

---

# Towards a metadata ecology for 21st-century education and research

David Kay, Paul Miller & Owen Stephens  
*on behalf of the*  
Resource Discovery Task Force  
<http://discovery.ac.uk/>

---

## OVERVIEW

The JISC and Research Libraries UK (RLUK) Resource Discovery Task Force (RDTF) vision<sup>1</sup> focuses upon maximising the usefulness of and access to scholarly resources. It proposes to enable aggregation across domains (notably libraries, archives and museums) in a manner that facilitates increased exposure and open service development to benefit all types of users – researchers and businesses, teachers and learners, curators and service managers.

This vision concerns reusing metadata in innovative and flexible ways to facilitate end users in discovery and also to support collection management. It relies on ensuring that it is simple (legally, technically and operationally) and economic to open up metadata from host institutions and then use them to build new and enhanced services as underlying information ecology evolves.

This paper sets out the practical steps being taken by JISC, RLUK and their partners to support the community in this journey and proposes how institutions might respond.

## A BRIEF HISTORY

With the active support of UK stakeholders, including SCONUL, the RDTF began to meet in 2008 to consider the changing patterns and opportunities in the ‘discovery’ of resources as the Web enters its third decade.<sup>2</sup> The result was a vision for enquiry and scholarship situated within a seamlessly navigable information ecosystem – a ‘web of data’ wherein the user and the service devel-

oper would be unencumbered by the source of a metadata record or its curatorial domain.

Library catalogue data could therefore be combined with archival collection descriptions and finding aids; an Open Education Resource might be surfaced side by side with a series of museum images; a map and location-based census data could both be invoked by a place name. Approaching the 450th anniversary of his birth, we might consider those possibilities in the context of Shakespeare scholarship, from the school classroom to the Hathaway family historian and the literary scholar.

Google gives end users a glimpse of this world, imperfect yet tantalising and simple to use. The linked data movement, championed by Sir Tim Berners-Lee, proposes in the semantic web an approach that could meet such expectations, combining ease of access with accuracy of recall.<sup>3</sup> However, librarians and other curators appear to be caught between the old world and the new, uncertain of how to respond to the challenges of entrusting discovery to search engines and/or committing scarce resources to semantic improvement.

Significant questions are raised by these developments in the wider world of the Web. Do they require ‘all-or-nothing’ revision of how we manage resource metadata and how the resources are made discoverable? Or are there steps along the way – opportunities to test value, low-hanging fruit to be harvested – ideally starting with a well-understood commitment, such as shared cataloguing or collection optimisation?

With these questions to the fore, the Resource Discovery Task Force has moved in 2011 from idea to implementation. In so doing it recognises that managers and practitioners need clear business cases (What? Why? When? Who? How much?), supported by relevant exemplars. In this context we focus on four essential developments: the levers for enabling change, the importance of understanding ‘open’, the options for harnessing technology and, not least, the momentum to be derived from developing exemplars.

## ENABLING CHANGE

These possibilities demand a change programme that addresses the realities of curatorial responsibilities and supply-chain relationships.

The RDTF programme is starting from a point where a wide range of stakeholders and existing practices are involved in making metadata available from host institutions and in aggregating it to build services (presently including such as Copac, the Serials Union Catalogue [SUNCAT], the Archives Hub and Culture Grid). Implementation therefore requires a balance between working with existing practices and developing new ones likely to fit future developments.

Furthermore, those who have a stake in information flow have different and sometimes conflicting needs. For example, a publisher's need for a return on the effort dedicated to creating metadata may conflict with the library's desire to enable those data to be reused in new services. Similarly an archive or museum's need to make metadata available according to new protocols may not fit with the development roadmap of their management software vendor. It is important to address these stakeholder interests to ensure that mutually beneficial and sustainable flows of information are developed.

Managers also need the internal institutional leverage and confidence to take on change. As evidenced in recent interviews for the RDTF programme, librarians, archivists and curators are keen that the resources they look after are used by as wide a community as possible and contribute as much as possible to the development of knowledge – there is therefore an *a priori* predication in favour of open access. Furthermore, local networks of libraries, museums and archives are already sharing data. However:

- Communication is an issue, as the language of resource discovery and aggregation does not help to express the vision and communicate what it means in practice
- There is a need for real examples and evidence of benefit.
- Current resource constraints are militating against large-scale developments; people need solutions, not works-in-progress.

## UNDERSTANDING 'OPEN'

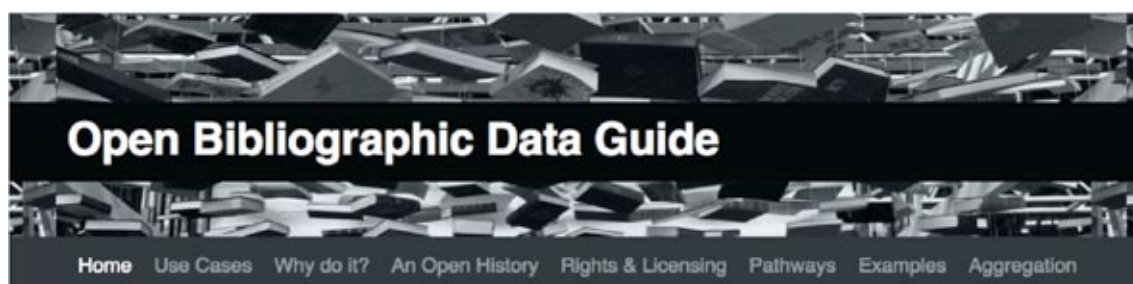
There are plenty of cases in which libraries – or more likely individuals within libraries – have released bibliographic data for others to use. Sometimes these releases have been carefully planned, strategic, and preceded by careful consideration of implications both good and bad. More often, though, data releases have taken place in an *ad hoc* fashion, driven by the persuasive powers of a dedicated enthusiast. Issues such as licensing, commercial use, long-term sustainability and the like were typically ignored, perhaps in the hope of avoiding hard questions.

Whilst this fits well with the operational style of enthusiastic early adopters, it is not a viable way of making significant and lasting decisions at the institutional level.

Meanwhile, open data, and therefore open bibliographic data, have come strongly into focus as the Web has transformed approaches to making a library catalogue (or an archival finding aid) 'discoverable' and as government has emphasised the principles of open publication.

In response, JISC commissioned an *Open bibliographic data guide*<sup>4</sup> on behalf of its RDTF partners. The guide was not intended to break new ground, nor to embark on exhaustive new research. Instead, it was meant to synthesise existing practice and offer this to the community in a fashion suitable for supporting library and institutional managers in making informed choices around real business requirements – activities such as sharing catalogue data with an aggregator such as Copac, SUNCAT or WorldCat.

The *Guide* takes a pragmatic approach, grouped around a set of seventeen typical use cases such as 'supply data for physical union catalogue' and 'supply holdings data for collection management'. Each use case describes the nature of the activity, the actors and data involved, and goes on to detail various motivations, practical issues, benefits, consequences and costs.



Importantly, it also links to working examples of practice. Rather than reading the *Guide* from beginning to end, stakeholders already considering a particular course of action are able to refer to the relevant use case in order to develop a fuller understanding of the issues. A set of pathways clearly demonstrates that many of the use cases are interconnected, and that institutional decisions made in one area have implications in others.

By using this *Guide* and leveraging a wealth of existing experiences across the community, decision makers can move beyond responding emotionally (positively or negatively) to the rhetoric of 'open', and to make sound business decisions on behalf of their institutions – decisions within which 'open' may well feature highly.

### **HARNESSING TECHNOLOGY**

Web-based publication platforms, from WordPress to Google Docs, have made publishing documents on the Web a 'one-click' process. The same goes for digital media, thanks to such as Flickr, YouTube and podcast services. However, publishing data (or metadata) on the Web are not always as straightforward, as recognised by Andy Powell and Pete Johnston of Eduserv, who have provided important insights into how we might think about the publication of metadata to address the RDTF vision.<sup>5</sup>

Running through all technical approaches is 'openness'. Whilst not in itself a technological issue, openness can have a significant impact on what can be achieved with the published data, no matter what mechanism is adopted.

Many institutions may find they are already publishing their metadata in a structured way, and sometimes the application of clear licensing may be all that is needed to achieve the RDTF vision. For others the technology to enable publication (and the associated effort and skills) is a significant barrier.

There are immediate opportunities for collections of descriptions. When institutions contribute to existing union catalogues or similar aggregations, this is most often achieved through export of large numbers of records from a catalogue in a standard format (for example, MARC21). Examples such as Copac, SUNCAT and the Archives Hub reflect this approach. In the RDTF context, it could be a significant step forward for aggregators to invite contributors to license and publish these 'bulk downloads' as 'open' metadata, so that reuse is

not limited (as currently) to a single aggregator with whom a specific agreement may have been made.

There are also opportunities for building better websites and joining the 'web of data'. Within the RDTF vision, the phrase 'build better websites' has been used to describe how metadata from museums, archives, libraries and related institutions might be made more discoverable by the major search engines and more reusable on the Web.

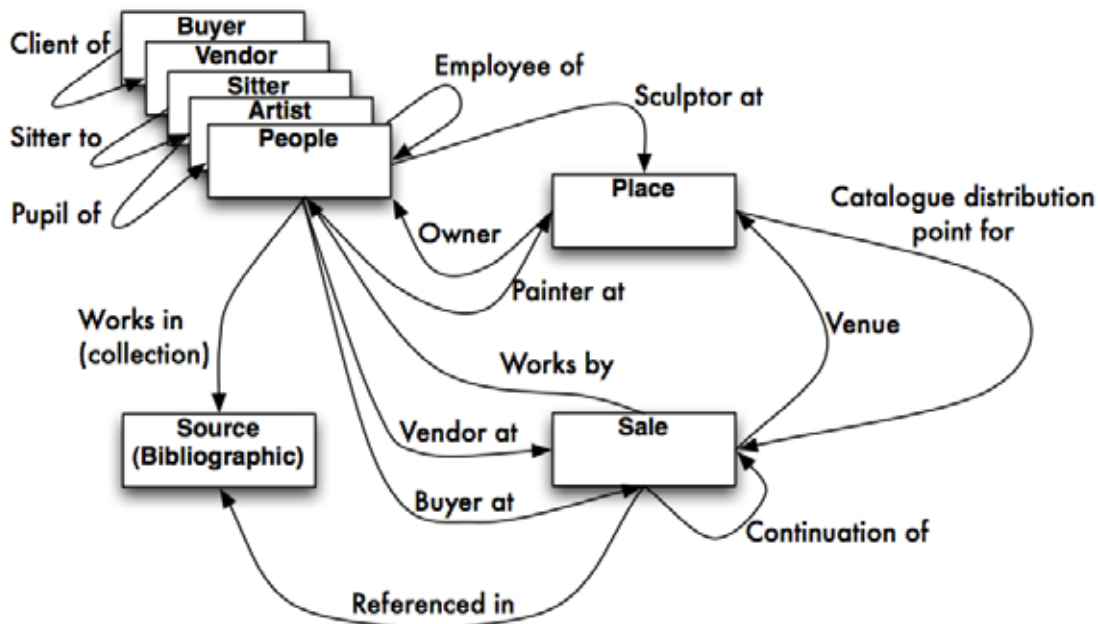
A wide range of activity could fall within this approach (such as described in a blog post by Paul Walk of UKOLN<sup>6</sup>). One of the principles underlying this approach is using reliable unique identifiers (i.e. URLs, also known as http URIs) for each entity, such as a book or an artefact, and to reuse these identifiers across datasets. This use (and reuse) of globally unique identifiers is part of the vision of the 'web of data' – the concept of interconnected sets of data linked through URLs, providing a parallel to the 'web of documents' that currently dominates the World Wide Web.

### **DEVELOPING EXEMPLARS**

In order to begin to test the challenges and the opportunities relating to the RDTF vision and thereby to develop both business cases and exemplar practice, JISC has funded eight institutional projects to publish metadata to the Web as part of the 'Infrastructure for resource discovery' strand of the 'Information environment' programme.<sup>7</sup>

These projects are exploring different approaches to publishing data to the Web, which will help inform decision making for other institutions. The projects include data not only from libraries but also from archives and museums – for example the AIM25 consortium and the Fitzwilliam Museum – and each project is committed to addressing a business requirement relating to the challenges of 'discovery'. The rich possibilities envisioned by the projects are thought-provoking.

**The OpenART project** involves the University of York working with the Tate to expose a linked open dataset entitled 'The London art world 1660-1735'. The project blog<sup>8</sup> highlights the project as 'very interesting, because it goes beyond a single layer of data about collections – it brings together a range of data sources including library, archive and museum collection and bibliographic information, along with original research... enabling users to explore this art world's lost networks, markets and geographies'. Nigel Thomas from the



York Digital Library team has illustrated the key relationships expressed in the data, highlighting a path running from contemporary records and works of art into current locations of and meta-data for collections.

Meanwhile, **Lincoln's Jerome<sup>9</sup> project** is exploring better ways of using library data within the wider institutional information ecosystem that underpins teaching, learning and research. One of the underlying requirements is to unify different collections into a single standardised searchable index. In addition to library catalogue and journal

subscription metadata, Jerome harvests the institutional repository and also uses data from Open Library<sup>10</sup> to boost the depth and accuracy of the local catalogue and for book-cover images. Once those core data are available, the challenge will be to determine what other sources of data might form part of a living web of data, potentially including activity data and the contributions of 'students as producers'.

In conclusion, we reference **Cambridge University Library's Comet project**.<sup>11</sup> Whilst focused simply on opening up regular library catalogue

## Beyond bibliography ...

UNIVERSITY OF CAMBRIDGE

data, albeit on a grand scale, the project has a much wider consequential set of opportunities on its radar.

Ed Chamberlain's presentation of Comet to the March 2011 RDTF 'Open data... open doors' event<sup>12</sup> concluded with considerations 'beyond bibliography', highlighting the mission-critical business case for the RDTF 'discovery' vision acting as a catalyst in support of UK teaching, learning and research in the 21st-century metadata ecology.

#### **MOVING FORWARD**

Whilst JISC, RLUK and other national stakeholders will play a part in realising this vision for 'discovery', especially in promoting principles and practice, the key rests with individual institutions and their partners in the aggregation and release of resource metadata.

This paper has highlighted important steps that are open to all. For example, in reviewing the four areas outlined above, you might consider:

- **enabling change:** work through the business case for enhancing discovery or, equally, for not doing it
- **understanding 'open':** consider whether making some or all of your metadata 'open' and free for reuse will add value to or harm that business case
- **harnessing technology:** engage your technology team or the technical partners you rely on in considering options for making more of your resources available on the Web
- **developing exemplars:** actively commit to testing these opportunities, principles and practices with some or all of your records, perhaps with an aggregator you already work with; and if you're already there, tell us about it.

Finally, please keep in touch with and play a part in the development of the metadata ecology for UK teaching, learning and research, by signing up at <http://rdtf.mimas.ac.uk/newsletter>.

#### **NOTES**

- 1 <http://rdtf.mimas.ac.uk/>
- 2 <http://rdtf.jiscinvolve.org/wp/>
- 3 <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>; see also [http://eprints.ecs.soton.ac.uk/12614/1/Semantic\\_Web\\_Revisted.pdf](http://eprints.ecs.soton.ac.uk/12614/1/Semantic_Web_Revisted.pdf)

- 4 <http://obd.jisc.ac.uk/>
- 5 <http://efoundations.typepad.com/efoundations/2011/03/rdtf-metadata-guidelines-next-steps.html>
- 6 <http://blog.paulwalk.net/2010/09/21/institutions-and-the-web-done-better/>
- 7 <http://www.jisc.ac.uk/whatwedo/programmes/inf11/infrastructureforresourcediscovery.aspx>
- 8 <http://yorkdl.wordpress.com/2011/02/06/openart/>
- 9 <http://jerome.library.lincoln.ac.uk/about>
- 10 <http://openlibrary.org/about>
- 11 <http://cul-comet.blogspot.com/>
- 12 <http://rdtf.mimas.ac.uk/newsandevents/events/2011/04/18/index.html>