



Date : 06/06/2006

“Tomorrow never knows”¹: the end of cataloguing?

Alan Danskin

Manager Data Quality & Authority Control
British Library
UK

Meeting:	102 IFLA-CDNL Alliance for Bibliographic Standards ICABS
Simultaneous Interpretation:	Yes

WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL
20-24 August 2006, Seoul, Korea
<http://www.ifla.org/IV/ifla72/index.htm>

Abstract

This paper reviews the perceived threats to the future of cataloguing posed by the increasing volume of publications in all media, coupled with a resource base which is declining in real terms. It argues that cataloguing is more rather than less important in such an environment and considers some of the ways in which cataloguing will have to change in order to survive.

Paper

The purpose of this paper is to review the challenges confronting cataloguing as we have known it and to consider how these challenges might be confronted and whether they may be surmounted. The main focus of this paper is on cataloguing rather than the catalogue, although it is obviously difficult to separate one from the other.

First of all, what does "cataloguing" mean? For the purposes of this paper I have adopted a broad definition incorporating the following activities:

- description of the resource sufficient for purposes of identification and for differentiation from other similar resources
- identification and control of access points
- identification and control of relationships with other resources
- subject analysis of the resource
- assignment of subject indexing terms
- assignment of classification numbers

The challenges facing cataloguing are all too well known. In no particular order, the major challenges are:

- Increasing inputs
- New kinds of information resource
- Competition from other mediation services.
- Perception that cataloguing is high cost and offers poor value for money.
- Fiscal constraints
- Declining workforce

This is a daunting list. We have a choice, we could, to paraphrase John Lennon, "Turn off our minds, relax and float down stream", until we retire, take voluntary redundancy, or retrain as marketing consultants; or, we can choose to confront these challenges and consider what they really mean for cataloguing.

It is estimated that during 2004 in the UK alone, 160,000 new monograph titles or editions were published². No figures are available for 2003, but this represents an increase of approximately 17,000 or 18% per annum. Looking on the bright side, there is no evidence that we will ever run short of material to catalogue. However bibliographic agencies do stand in danger of being swamped by the ever increasing tide of publication. Over the same period, the UK government's financial support to the British Library,

which accounts for 75% of the operational budget, rose by approximately 0.75%³. With the increase in intake exceeding income by approximately 17% per annum substantial efficiency gains are necessary just to keep processing up to date. Even worse, published monographs account for only a small part of the overall picture.

The web has fundamentally changed the nature of publishing. Historically the cost and complexity of printing, marketing and distribution confined the means of publication in relatively few hands, which served as a brake on new publications. That drag factor has been virtually eliminated and anyone with sufficient will and access to a pc and a browser can become their own publisher. Libraries find themselves charged with responsibility for archiving, storing and providing persistent access to these new publications, including preprints, Web pages and blogs, which either did not exist, or were not in the public domain. The simultaneous migration of serials to the web has created new requirements for providing access to articles. The sheer volume of these resources defies traditional cataloguing processes. The UK Web domain alone comprises over 4 million websites⁴.

New tools and services have in any case sprung up to mediate access to this information. Mediation between information seekers and information has been the traditional role of the library and the library catalogue. The catalogue is now perceived to be in competition with other mediation tools. This is most obviously manifested in the challenge from Google and other web search engines and services. On-line media stores, such as Amazon, provide intuitive, easy-to-use, friendly interfaces and rich descriptions of publications. By contrast catalogues are characterised as difficult to access, difficult to use, rather boring in appearance and circumscribed by their content⁵. There is evidence, particularly among undergraduates, that OPAC usage is declining, in favour of the web. The declining user base will make cataloguing increasingly hard to justify.

Cataloguing is expensive. Calhoun has reported that American research libraries spent an estimated \$239 million on technical services labour in 2004⁶. This equates to approximately one third of the total spend on new collection items in traditional media. Marcum⁷ noted that the Library of Congress spent a staggering \$44m on cataloguing every year. The British Library spent approximately £5.8 million on all cataloguing activities during 2005/6, around 5% of the total library budget.

Even assuming that we are able to pay for cataloguing, there is a perception that the profession needs an infusion of new blood. In the US it has been predicted that 33% of cataloguers will retire by 2010⁸. The recent wave of retirements from Library of Congress is perhaps the shape of things to come. In common with many other institutions around the world the British Library relies heavily on the high quality cataloguing produced by Library of Congress. A recent survey, by Leysen and Boydston, provides little ground for optimism that lost cataloguing posts are filled or that library schools are producing the next generation of cataloguers to fill them. Such evidence as there is suggests that there are fewer trained cataloguers out there. In the UK and the US the cataloguing component in Library education is diminishing and the faculty are aging⁹.

What can we do? "Lay down all thought" and, "surrender to the void". From this review of the challenges, two key questions arise:

1. Is cataloguing still relevant in the web environment and will it remain relevant in the medium to long term?
2. If cataloguing is still relevant, how does it need to change to meet these challenges?

Starting with the first question, is cataloguing relevant in the web environment? In the short to medium term the answer must be a resounding "yes", because the web has not yet superseded print as the carrier for certain types of information and certain types of entertainment. There is still no evidence for any decline in the output of printed books. Indeed, predictions point to substantial growth of the market in emerging economies such as China, India, Brazil and Eastern Europe, while established markets, such as USA and UK continue to enjoy significant growth¹⁰. If these new publications are to be added to library collections they will have to be catalogued.

However this could be no more than a stay of execution. The rate of technology obsolescence is increasing and although consumers have thus far resisted the electronic book, who is to say that the printed book will not go the way of the vinyl disk or the CD? Lord Young of Graffham, former minister in the Thatcher government, notorious for his love of gadgets, recently predicted that in 5 years we will all be downloading i-books from the Web¹¹. This seems slightly optimistic, but the tipping point is expected by some commentators within the next 10 years¹². Will there still be a future for cataloguing, when most new content has migrated to electronic formats and the content can be directly accessed by keyword? It

can (and it will) be argued that cataloguing is rendered redundant by direct access to content.

This argument ignores the fact that many resources are not textual and can't declare themselves without the addition of metadata. The answer patently is *not* within. Also, many resources will remain in print form in our collections for years to come. Calhoun has calculated that even on the most optimistic assessment it will take more than 20 years and millions of dollars to digitise the major print collections¹³. Let us assume, for the sake of argument, that through some miraculous convergence of public policy and technology this could be achieved and the world's knowledge really is digitised and available through the web. How would access to this vast resource be mediated? Will keyword searching and relevance ranking alone suffice? Neither Google nor Microsoft seems to think so. In their mass digitisation projects they are already reusing the catalogue records created for the printed originals.

Cataloguing is about much more than just describing resources. It is also about establishing a context for each resource. Cataloguing establishes a set of relationships that connect the resource being described to other resources; to entities involved in its creation and production; and to concepts contained within it. This is important information that is valued by real users. People want to know what else has been written by a given author, whose work they have enjoyed, even if written under different pseudonyms. For some people it is important to be able to distinguish between different editions of an early printed book; for others it is vital to find that samizdat translation into Czech of *Lord of the rings*. Cataloguers, working over the last century or so have constructed a map of the recorded knowledge and intellectual achievement of the human race. The map may be incomplete and imperfect, but if you want to see the lie of the land a map is much more useful than a gazetteer, however comprehensive.

It has been the failure to exploit the navigational potential of this rich metadata that has given the OPAC such a bad name. The guard book and the card catalogue offered a limited number of access points, but they traded this off in favour of effective presentation of the relationships between resources. The OPAC has tended to favour an increase in the number of access points over the effective presentation of the relationships between resources. The OPAC as we have known it is doomed, but far from rendering cataloguing redundant, web technologies offer the opportunity to integrate the strengths of printed catalogues and OPACs (On-line

Public Access Catalogues) and transform that two dimensional map into a powerful tool for navigating the world's knowledge.

The object of cataloguing is to save the time and money of the user of our collection. They may not set out to use our collection, but if our cataloguing can direct them there and has helped them to find what they want, we have delivered a public good. It is very difficult to measure these benefits because they are indirect. The British Library has undertaken some interesting research to try to quantify the financial benefit it delivers in return for Grant-in-Aid. This research did not specifically look at cataloguing, but there are lessons to be learned from the approach of converting qualitative into quantitative measures.¹⁴

There seems little question that the volumes of information available over the Web will be huge beyond imagining and the opportunity costs of discovering resources can also be expected to increase. A recent novel postulated that only through the manipulation of their physiology to alter their perception of the passage of time will the cataloguers and researchers of the future be able to keep up with the volume of information¹⁵. This seems unlikely to be a viable solution outside the realms of science fiction, so what hope is there for cataloguer stuck in the here and now?

There is a degree of hysteria attached to the idea that because it is impossible to catalogue the web it is no longer necessary to catalogue anything. The web does not need to be catalogued in its entirety. Historically libraries have been selective in what they have collected and still more in what they have catalogued. The challenge is not only a cataloguing challenge, but also a collection development challenge. What do libraries want in their collections? How can they sort the wheat from the chaff and how do they efficiently filter appropriate material for integration with the catalogue?

Material which is formally published will be relatively straightforward to identify, just as it is in print, because it carries standard identifiers and is promoted and acquired much like print. The vast majority of material on the web may be of cultural interest to future generations of scholars (or pornographers), or of no interest at all to anyone other than its creator. Libraries will become more like archives and the majority of web resources will of necessity be treated as archival rather than bibliographic resources. They still need to be organised but the approach is one of minimal intervention. Some pockets may attract the attention of scholars, and will receive more comprehensive description, but this will be the exception, just as it is with many existing archival collections.

Between these extremes there will be many sites and other web resources which warrant cataloguing in a more conventional sense. To a limited extent these resources may describe themselves, but it is unlikely that they will automatically slot themselves into the wider bibliographic context. Library of Congress has proposed an access level, "description-lite", record to facilitate the integration of "self describing" resource" into the syndetic structure of the catalogue¹⁶. The IFLA Working Group on (Electronic) National Bibliography is also working on selection criteria to identify resources which warrant intervention¹⁷.

Spreading the load has been the successful response of libraries to increasing demands on cataloguing resources. Libraries have co-operated with each other, but co-operation with other constituencies has been limited and constrained by sector specific standards. The challenge for cataloguing is to emerge from silos which are already becoming untenable. We must engage with partners in publishing, commerce and rights management to realise the "bibliographic continuum" so that we can reuse metadata created by other constituencies. Some work in this area is already underway, the Joint Steering Committee for the Revision of the Anglo-American Cataloguing Rules and ONIX "have launched a joint initiative to develop a common framework for resource categorization" with objective of agreeing, "a common framework that both RDA and ONIX can use to underpin the development of resource categories appropriate to their respective communities, and that can be the basis for interoperability of resource descriptions." This is a small beginning but it is an important engagement on content standards and reflects an increasing interest in metadata as an essential adjunct to identifiers.

The future of cataloguing depends on transforming the process from a craft into an industry. This requires unambiguous identification at different levels of granularity to facilitate repurposing of metadata created at the different stages of the process of creating and publishing resources. It also means we may have to be less precious about some of our cherished practices. The process of cataloguing is supported by a huge volume of documentation, which is constantly changing. This volume of rules and interpretations and exceptions is inefficient and confusing. The transition from *Anglo-American Cataloguing Rules* to *Resource Description & Access* points to a principles based system which, it is to be hoped, will move the focus from the minutiae of description to the principles of organisation. We cannot afford to control everything, so we must concentrate our resources on the things we need to control to deliver the core services of the catalogue: the facility to collocate

resources which share common attributes; the ability to distinguish between similar resources, according to the specific needs of the individual user.

It is incumbent on us to “listen to colour of our dreams”. If we are not clear about our core values, no-one else can be expected to understand the purpose of what we do. There are grounds for optimism. It is hard to imagine the web without catalogues: every successful on-line retailer has a catalogue promoting its wares; some of the most popular on-line resources, such Internet Movie Databases, have a strong bibliographic underpinning. There is even a service which enables anyone to catalogue their own library collection¹⁸. One day we will all be cataloguers. “It is not dying, it is not dying.”

References

-
- ¹ Lennon & McCartney. *Tomorrow never knows*. London: Northern Songs, 1966.
- ²Data from Publishers Association website: URL
<http://www.publishers.org.uk/paweb/paweb.nsf/0/3CEA683597F6BC6D80256BAC005034E7!opendocument> Site searched 5th May 2006.
- ³ British Library. *From knowledge to innovation: 29th annual report and accounts, 2001/2002* ;
British Library. *Making a measurable difference: annual report and accounts 2003/2004*
- ⁴ British Library. Collection development policy for UK websites. URL
<http://www.bl.uk/collections/britirish/pdf/modbritcdpwebsites.pdf> Website visited 5th May 2006
- ⁵ Schneider, Karen G. How opacs suck. Part 1: relevance rank (Or the lack of it) – Part 2: Checklist of shame. ALA TechSource.
<http://www.techsource.ala.org/blog/Karen+G./Schneider/100003/> Website visited 11th May 2006
- ⁶ Calhoun, Karen. The changing nature of the catalog and its integration with other discovery tools: final report. Ithaca, N.Y.: Cornell University Library, March 17th 2006. p.27
- ⁷ Marcum, Deanna B. *The future of cataloguing*. Address to the Ebsco leadership seminar Boston, Massachusetts, January 16, 2005 URL
<http://www.loc.gov/library/reports/CatalogingSpeech.pdf> Site searched 9th May 2006
- ⁸ Leysen, Joan M. and Boydston, Jeanne M. K.. “Supply and demand for cataloguers present and future.” LRTS 49(4) pp.250-265.
- ⁹ Morgan, John and Bawden, David. “Teaching knowledge organization: educator, employer and professional association perspectives.” *Journal of Information Science*, 32 (2) 2006, pp. 108-115
- ¹⁰ *The British Library’s content strategy consultation: meeting the knowledge needs of the nation.* <http://www.bl.uk/about/strategic/pdf/contentstrategy.pdf> p.15 .[Website accessed 15/5/05]
- ¹¹ *Midweek*. BBC Radio 4. Broadcast 5th April, 2006 21:30 BST; .
- ¹² McCrum, Robert . “E-read all about it”. *The Observer*. Review 15th January 2006, p. 23
- ¹³ Calhoun, Karen. The changing nature of the catalog and its integration with other discovery tools: final report. Ithaca, N.Y.: Cornell University Library, March 17th 2006. p.27
- ¹⁴ *Measuring our value: results of an independent economic impact study commissioned by the British Library to measure the Library’s direct and indirect value to the UK economy.*” <http://www.bl.uk/about/valueconf/pdf/value.pdf>
- ¹⁵ Banks, Iain M. *The algebraist*. London: Orbit, 2004. ISBN 1492299
- ¹⁶ Library of Congress. *LC implementation plans for access level MARC/AACR Records.* <http://www.loc.gov/catdir/access/accessrecord.html>
- ¹⁷ Minutes of the IFA Working Group on (Electronic) National Bibliographies, 18th-20th November 2004, Prague. In: IFLA Bibliography Section Newsletter, December 2004. <http://www.ifla.org/VII/s12/pubs/s12-newsletter-Dec04.pdf> Minutes of the 2006 meeting in Vilnius will appear in a future issue.
- ¹⁸ LibraryThing <http://www.librarything.com/>