

	<p style="text-align: right;">Date : 11/06/2007</p> <p>Description of collections in theory and practice</p> <p>Juha Hakala & Esa-Pekka Keskitalo The National Library of Finland Email: firstname.lastname@helsinki.fi</p>
Meeting:	125 Acquisition and Collection Development
Simultaneous Interpretation:	No
<p>WORLD LIBRARY AND INFORMATION CONGRESS: 73RD IFLA GENERAL CONFERENCE AND COUNCIL 19-23 August 2007, Durban, South Africa http://www.ifla.org/iv/ifla73/index.htm</p>	

1. Introduction

Compared with cataloguing of bibliographic resources, collection assessment and description – at least as a coordinated activity with a set of tools and common principles guiding the work – has started late. The practice of collection assessment started in the public libraries in the U.S.A. back in the 30s and 40s. But it was the information explosion of the 70s which led to the situation where the quality of collections became an important issue, and also one that was difficult to determine.

The pioneering collection description initiative, Conspectus, was launched at the beginning of the 80s. According to a Brief History of the RLG Conspectus¹,

"The RLG Conspectus Online, introduced in 1982, resulted from extensive, coordinated work by the university and college libraries who used it to assess their collections and collecting practices. For close to a decade, the RLG Collection Management and Development Program Committee used the file as a supporting tool in collective efforts to improve research collection management, development, and resource sharings."

Mary Bushing notes that the Conspectus Online database enabled libraries to create, store and manage conspectus data and to produce reports in both text and graphic formats². It is not surprising that the service became popular not only in the U.S.A. but also in some other countries including Australia and New Zealand. But while the Conspectus Online is still "alive" and being developed further by the current host organization, OCLC, it is not a global service like WorldCat.

¹ <http://www.rlg.org/conspechist.html>

² Bushing, Mary: "The Evolution of Conspectus practice in Libraries: The Beginnings and the Present Applications". <<http://klement.nkp.cz/Caslin/caslin01/sbornik/conspectus.html>>

During the recent years a number of national collection description initiatives have been launched in e.g. United Kingdom³ and Finland, but their technical basis is not Conspectus Online although they may share some of the best practices developed over years by the American colleagues.

2. Collection description metadata

Given that description of collections has a long history, the idea of defining an exchange format for this data occurred relatively late. We believe that there are at least three major reasons for this:

1. As long as RLG Conspectus was the only significant system containing significant amount of collection metadata, there was no need to share this information between systems.
2. The burden of designing a collection description metadata element set is non-trivial. There was no good platform upon which to build; for instance, MARC21 does not provide an ideal starting point.
3. While some other initiatives may have played with an idea of developing a metadata element set for description of collections, the RSLP project was the first one to develop such a set in the late 90s. The RSLP set⁴ proved to be very influential; many subsequent efforts such as the Michael project⁵ used RSLP as the starting point.

RSLP has two major strengths. It is based on a well designed analytical model of collections and their catalogues⁶. And it is built upon the Dublin Core Metadata Element Set, which can be extended to accommodate the resource description requirements of basically any community.

The need for developing a standard collection metadata element set could be easily seen from the eagerness with which the RSLP metadata element set was adopted. Pete Johnston from UKOLN took the initiative to develop RSLP specification into a Dublin Core Application Profile. This work, started in 2003, was completed in March 2007 after the approval of the profile by the DC Usage Board.

A Dublin Core application profile such as the DC Collections⁷ provides a good starting point for software developers. The profiles are a key part of the DCMI offering, and are usually well maintained. On the other hand, a formal analysis by the usage board guarantees that the profiles are “correct”; that is, aligned with the Dublin Core proper and other application profiles.

³ "RSLP Collection description". <<http://www.ukoln.ac.uk/metadata/rslp/>>

⁴ "RSLP Collection Description Schema". Last updated: May 2000.
<<http://www.ukoln.ac.uk/metadata/rslp/schema/>>

⁵ <http://www.michael-culture.org/en/home>

⁶ Heaney, Michael: "An Analytical Model of Collections and their Catalogues". Third Issue, Revised. Jan 14, 2000. <<http://www.ukoln.ac.uk/metadata/rslp/model/amcc-v31.pdf>>

⁷ "Dublin Core Collections Application Profile".

<<http://dublincore.org/groups/collections/collection-application-profile/2007-03-09/>>

The 15 Dublin Core base elements have been published also as a NISO standard⁸. When NISO launched its metasearch initiative, a decision was made to develop metadata element sets for description of both information retrieval services and collections available via them. Together, these two standards would define the metadata for metasearch engines (portals).

The developers of information retrieval service specification adopted ZeeRex⁹ as the starting point, whereas the collection specification was to build upon the Dublin Core profile, which was still under development when the NISO work started. Therefore the draft standard¹⁰ published in 2005 could not become a NISO standard prior to publication of the DC Collections. Now that the DC work has been completed, the intention is to finalize the NISO standard during Summer 2007.

Dublin Core and NISO specifications have been aligned, but the latter has some additional elements such as Completeness (subject and level) of collection. Since the guiding principle in creation of DC terms and application profiles is broad (resource description), adding new elements into the future versions of the profile should be easy, both from technical and political points of view.

As of this writing, all metasearch portals are based on proprietary metadata, and lack means of extracting or loading information retrieval and service specifications. Lack of established standards is a partial explanation to this problem, which makes it very difficult for libraries to cooperate in creation of portal metadata. In fact, libraries must rely on vendors providing e.g. the IR service specifications; a dependency that has in some occasions turned into a major shortcoming. Libraries should ask portal vendors to implement Z39.91 and Z39.92, so as to enable sharing of collection information in the same manner we share bibliographic information. This sharing will make it easier for us to provide to our customers access to the so called deep web, where search engines such as Google can not easily get.

3. The Finnish Collection-Level Description Project

In 2003, the Council for Finnish University Libraries launched a collection mapping project. The aim of the project was to create a coherent, uniform, and extensive general picture of collections in Finnish university libraries by means of quantitative and qualitative collection descriptions as well as collection evaluations. More details can be found in an article by Pentti Vattulainen.¹¹

The project was funded by the Ministry of Education. The project was managed by a steering group, and it had one employee as a planner and coordinator. Every participating library named a liaison who participated in a number of meetings that were held in order to increase awareness, interest and skills of librarians in charge of the collection maintenance and development. The Ministry extended its funding to early 2007. At the time of writing of this paper, the libraries are contemplating the options for the continuation of the work begun by the project.

⁸ "The Dublin Core Metadata Element Set".

<<http://www.niso.org/standards/resources/Z39-85-2007.pdf>

⁹ <http://explain.z3950.org/>

¹⁰ NISO Z39.91-200X: "Collection Description Specification. Draft Standard for Trial Use : Period: November 1, 2005 – October 31, 2006".

<http://www.niso.org/standards/resources/Z39-91-DSFTU.pdf>

¹¹ Vattulainen, Pentti: "Collection Map for the Finnish University Libraries - An Introduction". Signum, 3/2006. <<http://www.protsv.fi/stks/signumnew/200603/2.pdf>>

Practice of collection description was little known in Finland when the project was started. It was essential first to agree on common practices. Since we wanted to use standards whenever possible, the then latest Dublin Core Collections Application Profile¹² was chosen as the description format, with slight modifications. Later some of these modifications were adopted to the later versions of the Profile and of the NISO Z39.91-200X. Unfinished status of the DC profile was not a serious issue since we built our production environment locally. Anyone relying on purchased applications would be better off now that the DC profile is complete.

Collections descriptions are being collected to a database that has been built on DSpace, open source software that is mostly used as a platform of institutional repositories. Adapting DSpace to this usage was relatively easy. Cataloguing collections, however, was not that simple.

It was found necessary to establish a classification scheme that would define and group together the subject-based collections. The project built its own simple scheme. Existing classification schemes, including the one in Conspectus, were not suitable for us.

Although the decision to use DC Collections was easy to make, using the format is not that simple, since some of the data elements pose interesting challenges. For instance, there is no standard collection identifier. In such a situation, a project may develop an internal solution (in this case, an identifier system) or allow each participant do whatever they want. The Finnish solution was to develop a collection identifier in such a way that the system could be proposed as a new work item for ISO. In Spring the National Library of Finland sent to ISO TC46 a proposal for developing ISCI, International Standard Collection Identifier¹³. It consists of International Standard Identifier for Libraries and Related Organizations (ISIL), separating character (colon) and organization specific collection identifier string. For instance, the collection identifier of the national library's Slavic collection could be FI-H:Slavica.

Co-ordination of the identifier usage is not the only area where we have done our best to make it possible to exchange the collection descriptions internationally. There is a strong recommendation to provide relevant data elements such as the title of collection in multiple languages (say, Finnish, English and Swedish). In practice, however, libraries have not always followed this rule.

The absence of collection description rules is conspicuous, and potentially a more serious threat to data exchange than lack of multilingual collection descriptions. The task of creating "cataloguing rules" for collections was beyond the resources of the project, and even if there had been an appropriate person available, it would not have been appropriate to provide more than simple guidelines. Although such guidelines were written¹⁴, the cataloguers were not happy with this level of support. This shortcoming was strongly felt by the participating librarians who generally thought that the experience with bibliographic cataloguing and its rules does not easily lend itself to working with collections.

An individual library, if it happens to be active in ISO TC46, can launch an initiative to develop a standard identifier for collections. But no single library can take the responsibility to build cataloguing rules for collections. This is a task for IFLA. If we do not have such rules, then our

¹² "Dublin Core Collections Application Profile". 2003-11-10.

<<http://dublincore.org/groups/collections/collection-application-profile/2003-11-10/>>

¹³ Hakala, Juha: "Proposed ISO International Standard Collection Identifier (ISCI) – an overview". 2006.

<<http://www.collectionscanada.ca/iso/tc46sc9/docs/sc9n431.pdf>>

¹⁴ They are available, unfortunately only in Finnish, at

<<http://www.lib.helsinki.fi/kirjastoala/neuvosto/toiminta/Kokoelmakartta/271205Formaatti.pdf>>

collection descriptions will not be as compatible as they could be, given that most of us will use the same metadata element set.

In gathering data, possibilities of querying library databases could be exploited successfully¹⁵. The experiences varied, though. Necessary queries tended to become very convoluted. Where classifications are consistently used, extraction of data about collections on a certain subject was a more straightforward business than elsewhere. Unfortunately such consistency was found to be quite rare. - Complexities of database querying forced to use other methods to verify and complement the results gained that way. It could be demonstrated that as far as printed collections are concerned, very simple sampling methods yield perfectly acceptable results for the purposes of collection description.

¹⁵ Wilen, Raine & Ahtola, Anne: "Collection evaluation: micro and macro levels - preliminary guidelines and the results of a pilot study of two Finnish University Libraries". *Signum*, 3/2006.
<<http://www.protsv.fi/stks/signumnew/200603/7.pdf>>

4. Collection-Level Descriptions and the Digital Library - Incompatible?

At the beginning, the project shared the common understanding that subject-based collection-level descriptions would facilitate end-users' information searching in the digital and networked environment¹⁶. Since then, this vision has been challenged for a number of reasons that relate to the ways how libraries manage electronic resources – and how end users use them.

(1) Incorporating electronic resources into subject-based collection-level descriptions turned out to be a formidable problem that remains largely unsolved even today. The difficulties are mainly due to the fact that electronic resources are not really taken over by the libraries in the way the printed materials usually are. We illuminate the situation by the example of Finnish libraries.

The Nelli Portal¹⁷ is a Finnish library portal that is available in all university, polytechnic and public libraries. It is rapidly becoming the end users' principal gateway to library information resources. Nelli is a metasearch engine that enables simultaneous search in several databases. Most often, a user picks one of the predefined sets of databases in order to commit a search. Of course, these databases are not coherent collections. A predefined set, e.g. "behavioural sciences", is comprised of databases that contain *some* material relevant to the field. Inside our example set, we find "Academic Search Premiere", among others. Of course, that database covers a very wide range of subject matters, so that the end user is exposed to search results that are definitely outside the field of behavioural sciences.

A detailed description of a many-faceted, ever-changing licence package does not make sense.¹⁸ On the other hand, we do not yet have tools to re-arrange licensed resources into more meaningful sets. There are indications, though, that the trend is towards a more integrated approach of library resources management¹⁹ that has a bearing on the concept of collection, too.

(2) Even if we were able to gather data about the databases we probably could not translate that data into a service that would be obviously useful to the average end user. Certainly we cannot capitalize on it in the context of the Nelli, because the searches would still be based on the licensed databases and the result set is incoherent with the collection-level description. As electronic books are constantly gaining in popularity, even monographs are more and more often found through the publisher's or aggregator's databases, and not in the traditional context of library catalogues.

But is this a genuine problem? Should we stick to the concept of two-step searching process²⁰, where it's crucial that the user makes a careful choice of the resources he/she is going to use? Not necessarily; at least it is no more a practical imperative. Wide-ranging searches are not a problem of time, effort, and bandwidth like they used to be. On the contrary, the point of the Nelli Portal is to make it easy to use multiple resources simultaneously.

¹⁶ Johnston, P. & Robinson, B.: "Collections and Collection Description". Collection Description Focus Briefing Paper, No 1. January, 2002. <<http://www.ukoln.ac.uk/cd-focus/briefings/bp1/bp1.pdf>>

¹⁷ <http://www.nelliportaali.fi/>

¹⁸ For example, this is not a very useful description of a resource that in itself is most certainly useful: "The database includes full text over 4 000 publications as well as images, for nearly every academic field of study. It contains citations with abstracts to articles appearing in almost 4 000 periodicals."

¹⁹ Rochkind, J.: "(Meta)search Like Google". Library Journal, 2/15/2007. <<http://www.libraryjournal.com/article/CA6413442.html>>

²⁰ GALLER, A. M. et al. (1997) "Information Literacy: A Prototype to be Used in School Libraries" 63rd IFLA General Conference - Conference Programme and Proceedings - August 31- September 5, 1997. <<http://www.ifla.org/IV/ifla63/63gala.htm>>

(3) Finally, were it possible to rearrange the electronic resources into subject categories, or collections, then the easiness of searching does not seem to call for exhaustive metadata about those collections.

5. Collection-Level Description as a Surrogate

Special collections are a special category of collections, especially those that form a unity that is not apparent in the online catalogue, and of course those that are not catalogued online at all. The more digital the information environment is getting, the more these collections are in danger of becoming virtually invisible, even where they are a valid subject of study or a source of information. The imminent mass retirement of the post- World War II baby boomers also creates an urgent need to document their tacit information about the collections.

Collection-level description of special collections has been perhaps the most successful part of the project so far. The work with the relatively small collections has also been excellent training before undertaking more complex description tasks.

6. Information About Collection And Impact of Libraries

The Finnish higher education is facing yet more demands of efficiency and economy from the Government that is largely responsible for its funding. Deep-going structural changes in the field of universities and polytechnics are to be expected in the near future. In addition, during the recent years there has been a trend towards larger units in the library structure within the institutions of higher education.

Libraries have a twofold task in managing the change. First, they must be able to carry out necessary changes in library structures in an efficient way. Secondly, they must be able to assure that their position as a core success factor is understood and supported. We know that researchers and administrators do not always realize that the resources on their laptop do not appear there just like that. The libraries have to remind the users and decision-makers of their importance. They also have to be able to demonstrate their effectiveness and improve it. There is an obvious need for information that can be examined in relation to the aims and objectives of the institution as a whole, and also in comparison to other libraries.