Designing metadata for resource discovery

by Deirdre Kiorgaard
Presented at the ACOC seminar "RDA: next generation cataloguing standard" held at the Powerhouse Museum, Sydney, on Friday 24th October, 2008.
http://www.nla.gov.au/lis/stndrds/grps/acoc/program2008.html

Abstract. The cataloguing community is now preparing for a future beyond AACR, MARC, LCSH, DDC/LCC and local catalogue-based resource discovery. The focus is no longer limited to cataloguing and the use of common library standards. The resource description horizon now encompasses data re-use and interoperability with standards used in publishing, on the web, and in other resource description communities such as museums, archives and galleries. Resource Description and Access (RDA) will be an important building block in the creation of both better catalogues and other resource discovery services.

Introduction

As librarians and as cataloguers we are constantly aware of change in the environment in which we work. From digitisation to digital publishing; from the Internet and its search engines, through to Web 2.0 and its blogs, wikis and mash-ups; from *The Social life of information* and *The Long tail* and on to *The Big Switch*, we are seeing rapid changes to the way information is being created, accessed, shared, stored and owned.

There are several ways in which we can respond to these changes.

- React to them as challenges to our profession
- Use them as opportunities to exercise and refine our professional skills, or
- Plan for early retirement

Obviously, I would recommend that we see them as opportunities! Viewing the changing information environment as one full of opportunities will lead to the best outcomes, both for librarianship as a profession and for the users - who are the reason the profession exists in the first place.

Outline

Today I will begin by speaking about some of the myths that have sprung up around the need to change how we catalogue, and then talk about some of the things I think will actually have an impact on the data we provide. I'll talk about the value-adds that cataloguing offers. I'll describe the changes to the way data is being used and how this may affect the type of data we provide and also the type of standards we need to use. I'll also address some of the issues we face in data sharing.

The focus of this presentation is the impacts this new resource discovery environment has on the metadata we produce. Although I will mention RDA from time to time, it is not the main focus of my talk. Instead I hope to provide you with an overview of the broader context in which RDA has been developed and in which it will be implemented.

Cataloguing myths and legends

Recent discussion on the future of cataloguing is full of hyperbole.

- We can no longer catalogue everything

It is often said that we can no longer catalogue everything. The myth here is that we ever did. There was no golden age when cataloguers created full catalogue records for everything in the library's collection, let alone catalogued everything of potential interest to their users. The truth, as all of you will know, has always been more complex than that[1].

- The catalogue has lost its central place

And, although we may wish to convince ourselves otherwise, libraries and library catalogues have never been the centre of the information universe, and certainly never constituted the universe itself[2]. Even within libraries, although the catalogue has always played a central role, it has never been the only route into the library's collections[3][4].

Each of these myths fall within the category of "lies librarians tell themselves" as Stephen Abrahms has described them[5]. I think it is important to dispense with these myths so that we can look more clearly at the opportunities being offered to us.

Brave new world[6]?

Certainly the face of resource discovery has undergone a long overdue transformation.

- The power of the search engine

The advent of the internet has brought unprecedented resources (time, money, computing skills and research) to bear on the search process. Algorithms have been developed to interpret queries and optimise the results from keyword searching. Relevance ranking is constantly being improved. And many improvements that we asked for, but for various reasons our opac vendors never got around to providing, (such as synonym control and 'did you mean?') are now commonplace. All of this is nothing short of a revolution, particularly for access to text-based online resources[7].

And all of it should inform the development of our opacs[8].

---

[1] All types of library resources might be excluded from these sweeping statements –the books might have been all fully catalogued, but perhaps the maps were not; perhaps the course materials were processed with minimal records to make them available quickly.

[2] Many information inquiries were answered by the library, but they also were answered by the yellow pages, the local Citizen's Advice Bureau, the business or government information centre, the phone call to a friend or the poster at the local shops. Not to mention the museum, archive or gallery.

[3] Indexes, abstracts, bibliographies, reader's guides, and the reference collection are just a few of the other sources. And in many libraries audiovisual material (or any type of material that was considered 'special' by that particular library) may have been indexed in a separate database to the catalogue.

[4] There is also a fundamental misunderstanding at play. New common technologies do not usually replace earlier technologies, but augment them.

[5] In discussion. Libraries and Web 2.0 Discussion Group, World Library and Information Congress: 74th IFLA General Conference and Council, 10-14 August 2008, Québec, Canada.

[6] Huxley, Aldous. (1932) *Brave New World*.

[7] Relevance ranking is less suited to metadata than it is to full text resources; and none of these techniques is of use for non-textual resources unless metadata has been provided for them.

[8] For a quick 'wish list' of changes needed to our opacs see Karen Schneider's series of articles on "How opacs suck", and in particular the second part, subtitled "The checklist of shame" (Schneider,

- Next generation catalogues

Although the library catalogue as it presently exists is past its 'use by' date, to paraphrase Mark Twain "The reports of the death of the catalogue are greatly exaggerated"[9]. The library catalogue contains information tailored to the community it serves and so is a key tool in preventing information overload.

Today we are also seeing the development of the next generation catalogues. Librarians are adopting techniques developed in the context of the internet to create 'next generation' catalogues with improved interface design and search mechanisms; which allow users to tag resources, add reviews, and see recommendations; and which link to resources beyond those in the library's collection[10] and lots more.

All of this is fantastic and I for one am thrilled that we are now experimenting and exploring these possibilities to make the catalogue more relevant and to provide new navigational paths for our users.

People have the power[11]

The question is, to what extent, and when, do these advances remove the need for human intervention in resource description?

It is interesting to note that neither the internet search experts, nor the users on the ground, think that the search engine alone is enough – or at least not yet. As Danskin says:

> "Will keyword searching and relevance ranking alone suffice? Neither Google nor Microsoft seems to think so. In their mass digitisation projects they are already reusing the catalogue records created for the printed originals." (Danskin, 2006).

For a librarian this second quote is somewhat amusing for its naïveté:

> "Sure, Google is great. I use it everyday and there is a good chance you do too, but their algorithms are not perfect, and sometimes your results are not quite what you were looking for. Well, that's were people-powered search comes in. Search results that have been provided or filtered by humans. The idea is that if a person is deciding what results you see rather than a computer, your results will be closer to what you are looking for rather than a big list of all possible related links." (Gold, 2007)[12]

Although the sources used here are anecdotal, they are also backed by the available evidence (e.g. see Markey, 2007).

2006). Unfortunately catalogue vendors have been slow to respond, and libraries are now looking to open-source applications, for example see The eXtensible Catalog (XC) Project.

[9] The New Dictionary of Cultural Literacy, Third Edition. 2002. http://www.bartleby.com/59/6/reportsofmyd.html

[10] For a quick summary of what is in a next generation catalogue see Morgan (2007).

[11] "People have the power to redeem the work of fools" (Patti Smith, People Have the Power" © 1988 Druse Music)

[12] Before researching for this paper I was unaware of the extent to which people-powered search engines had taken off. Comments on the posting "Mahalo and Friends: 10 People Powered Search Engines" by Ben Gold quoted above provided many more names of people powered sites.

The forgotten thrill of cataloguing

Social tagging is another side to people power. This is a very curious phenomena: like most librarians I have been surprised by the sudden popularity of both social tagging and of cataloguing sites such as Library Thing. It seems that, just as many librarians seemed ready to consign cataloguing to the dustbin of history, the Google generation is discovering the thrill of cataloguing (Miksa, 2008) and the "miracle of organisation" (see "Tagging - People Powered Metadata for the Social Web (review)").

Some have suggested that social tagging could be a replacement for the subject descriptors devised by cataloguers. I don't see social tagging as a replacement for subject analysis by librarians, because it lacks all of the elements that make controlled vocabularies so useful. But we do need to harness the power of social tagging to enhance our catalogues and our build our controlled vocabularies using terms in current use.

To paraphrase Stephen Abrahms: we need to know when to use the mob and when not[13].

In the midst of all this change, both cataloguers and library managers need to stand back and think about what the changes in the resource description and discovery environment mean for the data we create and how we create it.

New basics

Although we still need to decide what needs to be described and create the data, change has affected the nature of even these basics.

- Decide what we want to provide access to

Our decisions about which resources need a description are affected by a changed understanding of our collections. With the increase in information which is freely available online we are no longer limited to describing resources that we hold as part of our physical collection. The resource that we wish to provide access to could be anything on the internet that is of value to the community which the particular library serves. Access to online resources via internet search engines may be enough, or we may wish to include a resource description for the online resource in our catalogues.

In determining the value of a resource we need to be wary, particularly if the community we serve is broad and our collection is designed for research value. Our judgements about what is of value have long been coloured by various applications of the 80/20 rule, e.g. that 80% of information needs can be met with 20% of the library's resources. But we also need to be aware of the flip side of that.

"As *Antiques Roadshow* demonstrates each week, you just never know what people will value in the future." McKinven (2002).

If we make information about our resources more widely available, those resources will be used more. We've often experienced this at the National Library - whenever

[13] In discussion. Libraries and Web 2.0 Discussion Group, World Library and Information Congress: 74th IFLA General Conference and Council, 10-14 August 2008, Québec, Canada.

we catalogue a collection that may have been lower down on our priority list, once the catalogue records are out there use of the collection increases, demonstrating a demand that we might have previously been unaware of. This is the effect of the long tail (Anderson, 2004; Boston 2007), and it applies to both recreational and research use of resources.

- Create {source, etc} the data

Once the decision has been made to provide access we need to decide the type and level of metadata to apply, for example full or brief record, access level record, AACR level one, two or three or in the future RDA core level, and so on.

Full original cataloguing is the most labour intensive and costly way to create resource descriptions. Librarians have long used sources of high quality data such as copy cataloguing data and CiP data to reduce the costs of original cataloguing. Although original cataloguing remains a vital activity in every library, because of the associated costs we may decide to reserve its use for resources with high value for our own library's users.

Today there are other sources of data that we can choose to use as well as copy cataloguing: text scanned from the resources, metadata from the creators of online resources, information from publishers, and so on. We can use this data as the basis for records which we then upgrade, or use the data with minimal changes. In RDA we have recognised the desire of some libraries to use text scanned from resources as the basis for descriptions, and have incorporated alternatives which allow this.

Later on I will talk about how to provide good quality, shareable metadata. But however valid, or not, the pursuit of the 'perfect record'[14] may be, we should not lose sight of the fact that even minimal data can allow resource discovery.

One of the benefits of the brave new world in which we are operating is that, once minimal data is made available, there are increased opportunities for our records to accrete more information over time, for example through tagging and linking, and also through machine intervention and enrichment.

Paradise lost or paradise regained?

Previously I talked about the myths and legends of cataloguing and said that I don't buy into the idea of the glorious past of the catalogue. However I do think there are some things which our users lost when we moved to the online catalogue, and which the new environment that we are working in now allows us to regain and build upon (see Danskin, 2006 and Markey, 2007, and Bade, 2007).

We need to pay attention to providing data that offers the biggest 'value add' to our resource descriptions. The next generation catalogue offers some new ways to derive order from our data, but there are some situations where order can't be derived from existing records but must be imposed.

[14] Bade (2008) notes that "Searching the literature for the 'perfect record' revealed no advocates of the 'perfect record but many denouncers." His survey concludes that the perfect record " … is simply a rhetorical strategy for dismissing all issues concerning quality by reducing the very complex and context dependent notion of quality to what is implied in the phrase 'the perfect record'."

To my mind the most important value-add to resource descriptions is the controlled names and vocabularies which provide context for resources, and navigational paths for their discovery. These provide power well beyond that offered simply by improved indexing of our databases.

Navigation and relationships

In traditional cataloguing, the cataloguer provided data which allowed the user to expand their search using links and vocabularies developed to provide navigational paths.

These included:

- the use of **forms of name** that allowed users to find all of the works of an individual, regardless of the name used on the resource;
- the use of **preferred names for works** or 'uniform titles' that allow the user to discover all the works with the same content, regardless of the title under which the are published;
- the carefully crafted **subject vocabularies** which allow the user to discover resources that meet their information need exactly, but which might contain not a single word in common with the terms used in their search query [15].

The use of these paths can be made as visible or invisible to our users as they, and we, prefer.

The failure of the opac

Although the opac has allowed access to any field we choose to index in the catalogue record, it has neglected navigation and relationships. As Danskin says:

"The OPAC has tended to favour an increase in the number of access points over the effective presentation of the relationships between resources. … It has been the failure to exploit the navigational potential of this rich metadata that has given the OPAC such a bad name." (Danskin, 2006.)

How many of us have accepted an online catalogue which has no links at all to authority data? Why have we accepted it?

Now we finally have the technologies to facilitate the use of our data in the way in which it was designed to be used – and this makes our data more valuable not less.

[15] Calhoun, 2006: "more than a third of records retrieved by keyword searches would be lost if subject headings were not present" summarised from Gross and Taylor (2005)

RDA and relationships

I'd like to say a few words about RDA at this point. Ebe Kartus will be expanding on some of these points later this afternoon. Although RDA will not cover subject description and access when it is released, it will offer some improved mechanisms for providing navigational paths for our users. Some examples are:

- Preferred titles for works and expressions

The AACR concept of uniform titles has been expanded to incorporate preferred titles for both works and expressions.

- Links between the FRBR group 1 entities

You will be able to create explicit links between resources related at the work, expression, and manifestation levels.

- Relationships among works, etc

You will be able to provide generic information about the nature of the relationship between works and expressions using specific data elements, or more specific information about the nature of the relationship using relationship designators such as 'Translation of', "Sequel to' and so on. For example, you could specify that 'The fellowship of the ring' has a sequel called 'The two towers'.

- Relationships between works etc, and their creators, etc

You will be able to indicate relationships between a creator and a work, or between a contributor and an expression. You can also be more specific about the nature of the relationship. For example, you could choose to specify that Vivaldi is the composer of 'The four seasons'.

- Relationships between persons, families and corporate bodies

You will also be able to deal more explicitly with relationships between persons, families and corporate bodies, for example to record that Frank Seiberling is the founder of the Goodyear Tire and Rubber Company.

These are the types of relationships that it is difficult if not impossible for a machine to derive, although new technologies can facilitate their creation and make them cheaper to provide.

The introduction of these concepts into RDA is an important step. They go beyond what we were able to provide with AACR, and will allow the user to better navigate the catalogue or resource discovery system. For example, they allow resources to be grouped to show they belong to a particular work or expression. This can be used to allow users to move between related works, or for systems to organize large results sets in a way that is more meaningful to users.

The (not so) secret life of catalogue data

While many have focussed on the changes that the internet has brought to the interfaces to our resource discovery systems, there has also been a more quiet revolution in the life of catalogue data and the contexts in which it is being used.

"metadata increasingly appears farther and farther away from its original context" Shreeves, Riley and Milewicz (2006).

- Library catalogues

In times past, the focus was on creating resource descriptions to serve the needs of your users: the local community for public libraries, the particular firm or government department for special libraries, and for the university library, the university community: academics, students and researchers. These resource descriptions were made available locally in the catalogue, first the card catalogue then the opac. The scope of the library's catalogue was closely linked to the scope of the library's collections.

- Shared library databases

Over time the focus of resource description has broadened as increasingly librarians have realised that their resources and records were valuable outside the immediate community they served. Now catalogue records are made available through union catalogues, through national databases like Libraries Australia and through international databases like OCLC WorldCat.

- Digitisation projects

More recently, libraries may engage in small and large scale digitisation projects. In some cases libraries have treated these digital library collections as separate collections, used different standards to describe them, and excluded them from their catalogues[16], providing access through alternative discovery paths.

- Institutional repositories

Also over recent years university libraries may have become responsible for setting up institutional repositories for their institutions. These repositories often begin their lives as stand-alone services and systems, but increasingly the metadata is being harvested and included in aggregated services such as Arrow.

- The GLAM sector

Another context which data from libraries is increasingly associated is with aggregations of data from other cultural heritage institutions (Elings and Waibel, 2007) – also known as the GLAM sector. GLAM of course refers to galleries, libraries, archives and museums. Local examples of such aggregations include Picture Australia and Music Australia.

- The Internet

In the last few years catalogue records have also 'jumped the fence' of the library world as libraries have made their data available to web harvesters. Resource descriptions from library catalogues now appear in results sets in Google, and photographs held in library collections can be accessed on Flickr.

So over the last decade or more there has been an increase in services based on data aggregations, greater sharing of library data with other sectors, and increasing exposure of catalogue data to the internet. In turn this has led to a focus on what

[16] In part, particularly for overseas libraries, this has been driven by these activities being treated as projects and being funded separately.

makes data shareable. Today we need to be aware of this broader context when we design our metadata for resource discovery.

Making data shareable

We are no longer designing metadata just for our own library catalogue but instead for any service in which the metadata might be used now, and for any future service. We can't and wouldn't want to re-catalogue our resources every time they are used in a different context or different service[17]. When we catalogue our resources we need to "catalogue once for all" or "catalogue once, use many times", and for that we need to describe our resources in a way which maximises the value that can be extracted from the data.

So what does that mean in practical terms?[18] Shreeves (2006) gives a very readable overview of what makes metadata interoperable.

To be shareable, our data needs to be:

Humanly understandable

- Understandable outside of its original context

Too much of our data is only understandable within the context of the library or catalogue for which it was created.

Wendler (2004) has called this the "on a horse" problem. The example she uses is of a photograph in the Theodore Roosevelt Collection which has been simply titled "on a horse". Within the original context, the additional detail of who was 'on a horse' may have been seen as unnecessary. But that contextual information is vital once a record leaves its original context. It wasn't hard to find an example of this in the National library catalogue – catalogued I hasten to say, before we developed guidelines to prevent this type of problem.

Another example: although cataloguers understand the value of indicating that a resource is a music score or a map, they may not provide a GMD or General Material Designator of text for their books. In the context of a library where the vast majority of resources in the catalogue may be books, you could argue that this does not matter. But, once the record leaves those confines, it can affect the ability of both systems and users to interpret the record.

- Understandable outside of its original language

Wherever possible, our data needs to be language neutral. One way to achieve this is to label the data elements, for example use a language neutral encoding schema like MARC which allows machines (and people) to process or read a record without understanding the language. Another way is to use coded values instead of, or in addition to, language terms to describe aspects of the resource.

---

[17] The issues are a little different depending on whether the data is to be re-purposed for use in a specific aggregated service, or just making it available for harvesting. If the former, the aggregator is able to apply normalisation and use data mining techniques.

[18] An important part of data sharing is the use of protocols such as the OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). The use of such protocols and the technical issues they raise is beyond the scope of this paper.

Machine processable
We also need to make our data understandable to machines, and to do this it must :

- Free from errors

Data quality is an issue even in standalone discovery services although it is probably easier to 'recover' from errors in a contained system. Beall (e.g. Beall and Kafadar, 2007) has done some interesting work on the effect of typographical errors on retrieval.

- Clean, consistent and appropriately granular

When aggregating data from multiple sources consistency becomes especially important. If all of the data from a single source is consistent, the aggregator is better able to map to the appropriate fields in the aggregated service. They are also better placed to make assumptions when needed to bring the data into the aggregation.

Some common problems are if:

- An encoding schema has two possible fields where similar information might be recorded, and the data has been sometimes encoded in one and sometimes in the other.
- Multiple instances of values for an element are encoded within a single instance of that element, instead of repeating the element.
- Multiple concepts are packed into a single element, instead of into separate elements.
- Multiple resources are described in a single record, instead of in separate records.

In each of these cases no amount of machine manipulation afterwards can 'unpack' the data.

- Use identifiers

A quick word about identifiers: identifiers are neutral, independent and reliable ways of linking that are ideal for use by machines.

RDA and sharing data
RDA is addressing some of these issues. The RDA element set has clearly defined elements for both attributes and relationships. Some AACR elements have been split if they covered more than one concept, and new elements have been added to parallel elements use din related schema, e.g. MARC 21. Multiple instances of elements are provided for. And RDA will make greater use of identifiers.

"The successful use of information technologies used for purposes of communication requires far more standardization than human beings need for interpretation and use." Bade (2007)

In this presentation I have purposely avoided going into issues related to the semantic web, although these are clearly also relevant. The semantic web will be covered in Philip Hider's presentation this afternoon.

Whose standards?

Standards for data content are also part of how we make our data shareable. The question is, whose standards?

"Standards are like toothbrushes; everyone agrees they are a good idea, but nobody wants to use anyone else's." Baca (2008)

- Library standards

In the local library catalogue, standardisation of descriptions was less important than providing relevant and timely descriptions to suit the users of the particular catalogue. With the advent of shared databases, using shared standards like AACR, MARC, LCSH, and DDC/LCC has been essential to the ease of both record and resource sharing within the library sector.

- Digital library standards

For digital resources, over time we have seen a steady move towards using core library standards for the description of these resources[19], supplemented as necessary with administrative metadata and metadata for digital preservation and rights management. And as the metadata from individual repositories is aggregated with metadata from other repositories standardisation soon becomes an issue, and groups such as MACAR are formed to advise on appropriate standards.

- Cultural institutions

Cooperative projects involving data aggregation across the GLAM sector has also highlighted the similarities and differences in description practice in these communities[20], and a variety of efforts are underway to find and build upon commonalities between the standards used in these communities[21].

- Publishing

Both the publishing sector and libraries can see a lot of value in sharing standards to allow better interchange of data. Apart from the obvious success of the ISBN which has been with us since 1966, the lack of standards in the publishing sector, combined with the different imperatives which drive the two sectors, has meant that not as much has been achieved as either side would like.

Sharing standards

To create effective mechanisms for resource discovery on the web and in data aggregations we turn to standards. Then we find that the multiplicity of standards in the resource description community is itself a barrier to semantic interoperability. Different methods have been developed to address this.

- Mapping and crosswalks

---

[19] For examples of the use of standards in the digital library see the report of the IFLA working group on digital library guidelines (forthcoming).

[20] See Elings and Waibel (2007) for an overview of standards used in the library, archives and museums sector.

[21] For example, FRBRoo (FRBR-object oriented) being developed by CIDOC Conceptual Reference Model (CIDOC-CRM) and the Functional Requirements for Bibliographic Records (FRBR) working groups.

At the most simple level we need to be able to map the elements used in one standard with equivalent elements used in other standards[22].

Mappings allows us to answer questions such as "Does the 'DC: title' element correspond to AACR 'title proper'?". However, if the elements are at different levels of granularity (and in this example they are, because 'DC: title' covers a broader concept than AACR's 'title proper') then the mapping is only an indication of equivalence. And if the scope of the schema is different the elements will only partially overlap and each standard will include elements not covered by the other schema.

Despite these drawbacks, creating mappings has been a common activity over recent years and on the MARC 21 website[23] alone there are mappings from (and sometimes also to) seven different standards: MODS, Dublin Core, Digital Geospatial Metadata, GILS, UCS/Unicode, ONIX and UNIMARC.

When it is released RDA will include mappings to ISBD, MARC 21 and Dublin Core, and mappings to other schema may be added in the future.

I see all these crosswalks as the equivalent of the adaptors on the screen – we need one for every schema we wish to interact with. So in a way we are duplicating the problems associated with multiple schema.

- Switching schemas and translators

One possible solution to this problem is to use a switching-across schema[24] as described by Chan and Zheng (2006). And there are other possibilities such as the translator proposed by Godby, Smith and Childress (2008).

"Crosswalks, derivatives, hub and spoke models, and application profiles respond to the need to identify common ground in the complex landscape of resource description. But these objects also imply an unresolved tension between the need to minimise proliferation of standards and the need to create machine-processable descriptions of resources." (Godby, Smith and Childress, 2008).

These solutions are all a bit like the universal adaptor on the right of the screen. Although it is better that the large number of adaptors on the previous page, it is still a complicated object.

All of these solutions to the multiplicity of standards create another new issue: how to keep them current and correct as each of the standards or schema are revised and enhanced independently and to different schedules.

For data from other resource description communities who have already established their own schema and standards we will need to rely on these types of solutions.

---

[22] In this presentation I'm grouping content standards and encoding and exchange schema together: although they are different things we do need them to be able to work together.
[23] http://www.loc.gov/marc/marcdocz.html
[24] Chan and Zeng, 2006.

Achieving commonality

Within the library sector itself there is more room to influence how we develop our standards and schema now and in the future to maintain as much interoperability as we can.

When choosing the standards to use within the library sector, we should:

- use existing standards where they exist
- influence the development of existing standards to cover any perceived gaps or to address any issues

When working with other communities we should:

- use elements from existing standards where needed, rather than re-inventing the wheel (i.e. if another standard can fill the gap – use it)
- use and/or develop common vocabularies wherever possible
- use or build upon common models and principles
- make our element sets available on the web for others to use

RDA and achieving commonality

These types of issues have been recognised as we have developed RDA. There are a number of times within RDA when we could have developed an RDA vocabulary but instead we have chosen to specify an external vocabulary (e.g. names of languages). There are times when we have worked with other communities to develop vocabularies where no acceptable ones existed (e.g. the RDA-ONIX joint framework for resource categorization (see Dunsire, 2007)). We have drawn on standards used in the cultural heritage community, including archives and museums and have started the process of building relationships with these communities. With the DCMI community we are looking at areas where the development of joint vocabularies would be useful to both communities. We are building RDA using the FRBR entity relationship model and the Statement of International Cataloguing Principles. And we intend to make the RDA element set and some or all of its controlled vocabularies available for free on the internet.

As well as improving interoperability, these measures will also help to keep down costs for libraries. My involvement with RDA has brought home to me how costly standards are to develop and maintain.

Born free?

Speaking of costs, there are just a few more things I would like to mention in relation to sharing data.

Libraries are used to sharing data amongst ourselves. Sharing of data within the library world is one of the ways that we have managed shrinking library budgets alongside unabated increases to publishing output[25]. Everyone recognises the public good that comes from making our data widely shared and widely available. But

[25] The other way is through continuous improvements to efficiency – there has been a low rate of rise in cataloguing unit costs.

sharing data is not without costs and it comes with a few new risks in the web environment.

There was an interesting discussion on this topic at a Web 2.0 Discussion Group meeting at IFLA in Quebec in August this year, and I'd like to share some of the issues that were raised in that forum.

- Shared library databases

The perception that information should be freely available is an issue for providers of shared databases and clearly an issue of hot debate in the library world if this quote from Richard Wallis from TALIS is anything to go by:

"… OCLC are trapped in an increasingly inappropriate business model. A model based upon the value in the creation and control of data. Increasingly, in this interconnected world, the value is in making data openly available and building services upon it. When people get charged for one thing, but gain value from another, they will become increasingly uncomfortable with the old status quo." (Wallis, 2007)

But organisations such as OCLC and Libraries Australia need to achieve economic viability or cost recovery.

Karen Calhoun recently reported on the work of an OCLC study group that has been looking at data sharing:

"One lesson we took away from the analysis was that the prevailing opinion in the blogosphere is that data should be free and open. The reality is that nearly every organization has terms and conditions for data sharing" (Calhoun, 2008, slide 7).

She also noted that there is a need to transition from a cost recovery method based on the value in "the creation and control of data" to one based on "the value in exchange and linking of data".[26]

- Data is not free to produce

Obviously the process of data creation is not cost free. But as well as costing money to produce, there is an opportunity cost within an organisation. The money spent on data creation is money not spent on other library activities. In recent years it has become obvious that even the Library of Congress needs to take into account whether the provision of certain types of value-added services fit within their core role. Those that do might be offered cost free to the library community, those that don't become chargeable[27].

According to Stephen Abrahms, to ask for everything to be made freely available is a symptom of the "fiscal illiteracy of librarians"[28]. Data itself can never be cost free, but the provision of unfettered intellectual access to information is still a worthy goal.

---

[26] Calhoun, 2008 : slide 6.

[27] Sally McCallum in discussion: *Libraries and Web 2.0 Discussion Group*, World Library and Information Congress: 74th IFLA General Conference and Council, 10-14 August 2008, Québec, Canada.

[28] Stephen Abrahms in discussion: *Libraries and Web 2.0 Discussion Group*, World Library and Information Congress: 74th IFLA General Conference and Council, 10-14 August 2008, Québec, Canada.

As an aside I'd also like to mention that standards are not free to produce either. You may have heard of calls to make RDA freely available, but this is simply not possible – it is also run on a cost-recovery basis. However, as I mentioned before there is an intention to make the RDA element set and some or all of its controlled vocabularies available for free.

The problem of invisibility

In an environment where libraries need to justify their existence to funding bodies we need to be wary of rendering invisible the contribution of libraries to the availability of information. In *The Social life of information* [29] the authors talk of a colleague who was singing the praises of the digital world where he can get direct access to information. As the authors note:

"His enthusiasm had screened out an enormous array of people, organizations, and institutions involved in this 'direct' touch. The university, the library, publishers, editors, referees, authors, the computer and infrastructure designers, the cataloguers and library collection managers, … had no place in his story. When they do their job well, they do it more or less invisibly." (Brown & Duguid, 2008. p. 5-6.)

Sharing in a commercial environment

We also need to be aware that different rules apply in a commercial environment. Commercial use or re-use of free library data may be an issue. For example, data given by libraries freely to a non-commercial site may be on sold to other sites.

Libraries may be burned when making agreements with commercial enterprises, for example for digitising of their resources, if they do not continue to hold the rights in the data they have shared. Rights management for metadata is also complicated by differences in laws between different countries, and in the seemingly nation-less state of the internet.

Also, Libraries need to be savvy about the different way businesses can become established on the internet. Social cataloguing sites may begin their life as non-commercial entities with whom we gladly share data, becoming commercial only once they have established a niche.

Conclusion

Although there are some risks inherent in the evolving resource discovery environment, I think that overall we are experiencing a time of exceptional opportunity. There are opportunities to use our skills in new ways, to use new technologies to develop the catalogues we have always known we should provide, to find new audiences for our information, and to extract maximum value from the data we create.

[29] Also quoted by Calhoun (2008).

Bibliography

Anderson, Chris (2004) "The long tail" in *Wired* Issue 12.10, October 2004. http://www.wired.com/wired/archive/12.10/tail.html. Viewed 13 October 2008.

Arrow: Australian research repositories online to the world. http://www.arrow.edu.au/. Viewed 13 October, 2008.

Bade, David (2007) "Structures, standards, and the people who make them meaningful" [Paper submitted to the Working Group on the Future of Bibliographic Control]. http://www.loc.gov/bibliographic-future/meetings/docs/bade-may9-2007.pdf. Viewed 1 June 2008.

Bade, David (2008) "The Perfect Bibliographic Record: Platonic Ideal, Rhetorical Strategy or Nonsense?" " in *Cataloging & Classification Quarterly* Vol. 46, No. 1, pp. 109-133. doi:10.1080/0163937982183081

Beall, Jeffrey and Kafadar, Karen (2007) "Measuring Typographical Errors' Impact on Retrieval in Bibliographic Databases" in *Cataloging & Classification Quarterly* Vol. 44, No. 3/4, pp. 197-211. doi:10.1300/J104v44n03_03

Boston, Tony (2007) "The 'long tail': implications and opportunities for union catalogues and resource sharing" in *Wagging the long tail: managing print collections in a digital age" held on 3 May 2007* http://www.caval.edu.au/assets/files/Research_and_Advocacy/Wagging_the_Long_Tail/boston.pdf. Viewed 13 October 2008

Brown, John Seely and Duguid, Paul (2000)*The Social Life of Information.* Boston, MA: Harvard Business School Press.

Calhoun, Karen (2006) *The Changing Nature of the Catalog and its Integration with Other Discovery Tools: Final report*, March 17, 2006. http://www.loc.gov/catdir/calhoun-report-final.pdf. Viewed 19 April, 2006.

Calhoun, Karen (2008) "Some rights reserved: the environment for data sharing" part of Libraries and Web 2.0 Discussion Group, World Library and Information Congress: 74th IFLA General Conference and Council, 10-14 August 2008, Québec, Canada. Slides available from http://www.slideshare.net/amarintha/calhoun-data-sharing-panel-ifla-aug-2008-presentation. Viewed 21 October, 2008.

Carr, Nicholas (2008) *The big switch: rewiring the world from Edison to Google*. New York: W. W. Norton.

Chan, Lois Mai and Zheng, Marcia Lei (2006) "Metadata Interoperability and Standardization – A Study of Methodology Part I: Achieving Interoperability at the Schema Level" in *D-Lib Magazine*, Volume 12 Number 6, June 2006. http://www.dlib.org/dlib/june06/chan/06chan.html. Viewed 13 October 2008.

Danskin, Alan (2006) “Tomorrow never knows: the end of cataloguing?” http://www.ifla.org/IV/ifla72/papers/102-Danskin-en.pdf . Viewed 13 October 2008. [Paper presented at 102 IFLA-CDNL Alliance for Bibliographic Standards ICABS meeting, IFLA 72, Seoul, Korea. Also published in IFLA Journal, Vol. 33, No. 3, 205-209 (2007) DOI: 10.1177/0340035207083301]

Dempsey, Lorcan (2006) “Libraries and the Long Tail : some thoughts about libraries in a network age” in *D-Lib Magazine*, Vol. 12, No. 4, April 2006. http://www.dlib.org/dlib/april06/dempsey/04dempsey.html. Viewed 13 October 2008.

Dunsire, Gordon (2007) “Distinguishing Content from Carrier: The RDA/ONIX Framework for Resource Categorization” in *D-Lib magazine* Vol. 13, No. 1/2, January/February 2007. http://www.dlib.org/dlib/january07/dunsire/01dunsire.html. Viewed 13 October 2008.

Duval, Erik; Hogins, Wayne; Sutton, Stuart and Weibel, Stuart L. (2002) “Metadata principles and practicalities” in *D-Lib magazine* Vol. 8, No. 4, April 2002. http://www.dlib.org/dlib/april02/weibel/04weibel.html. Viewed 13 October 2008.

Elings, Mary W. and Waibel, Günter (2007) “Metadata for all: descriptive standards and metadata sharing across libraries, archives and museums” in *FirstMonday*, Vol. 12, No. 3, 5 March 2007. http://www.firstmonday.org/issues/issue12_3/elings/. Viewed 13 October 2008.

The eXtensible Catalog (XC) Project http://www.extensiblecatalog.org/?page_id=2. Viewed 13 October 2008.

Flickr. http://www.flickr.com/. Viewed 13 October 2008.

Gardner, Sue Ann “The changing landscape of contemporary cataloging” ” in *Cataloging & Classification Quarterly* Vol. 45, No. 3, pp. 81-99. doi:10.1300/J104v45n04_06

Godby, Carol Jean; Smith, Devon and Childress, Eric (2008) “Toward element-level interoperability in bibliographic metadata” in *Code4Lib Journal* Issue 2, 2008-03-24. http://journal.code4lib.org/articles/54. Viewed 12 May 2008.

Gold, Ben (2007) *Mahalo and Friends: 10 People Powered Search Engines* June 1, 2007 - 7:18 pm PDT. http://mashable.com/2007/06/01/10-people-powered-search-engines/. Viewed 13 October 2008.

LibraryThing http://www.librarything.com/buzz. Viewed 13 October 2008.

Markey, Karen (2007) “The online library catalog : paradise lost and paradise regained?” in *D-Lib magazine*, Vol. 13, no. 1/2, January/February 2007. http://www.dlib.org/dlib/january07/markey/01markey.html. Viewed 13 October, 2008.

Metadata Advisory Committee for Australian Repositories (MACAR) http://www.arrow.edu.au/macar.php. Viewed 13 October, 2008.

Miksa, Shawne D. (2008) "Educators: What Are the Cataloging Issues Students Get Excited About? – Professional and Intellectual Appeals of Cataloging and Students' Misconceptions of Cataloging" in *Cataloging & Classification Quarterly* Vol. 45, No. 3, pp. 17-24. doi: 10:1300/J104v45n03_03

Morgan, Eric Lease (2007) '"Next Generation" library catalogs in 15 Minutes" in *Infomotions' Musings on Information and Librarianship* http://infomotions.com/musings/ngc-in-fifteen-minutes/ngc-in-15-minutes.pdf . Viewed 13 October, 2008

Music Australia. http://www.musicaustralia.org/. Viewed 13 October, 2008.

Open access repositories in New Zealand http://www.oarinz.ac.nz/ Viewed 13 October, 2008.

Open Archives Initiative Protocol for Metadata Harvesting http://www.openarchives.org/pmh/. Viewed 13 October, 2008.

Open library. http://openlibrary.org/ . Viewed 13 October 2008.

Picture Australia. http://www.pictureaustralia.org/index.html. Viewed 13 October 2008.

Schneider, Karen G. (2006a) "How OPACs Suck, Part 1: Relevance Rank (Or the Lack of It)" in *ALA Techsource*, http://www.techsource.ala.org/blog/2006/03/how-opacs-suck-part-1-relevance-rank-or-the-lack-of-it.html. Viewed 13 October 2008.

Schneider, Karen G. (2006b) "How OPACs Suck, Part 2: The Checklist of Shame" in *ALA Techsource*, http://www.techsource.ala.org/blog/2006/04/how-opacs-suck-part-2-the-checklist-of-shame.html. Viewed 13 October 2008.

Schneider, Karen G. (2006c) "How OPACs Suck, Part 3: The Big Picture" in *ALA Techsource*, http://www.techsource.ala.org/blog/2006/05/how-opacs-suck-part-3-the-big-picture.html. Viewed 13 October 2008

Shreeves, Sarah L., Riley, Jenn and Milewicz, Liz (2006) "Moving towards shareable metadata" in *FirstMonday* Vol. 11, No. 8, 7 August 2006. http://www.firstmonday.org/issues/issue11_8/shreeves/index.html. Viewed 13 October 2008.

Shreeves, Sarah L. (2007) *Creating rich shareable metadata: The DLF Aquifer MODS implementation guidelines* paper presented at ALA Annual – Washington, D.C. June 25 2007. http://www.loc.gov/standards/mods/presentations/shreeves-ala07/ShreevesALAAnnual2007.ppt. Viewed 13 October 2008.

"Tagging - People Powered Metadata for the Social Web (review)" from the blog *Epublishers Weekly*, http://epublishersweekly.blogspot.com/2008/09/tagging-people-powered-metadata-for_23.html. Viewed 13 October 2008.

Tennant, Roy (2007) "Digital libraries: demise of the local catalog" in *Library Journal*, 7/15/2007. http://www.libraryjournal.com/article/CA6457238.html. Viewed 13 October, 2008.

Wallis, Richard (2007) "OCLC and ROI" in Panlibus Blog (TALIS) http://blogs.talis.com/panlibus/archives/2007/12/oclc_and_roi.php . Viewed 20 October 2008.

Wendler, Robin (2004) "The eye of the beholder: challenges of image description and access at Harvard" in Hillman, Diane I and Westbrooks, Elaine (editors) *Metadata in Practice*. Chicago: ALA Editions, pp.51-69.

Yee, Martha M. (2007) "Cataloging Compared to Descriptive Bibliography, Abstracting and Indexing Services, and Metadata" in *Cataloging & Classification Quarterly* Vol. 44, No. 3/4, pp. 307-327.

Zheng, Marcia Lei and Chan, Lois Mai (2006) "Metadata Interoperability and Standardization – A Study of Methodology Part II: Achieving Interoperability at the Record and Repository Levels" in *D-Lib Magazine*, Volume 12 Number 6, June 2006. http://www.dlib.org/dlib/june06/zeng/06zeng.html. Viewed 13 October 2008.

Images from Flickr (used under Creative Commons licenses)

Montage of shelves etc by Brungrrl
http://www.flickr.com/photos/brungrrl/2246691272/

“Seattle Public Library” by Jeff Maurone
http://www.flickr.com/photos/jeffmaurone/2505572445/

“Harry Potter Books being brought into Barnes and Nobel today” by Snappy.joneS’
http://www.flickr.com/photos/swift/839373017/

“Order of the Phoenix Premiere @ Leicester Square 07.03.2007” by Melanie Mcdermott. http://www.flickr.com/photos/pierrotsomepeople/710443488/

“Harry Potter's platform!” by tripu http://www.flickr.com/photos/tripu/267155109/

“Just full of ideas” by Cayusa http://www.flickr.com/photos/cayusa/981372736

Images from the Internet

Image of Earth from space on website by Liz Strauss
http://www.successful-blog.com/2007/09/

Image of cans with no labels on website by United States Container Corporation
http://www.uscontainer.com/cans.html

Image of cans with labels on website by The Vending Revolution
http://www.vendingrevolution.com/pages/Snack-a-Can.aspx

Image of stars on red carpet on website by PopSugar
http://popsugar.com/150953

Image of toothbrushes in a cup on website by Wikimedia Commons (image by Henrik Abelsson)
http://commons.wikimedia.org/wiki/Image:Toothbrush-20060209.JPG

Image of travel adaptor plugs on website by Power-Plugs.com
http://www.power-plugs.com/adaptors.html

Image of single travel adaptor on website by Answers.com
http://www.answers.com/topic/adapter