

Collections and Collection Description

Pete Johnston & Bridget Robinson
Collection Description Focus

Introduction

The managers of the valuable resources held by museums, archives and libraries compile descriptions of their holdings both so that they can **administer** and control those resources more effectively and in order to enhance **access** to those resources by disclosing information about their existence and availability to potential users. The channels through which resource managers provide that information, and through which potential users seek information about resources are, increasingly, digital.

Although librarians, archivists and museum curators have all considered the items within their custodianship as forming groups or “collections” of some form, the criteria by which they define these groupings, and the emphasis placed on the creation and use of descriptions of those groupings, has varied widely. Different ideas about “collections” have led to different approaches to “collection description”. For archivists, the individual item is an integral part of a group of items that forms the record of an individual or organisation, and the description of such aggregates forms a fundamental (and standardised) element of descriptive practice. Traditionally, librarians have concentrated on the description of individual items. The notion of the “collection” is certainly present, with aggregates defined perhaps by various criteria including location, subject, form or use, but the descriptions of these aggregates tend to be more informal and less structured than those of their component items. Museums too employ the concept of the “collection”, and use a range of criteria (form or type of object, subject, the objects donated by an individual benefactor) to delimit the aggregates they describe and manage.

The practice of creating descriptions of sets or aggregations of the items or objects held within their repositories instead of, or in addition to, descriptions of the individual items is not new to the different curatorial traditions. Such an approach, however, is receiving renewed attention as a means of improving the effectiveness of **digital** resource discovery procedures for these **physical** items, particularly where users wish to search across the distributed holdings of several repositories. At the same time, there is growing recognition of the value of describing aggregates of **digital** resources. This paper presents a brief overview of what is meant by a collection and how collections are described. It explores briefly the differences in approach to collections and their description, and suggests that a

consistent approach to description at collection level is an important factor in initiatives which seek to provide integrated access to distributed resources, whether those resources are physical or digital.

What is a collection?

At the simplest level, one can conceive of a collection as any aggregation of individual items (objects, resources). This definition says nothing about the form or nature of those items: they may be physical or digital, and digital items may be surrogates of physical items or they may be “born-digital”, the primary manifestations of a work. A “catalogue” may be thought of as a collection, where the items are the catalogue records [1].

The definition is also neutral on the size of a collection: in theory at least, it is possible to have a collection containing only one item!

Collections may have varying degrees of permanence or transience. A collection of digital items may exist only while it is transferred between applications or for the duration of a user query; even for physical items, aggregates might be created for a limited period only. And the process of aggregation does not necessarily imply a physical juxtaposition: collections may be distributed, with the items dispersed across multiple physical locations.

Collection description and collection-level description

A description of a collection may include information about the aggregate as a whole, information about the individual items which make up the collection, or indeed information about some groupings of the items which form a subset of the whole.

On this basis, Heaney [2] suggests that “collection descriptions” (or “finding aids”) may be classified as belonging to a small number of types. The principal distinction is between an “analytic” finding aid, consisting of information about the individual items only, and a “unitary” finding aid, which only describes the collection as a whole. A “hierarchical finding aid” provides information about both the whole and the items, including contextual information about the relationship of the items to the whole. In practice, as Heaney acknowledges, even an analytic finding aid may have some structure that dictates that meaning is conveyed by the relationship between the descriptions of individual items.

Different conceptions of “collections” result in different approaches to describing those collections, so that individual “collection descriptions” can be classified as belonging to one of Heaney’s ideal types.

A further word of warning is perhaps necessary: the terms “collection description” and “collection-level description” are sometimes used interchangeably. The term “collection description” might be applied to any of Heaney’s types of finding aid; but a “collection-level description” supplies information (at least primarily) about the aggregate as a whole i.e. a collection-level description is, in Heaney’s typology, a unitary finding aid.

Why create collection-level descriptions?

Perhaps the most obvious benefit of collection-level description is that it can provide an overview (albeit a necessarily superficial one) of groups of otherwise uncatalogued items.

Even where item-level descriptions already exist, collection descriptions may be useful. For resource **discovery**, the existence of collection-level descriptions supports the high-level navigation of a large (and perhaps distributed and heterogeneous) resource base. For example, a researcher may make use of collection-level descriptions firstly to **discover** the existence of collections, and then to **target** their (item-level) queries to selected collections on the basis of their characteristics, or a software agent may perform these functions on behalf of a human user. Collection-level descriptions may be used to support controlled searching **across** multiple collections, and to assist users by reducing the number of individual hits returned in an initial response to a query [3].

Collection-level description has a potentially important role in supporting **cross-domain** resource discovery. Researchers want to discover and access resources drawn from across the collections of diverse institutions, and the technical infrastructure to support this is maturing. One of the challenges to be met is that (for good reasons) the managers of different classes of resource describe their items using different standards. Initiatives such as the Dublin Core seek to address this by defining a small set of elements, the semantics of which are commonly understood: they help to overcome the problems of differences in descriptive practice and terminology by serving as a “metadata pidgin” for the non-specialist user [4, 5]. Even with such support, however, a researcher may face the problems of managing large numbers of item-level “hits” in response to a query, where those hits describe **heterogeneous** resources. Description at collection level, using a common set of properties and some consensus on the criteria for defining collections, offers the possibility of comparing broadly **similar** high-level objects. Powell, Heaney & Dempsey employ a geospatial metaphor: “the scholar is concerned at the initial survey to identify areas rather than specific features - to identify rainforest rather than to retrieve an analysis of the canopy fauna of the Amazon basin” [6].

In addition to these benefits for resource **discovery**, description at collection level is an important component in collaborative approaches to resource **management**.

Archival collections

The archival community has not traditionally used the term “collection” to label the aggregates of material they typically describe. Archivists make the distinction between an archival fonds, where the items are of known provenance and their arrangement reflects their original working order as the records of an organisation or individual, and an “artificial collection”, where the items are associated but lack the coherence of a fonds [7, 8]. The archivist recognises the fonds as the set of items that have been created and accumulated by an identifiable individual body (or bodies). However, it should be emphasised that both these classes of aggregates (the fonds and the artificial collection) are “collections” in the more general sense in which the term is used here. Within an archival fonds, an item can be fully understood only within the context of its relationship with other items and aggregates in the fonds, and descriptive practice reflects this.

Description at the level of the aggregate (or rather at various levels, since descriptions of archives are usually arranged hierarchically) is a fundamental part of archival descriptive practice, and the archival community has well-established national and international standards for such “collection description”. Indeed the level of description provided by archival catalogues often stops short of the description of individual items, particularly where there are multiple instances of the same type of item.

The General International Standard for Archival Description (ISAD(G)) is a permissive standard which defines a set of data elements for archival description, to be deployed within a framework of multi-level description from the general to the specific i.e. the ISAD(G) element set may be applied to any unit of description from the whole to the item (though in practice some elements are more applicable at some levels of description than others). In Heaney’s typology, then, the descriptions of archives are typically hierarchic finding aids. These same principles underpin the design of the Encoded Archival Description (EAD) standard, which defines an element set and an SGML/XML DTD for the encoding of archival finding aids [9]. Like ISAD(G), EAD supports description either at collection level or at lower levels of detail i.e. it can represent both unitary and hierarchic finding aids. EAD was designed to capture a broad range of descriptive practice, and is sufficiently flexible to permit the encoding of a wide variety of catalogues and inventories. It is a measure of its designers’ success that it has been widely adopted as a means of digital data exchange for archival description.

Library collections

Libraries have used many different criteria to define the scope of, or delineate, their “collections” [10]. The first, which is often used implicitly, is that of **institution** or **location**: a collection is the totality of the holdings of a named library or repository.

Collections are often defined by the **subject** or coverage of the content of the items. A subject-based collection may coincide with the entire holdings of a library if the library is dedicated to collecting materials in a specific subject area, but more commonly it will be a subset of the larger (institution/location-based) collection. The items of a subject-based collection might be physically located

together (and this is more likely if the items are also related through some other association, perhaps with a collector or donor), but it is quite probable that they are dispersed throughout the library. Since subject schemes may be hierarchical, collections defined using such schemes may also have hierarchical relationships. However, subject-based collections are not discrete units in the way that archival fonds are. Instead, they form a set of overlapping “windows” on the holdings of one or more libraries: an item may form part of multiple subject collections, and the specialists in different subjects may present different perspectives on the collections of an institution. The relationships between subject collections may be complex.

Collections may also be defined by the **form** of the items (e.g. a video collection) or by some aspect of their **use** (e.g. items for the partially sighted, or items to which access or use is restricted).

Library collections may be distributed across several physical institutions. That distributed collection may comprise the entirety of the institutions holdings, but it is more likely to be a subset defined by one of the criteria noted above, such as subject.

Although libraries have recognised collections as units that they define and manage, their primary focus for resource description has been the item. Perhaps in part because of the diversity of criteria applied in defining library collections, collection-level description has tended to be informal, shaped by local conventions, and relatively unstructured. The description of library collections has not been subject to the standardisation which has been applied to item-level description in the form of MARC and AACR2, and standards for machine-readable collection-level descriptions have not been widely deployed.

Museum collections

The notion of the collection is a familiar one to museums and galleries, and the physical arrangement of objects within a museum may be based around collections and their curators. As in the library case, the criteria used to establish what constitutes a collection may vary. A collection may be the entire holdings of an institution, or it may be a subset of those holdings defined according to some other common attribute (subject, type or form of object, medium or technique etc.). In this context, one particularly important criterion is that of an aggregate of material collected by an individual and donated to the museum or gallery. And as with libraries, there may be complex relationships between these subsets: a collection defined by subject or object type may include subsets of items from the collections of several donors.

Museum collections may span the holdings of several institutions. This is particularly true as museums construct collaborative “virtual” collections composed of digital representations of physical objects housed in many different locations. Within this framework of the virtual collection, the researcher may wish to construct “collections” corresponding to criteria that match their own specific research interests. However, the “collection management systems” used within museums have tended to focus on the description of individual items or objects. There is a practice of describing aggregates of these objects, and there have been successful examples of

digital “guides” using collection-level descriptions as a gateway to the object-level databases and catalogues which describe the items within those collections [11]. As in the library case, however, there has been little effort to develop a standardised approach to collection-level description.

Digital collections

The discussions above focused on the description of collections of physical items. However, information managers from all three of the domains above are also facing the challenge of managing collections composed of **digital** items, which may be digital representations of physical resources or may be primary “born-digital” resources. Some collections may be “hybrid”, in the sense that they are made up of both digital and physical items. Some of these collections are made up of digital items that are descriptions of physical resources e.g. library OPACs or bibliographic indexes.

More broadly, the managers of Web resources need to describe aggregates of digital resources. Powell makes a high-level distinction between collections of Web accessible-items and collections of information about such items i.e. collections of metadata records. The diversity of the nature and use of Web-accessible resources means that the criteria for defining aggregates of these resources vary widely. As a consequence, there have been various attempts to provide ways of describing aggregates of Web resources, but often they have been shaped by a particular use requirement and none have been widely adopted [12, 13].

Collection description & the Research Support Libraries Programme (RSLP)

The Research Support Libraries Programme aims to facilitate arrangements for research support in UK libraries [14]. Two major strands of the programme emphasise collaborative arrangements for the management of collections and the improvement of information about collections in order to enhance discovery and access. The collections with which RSLP is concerned are primarily, but not exclusively, collections of **physical** items held by libraries, archives, museums and other specialist repositories. A consistent approach to the description of collections was considered important to the success of both of these goals, and RSLP supported a project to develop a model of collections (and their catalogues) and a metadata schema for the description of collections based on that theoretical model [6, 15]. The RSLP Collection Description schema provides a basis for creating relatively **simple** descriptions of collections of many different types. Just as the Dublin Core metadata element set is not intended to replace richer standards for resource description at item level, the RSLP CD schema is not a substitute for existing collection description schemas such as the archival description standards mentioned above. Like Dublin Core, however, it offers a simple set of attributes with commonly understood semantics which allows resource managers to disclose and exchange information about their collections.

Collection description & the JISC Distributed National Electronic Resource (DNER)

The Distributed National Electronic Resource (DNER) is a managed information environment that allows users in UK higher and further education to access quality assured digital resources from many sources and of many different types [16]. Collections of resources may be held and managed within local institutions, at central JISC services, or by external agencies. Although the collections with which the typical DNER user interacts in the first instance may be digital, they may use these collections as an intermediate step to accessing a physical resource. So a researcher might use (digital) bibliographic indexes and library catalogues in order to locate a (physical) copy of a book or journal. In this sense, the DNER is concerned with both **digital** and **physical** collections: it is a "hybrid" information environment. Within the technical architecture envisaged for the DNER, a collection description service will provide information about the collections available [17]. The users of this service may be human researchers, but they may also be software agents acting on behalf of users, perhaps according to preferences or restrictions specified in a personal or institutional profile. Such a service requires machine-readable collection descriptions that are consistent in their structure and semantics, and the RSLP CD schema is being evaluated for its suitability in this context.

Summary

The idea of describing aggregates of resources as a unit is not new. However, recent emphasis on the collaborative management of distributed collections and on providing integrated access to distributed collections of heterogeneous items has generated renewed interest in description at collection level. Initiatives such as the RSLP and the DNER highlight the value of a shared and consistent approach to the creation of collection level descriptions that are both human- and machine-readable, for resource discovery and for resource management. If collection level descriptions are to support the discovery of resources from across the holdings of diverse institutions, a common approach to the creation of those descriptions is essential.



Collection Description Focus is a national post, jointly funded for a twelve-month period by the Joint Information Systems Committee/Distributed National Electronic Resource (JISC/DNER), the Research Support Libraries Programme (RSLP) and the British Library. The Focus aims to improve co-ordination of work on collection description methods, schemas and tools, with the goal of ensuring consistency and compatibility of approaches across projects, disciplines, institutions and sectors. The Focus provides support both for projects actively involved in collection description work and for those investigating or planning such work. The Focus is located within UKOLN, and physically based at the University of Bath.

References

- [1] Powell, Andy. Introduction, in Powell, Andy, ed. Collection Level Description: a review of existing practice. An eLib supporting study. UKOLN, August 1999. HTML: <http://www.ukoln.ac.uk/metadata/cld/study/>
- [2] Heaney, Michael. An Analytic Model of Collections and their Catalogues. UKOLN/OCLC. Third edition, revised. January 2000. HTML: <http://www.ukoln.ac.uk/metadata/rslp/model/>
- [3] Miller, Paul. Collected Wisdom: Some Cross-domain Issues of Collection Level Description, Dlib, Vol 6 No 9. September 2000. HTML: <http://www.dlib.org/dlib/september00/miller/09miller.html>
- [4] The Dublin Core Metadata Initiative. HTML: <http://dublincore.org/>
- [5] Baker, Thomas. A Grammar of Dublin Core, Dlib, Vol 6 No 10. October 2000. HTML: <http://www.dlib.org/dlib/october00/baker/10baker.html>
- [6] Powell, Andy, Michael Heaney and Lorcan Dempsey. RSLP Collection Description, Dlib, Vol 6 No 9. September 2000. HTML: <http://www.dlib.org/dlib/september00/powell/09powell.html>
- [7] Methven, Patricia. An Archival Perspective, in Powell, Andy, ed. Collection Level Description: a review of existing practice. An eLib supporting study. UKOLN, August 1999. HTML: <http://www.ukoln.ac.uk/metadata/cld/study/>
- [8] International Council on Archives. ISAD(G) : General International Standard Archival Description. Second Edition. ICA, 2000.
- [9] Encoded Archival Description. Official Web Site. HTML: <http://www.loc.gov/ead/>
- [10] Murray, Jim. A Library Perspective, in Powell, Andy, ed. Collection Level Description: a review of existing practice. An eLib supporting study. UKOLN, August 1999. HTML: <http://www.ukoln.ac.uk/metadata/cld/study/>
- [11] Dunn, Heather. Collection Level Description - the Museum Perspective, Dlib, Vol 6 No 9. September 2000. HTML: <http://www.dlib.org/dlib/september00/dunn/09dunn.html>
- [12] Powell, Andy. An Internet/Web perspective, in Powell, Andy, ed. Collection Level Description: a review of existing practice. An eLib supporting study. UKOLN, August 1999. HTML: <http://www.ukoln.ac.uk/metadata/cld/study/>
- [13] Hill, Linda L., Greg Janée, Ron Dolin, James Frew and Mary Larsgaard. Collection Metadata Solutions for Digital Library Applications, Journal of the American Society for Information Science, 50 (13): 1169-1181. November 1999
- [14] The Research Support Libraries Programme. HTML: <http://www.rslp.ac.uk/>
- [15] RSLP Collection Description project. HTML: <http://www.ukoln.ac.uk/metadata/rslp/>
- [16] The JISC Distributed National Electronic Resource. HTML: <http://www.jisc.ac.uk/dner/>
- [17] Powell, Andy and Liz Lyon. The DNER Technical Architecture: scoping the information environment. HTML: <http://www.ukoln.ac.uk/distributed-systems/dner/arch/dner-arch.html>

<http://www.ukoln.ac.uk/cd-focus/>
<mailto:cd-focus@ukoln.ac.uk>



The University of Bath
Bath, BA2 7AY
ukoln@ukoln.ac.uk
tel 01225 386580 -
<http://www.ukoln.ac.uk>

UKOLN is funded by Resource: The Council for Museums, Archives & Libraries, the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils, as well as by project funding from the JISC and the European Union. UKOLN also receives support from the University of Bath where it is based.

