

Minutes of RSLP Collection Description Working Group held on 23rd March 2000 at The Charity Centre, London

Present: John Byford (British Library), Peter Clinch (Collaborative Collection Management of Foreign Legal Materials project), Gordon Dunsire (SCONE project), Lesley Forbes, Susan Jephcott (Mapping Asia project), Ronald Milne (RSLP, Chair), Lesley Pitman (COCOREES project), Andy Powell (UKOLN), Alason Roberts (CURL pamphlets project).

Contents

1. Preface to the meeting

- 1a. CURL interest in RSLP/UKOLN Collection Description work**
- 1b. Report from UKOLN**
- 1c. Details of new Mailbase discussion list**

2. Aims of the meeting

3. Issues raised in discussion document previously circulated

- 3a. Search Tool and Database issues**
- 3b. Standards issues**
- 3c. Pitching (functional granularity) issues**
- 3d. Directory information issues**
- 3e. Policy issues**

4. Follow-up actions

Minutes

1. Preface to the meeting

1a. CURL's interest in RSLP/UKOLN Collection Description work

Ronald Milne (RM) noted a need to publicise progress with the Model, Schema and Tool in order to provide clarification for the CURL community and to highlight the differences between the RSLP/UKOLN work and the NPO Register of Collection Strengths.

It was important to stress:

- that the work is being done to recognised metadata standards
- the special facilities of the Web tool (e.g. HTML and RDF generation etc.)

RM suggested that one or two project members might make a presentation about RSLP/UKOLN work at a future CURL board meeting. The next such meeting was likely to be in early June and it may be possible to have something to demonstrate by then.

Andy Powell (AP) explained that the true benefit of the Model, Schema and Tool was to provide a structured way of storing and exchanging records created by projects, in contrast to the NPO approach, which did not appear to allow for structured interchange of data. This meant that third party services could use the RDF output from the UKOLN Model, Schema and Tool as they wished, without being tied into RSLP HTML designs and layout. More information on RDF could be found at:
<http://www.ukoln.ac.uk/metadata/resources/rdf/>

Another advantage of RSLP/UKOLN work was to show how HEIs are working *together* on such a project. Additionally, RSLP/UKOLN Collection Description work would definitely mesh with other work being conducted at UKOLN and this would also extend the collaborative nature of the Programme.

It was **agreed** to take up the opportunity to present the aims of the RSLP/UKOLN work to CURL. Lesley Forbes (LEF) **agreed** to clarify the timetable of the NPO Project and advise on the list where RSLP project members can get full information on NPO and archival Collection Description work.

1b. Report from UKOLN

AP reiterated that the UKOLN work has three layers – the Model, Schema and Tool. The Tool may not necessarily implement the Schema in full – there are some limits on data entry, but the Schema should be correct – changes to the Tool could be made more easily. AP noted further that there were particular issues for projects dealing with dispersed collections within libraries which do not have a title as such (e.g. a “collection” of materials on *war studies* which may be scattered physically in a library amongst history, politics and psychology divisions etc.). It was not clear whether this could be incorporated into the Model and Schema. He explained that the Tool was currently suitable for entry of a single Collection Description; the resulting output in RDF or HTML form could be saved locally and used for whatever purpose. However, the Tool was not at present really suitable for multiple entries and records were not yet added to an underlying database.

AP also noted that the use of the term “cross-searching” had been misleading. Any gathering of data from different projects would be collected into a single central database which could then be searched as another resource itself; there was **no** capability for the kind of cross-searching of data from distributed projects such as was provided by a Z39.50 type mechanism. AP could work on developing a central Search Service using a set of sample records.

The Group **agreed** to provide some examples of Collection Descriptions through the Tool which could then be used by UKOLN to develop the demonstrator Search Tool. Examples would be provided by Mapping Asia and COCOREES projects – (and other projects if desired) within three weeks if possible. These examples might not include final implementations of the decisions by projects as to their own standards within each field, but would be the best possible in the time frame. Depending on decisions made by each

Project regarding use of standards, the process of providing examples might take longer but could be discussed through the new Mailing list (see below).

AP also **agreed** to make the current work on an Access database available through the UKOLN/RSLP web pages as soon as possible. Progress so far included identification of the tables required to implement the Schema and these could be made available for project members to adapt for their own database design, if so required. As yet, the mechanism for delivering RDF from the database was not available. Again this development would be notified through the new discussion list.

Gordon Dunsire (GD) also requested that if possible AP should include a description of where divergence might be permitted in the Model. He mentioned that SCONE's work would complement (rather than overlap) with the UKOLN work. For example, SCONE was investigating how links could be made from search results into relevant *subsets* (rather than just the top-level home page) of a local OPAC, as a Finding Tool deliverable. Further details on the SCONE project can be found at:
<http://scone.strath.ac.uk/>

This kind of information could also be shared on the list.

1c. Details of new Mailbase discussion list

AP mentioned that a Mailbase discussion list had now been set up for RSLP project members to discuss any issues relating to Collection Description work. This is rslp-cd@mailbase.ac.uk; members can join through the Web page at: <http://www.mailbase.ac.uk/lists/rslp-cd/join.html> or by sending an email message to mailbase@mailbase.ac.uk with this command in the body of the message:

join rslp-cd firstname lastname

where firstname and lastname are members' personal names.

2. Aims of the meeting

The Group agreed that the main aims of the meeting were:

- to be clear about the extent of UKOLN's remit and to identify any other mechanisms which might need to be put into place to advance the UKOLN/RSLP product.
- to see if any help could be provided to projects which needed to develop their own databases in addition to the tool provided by UKOLN.
- to make decisions about some elements of the Schema, still to be implemented (e.g. enumerated type listing), and to consider standards.

The Group then worked through the items on LEF's discussion document.

3. Issues raised in discussion document previously circulated

3a. Search Tool and Database issues

Based on **Question: Can more information be provided about the Search Tool for users, which has a crucial bearing on the input of data?** (Section B Item 1)

AP responded that he could develop this with sample data as had been discussed, but that each project had to take a view about how their own end-users may wish to access their own Collection Descriptions. It had always been envisaged that each project would want its own project search tool with individual functionality. In addition, management of data was more easily achieved through a local database. A subject based approach (such as Law, Asian and Eastern European Studies) with project Search Tools would fit into the RDN (Resource Discovery Network model). The potential central database would be a demonstrator NOT a service.

If a project needed to display certain information to the end users then corresponding fields needed to be available at data entry and search stages.

LEF described an example of users wishing to see a list of the ten largest Arabic collections in the country with some indication of numeric size of the collections. *Size* as an attribute of collection is a level of functionality which is currently not present in the Tool and which had not been envisaged. This led into a mini-discussion of strength of collections being measured by objective rather than subjective means. Other members felt that, for example, *rate of addition to collection* would be a more pertinent measure than size of the strength of a collection. This illustrated clearly the fact that each project would require **additional but differing** functionality.

However, AP said that if enough projects were adding the *same* extra functionality then this would be a candidate for inclusion into the central Tool. The Schema is extensible within each field. The Group agreed that the new rslp-cd list would be the best forum for projects to register their chosen additions to functionality, so that other projects could make comparisons.

AP also noted some database design considerations. Projects might choose to use MS Access or a ROADS SQL facility, for example. A benefit of ROADS was that a Web search interface was included in the package, but a disadvantage was that underlying databases could not be relational in this scheme, so there was much replicated data. The disadvantage of using MS Access was that providing a web interface is more tricky.

Again, GD mentioned that SCONE may be able to help with this and the group **agreed** that once again these issues could be discussed on the mailing list rslp-cd.

3b. Standards issues

Based on **Question: Should underlying standards be defined (and adhered to) in order to achieve consistency?** (Section B Item 4)

After discussion the Group **agreed** that the most that could be done centrally was that RSLP should encourage projects to use the following standards where possible, but it was noted that these have to remain optional. It was **agreed** that we should ask all projects to “register” their intended inclusion of standards on the rslp-cd mailing list by an agreed deadline.

- **Names** - AACR2 Chapters 20-24 suggested.
Some discussion arose about the size of the task at Collection Description level, compared to item level records. The Group **agreed** that projects should be encouraged to use Anglo-American Authority Files for the fullest form of the name.
- **Places as subjects** – Should be treated the same as under “concept” (see below)
- **Languages** – ISO three-letter code agreed on 18 February 2000
- **Time** – coverage of items in the collection – must be able to search date ranges, e.g. 12th-16th century. In MARC fixed fields the date of publication was essentially designed to describe single or multiple dates rather than formatted date ranges. Another possible suggestion was to use the date range encoding as recommended by the Dublin Core schema. There seemed to be no easy solution to this, and it was agreed that LEF should investigate further decisions made by the archiving community.
- **Material types** - it would be possible to enter the codes from the MARC leader field into the Tool’s *Physical characteristics* field. There was some discussion as to whether it was rare for a library ‘collection’ to mix up physical types to any extent, and that therefore, rather than use a coding system which might retrieve skewed results, (e.g. a monograph collection of several thousand texts with only two videos), the type could be reflected in the title of the ‘collection’.
- **“Concepts” (subjects)** – LCSH level 1 (as used in the Conspectus methodology) was suggested, but this was felt too prescriptive. A long discussion ensued but it was finally agreed that projects should be encouraged to take terms from LCSH, but alternatively, or additionally, projects may wish to use other controlled vocabularies (such as specialised thesauri, the Getty thesaurus etc.)
- **Other standards (e.g. for transliteration)** – these would have to be decided on a project by project basis.

- **Enumerated list of Collection Description types** –There was some concern that Dublin Core based types did not meet libraries needs. It was **agreed** that AP would edit the Types discussion document to mirror more closely the descriptions provided in the “What is a collection – a library perspective” from the eLib supporting study. This meant including the definitions for **dispersed, distributed** and **virtual** collections.
- **Establish a minimum level record (MLR)**
There was some discussion as to whether a minimum level record (MLR) should be established. It was felt that this would have to be decided on a project by project basis. Title (and/or) name of location and description were considered to be the basic fields.

3c. Pitching (functional granularity) issues

Based on **Question: The Tool is well adapted to describe discrete archival type collections, but it is more difficult to envisage the boundaries (or functional granularity) for library collections, particularly dispersed subject collections within libraries. Can guidelines to enable some uniformity/consistency of “pitching” of descriptions of dispersed subject collections be devised?** (Section B Item 3)

This was felt to be entirely up to each project. There was some discussion of display of search results and differing amounts of information provided, but it was not felt desirable to define these further.

3d. Directory information issues

Based on **Question: Can UKOLN/RSLP set up or provide (and maintain) a central database table for institutional directory-type details of libraries, which are common to all projects?** (Section A Item 3)

There were differing opinions on this issue. Some felt that there was a clear need for a central database which could be used not only by RSLP but by all Library, Archive, HEI projects. Others felt that directory information should and could only be provided and maintained locally, and that RSLP projects should only point to that information.

If a central service was possible, who would create and maintain such a resource? Some possible contenders were mentioned as bases from which to start such a compilation: e.g. SCONUL, HEFCE databases, but no-one knew of an up-to-date comprehensive service for the UK as a whole. GD mentioned that a central directory system existed in Scotland for Scottish institutions, (as part of the Digital Scotland initiative?) but the issue of maintenance had not been fully addressed.

It was **agreed** that because this information was needed now, each project would have to compile their own contact information, but with as much inter-project co-operation as possible.

The group also discussed using ILL Codes to provide a unique key number for each institution within the central database, which could act as a future lead into a proper relational database. AP agreed to look into this.

3e. Policy issues

Based on **Question: Is establishing a description of collections framework for collections of all sorts, e.g. libraries, archives and museums (particularly dispersed subject collections within libraries) a national issue?** (Section A Item 2)

The group **agreed** that a **Collection Description framework** was a national issue and recommended that the relevant bodies should be urged to promote better co-ordination of the different programmes working in this area as a matter of urgency. The British Library was already working in this area through its Co-operation and Partnership programme.

Please see next page for Follow-up actions to meeting

4. Actions

Andy Powell/UKOLN

- Work on sample records provided by Projects as agreed, to develop Search Demonstrator Tool for mini demonstration purposes, e.g. to CURL in early June.
- Publicise existence of rslp-cd Mailbase list through rslp-announce list.
- Make the provisional Access database available on the RSLP web pages, and notify this development through the rslp-cd and rslp-announce lists.
- Note additional functionality 'requirements' posted to list by project members and revise Schema where possible – the proviso being that not all of these additions will necessarily be adopted.
- Look into ILL codes to provide a unique identifier for each institution.
- Look into incorporating dispersed, distributed and virtual collections types into Model and Schema.

Ronald Milne/RSLP

- Approach CURL offering to describe the aims of the UKOLN/RSLP Collection Description work and arrange a demonstration of progress with the Search Demonstrator Tool.
- Investigate ways of advancing issues of national policy on Collection Description work within the DNER, and also the matter of a common directory information service for libraries and archives, with the relevant bodies.

Project members

- Join **rslp-cd** mailing list.
- Use **rslp-cd** list to register their use of:
 - chosen additions to functionality of schema in local database design, to assist other projects in sharing information where required, and for possible inclusion by UKOLN.
 - chosen standards (as discussed in section 3b above) by a certain date (2 May 2000 is suggested).
- Clarify timetable of NPO Project and advise on the list where Project members can get full information on NPO and archival Collection Description work (LEF).
- Approach archival projects to find specific information on standards chosen for data entry guidelines, especially with regards to date range and subject standards (LEF).
- Work as much as possible with other projects to co-operate on sharing directory type info, using the rslp-cd list to share progress.
- Use the rslp-cd list to discuss issues of database design and software choices etc.
- Provide examples of Collection Description data (up to 30 examples in total from **Mapping Asia, IALS and COCOREES**) for UKOLN to enter into the Tool, within the next three weeks or as soon as possible.