Collection-level Description: joining up the domains

ANN CHAPMAN, *University of Bath*

Abstract. This article focuses on a new tool for information retrieval–collection-level description. It allows users to search for resources across archive, library and museum domains, enabling them to identify appropriate collections for visits or item-level searching. Collection-level description originated with the Heaney entity-relationship model of collections; the metadata schemas based on the model can be used to structure searchable relational databases in which descriptions are held and a Dublin Core Application Profile is being developed. Examples are given of existing collection description databases. The final section of the article considers the relevance of collection-level description to the archives community.

The information landscape

The past few years have seen many changes in the information landscape. Those navigating the landscape used to fall into well-understood groups (eg academic researchers, employment-based researchers, the serious amateur researcher), who used, studied and viewed collections of physical items in archive repositories and record offices, libraries, museums and art galleries. The routes to these collections were sometimes recorded in subject guides and directories, but often depended on a reference in a publication or a personal contact—the 'native guide' as it were—to point the user in the right direction.

Now the landscape has changed dramatically. The old groups of users are still around but they want access to a wider range of resources, both nationally and internationally. New groups of users are emerging. School children are working on curriculum-based projects and an increasing number of people are tracing family trees and local history, while temporary interest of varying duration is stimulated by the media, witness current or recent interest in archaeology (*Time Team*), Clarice Cliff pottery (*Antiques Roadshow*, etc), the Titanic (films) and 'life in past times' (through 'historical-reality' shows).

The collections have changed too; the physical items are still there, but alongside them are the digital copies and the new electronic resources, which allow users to do some of their work remotely. There is better signposting these days, but many users, especially the newer groups, still find themselves on the bit of the road without any signposts or any 'native guides' around to help them. Collection descriptions are a

Correspondence: Ann Chapman, UKOLN, University of Bath, Bath BA2 7AY, UK.

way of providing signposts or roadmaps to enable users to navigate the landscape more effectively.

The impetus for collection description

In the past academics followed up references and personal recommendations and then visited archives, museums and libraries to consult relevant materials. While actual visits are still necessary, time is also spent in desk research to locate resources using electronic catalogues and other findings aids, with the aim of avoiding wasted journeys, identifying the nearest repository for non-unique and representative-type items, and making the best use possible of limited time at particular repositories.

Other people also need to find information prior to visiting. For example, where people are following up leisure interests or curriculum-based projects, they are often interested in materials from all three domains—books about the development of canals, examples of decorated canal-ware, posters, paintings and photographs of canals and canal workers, archives related to the building of specific canals—and they are often limited to visiting resources within a specific locality, because of time and financial restraints, and/or to using resources to which they are allowed access.

The *Full Disclosure* report, which focused primarily on a strategy for retrospective conversion of item-level descriptions, drew attention to the amount of material not catalogued in machine-readable form and—in some cases—not catalogued at all.[1] This resulted in searches not finding relevant items. In libraries, it also resulted in people visiting (and overloading) one collection because it was not known that other copies existed in other collections, and some material risked disposal simply due to a lack of awareness of its contents. The follow-up *Full Disclosure Prioritisation Study* identified 'a widely held view across all three domains that collection level description should be regarded as an essential first step in identifying priorities for more detailed retrospective conversion, cataloguing and documentation work'.[2]

In addition, the current national focus on 'joined-up' delivery of services and greater collaboration to make the most effective use of scarce resources means that:

- There is a demand to improve the disclosure of holdings in archives, libraries and museums where, in many cases, comprehensive description at item level is not available.[3]
- There is increasing pressure on (library) resources and a recognition of the benefits of collaboration and resource sharing (see eg the work of Higher Education Consultancy Group/CHEMS Consulting and the Higher Education/British Library Task Force).[4]
- Users now want to compare broadly similar entities across the information domains where the nature of 'items' may vary widely.[5]
- Digital services now seek to deliver integrated access to the descriptions of resources (both physical and digital) made available by multiple resource owners/providers.[6]
- There has been a rapid increase in the number of collections of digital resources, especially through large-scale digitisation programmes, with the consequent recognition that the value of these collections is not through their use as 'stand-alone' resources but as components that are (re-)used in many different services.[7]

Creating collection-level descriptions provides an additional resource for users. It enables searching across domains and high-level initial searches. The data contained

in the descriptions allows a user to judge the potential usefulness of a number of collections. In addition to their use in open access databases, collection descriptions can be used in collection management, either at an individual repository level or at a multi-repository or multi-partner project level such as a digitisation project involving several collections.

The search process

Collection descriptions support users in their search for information. Since users are now looking for information at a range of levels from the very specific (an entry in a marriage register) to the very general (anything about Clarice Cliffe), the entity or resources that satisfies their search criteria might be a specific item, a representative type item or a collection of items covering a particular area. But irrespective of what they are searching for, the process is always the same—they want to discover, identify, select and obtain an entity.

- *Discover*: to find one or more entities corresponding to the user's search criteria.
- *Identify*: to confirm that the entity described is the entity that is sought, or to distinguish between two or more entities.
- *Select*: to choose an entity that is appropriate to the user's needs with respect to content, physical format, etc or to reject an entity as inappropriate.
- *Obtain*: to gain access to the entity described, through a method appropriate to the entity in question (purchase, loan, study on site, access to database, or online access to an electronic resource).

However, while the same underlying process is needed for all searches, the tools that support user searches will vary according to the type of search target. If a user is searching for a specific item, archives, museums and libraries have all developed item-level descriptions in the context of the specific material they hold and the services they offer to users. To an ever-increasing extent these descriptions are available in electronic form.

When the search is for relevant collections, then collection-level descriptions enable users to search across domains. Old-style directories are fine for in-domain searching, but cross-domain searching is better served by electronic collection-level descriptions. Collections can be described at different levels: a description may include information about the collection as a complete unit, information about individual items that make up a collection, or information about some groupings of the items that form subsets of the whole.

Collections and finding aids

As part of the Research Support Libraries Programme (RSLP), Michael Heaney proposed a model that described collections and their relationships, irrespective of domain.[8] In his model collections contain items which contain intellectual or artistic content, and the collections themselves can contain or be contained by other collections. They have a physical and/or electronic location, and associated agents—collectors, owners, administrators, producers.

Heaney's model also identifies four types of collection description or finding aid, though he notes that, in practice, the distinctions may be blurred. Of course, many

users who begin their search at collection-level will want to move down into item-level finding aids once they have identified an appropriate collection.

1. A unitary finding aid only describes the collection as a whole, eg a directory of libraries.
2. An analytic finding aid consists of information about individual items and their content, eg a library catalogue.
3. A hierarchic finding aid has information about the collection as a whole, the individual items it contains and the relationship between them, eg an archival finding aid.
4. An indexing finding aid consists of information derived from the items in a collection, regardless of the content of the items, eg an automatically generated index of words occurring in a collection of digital documents.

Heaney's model was used to develop the RSLP schema and SCONE schema. These were developed independently, though with some contact between the developers, during the same time period. Since both schemas draw on the Heaney model, it is possible to map from one schema to the other in order to exchange data.

Developing collection-level descriptions

The RSLP schema was designed to provide a consistent form of collection-level description for collections which were the focus of RSLP project work. While it implemented most of the Heaney model, it did not do so in full. The RSLP schema was subsequently used as the schema for a number of collection-level databases, including:

- RASCAL (Research and Special Collections Available Online), an electronic gateway to research humanities and social science resources in Northern Ireland.[9]
- Backstage, a cross-domain database of performing arts collections in the UK. The database also contains item-level descriptions (based on the General International Standard Archival Description (ISAD(G)) for around half of the collections.[10]
- Cecilia, an online guide to music collections in archives, libraries and museums in the UK and Ireland.[11]
- 'Find it in London', a pilot project, with partners from the public, higher education and further education library sectors as well as archives and museums, to design a database of collections in London.[12]
- Collections Wales, a bilingual online database of descriptions of research collections in Wales.[13]
- Crossroads, a prototype website that contains collections related to the potteries industry in north Staffordshire.[14]
- EnrichUK, a database providing the gateway to the collections of 150 sites developed through funding from the New Opportunities Fund (NOF) Digitise programme.[15]
- Revealweb, a database of collections of materials that can be used by visually impaired people. Collections include material in Braille and Moon, spoken word recordings, large print, tactile maps, diagrams and other images, tactile and audio music scores and electronic texts.[16]
- Cornucopia, a database developed by the Museums, Libraries and Archives

Council (MLA). Currently it lists only museums but the intention is to extend the database to cover libraries and archives.[17]

The RSLP schema is also the basis for the Dublin Core Collection Description Application Profile (DC CD AP). This is currently under development by the Dublin Core Collection Description Working Group and a draft version will be submitted to the Dublin Core Usage Board during 2004. The application profile records only the elements of the schema that directly describe collections. It currently omits the elements that describe associated agents (eg owner, collector) since these are within the remit of other Dublin Core working groups.

The SCONE project, a database of descriptions of collections held in Scottish archives, museums and libraries, and collections about Scottish issues held elsewhere, was a fuller implementation of the Heaney model.[18] Additionally, the Joint Information Systems Committee (JISC)-funded CC-interop project[19] is investigating collection description schemas in relation to both clumps (groups of metadata resources (eg electronic catalogues) which can be searched together) and COPAC, and is examining how the SCONE database can be extended to meet requirements for incorporating COPAC into clumps and participating in the JISC Information Environment.

The RSLP Collection Description schema

The RSLP Collection Description schema[20] is a metadata schema that identifies specific elements or attributes of a collection. Each attribute has a label (eg *Title*) and an Resource Description Framework (RDF) property (eg dc:title). The schema can be used to design a relational database to hold the descriptions but is not tied to any specific software.

The first group of elements describes the collection or sub-collection. The *title* element records the official name of the collection, while the *description* is a piece of free text prose identifying the major features of the collection. The *strength* element records the subject focus and depth, while *physical characteristics* identifies the format(s) of items in the collection. The range of dates over which a collection was accumulated is recorded in *accumulation date range* and the range of dates of individual items in the collection is recorded in *contents date range*.

The second group of elements provides searchable access points. The *Concepts* element holds keywords or subject terms relevant to the collection (eg topics, objects, places and time periods) and *language* records the language (when relevant) of items in the collection. The *names* element records personal and corporate names associated with the collection.

A third group of elements records information about relationships between the collection being described and other resources. These elements record whether there is a *sub-collection* or a *super-collection* (ie parent collection), a *catalogue*, and any *associated collections* and/or *associated publications*.

The fourth group of elements is concerned with the management of the collection. The *accrual status* element records whether material is still being added to the collection and whether this is by purchase or donation, etc. Other elements record the *legal status* of the collection and its *custodial history* or provenance. Details of who is allowed to use the collection and under what conditions are recorded in *access control*, while *location* attributes hold details of the physical or digital location, including postal address and access conditions (hours of opening, disabled access).

The fifth set of elements records details about agents connected with the collection—the *owner(s)*, the *collector(s)* and the *administrator(s)*. For each agent a number of attributes can be recorded: name, organisation, role and contact details.

The RSLP schema provides only the set of elements. It does not prescribe whether any element should be mandatory, or whether it can be repeated, or the exact way that data should be recorded in any particular element. This level of specification is left to the implementer. So an implementer may specify that controlled vocabulary must be used in a particular element; eg using the Art & Architecture Thesaurus (AAT) in choosing terms for the concept element. Implementers often also include additional elements, usually those relating to collection management.

The schema also does not prescribe how the elements comprising a description are displayed to the database user. Implementations may display title and description initially, with an option to see the full record. It is possible to display different combinations of elements to different users by setting up a number of search pages (so that the children's search page would not display legal status, custodial history, etc) and by providing alternative description elements for different users (children, adults, researchers).

What is a collection?

There is no set size for a collection and, in theory at least, a collection could comprise a single item. Collections can be located in a physical or digital location, or a set of related physical or digital locations. Collection contents can also be either physical and/or digital in format. They also need not be permanent—eg exhibitions are temporary collections—though if descriptions of such collections are included in a database care would need to be taken that these are removed either from the database or at least from public view once the exhibition is over.

Descriptions can also be created for sub-collections (and sub-sub-collections) and for catalogues and finding aids, which are also types of collection. These descriptions can all be linked to one another in a database allowing the user for example to navigate from the description for the University of Bath Library to the description for its sub-collection Pitman Collection and from both of these to the description for the library's online public catalogue (OPAC), which may contain a link enabling the user to then enter and search the OPAC.

Collection descriptions and archives

So what relevance does collection-level description have for archives? From an archivist's perspective, collection description is already being done. Archival finding aids already relate items to collections, and describe collections. If they are in electronic form, using ISAD(G) and Encoded Archival Description (EAD), they can be cross-searched with other catalogues and finding aids in the same format. Indeed, archivists are already submitting descriptions of their holdings to the National Register of Archives, contributing to one or more of the Archives Hub, the Access to Archives initiative or the Research Libraries Group's (RLG) Archival resources.

However, this only supports the user within the archives domain. Collection-level descriptions provide a higher-level route of access to the user, particularly at initial stages of research or enquiry and where resources from more than one domain may

be of interest. So it is important that archival resources are described alongside museum and library resources in databases such as Cecilia, RASCAL and Backstage.

Collection-level description is not a substitute for item-level description but an effective complement to it. It may be useful in providing an initial description when item-level catalogues and finding aids are not available, but there remains a need to complete the task and create item-level descriptions when resources are available.

While collection-level descriptions are a valuable resource, that does not imply that archives should start to use the RSLP metadata schema in place of their existing documentation formats. But by mapping the RSLP schema to ISAD(G) and EAD, archivists, unlike museums and libraries, can contribute collection descriptions to cross-domain databases without having to create them from scratch.

NOTES

1. A. Chapman, N. Kingsley and L. Dempsey, *Full Disclosure: realising the value of library and archive collections* (1999), available at <http://www.ukoln.ac.uk/services/lic/fulldisclosure/> (accessed 20 July 2004).
2. Cultural Heritage Consortium, *Full Disclosure Prioritisation Study. Final Report* (2002), p 21, available at <http://www.bl.uk/concord/pdf_files/fdigpriorityfinal.pdf> (accessed 20 July 2004).
3. Chapman et al, *Full Disclosure* (1999).
4. Higher Education Consultancy Group/CHEMS Consulting, 'A report to the RSLP on: Barriers to resources sharing among higher education libraries' (2002), available at <http://www.rslp.ac.uk/circs/2002/barriers.htm> and Higher Education/British Library Task Force, 'Coordinating the distributed national collection of research resources: Report to the Research Support Libraries Group' (June 2001), available at <http://www.bl.uk/concord/pdf_files/blhe-dnc.pdf> (both accessed 20 July 2004).
5. P. Miller, 'Collected wisdom: some cross-domain issues of collection-level description', *Dlib*, vol 6 (September 2000), available at <http://www.dlib.org/dlib/september00/miller/09miller.html> (accessed 20 July 2004).
6. See the Joint Information Systems Committee (JISC), *Information Environment Development Strategy 2001–05 (draft)*, available at <http://www.jisc.ac.uk/dner/development/IEstrategy.html> and A. Powell and L. Lyon, *The DNER Technical Architecture: scoping the information environment* (UKOLN, 2001), available at <http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/dner-arch.html> (both accessed 20 July 2004).
7. IMLS Digital Library Forum, *A Framework of Guidance for Building Good Digital Collections* (2001), available at <http://www.imls.gov/pubs/forumframework.htm> (accessed 20 July 2004).
8. M. Heaney, *An Analytic Model of Collections and their Catalogues*, 3rd revised edn (UKOLN/OCLC, 2000), available at <http://www.ukoln.ac.uk/metadata/rslp/model/> (accessed 20 July 2004).
9. Available at <http://www.rascal.ac.uk/> (accessed 20 July 2004).
10. Available at <http://www.backstage.ac.uk/backstage/> (accessed 20 July 2004).
11. Available at <http://www.cecilia-uk.org/> (accessed 20 July 2004).
12. Available at <http://www.fiil.org.uk> (accessed 20 July 2004).
13. Available at <http://www.mappingwales.ac.uk> (accessed 20 July 2004).
14. Available at <http://www.crossroads-wm.org.uk> (accessed 20 July 2004).
15. Available at <http://www.enrichuk.net/> (accessed 20 July 2004).
16. Available at <http://www.revealweb.org.uk> (accessed 20 July 2004).
17. Available at <http://www.cornucopia.org.uk> (accessed 20 July 2004).
18. Details of the SCONE project are available at <http://scone.strath.ac.uk/> (accessed 20 July 2004).
19. Details of the CC-interop project are available at <http://ccinterop.cdlr.strath.ac.uk/> (accessed 20 July 2004).
20. Available at <http://www.ukoln.ac.uk/metadata/rslp/schema/> (accessed 20 July 2004).