

eLib Concertation day for Collection Level Description.

ULCC, 7th March 2000

Twenty eight people attended this event at the University of London Computing Centre. The event began at 10.30 with introductions from those attending and a brief outline of their interest in collection description. This ranged from a general level of interest to developing a scheme for use in a particular project.

Chris Rusbridge, the eLib Programme Director, outlined three issues which it was hoped could be discussed throughout the day:

- Why do we have collection descriptions? What are they for and what do they do?
- What relation do collection level descriptions bear to Dublin Core?
- What happens next? Do activities have to be separate / private or should they be shared / public?

Andy Powell noted that the principle of a Dublin Core Collection Description Working Group had been agreed at the last Dublin Core workshop in Frankfurt. In addition, an eLib Working Group on this topic had been set up following a UKOLN-organised meeting in 1998 and had produced a draft proposal <http://www.ukoln.ac.uk/metadata/cld/simple/>. Much of this work had been based on a review of other standards, e.g. Dublin Core and ISAD(G) and had been deployed in particular within the eLib projects RIDING and Agora.

The event was broken into four sections: eLib projects, JISC/HE activities, work outside HE and a final discussion section.

1. eLib work

The first section of the event consisted of presentations of developments from four eLib projects, one clumps project (RIDING) and three hybrid library projects (Agora, MALIBU and HeadLine).

RIDING <http://www.shef.ac.uk/~riding/>

A RIDING scheme <http://www.shef.ac.uk/~riding/cld/eldschem.html> based on the eLib Working Group's draft proposal has been developed as one of the project's deliverables. It is a recognition of the presence of uncatalogued material in many libraries, and also the difficulty of retrieving items as part of a collection. The need for a controlled list for subject descriptions has been recognised and is a possible future development.

There are currently 60 existing collection descriptions at the RIDING site.

Agora <http://hosted.ukoln.ac.uk/agora/>

Agora sees a user's individual information landscape as an aggregation of resources, and that collection level descriptions are crucial in building this aggregation. There are approximately 50 descriptions in Agora, which cover a wider range of resources, but in less depth than RIDING. They are also based on the Working Group's proposal.

MALIBU <http://www.kcl.ac.uk/humanities/cch/malibu/>

Malibu's interest in collection description stems from a focus on resource discovery, metadata

and building an integrated catalogue. Their focus is on electronic materials, and they have identified a number of problems specific to this type of material including describing the same content in different formats, and the definition of different permissions and licensing arrangements.

HeadLine <http://www.headline.ac.uk/>

HeadLine's model of the hybrid library is based around a user's Personal Information Environment (PIE) and is designed to ease the transition between resource discovery and resource access, whether the materials held are electronic or print. HeadLine's descriptions include data type, date range and access restrictions. The project's work is not based on that of the eLib Working Group but comes from a relational database background. Some difference in usage of terminology was noted, for example, collection = service entity, resource = service instance.

2. JISC/HE work

The second section of the day focused more closely on work ongoing outside of eLib, but within the Higher Education community. Andy Powell from UKOLN led this section of the day, with an outline of the Research Support Libraries Programme (RSLP) collection description project <http://www.ukoln.ac.uk/metadata/rslp/>. This aims to ensure consistent collection description across RSLP as a whole to minimise duplication of effort. It is functionally concerned with finding and identifying collections in print or digital format, and is largely based on a library/archive view of collections. The scheme is based on Dublin Core wherever possible. The project aims to produce a central database of RSLP collection descriptions and a web-based tool to enable projects to add to this.

The Resource Discovery Network (RDN) <http://www.rdn.ac.uk/> is considering collection descriptions as a tool for building value-added services, such as the RDN ServiceFinder, the RDN ResourceFinder and the JISC Current Content Collection. These types of services are fundamental to the development of the DNER so that it seems sensible to begin thinking about a DNER-wide collection description service. This would require schema, tool(s) for creating descriptions and an architecture for sharing descriptions. The RDN proposes a centralised collection description database, hosted by the RDN for the DNER and based on the work done for RSLP. This will include a Web-based tool for creating RDF-based descriptions, which can also be gathered using a robot and accessed using LDAP.

In the general discussion session which followed, it was agreed that for the time being at least, there is no single correct answer as regards a collection description scheme, and plenty of room for experimentation. However, there is also room for sharing descriptions more widely. It was suggested that D-Lib magazine might be approached to host a collection of articles on the topic. This would need editing and a common vocabulary to be deployed.

It was also acknowledged that it is important to study how users use these kinds of services. What questions are conceptualised? It was felt to be important for users to be able to browse a hierarchical structure in order to facilitate information retrieval. The 'Mona Lisa' question was raised to illustrate the importance of appropriate granularity, or of collection descriptions linking through to item descriptions.

3. Work in other sectors

After lunch, the presentations continued, focusing on three initiatives which are ongoing outside of the HE sector.

Cornucopia <http://www.cornucopia.org.uk/>

Cornucopia is funded by the Museums and Galleries Commission. It is a database which will eventually provide a complete picture of the range of museum collections in the UK. Cornucopia will provide good quality information on museum collections for the public, and will contribute towards lifelong learning in museums. The database provides a data structure and preferred terminology which will feed into initiatives such as the 24 hour museum.

National Preservation Office: Register of Collection Strengths
<http://www.bl.uk/services/preservation/>

The register focuses on printed material and has developed from a CURL / British Library Working Group on retention. This will provide a register of collection strengths, retention and preservation intentions to inform librarians and end-users. A pilot database will be developed using CURL libraries as a testbed.

PRIDE <http://lirn.viscount.org.uk/pride/>

PRIDE is a European-funded demonstrator project to raise issues and questions. The project's directory holds collection descriptions and an RDF harvester gathers these. Integrated services provided include service to service authentication, profiles for groups of users, and collection descriptions. A live demonstrator is intended to be put into operation at the beginning of April.

4. Final discussion and actions

The final section of the day was a general discussion session, intended to look at collection description in the context of the DNER, and to devise a list of action points.

Reasons for building a collection description scheme might be:

- Disclosure
- Development of broker services
- Management of collections – identifying strengths and gaps

Differences between schema might arise depending on whether the descriptions are aimed at humans or machines. The core set needs to be simple but not simplistic.

There is a question of who should describe resources which are not owned but licensed. In the current situation this would obviously lead to multiple descriptions at different institutions.

Should descriptions be provided more centrally? e.g. should MIMAS provide a JSTOR description? (or should JSTOR provide a JSTOR description?)

This could be useful, but it was also recognised that individual institutions would need the facility to modify records and provide contextual information.

There is also a question about maintenance – how/how frequently will records be updated? Who will take on a quality control role?

There was also some discussion about whether it is too early to attempt to define a set of core fields. So far, projects have produced very small numbers of descriptions and there is a need to examine how users use this facility e.g. through search logs. However, there is some level of synergy between projects and core fields which can be identified.

Because of resource limitations, the RSLP demonstrator can't be expanded at this time, but the schema will be used to develop DNER portals which will be widened out to a full DNER collection description service.

It was suggested that these schema should be registered for comparison and to provide a central access point. We should also look beyond libraries.

Actions

1. Complete and publish the eLib supporting study on collection description (Andy Powell / eLib) <http://www.ukoln.ac.uk/metadata/cld/study/>
2. Propose a study into user search strategies (eLib/UKOLN)
3. Map existing approaches, looking beyond libraries alone (eLib/UKOLN)
4. Take forward the idea of subgroup of the Dublin Core Working Group, looking at collection description (UKOLN).
5. Suggest a themed issue for D-Lib with a collection of articles looking at collection description (UKOLN/eLib to co-ordinate)
6. Register of 'locations' across DNER/MLAC etc? (UKOLN/eLib)
7. Build a larger body of data for test purposes. The DNER Programme Director should take forward the idea of a central collection description service based on RSLP, NPO, JISC work etc....