



# discovery

A metadata ecology for  
UK education & research

## open data open doors

Feature edition - April 2012

### Articles in this edition:

- » **The Business Case for Resource Discovery** – feedback from the SCONUL and Discovery workshop for Library Directors
- » **Special Feature:**  
DPLA and the Discovery Programme – commentary by JISC Deputy Chair Professor David Baker
- » **Special Feature:**  
A Publisher's Perspective – an interview with Robert Faber, Director of Discoverability at Oxford University Press
- » **Signatory Profile** – the UK Data Archive
- » **Looking to the 2020 Horizon** – Warwick workshop highlights Discovery

## Latest steps in defining the business case for resource discovery

Library Directors and Assistant Directors from 18 institutions gathered in London recently to discuss what shape a business case for investment in resource discovery should take. The joint SCONUL and Discovery workshop started with a review of each institution's strategy which revealed significant uptake of new discovery layer products with associated increase in usage of institutional resources – a business case in its own right.

David Ball, Head of Academic Development Services, Student and Academic Services at Bournemouth University, opened proceedings by sharing how his institution has focused on putting users in control: "We asked ourselves what users do in practice and tried to deliver that as simply as possible. Our philosophy is to aim for what is good enough." At Bournemouth the library catalogue and the repository are integrated into the VLE and business systems and they have implemented EBSCO Discovery service.

A similar user-centred approach is the driver at University of Lincoln, as explained by Paul Stainthorp, Electronic Resources Librarian: "We see the role of the student as collaborator in the production of knowledge and focus on improving the student experience through their active engagement." Lincoln has an open development group which is committed to exposing library and other institutional data through APIs as the basis for agile and innovative development. "We are starting to see the results in terms of new cataloguing workflows, knowledge share, staff development, and new partnerships."

As the workshop progressed there was strong consensus that resource discovery is in fact an incremental development of the core business of the library in serving the institution. Likewise, that whilst the current investment in discovery layer products is a sign of the growing priority given to this agenda, Library Directors see these services as part of an evolving picture – a point picked up by Professor David Baker and OUP's Robert Faber overleaf who both emphasise the importance of focusing on data and content as a route to future-proofing. Read the full articles inside...

# Common ground: the discovery programme and the DPLA

To see full details of Robert Darnton's Lecture visit: [JISC Inform/Issue 33](#).



**By Professor David Baker, Deputy Chair of JISC and Chair of the JISC Transition Group**

The Royal Society's Kohn Centre was the venue on 17 January 2012 for a lecture by historian and director of the Harvard Library, Robert Darnton, a pioneer in the field of the history of the book and founder of the [Gutenberg-e program](#). The lecture was hosted jointly by JISC and SCONUL and the main focus was on the recently-founded [Digital Public Library of America \(DPLA\)](#). I was keen to chair the lecture, given my role within the Discovery Programme and its predecessor body the Resource Discovery Task Force (RDTF) as the event gave attendees the opportunity to hear more about the DPLA and to start to think about what such a concept would mean within the UK.

It became clear during the lecture and the subsequent question and answer session that there are many parallels between the DPLA and the RDTF's work, and there is much to be gained from further interactions as the respective projects develop. The aim of the DPLA is nothing less than 'to realize the great promise of the Internet for the advancement of sharing information and of using technology to enable new knowledge and discoveries in the United States'. The concept is not new, but it has been the recent work of organisations such as the Internet Archive, the Hathi Trust and the Library of Congress that has moved the idea forward in tangible ways, with a two-year programme of work between organisations across the US to help scope and understand the requirements of a potential DPLA. The programme will also unveil a prototype of the "system" with some specially digitised material.

Similarly, the UK has long had a vision of 'a borderless digital library; fully integrated into research and learning, underpinning the creation and use of knowledge'. This vision is now called the UK Digital Library (UKDL) and is seen as a compelling way to highlight the strategic approach taken to delivering a 'digital library' for teaching and research. The concept is a broad one: as well as making resources available for access and use, the UK digital library model supports the creation, deposit, curation and sharing of resources. Licensing frameworks, common standards and supporting policies and practice are developed and made available in order to enhance access and re-use of resources. Both the DPLA and the UKDL visions are all encompassing in terms of information; tools to use and create research and learning and the infrastructure services that underpin the delivery of a world class 'digital library' for education, research and public enquiry.

Making resources discoverable, whether books, journals or archives, has been a concern of librarians and information specialists for many years. It has become clear that library services need to work with the Web and search engines, therefore ensuring catalogues are indexed by Google and other search engines as a route to users; otherwise they will be left behind. Hence the formation of the RDTF and the Discovery Programme. The driver, as with the DPLA, has been to ensure access to knowledge is maximised. The benefits for end users in both initiatives can be summarised as the ability to have access to resources anytime and anywhere; to be able to re-use and annotate easily; and to recombine and bring a rich range of resources together to meet a wide variety of needs that range from a general search to a specific research question.

Both visions are ambitious; and supporting programmes need to investigate the key issues that stand in the way of implementation. These include licensing, the business case for open approaches and technical issues. Key to both the DPLA and Discovery is the move towards 'open data'. This movement has covered many data types; and in terms of this discussion its application to bibliographic data is relevant. Changing the business and rights models around bibliographic data will be necessary if maximum re-use is to be achieved. But arguably the most important element of the two programmes is ensuring that the many people with a stake in resource discovery - including publishers - are fully engaged.

Clearly there has been a significant shift in the resource discovery landscape, one which might best be classified as a move from 'service orientated' to 'resource orientated'. What this means is that data and content is the key and can be seen as a way to support a number of different use cases and can be made available so a variety of services can be developed. That is not to say that there are no service driven requirements, there are; but it seems the best way to ensure resources are accessed and used is to make those resources re-useable in a granular way so they can be surfaced via many routes.

Resource discovery is global, as evinced by the dominance of Google. Of course there are local needs that must be served; however it is important that we work within a global context not only so we can learn from each other but also because this can lead to less duplication and further potential for economies of scale and meeting research and learning needs. In this context, it will be interesting to see what choices the DPLA makes, and it will be important to share lessons so good practice across borders emerges.

So what is beyond 2012? The digital environment is not static and resource discovery services must be agile in order to respond. A key lesson that has emerged when considering the academic resource discovery infrastructure that JISC provides in the context of the new directions is that we can perhaps split the roles in the provision of resource discovery between 'service providers' and 'data providers'. Data providers do not have to provide services, and likewise, service providers do not have to be data providers.

Libraries, archives, museums and infrastructure providers do not have to do everything, when participating in the provision of resource discovery they should choose what they do well and focus on that. The critical aspect to a responsive resource discovery infrastructure is about a commitment to making data (descriptions and content) available in ways that it can be re-used by many.



**(Left) Professor David Baker who chaired the Robert Darnton lecture.**

# An Interview with Robert Faber, Director of the Oxford University Press Discoverability Program

OXFORD  
UNIVERSITY PRESS

## Why does OUP have a Discoverability Program?

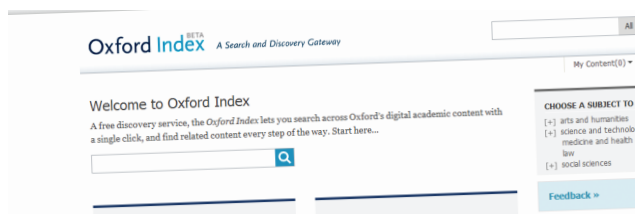
Two years ago we decided to audit how easy it was to find our content. OUP has a great breadth of content. If you're a niche publisher you can assume a standard way of doing things for a specific set of users, but not when you have been publishing a great variety of material across a wide range of sectors for a long time. We found very good practice in some areas but not all and we wanted to consolidate that good practice so that we could offer a consistent approach to finding our resources to our external search partners, and ultimately improve the journey for our users.

We know customer and user behavior is changing. Typical searches are now by topic not by publisher or by journal or book. For example, 50% of traffic to Oxford Journals is direct to the article from Google. There are lots of places where users can get our content from and there are lots of different levels of search. We have to think how we can best ensure that a link to our content appears in as many places as possible – so consistent and systematic description of our content available is vital.

This is not just about increasing sales; it's about increasing usage because that's the lifeblood of digital scholarship. Use of books and subscriptions shows the value of the content. Discovery reveals interest and that can lead to demand for new content. But the journey begins with discovery.

## What is the Oxford Index?

The Oxford Index is about improving metadata, discoverability and connections – three big projects in one. The first is about assembling a standardised description of every item of Oxford content in one place. We can then improve the consistency of published metadata that we export to all our search partners – libraries, research hubs etc.



The second element is the Oxford Index service. As we have to produce an index of metadata for our search partners, we realised it made sense to build our own index that would be searchable from the web and cross-searchable across Oxford products. At the very least it provides a backstop for our users, offering landing pages for every item we produce and quick pathways to full text.

Our third project is about linking content. With a consistent set of data describing all our content, we can plan smart, integrated linking across products. We're building infrastructure at the moment – the data and web interface – but ultimately we're interested in better linking.

## How far have you got with this vision?

The Oxford Index is currently available as a live 'beta' site. First we put the basics in place such as a landing page for each item with standard metadata and the facility to cross search content types. We have just added a lot more content, and extra facilities such as filtering results by access rights, or the option to search library resources. We're now building a set of overview pages with at-a-

glance views of single topics – and will use these to build links to related content.

It's been interesting to develop. Our goal is to ensure that our users spend as little time as possible on our site! The idea is to get them straight off to the primary full-text content.

## Why are you collaborating with Discovery?

OUP is part of a university and we see things from the point of view of scholarship. We have similar objectives to Discovery, to improve the research process. We want to join up with another hub of expertise working on approaches to discovery in academic institutions. For example, we need to take decisions on what standards to adopt for open metadata. It's not down to OUP to choose which standard will become the most common so we need to work with people out in the research community to help us choose which standard to adopt. The Discovery programme and the lead partners, in particular JISC and RLUK, are already engaging with our core constituency on these issues.

We are looking at participating in a Discovery developer forum where we would provide open metadata for developers to try out new service ideas. Having OUP metadata going out and having new ideas coming in seems a very good idea to us.

## Do you have a sense yet of whether the Index is going to bring benefit?

It's very early days and the service is still being built out. That said, we are very encouraged by some initial results. The service is being vigorously indexed by Google bots – in effect, it gives Google a super-sitemap across OUP content. The traffic for our monograph series is already getting a measurable increase from the Index.

But ultimately we are looking for benefits beyond the specific Oxford Index website and we recognise that we are nowhere near a stable point in the evolution of discovery type services. The advent and uptake of new web-scale discovery systems has added impetus to the need for good metadata, but discovery systems are part of a bigger picture. We don't yet know the impact of semantic web development, social pathways to content etc. OUP will not control what users adopt, so our philosophy is that we need to be able to take advantage of whatever new methods emerge. Having metadata in a systematically re-usable form is a key part of that strategy. We need our content descriptions to be easy to update, distribute and re-purpose.

## Do you think there's a business case for investing in Discovery?

There are good business reasons for focusing on discoverability of your content but you have to take the longer view. What matters to OUP is dissemination of knowledge and scholarship. We can measure that in various ways, of which sales are just one. There are other business reasons too. Investing in consistent methods of description and indexing creates efficiencies downstream. Consistency across products also provides a better user experience and a better platform for developing services in the future.

This is not just a benefit to OUP – it's equally relevant to trade publishers, for example. The publishing business routinely produces content descriptions but it tends to be restricted to titles. The tendency is to think at the product rather than the content level. Our view is that users increasingly want access to specific content rather than products. So when we make it easy for researchers to find and use content in our medical journals, say, we want to ensure that we link it to other types of content across our range and elsewhere so it opens up new possibilities for research.

# Book your place on the Discovery Licensing Clinic

M	T	W	T	F	S	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

Wednesday 9th May,  
Birmingham

An opportunity to demystify licensing options and come away with a feasible action plan.  
[Register here.](#)

## Warwick workshop highlights Discovery

In January, JISC and SCOUNL convened a workshop for Library Directors and Senior Managers to review the evolving requirements for institutional Library Management Systems. The workshop focused on developments impacting the local LMS footprint, given evolving approaches to Resource Discovery and shared service developments in the management of subscription resources.

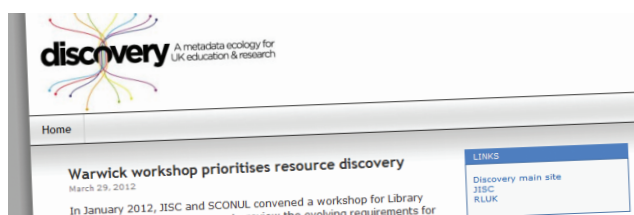
The workshop reviewed a catalogue of over 60 potential library service and institutional knowledge management objectives in terms of desirability, feasibility and their potential to act as drivers of mission critical change.

It was striking that the Discovery agenda represented a very large proportion of the items ranked as high priority looking to the 2020 horizon. The headline priorities are detailed in [David Kay's Discovery blogpost](#).

## The Discovery Blog

Recent items on the Discovery blog have included:

- » A useful summary of the [latest Discovery open metadata projects](#). We'll be closely following the projects' progress and sharing their achievements over the coming weeks and months.
- » Our regular [Discovery news digest blogposts](#) which offer a roundup of what's happening across within the Discovery initiative, the world of open metadata and beyond.



Many of the items in the news digest are from the knowledge being shared via the [#ukdiscovery](#) twitter hashtag which you can view at any time by opening up this [FiveFilters 'newspaper' pdf](#).

# Signatory to the Discovery Open Metadata Principles: UK Data Archive

The [UK Data Archive](#) holds the country's largest collection of digital data in the social sciences and humanities, with more than 5,200 datasets relating to historical and contemporary society, economics and politics. The Archive has been in operation since the 1960s, curating and preserving data from bodies such as the Office for National Statistics, the Home Office, the Ministry for Justice and the Department for the Environment. Since much of these agencies' data have been gathered using public funds, the Archive supports government accountability in making the data available free of charge to support high-quality academic research and teaching.



The Archive provides several services, the primary one being the [Economic and Social Data Service \(ESDS\)](#) funded by the ESRC and JISC; it is through the ESDS that these data are preserved long-term and disseminated to academia, policy-makers, the private sector and beyond. ESDS consults with depositors across the data lifecycle, from acquisition through access. Metadata describing the collections are created using the standard DDI schema and rigorous controlled vocabularies, and made available for each collection. ESDS also provides representative information for each of the collections that it holds, as a separate data entity.



While a policy of openness presides, licensing restrictions currently mean that access to most data collections are not freely available to everyone for any purpose. However, metadata is generally published under a Creative Commons licence. A recent resource discovery strategy review has prompted some streamlining of search functions, and identified further developments linking ESDS resources with other repositories.

## Stay in touch

To endorse the Discovery Open Metadata Principles visit:

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

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:

[rdtf-discovery@sero.co.uk](mailto:rdtf-discovery@sero.co.uk)

Discovery website: <http://discovery.ac.uk>

Discovery blog: <http://blog.discovery.ac.uk>