings.

The afternoon, chaired by Nick Poole, addressed the supply side - opportunities to expand data access and visibility. Veronica Adamson identified examples from Special Collections, and Peter Burnhill outlined the potential of aggregation as a tactic in support of the 'Art of the Possible'.

Presentations inspired lively debate with contributions from many different standpoints, while a 'tweet wall' highlighted key points, challenges and more subversive comments from the community. A key theme was 'just do it!' - calling for practical examples of open data in action to develop momentum and strengthen the business case.

Delegates left the event inspired with optimism, and encouraged to make open data happen. While there is still a long way to go on the Voyage of Discovery, the vision is clarified, the map is drawn, and the course is being set.

For a discursive summary of the keynotes, whole conference debates and twitter feed visit the Discovery blog at <http://blog.discovery.ac.uk>

A photograph of a museum interior. In the foreground, a white sculpture of a seated figure is on a pedestal. In the background, another sculpture is visible. The lighting is warm and focused on the art.

## Focus on Archives & Museums

From its origins in the Resource Discovery Taskforce (RDTF), Discovery has most readily been associated with the world of libraries — and Higher Education's libraries in particular. This early focus should perhaps be characterised as an accident of history rather than an explicit intention to narrow scope, and non-library organisations such as the Collections Trust have long been involved.

In June, representatives of the museum and archival communities came together with members of the Discovery team for a day of discussion in search of common ground. At a high level, many of the familiar themes resonated strongly; licensing is an often thorny issue, funding is tight, and there are a plethora of worthwhile initiatives competing for limited attention.

For the majority of institutions, the faith-based belief in the value of "open" that drove so many early adopters is proving insufficient. This next wave of potential adopters need to see a business rationale for opening access to data that is clear, robust, and defensible.

In exploring more tangible next steps, discussion frequently turned to the topic of controlled terminologies, thesauri, or name authorities. A small number of key resources from organisations such as the Getty and The National Archives are in widespread use, and there was a shared enthusiasm for exploring ways in which these might be opened up online.

Openly available over the web, with every concept or topic assigned a citable and unchanging web identifier, might these resources take their place as pivotal hubs, capable of connecting content from across institutions and domains? The team at Discovery is looking at the best way to proceed, and welcomes comments from the community.

# Recent News and Events in brief

26 May 2011:

The **Discovery initiative launch event** was held at the Wellcome Collection building in London. As well as interesting presentations and discussions the Discovery identity was unveiled (including the new website: <http://discovery.ac.uk>) and the Open Metadata Principles were launched (<http://discovery.ac.uk/principles>).

---

8 June 2011:

**Day two of the OpenCulture 2011 conference** was attended by members of the Discovery team. Lots of overlapping areas of interest. BBC's Bill Thompson raised a cheer for the idea of 'Linked Data' as the next Killer App for Culture. Resources from both days of the conference are available on the event website: <http://www.openculture2011.org.uk/resources-from-2011>

---

14 June 2011:

The **Discovery Museums and Archives Workshop** took place in London. This event brought together representatives from the Museums and Archives sector to explore common ground and broaden the conversation beyond HE academic libraries.

---

4 July 2011:

The **Discovery competition** for open software developers and open data was announced. Full details of the competition are available online: <http://discovery.ac.uk/developers/competition>  
Visit <http://discovery.ac.uk/news> for all the latest news and developments.

---

## Stay in touch

To make contact with members of the project team: <http://discovery.ac.uk/contact/>

To sign up to this newsletter simply send an email to: [rdif-discovery@sero.co.uk](mailto:rdif-discovery@sero.co.uk)

# Discovery and Special Collections

RLUK

By Mike Mertens, Deputy Director - RLUK

Over the past decade or more, despite or perhaps because of seeming digital preponderance, special collections, how they are constituted and how to frame them, have come back into something now approaching sharp focus [see [references](#)]. Unique and distinctive collections (UDCs) – traditionally termed 'special collections' – have always been a common and prominent feature of collection holders. Indeed, they have helped to define the identity, as well as patterns of use, of libraries, archives and museums.

UDCs may consist of all types of documentary material: manuscripts and archives, books, pamphlets and periodicals, maps, graphic materials, sound and moving image material, as well as increasingly 'born-digital' collections.

A UDC may consist of entirely unique or rare materials; more commonly its uniqueness or distinctiveness derives from the totality of items rather than individual items that may not be special themselves.

Indeed, one might go further. As the electronic displaces print and the rate of addition to print stocks slows down, many institutions will hold material that may not be considered initially to be 'special collections' at all but which could be exploited more closely for research purposes.

This would be especially true where large-scale retroconversion or conservation projects are enabling greater access to extant print collections and bringing into sharper relief the question of how stock, its use and presentation, can be shaped by academic choices.

The context of the material and the research contour of the institution where it is held mean it can be exploited to create a distinct local base for creating research opportunities, and attracting funding and staff.

As non-unique and common research material becomes more accessible online, often without the need for library mediation (e.g. via Google Books), the salience of UDCs arguably grows stronger still. They can attract researchers and research funding, and enhance the institution's extramural reputation. As well as serving external needs they could just as easily benefit the institution, for example they led to an external impact required by a research council.

This article is continued on the next page »

UDCs are therefore a valuable asset. They can also be a liability to the holding institution: they take up space, they require specialist care and servicing, their use may be slight (or by externals only). They may find it difficult to compete for institutional attention with high-performance computers or stem cell laboratories, especially in a funding environment less friendly to the humanities subjects that gain most benefit from them.

### **Documentation, discovery, digitisation and promotion**

New technologies, and especially digitisation, have however made UDCs much more visible and accessible, and opened them up to new methods of exploration by researchers. Digital technologies – digitisation and beyond – offer UDCs the opportunity to escape far beyond the walls of their physical homes: to reach new audiences, to connect with other, complementary, UDCs to harness the contributions of countless co-operators around the world. They also help libraries connect with wider digital humanities research and development.

It's commonplace but essential to state that to be used effectively UDCs must be visible and their contents discoverable. Their existence needs to be promoted. At one time, collection level descriptions would suffice to make them widely known. Now, however, full item level online cataloguing is usually essential; often more detailed indexing may be necessary in the case of specialist material.

UDCs have been a prime target for digitisation programmes, funded publicly or privately, managed by the owning institution or a partner (commercial or otherwise).

Over the past six months, JISC, RLUK and many other stakeholders, have been sponsors of a concerted set of activities built on the vision of a new metadata infrastructure and discoverability regime for the UK - Discovery.

As projects designed to increase the openness and reuse of resource descriptions within this initiative have begun to bear fruit, special collections have emerged across the community as a direct, tangible vector for engagement with the principles and ideals that stand behind and propel Discovery forward.

As Jane Plenderleith and Veronica Adamson established

in their recent Discovery advocacy work with research and public libraries, museums and archives, special collections can make sense of open data and aggregation, since it can be difficult to understand the language and concepts of resource discovery, and its applicability to institutions, of whatever size.

Within the Discovery schema, special collections can give profound impetus and shape to the kinds of aggregations that Discovery is intended to produce; whether based around a person, or an event, collations of relevant special collections can concretely foster collaboration between collection holders, and bring apposite collections that are spread across the UK into sharp focus.

We need to attend to such collections, which are widely kept and managed, in order to make sure they do not collect 'digital dust'. They are our jewels, meant to adorn research and engagement.

RLUK is working hard on this issue; alongside JISC and Discovery, through which we are part of a collaborative collections management overlay project built on Copac data (<http://www.rluk.ac.uk/content/copac-collections-management-project>). We are also joining up with OCLC Research to conduct a large-scale survey of special collections within the UK and Ireland, the data from which it is hoped will help establish norms within the special collections community, and effectively support decision-making for strategic priorities and collaborative projects.

In addition, as part of the RLUK Strategic Plan 2011-2014, we have the Unique and Distinctive Collections strand, which is designed to maximise the value of UDCs, through external engagement, fundraising and promotion, as well as augmenting the now broader staff skills necessary to make them successful.

We very much look forward to working further with Discovery, as well as the wider special collections community, to give special collections more of the regard they rightly deserve.

### **RLUK Unique and Distinctive Collections Group:**

*Chris Banks, University of Aberdeen*

*Andrew Green, National Library of Wales*

*Sarah Thomas, University of Oxford*

*Mike Mertens, RLUK*

### **References**

- Special Collections in ARL Libraries: a discussion report from the ARL Working Group on Special Collections (March 2009) <http://www.arl.org/bm~doc/scwg-report.pdf>
- Jackie M. Dooley and Katherine Luce, Taking our pulse: the OCLC research survey of special collections and archives, OCLC, 2010 <http://www.oclc.org/research/publications/library/2010/2010-11.pdf>
- LIBER, Report of the LIBER Working Group on the representation of Special Collections at a European level within the framework of LIBER given by Ivan Boserup (Chairman) at the LIBER Annual General Assembly in Koç University on Friday 4 July 2008. [http://www.libereurope.eu/files/LIBER%20Members%27%20Mailing%20\(August%202008\)%20%20\(Report%20on%20Special%20Collections,%20July%20%202008\).DOC](http://www.libereurope.eu/files/LIBER%20Members%27%20Mailing%20(August%202008)%20%20(Report%20on%20Special%20Collections,%20July%20%202008).DOC)
- 2010 RLG Partnership European Meeting: Moving the Past into the Future: Special Collections in a Digital Age, 12-13 October 2010, St Anne's College, Oxford University <http://www.oclc.org/research/events/2010-10-12.htm>
- Daniel H. Traister, Is there a future for special collections? And should there be? A polemical essay, 2000. <http://cocomaan.net/Traister%20-%20Is%20there%20a%20future%20for%20special%20collections.pdf>
- Daniel Traister, 'Some aspects of the relationships between special collections and academic theory': a paper presented at the Rare Book and Manuscript Section Annual Preconference, Bloomington, Indiana, 23 June 1995. <http://www.sas.upenn.edu/~traister/rbtheory.html>
- Frank M. Turner, 'Meditations on the Beinecke Rare Book and Manuscript Library', 2008 [http://www.library.yale.edu/mtblog/ulibrarian/archive/2008/03/meditations\\_on.html](http://www.library.yale.edu/mtblog/ulibrarian/archive/2008/03/meditations_on.html)
- Barbara M. Jones, 'Hidden collections, scholarly barriers: creating access to unprocessed special collections materials in North America's research libraries: a white paper for the Association of Research Libraries Task Force on Special Collections'. <http://www.arl.org/bm~doc/hiddencollswhitepaperjun6.pdf>

# RDTF PROJECT PROFILES

In this issue we focus on two of the RDTF projects that are using linked data to maximize the benefits available from the integration and aggregation of archival and museum collections and from shared authority files.

## OpenART: A seamless connection from York to London

OpenART is a partnership between the University of York, the Tate and technical partners, Acuity Unlimited, which builds on prototyping carried out under the JISC-funded YODL-ING project.



The project will design and expose linked open data for a multi-faceted research dataset covering "The London Art World 1660-1735", which was originally created by the AHRC-funded 'Court, Country, City: British Art 1660 – 1735' project.

Drawing on metadata about artists, places and sales from a defined period of art history scholarship, the dataset offers a complete picture of the London art world during the late 17th and early 18th centuries.

Links to the Tate collection will allow exploration of works in their current locations. OpenART is re-using existing authorities and vocabularies, ontologies and metadata application profiles to normalise and structure the metadata.

OpenART is producing a blueprint for an open, re-usable data store designed to be scalable to benefit richer and more varied datasets and the wider exposure of open metadata by cultural institutions.

<http://yorkdl.wordpress.com/category/openart>

## Open Metadata Pathway: The M25 without bottlenecks

This project is a collaboration between King's College London Archives, King's Centre for e-Research and the University of London Computer Centre (ULCC). It utilises data held by partners in the AIM25 (Archives in London and the M25) archival description aggregation.



The Open Metadata Pathfinder demonstrates the effectiveness of opening up archival catalogues to automated linking and discovery through embedding RDFa metadata in the collection level catalogue descriptions. Key deliverables from the project include:

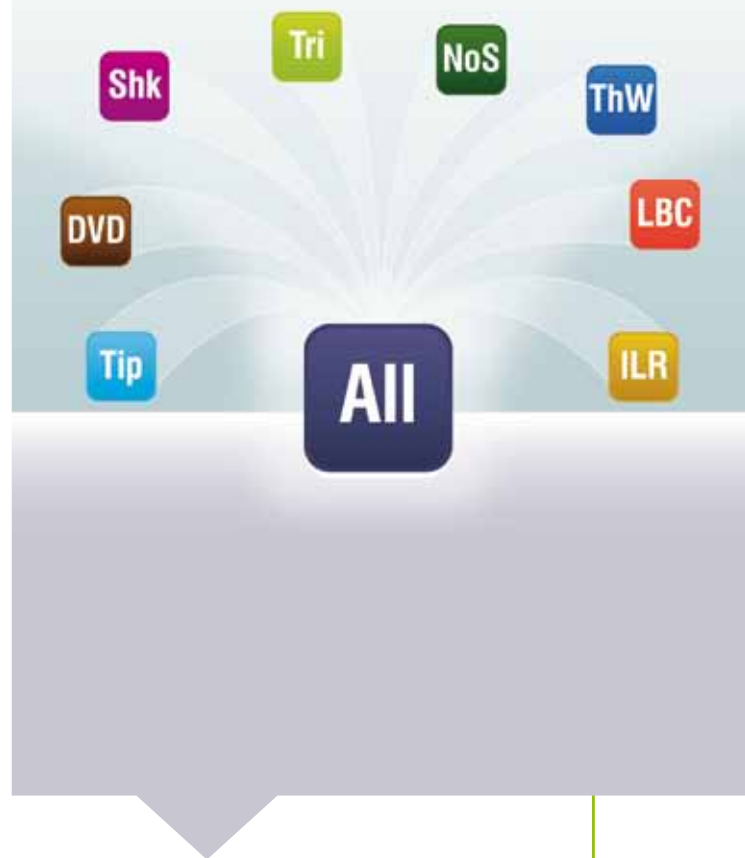
- » Reimplementation of AIM25 data creation tools to include RDFa creation, assisted by automated natural language processing of catalogues via the GATE service.
- » Adoption of the guidance published by the LOCAH Project regarding URIs based on the Cabinet Office's public sector guidelines.
- » A data profile based on the public schemas and ontologies identified for each domain and a URI scheme for entities in the AIM25 namespace.
- » An enhanced web application including SPARQL APIs and RDF query tools.
- » Openly published authority metadata in linked data format, with SKOS implementation of the AIM25 thesaurus data.

<http://openmetadatapathway.blogspot.com>

# Discovery resources

You can find a growing body of information resources reflecting on the work of Discovery as well as resources designed to help facilitate tangible next steps:

- » To see presentations from the Opening Doors event and UK Discovery conference, check out the [Discovery Slideshare channel](#).
- » To learn more about the possible use cases supported by open metadata and to learn about practical concerns such as business cases, visit the [open bibliographic data guide](#).
- » To see what librarians, archivists and curators think about resource discovery, go to the [Discovery YouTube channel](#).
- » To assess licensing options for your data, consult the [Guide to Licensing Open Data](#).
- » To respond to the UK Discovery call for more open metadata, sign up to the [Discovery Open Metadata Principles](#).
- » Lastly if you're interested in reusing open metadata to develop new applications, visit the [Developers zone](#).



## No more channel hopping

Increasing quantities of archive film, television and radio content are available, but usually delivered as stand-alone collections, with users needing to know where to look before they can begin their research.

A new federated search environment has been launched by Royal Holloway, University of London (RHUL) and the British Universities Film & Video Council (BUFVC), which could transform moving image and sound resource discovery by replacing the need for researchers to locate and access various databases and collections through multiple channels.

The BUFVC federated search environment has launched in 'beta' (<http://beta.bufvc.ac.uk>) and allows researchers to search nine online databases, containing more than 13 million records relating to film, television and radio content from a single entry point and view collated results with intelligently generated related records and searches'.

