This paper was originally published in Library Review, Vol.52 No.6, pp.247-250 and should be referenced accordingly.

Collection-level descriptions: metadata of the future?

George Macgregor,
Centre for Digital Library Research,
Department of Computer & Information Sciences,
University of Strathclyde, Glasgow, UK
Email: george.macgregor@strath.ac.uk

Abstract
The potential for digital library growth has recently drawn into question the ability of users to navigate large distributed and heterogeneous collections. This column attempts to summarise some of the potential benefits to be derived through the implementation of collection-level descriptions for both user resource discovery and institutional collection management. In particular, the concept of "functional granularity" is introduced and some related issues are briefly explored.

Introduction

It is now almost farcical to think that the accommodation of the newer formats, such as films or sound recordings, were considered as a "revolution" in the 1970s and 1980s. Since then, of course, libraries and information services have undergone developments of near seismic proportions as they attempt to tame what Manoff (2000, p. 861) refers to as "the information monsters". The proliferation of electronic information, mainly via the Web, has forced information professionals to swallow that bitter pill: "access vs. ownership". Many may have swallowed it, but few have truly digested it. We still find ourselves trying to exert the same degree of control over electronic information resources as that of print based resources. In the past decade we have discovered that libraries increasingly provide access to highly volatile information, information ith an apparent lack of fixity, and information that is often bereft of permanence. The transient nature of this information is such that users are directed to a plethora of information (e-journals, Web sites, related collections, and suchlike) held outwith the traditional confines of the "collection".

Whilst the arguments continue to rage as to whether this constitutes a desirable model of information provision, what is certain is that such arrangements thwart efforts to assimilate them into traditional bibliographic forms. This perception of library collections, particularly in the realm of digital libraries, has been changing. The use of collection-level descriptions has become an increasingly topical and relevant issue in recent years, especially since digital libraries represent a more heterogeneous manifestation than traditional libraries (Hill *et al.*, 1999, p. 1169). The emergence of digital and hybrid libraries, maturing library catalogue systems, the exponential gathering of digital resources into "collections" and the further aggregation of these collections has renewed interest in the use of collection-level description as a means of enhancing resource discovery and collection management. The purpose of this issue's column is therefore to illuminate the emerging potential of collection-level description (CLD) and to perhaps raise some issues worthy of further thought.

CLD: the scenario

CLDs are nothing particularly new. Archives, for instance, have long been using such resource discovery tools. Items within archival fonds can only be understood and appreciated within the context of those other items belonging to the fonds, and the descriptive practice employed by archivists reflects this approach. Yet, in the library and information science domain the collection has not traditionally been at the forefront of resource discovery. Obviously special collections and other significant collections have always existed, but such a view of libraries and information services has not traditionally underpinned the delivery of services in the same way that it has in the world of archives and museums. Emphasis has been on item-level activities such as cataloguing and circulation, while collection-level activities have been implicit in the local service environment. Nevertheless, the rampant march of the digital libraries has perpetuated the rise of information repositories of unforeseen magnitude and of tremendous diversity, often spanning a variety of domains. More importantly, the potential for digital library growth far exceeds the

humble parameters established by the print based library. It is thus appropriate to aid user navigation of such "information landscapes" in order that information contained therein is not rendered wholly redundant by its apparent abundance.

The brief nature of this column prohibits any detailed explanation of CLDs; references are provided for that purpose. However, for our interim purposes we can consider a CLD to be a structured, open, standardised and machine-readable form of metadata providing a high-level description of an aggregation of individual items. Such descriptions disclose information about their existence, characteristics and availability, and employ the use of implicit item-level metadata and, more particularly, contextualise that aggregation of item-level descriptions. CLDs are clearly desirable since they can enable the discovery of collections of interest, particularly prior to item-level discovery or data mining. Providing us with an eloquent analogy, Heaney (2000, p. 3) states that the:

> *information landscape can be seen as a contour map in which there are mountains, hillocks, valleys, plains and plateaux…The scholar surveying this landscape is looking for the high points. A high point represents an area where the potential for gleaning desired information by visiting that spot is greater than in other areas.*

We are therefore able to harness the potential of CLDs to provide an overview of groups of items, perhaps even uncatalogued items or those items where item-level details are inappropriate. Such an approach is conducive to the "high-level" navigation of large and often distributed or heterogeneous resource bases. A scholar, for instance, may wish to utilise CLDs to discover the existence of collections spanning numerous domains but with a common characteristic such as subject or collector, and then to subsequently rationalise and direct their item-level queries on the basis of the characteristics intrinsic to that collection. In essence, we can deliver improved distributed networked services for users with the uptake of such metadata, particularly when clear opportunities arise to augment interoperability ± for example, the implementation of agreed schemas such as the Research Support Libraries Programme (RSLP) Schema (Powell *et al.*, 2000).

"Functional granularity"

Leaving aside the contentious issue of what actually constitutes a "collection", an institution that agrees upon the particular aggregations that form its collections will invariably discover that these collections are related on a variety of levels. Thus, relationships could be applied to collections of varying sizes and granularity so that, for instance, a "collection" may contain numerous "sub-collections", and vice versa ("super-collections"). The use of granularity is of obvious importance in the context of CLD resource discovery, and Geisler *et al.* (2002, p. 216) have already commented that the relational attributes will be essential, not only for discovering resources within single repositories, but also across libraries of all types, and across different domains:

> *By explicitly representing not only a wealth of collections, but also the relationships among them, regardless of their physical location, a collection level metadata schema should greatly improve the navigability of the [digital library].*

Such conceptualisations of resource navigation have already been instantiated by CLD projects such as SCONE (2002). In SCONE rich forms of CLD are capable of being "drilled down" from the highest level of granularity through related sub-collections (many of them distributed) until the desired degree of specificity is reached. The ability to exploit descriptions created by institutions for practical and functional reasons means that the user is capable of surveying the landscape for the elusive ''high points''. Not only is this approach functional for the user, but it is also functional for the institution. A CLD, based on a recognised schema or standard, can provide a simple, less labour intensive and standardised means of disclosing an institution's curatorial responsibilities. Disclosing such responsibilities can underpin collection management duties and related initiatives by providing a convenient tool for coordinating collection development, bibliographic access, storage, and preservation, and by enabling informed strategic planning at institutional, cross-

institutional, regional, sectoral and national levels. The CLD then assumes a brand of re-usable or recyclable metadata.

Functional for whom?

The concept of functional granularity is unquestionably an intriguing proposition. Heaney' s paper, "An analytical model of collections and their catalogues", which has informed the work of UKOLN's collection description focus (UKOLN, 2003) and CDLR projects like SCONE and CC-interop (CC-interop, 2003), suggests that a functional granularity approach should be adopted by an institution in the description of its collections in order to:

> *make explicit those elements of the collection descriptions which the institution deems to be useful or necessary for the purposes of resource discovery or collection management (i.e. should adopt a "functional granularity" approach) (Heaney, 2000, p .5).*

Clarifying this supposition, Dunsire (2002) states that, "if intellectual or administrative effort has gone into the definition of the collection, then it is probably worth recording". The concept of functional granularity is therefore entirely flexible and relies on the judgements of the administrators of that collection (or those closely associated with the collection) to make informed decisions over what they consider to be a relevant and useful aggregation of items. In doing so they provide flexible tools for collection management, breaking a collection down into manageable sub-collections. More importantly they create levels of granularity that are enlightened by the information professionals' unique knowledge of these collections. These are capable of supporting navigation of the chaotic information environment to which we wish to restore order, as well as providing an efficient filtering mechanism. More succinctly:

> *If records are created for both a collection and its significant "sub-collections", then it is possible to choose between presenting only the "supercollection" record (while filtering out the more detailed "sub-collection" records), or presenting the more detailed "hierarchical" view (Johnston, 2002).*

The flexibility and functionality of this concept is exemplified yet further when one realises that it is possible to be embraced not just by collection administrators for users (and themselves), but also by users for users. SCONE has collections defined by special groups of users, such as the Scottish Working Group on Official Publications. Such user groups have created their own CLDs expressly for enhancing their own activities. As a collection description service, SCONE has similarly created specific collections to improve the ''functions'' of general retrieval and data cascade/inheritance.

Yet, there remains the question of whether locally dictated choices are conducive to a globally accessible information infrastructure. This is, after all, the age of "think globally, act locally". Employing the use of functionally granular techniques, especially for resource discovery across distributed networked services, is undoubtedly useful and it does provide a trajectory worth pursuing in digital library research. The success of SCONE bears testament to this. However, a project like SCONE remains within the confines of a distinct geographical and networked area, where the attributes accorded to collections are derived from similar socio-political and cultural perspectives, not to mention similar information science perspectives. Users from outwith these areas may not be so informed by such peculiarities, and nor should they be expected to be. This is particularly true since they are likely to be driven to interrogating the said repository as a result of Belkin's now legendary "Anomalous State of Knowledge" (ASK) conundrum[1] (Belkin *et al.*, 1982). Of course, we can always profess that local expertise informs global expertise, but such an approach smacks of arrogance and few presumptions should ever be made over who your clientele are.

[1] Belkin famously remarked that it is "unrealistic to ask the user of an IR system to say exactly what it is that she/he needs to know, since it is just the lack of that knowledge which has brought her/him to the system in the first place" (Belkin *et al.*, 1982, p. 66).

Concluding remarks

It is clear that functional granularity, and the use of granularity generally, is an area of digital library research that should be pursued further and one that should find further applications, especially in conjunction with emerging CLD schemas being championed in the UK and, to a lesser extent, the USA. The real question, however, remains as to whether such an approach to resource discovery lends itself to applications outwith the locality from whence it originated. If not, how can we best tinker with functional granularity so as to maximise its relevance to those disparate communities that do not share common views on what is functional and what is not? What is indisputable is that there is an evident paradigm shift afoot: item level description to collection level description. Digital libraries, in their various permutations, have ushered in an era whereby the importance of item level description is diminishing. It will always be significant (can you imagine a world without it?), but the gargantuan size of digital libraries and their potential for growth have emphasised its limitations and demonstrated the untenable and unwieldy nature of item level description for searching large distributed and heterogeneous collections. So is it fair to say that item level description has had its day? Not quite. It's just not the answer to everything anymore.

References

Belkin, N.J., Oddy, R.N. and Brooks, H.M. (1982), "Ask for information retrieval: Part 1. Background and Theory", *Journal of Documentation*, Vol. 38 No. 2, June, pp. 61-71.

CC-interop (2003), available at: http://ccinterop.cdlr.strath.ac.uk / (accessed 12 March 2003).

Dunsire, G. (2002), "Guide to the SCONE database", Glasgow, CDLR, available at: http://scone.strath.ac.uk/service/Guide/gContents.cfm (accessed 12 March 2003).

Geisler, G., Giersch, S., McArthur, D. and McCelland, M. (2002), "Creating virtual collections in digital libraries: benefits and implementation issues", *Joint Conference on Digital Libraries 2002*, Portland, OR, pp. 210-218, available at: www.ukoln.ac.uk/metadata/rslp/model/amcc-v31.pdf (accessed 12 March 2003).

Heaney, M. (2000), *An Analytical Model of Collections and their Catalogues*, UKOLN, Bath, available at: www.ukoln.ac.uk/metadata/rslp/model/amccv31.pdf (accessed 12 March 2003).

Hill, L.L., JaneÂ e, G., Dolin, R., Frew, J. and Larsgaard, M. (1999), ``Collection metadata solutions for digital library applications'', *Journal of the American Society for Information Science*, Vol. 50 No. 13, pp.1169-81.

Johnston, P. (2002), *Creating Reusable Collection Level Descriptions ± Collection Description Focus, Guidance Paper 1*, UKOLN, Bath, available at: www.ukoln.ac.uk/cd-focus/guides/gp1/ (accessed 12 March 2003).

Manoff, M. (2000), "Hybridity, mutability, multiplicity: theorizing electronic library collections", *Library Trends*, Vol. 49 No. 10, Summer, pp. 857-76.

Powell, A., Heaney, M. and Dempsey, L. (2000), "RSLP collection description", *D-Lib Magazine*, Vol. 6 No. 9, available at: www.dlib.org/dlib/september00/powell/09powell.html (accessed 12 March 2003).

SCONE (2002), available at: http://scone.strath.ac.uk/ (accessed 12 March 2003).

UKOLN (2003), available at: www.ukoln.ac.uk/cd-focus/ (accessed 12 March 2003).