



The RSLP Collection Description Schema

Pete Johnston (UKOLN)

Introduction

The **Research Support Libraries Programme** (RSLP) aimed to facilitate arrangements for the support of academic research in UK libraries [1]. Two major strands of the programme emphasised 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 was concerned were primarily, but not exclusively, collections of **physical** items such as books, periodicals, manuscripts, archival records, held by libraries, archives and other specialist repositories.

Libraries have long recognised “collections” as units that they define and manage, and have applied a range of criteria to delineate those aggregates. It is common to find library collections defined either by **institution** or **location** (i.e. a collection is the totality of the holdings of a named library) or by the **subject** of the content of items. A subject-based collection may coincide with a library’s entire holdings if the library is dedicated to collecting materials in a specific subject area, but more commonly it will be a subset of that larger collection. The items of a subject-based collection might be physically located together, but it is more likely that they are dispersed throughout the library. Since subject schemes may be hierarchical, collections defined using such schemes may also have hierarchical relationships [2].

Outside the archive community, until recently at least, collection-level description had tended to be informal, shaped by local conventions, and relatively unstructured. The RSLP programme coincided with a growing interest in the role of collection-level description in enhancing resource discovery in a networked environment [3]. Collection-level description can provide an overview of groups of otherwise uncatalogued items, but collection-level description can also **complement** item-level description by supporting the high-level navigation of a large (and perhaps distributed and heterogeneous) resource base. Particularly in the context of cross-domain resource discovery, item-level metadata records may describe diverse resources, using different metadata schemas appropriate for those items. 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 as a first step in delivering integrated access to distributed resources [4]. Heaney employs a geospatial metaphor for this phase of the resource discovery process: “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” [5]

Within RSLP, projects were typically describing collections of physical items and collections of digital metadata records describing those physical items (catalogues). A consistent approach to the description of collections was considered fundamental to the success of RSLP’s objectives regarding the management of and access to research collections. RSLP supported a project to develop a model of collections and their catalogues (developed by Michael Heaney of the University Library Services Directorate, University of Oxford, with some additional funding from OCLC) and a metadata schema for the description of collections based on that theoretical model (developed by Andy Powell of UKOLN) [5, 6, 7].

The RSLP Collection Description model

The RSLP collection **model** adopts the view that the term “collection” can be applied to any aggregation of individual items, where those items may be physical or digital. Some digital items may be **representations** or “surrogates” of physical items; some items may be **descriptions** of other physical or digital items (i.e. a catalogue may be considered a collection where the constituent items are metadata records).

The model is silent on the criteria for delineating a collection and on the size of a collection. Inevitably, Heaney suggests, the decision on what constitutes a collection is conditioned by pragmatism:

Where an institution can choose between different degrees of aggregation in determining what are its Collections, there is no structure inherent in the model that requires or predisposes a particular level of aggregation. The institution should base its choices on its own pragmatic grounds, such as the level of detail required to make explicit those elements of the Collection-Description that the institution deems to be useful or necessary for the purposes of resource discovery or collection management, i.e. institutions should adopt a *functional granularity* approach [5].

If it serves the purposes of resource disclosure or resource management to treat this aggregate of items as a unit in order to describe characteristics of this unit, then it is a collection.

The model presents a view of the collection as an entity that has relationships with a number of other entities. A simplified view of the primary entities and their relationships is presented in Figure 1:

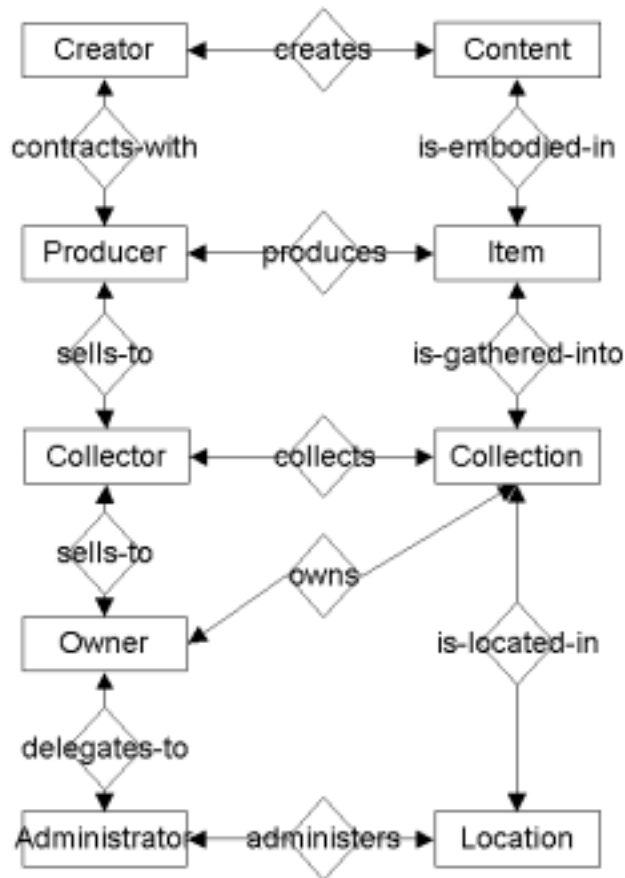


Figure 1 : RSLP Collection Description model

For each of these classes of entity, the model identifies the "attributes" which characterise each of the entities, based on an analysis of a number of widely used metadata schemas used to construct descriptions of those entities (or of comparable entities). In most cases attributes are associated with a specific subtype of the entity: for example, an instance of the Location class may be a "Physical repository" or an "Electronic repository", each with different attributes. Since the model is concerned primarily with the description of collections rather than the items that make up those collections, it considers the attributes of the "Content" and "Item" entities only in so far as their attributes impinge on the description of the Collection.

The model treats the relationships between entities as objects which themselves have attributes. For example, an "Administers" relation exists between the Administrator and the Location entities, and this relation carries an "Access conditions" attribute (a description of the hours of access, and any other restrictions). This means that if necessary it is possible to describe variations in the access conditions to the same Location according to the availability of a particular Administrator.

The model focuses primarily on the description of a single instance of a Collection. However it also identifies relationships that may exist between instances, and which are not part of this "internal" structure, for example,

- one collection is a version of a second collection (has Version/is Version Of);
- one collection is part of a second collection (has Part/is Part Of);
- two now separated collections were formerly part of a cohesive whole (has Complement);
- two collections are related through some other common association (has Association);
- one Collection is described by a second Collection, where that second Collection is a Catalogue (is Described By).

The RSLP Collection Description schema

Using this entity-relation model as a basis, the RSLP Collection Description (CD) **schema** provides a clearly defined set of metadata elements or properties that can be used to create relatively simple descriptions of collections of many different types. Each of the properties is uniquely identified [7]. Following the typology of collection descriptions (or finding aids) developed by Heaney, the schema enables the creation of a **unitary finding aid**, i.e. one which provides information about the collection as a whole, a "collection-level description".

Like the model, the schema addresses the description of a number of related entities, and provides metadata elements to describe a number of different types of resource. The RSLP schema is both narrower in scope and less expressive than the full RSLP model. Firstly, the schema provides properties to describe only a subset of the entities identified by the model. Secondly, although the relationships between those entities are as described in the model, the schema does not represent those relationships as resources that have attributes of their own. So, taking the Location-Administrator example above, in the schema, a Location has a property `rslpcd:administrator` which relates the Location to its Administrator(s), but the `rslpcd:accessConditions` property is simply another property of the Location. Thirdly, for those entities covered by the schema, it does not explicitly distinguish all of the "subtypes" described in the model.

The schema includes properties to describe:

- attributes of the Collection
- attributes of the Location of the Collection
- attributes of the Collector of the Collection
- attributes of the Owner of the Collection
- attributes of the Administrator of the Location of the Collection
- the relationships between these different entities
- the "external" relationships between multiple Collections

Wherever possible, the schema uses properties from existing metadata schemas, particularly the Dublin Core Metadata Element Set, including the use of element refinements and encoding schemes [8]. The "agent" properties are taken primarily from the vCard vocabulary [9]. Generally, properties are repeatable, e.g. a Collection can have multiple types, or have multiple subject terms associated with it. It should be noted that in some cases a single property is associated with the description of multiple resources: for example, occurrences of the `dc:title` property may be associated with both the Collection and the Location.

The data model for the schema is represented in Figure 2:

The schema as defined is permissive. The data entry guidelines for the schema suggest that a minimal description should include either the title of the collection and a short description, the name of the location, and the address of the location (for a physical location) or its locator (for a digital "location") [10]. The guidelines also make some recommendations for the control of the values of some of the descriptive properties, in the form of formats for dates and some controlled vocabularies. Within the context of a particular application, implementers supplement these high-level guidelines with more specific rules to specify how the schema should be deployed to meet the functional requirements of that particular application. For example, implementers would probably wish to mandate the use of specific controlled vocabularies, particularly for properties that are used as access points for browsing and searching.

A collection-level description conforming to the schema can be represented in many different forms. The RSLP Collection Description project recommended an XML **encoding** for the schema that made use of the syntactic conventions specified by the Resource Description Framework (RDF), the W3C's recommended language for expressing metadata, and the project created a simple form-based tool for the creation and editing of descriptions using this syntax [12].

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 established standards for archival description. Like Dublin Core, however, it offers a fairly **simple** set of attributes with commonly understood semantics that allows resource managers to disclose and exchange information about their collections. Mappings from two of the primary standards for archival description (ISAD(G) (2000) and EAD (1.0)) to the RSLP CD schema are available [13, 14].

A simple Dublin Core metadata record describing a collection could also be derived from the collection description part of an RSLP CD schema "collection description" using the DC "dumb-down" rule [15]. For such a transformation to be effective, it is critical to remember that an RSLP CD schema "collection description" is a description of multiple resources.

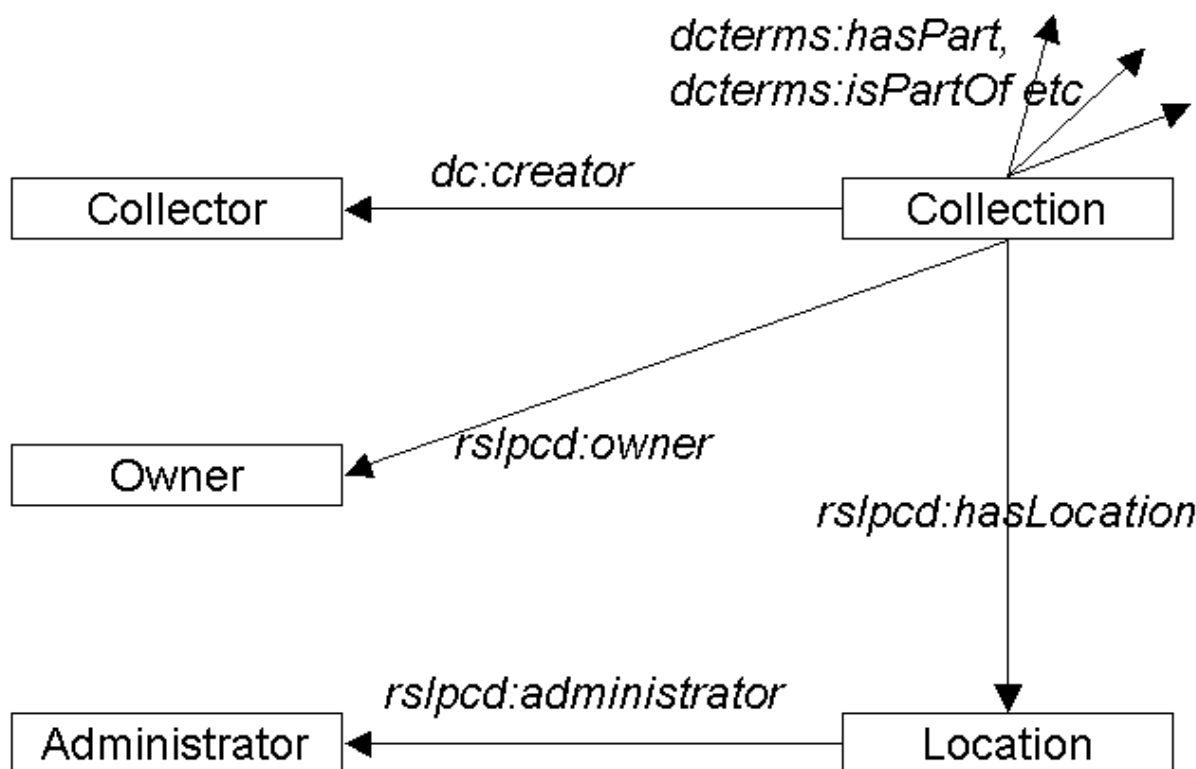


Figure 2: Data model for RSLP CD Schema

Describing the Collection

The Collection description properties are intended to be generically applicable to a wide range of collection types. Some properties are directly related to the content of the items in the collection and/or the process of their aggregation into a collection, others (like “Description” and “Strength”) represent attributes assigned by the metadata creator to support the disclosure/discovery, management and use of the collection.

Subject/coverage properties, and also language, type and date properties, are typically used as access points for indexing and the values of these attributes are usually drawn from controlled vocabularies or subject to the use of formatting rules so that the values can be interpreted precisely.

The schema distinguishes between access controls on the Collection itself (perhaps because of the status or physical condition of the items), and conditions of access to the Location of the Collection.

General descriptive properties of the Collection

Identifier	dc:identifier	A formal identifier for the collection.
Title	dc:title	The name of the collection.
Description	dc:description	A description of the collection.
Strength	rslpd:strength	An indication (free text or formalised) of the strength(s) of the collection.
Physical Characteristics	dc:format	The physical or digital characteristics of the collection.
Language	dc:language	The language of the items in the collection.
Type	dc:type	The type of the collection.
Access Control	rslpd:accessControl	A statement of any access restrictions placed on the collection, including allowed users, charges, etc.
Accrual Status	rslpd:accrualStatus	A statement of accrual policy (closed, passive, active, partial/selective), accrual method (purchase, deposit) and accrual periodicity (closed, irregular, periodic).
Legal Status	rslpd:legalStatus	A statement of the legal status of the collection.
Custodial History	rslpd:custodialHistory	A statement of any changes in ownership and custody of the collection that are significant for its authenticity, integrity and interpretation.
Note	rslpd:note	Any general information about the collection.
Accumulation Date Range	rslpd:accumulationDateRange	The range of dates over which the collection was accumulated.
Contents Date Range	rslpd:contentsDateRange	The range of dates of the individual items within the collection.

The RSLP Collection Description project defined a short enumerated list of collection types that can be used as a controlled vocabulary for the dc:type property of the Collection [16].

Subject/coverage of items in Collection

Concept	dc:subject	A concept (keyword) of the items in the collection.
Object	rslpd:objectName	An object name associated with the items in the collection.
Name	rslpd:agentName	A personal or corporate name associated with the items in the collection.
Place	dcterms:spatial	The spatial coverage of the items in the collection.

Relationships between Collection, Location, Collector and Owner

Location	rslpd:hasLocation	The identifier for the physical or online (digital) location of the collection.
Collector	dc:creator	The identifier for an agent who gathers (or gathered) the items in a collection together.
Owner	rslpd:owner	The identifier for an agent who has legal possession of the collection.

External Relationships

Sub-collection	dcterms:hasPart	The identifier or name of a second collection contained within the current collection
Super-collection	dcterms:isPartOf	The identifier or name of a second collection that contains the current collection.
Catalogue or description	rslpd:hasDescription	The identifier or name of a second collection that describes the current collection (for example, the catalogue for the current collection).
Described collection	rslpd:isDescriptionOf	The identifier or name of a second collection that is described by the current collection.
Associated collection	rslpd:has Association	The identifier or name of a second collection that is associated by provenance with the current collection.
Associated publication	rslpd:hasPublication	The identifier or name of a publication that is based on the use, study, or analysis of the collection.

Describing the Location of the Collection

While the Collection description properties are intended to be generically useful, the descriptive properties of the Location of the Collection are rather different for the two cases of a physical collection and a digital collection. For the user of a digital collection, it is (usually) irrelevant whether the networked file server on which the digital objects are stored is located in the same city or the same country; for the user of a physical collection, however, geographical location may be an important consideration in the process of selecting a collection.

General descriptive properties of the Location of the Collection

Identifier	dc:identifier	A formal identifier for the location.
Name	dc:title	The name of the location.
Access Conditions	rsrpdc:accessConditions	Hours of access, classes of permitted user, etc.

Relationships between Location, Collection and Administrator

Held collection	rsrpdc:isLocationOf	The identifier for a collection held at this physical or online (digital) location.
Administrator	rsrpdc:administrator	The identifier for an agent who has responsibility for the physical or electronic environment in which the collection is held.

External Relationships

See also	rsrpdc:seeAlso	The identifier of a resource that provides further information about this location (typically the URL for an organisational home page).
----------	-----------------------	---

The Location of a physical collection is usually one or more repositories, and the description of each location should include a postal address. Note that the `rsrpdc:locator` property is not used in the description of the location of a physical collection. The schema also distinguishes between the repository, and the organisation responsible for the place: the latter is the Administrator of the Location.

Properties of Physical Location

Postal address	rsrpdc:address	The full postal address for the physical location of the physical collection.
Post/zip code	rsrpdc:postcode	The post code or zip code for the physical location of the collection.
Country	rsrpdc:country	The country in which the collection is physically located.

For a digital collection, the Location is perhaps better conceptualised as the digital service which makes the collection available. It is common for digital collections to be made available through multiple digital services. For example, a catalogue (a collection of metadata records) may be available to a human user via a Web site (probably through a search or browse application), and it may also be available to software applications through a Z39.50 target or an OAI repository.

Properties of Digital Location

Locator	rsrpdc:locator	The online location (URL) of a digital collection.
---------	-----------------------	--

Describing the Collector and Owner of the Collection and the Administrator of the Location

The schema uses a common set of properties to provide simple metadata to describe these three types of agent:

Identifier	dc:identifier	A formal identifier for the agent.
Name	vcard:fn	The name of the agent.
Organisation name	vcard:org	The organisational name of, or affiliated with, the agent.
Role	vcard:role	The role (typically an organisational role) fulfilled by the agent.
Telephone number	vcard:voice	The telephone number of the agent.
Fax number	vcard:fax	The fax number of the agent.
Email address	vcard:email	The electronic mail address of the agent.

Where information on a named individual is likely to change, it may be more appropriate to describe an organisational role rather than a specific named individual. For the Collector of the Collection only, an additional property is available:

Agent History	rsrpdc:agentHistory	An administrative history of, or biographical details on, the agent.
---------------	----------------------------	--

The biographical or administrative history of a collector may be particularly important in providing context for the interpretation of certain types of collection.

Summary

The description of resources at collection-level is being recognised as an important component of information services that seek to provide integrated access to distributed resources. The RSLP CD schema supports the creation of simple collection-level descriptions that can cover a wide range of collection types and can be used in a variety of contexts. It is closely aligned with other metadata schemas for resource discovery, particularly the Dublin Core Metadata Element Set. It is intended to complement rather than replace other, richer, domain- or community- specific approaches. The schema has been deployed within the RSLP programme to describe collections of physical items held in libraries, archives and museums and more recently within the Enrich-UK portal to provide a single point of access to the distributed digital collections created within the NOF-Digitise programme [17].

The creation of the RSLP CD Schema has provided a starting point for the development of a Dublin Core Collection Description Application Profile (DC CD AP) by the Dublin Core Collection Description Working Group. At the time of writing (September 2003), this work is in progress. It is expected to be completed in 2004.

Collection Description Focus is a national post, jointly funded by the Joint Information Systems Committee, the British Library, and Resource. 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] The Research Support Libraries Programme. HTML: <<http://www.rslp.ac.uk/>>
- [2] Johnston, Pete and Bridget Robinson. *Collections and Collection Description*, Collection Description Focus Briefing Paper, No 1. January 2002. PDF: <<http://www.ukoln.ac.uk/briefings/bp1/bp1.pdf>>.
- [3] 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>>
- [4] 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>>
- [5] 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/>>
- [6] RSLP Collection Description project. HTML: <<http://www.ukoln.ac.uk/metadata/rslp/>>
- [7] RSLP Collection Description schema. HTML: <<http://www.ukoln.ac.uk/metadata/rslp/schema/>>
- [8] *DCMI Metadata Terms*. HTML: <<http://dublincore.org/documents/dcmi-terms/>>
- [9] *vCard MIME Directory Profile*. TXT: <<http://www.ietf.org/rfc/rfc2426.txt>>
- [10] RSLP Collection Description Data Entry Guidelines. HTML: <<http://www.ukoln.ac.uk/metadata/rslp/tool/?mode=printGuide>>
- [11] Beckett, Dave. *RDF/XML Syntax Specification (Revised)*. W3C Working Draft 23 January 2003. HTML: <<http://www.w3.org/TR/2003/WD-rdf-syntax-grammar-20030123/>> Latest version: <<http://www.w3.org/TR/rdf-syntax-grammar/>>
- [12] RSLP Collection Description tool. HTML: <<http://www.ukoln.ac.uk/metadata/rslp/tool/>>
- [13] ISAD(G) Second Edition to RSLP CD Schema mapping. HTML: <<http://www.ukoln.ac.uk/cd-focus/tools/isadg2rslp.html>>
- [14] EAD 1.0 to RSLP CD Schema mapping. HTML: <<http://www.ukoln.ac.uk/cd-focus/tools/ead10rslp.html>>
- [15] *DCMI Grammatical Principles*. HTML: <<http://dublincore.org/documents/principles/>>
- [16] RSLP Collection Description: CLDT enumerated list of collection types. HTML: <<http://www.ukoln.ac.uk/metadata/rslp/types/>>
- [17] EnrichUK. HTML: <<http://www.enrichuk.net/>>



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.