PART 1
FUNDAMENTALS

Cataloguing in an Electronic Age

Michael Gorman

SUMMARY. Examines the achievements in bibliographic control of the last thirty years and the strides made toward Universal Bibliographic Control. Discusses the intended and unintended effects of three standards–the MARC format, ISBD, and AACR. Analyzes the types of resources in cyberspace to be organized and their similarities to and differences from documents librarians already know. Suggests strategies for solving the seemingly insoluble problems of cataloging the Internet and predicts how metadata will evolve.

KEYWORDS. Cataloging standards, MARC21, AACR2R, ISBD, Universal Bibliographic Control, cataloging Web resources, metadata standards

Michael Gorman is Dean of Library Services, California State University-Fresno, 5200 North Barton, Fresno, CA 93740-8014 (E-mail: michael_gorman@csufresno.edu).

[Haworth co-indexing entry note]: "Cataloguing in an Electronic Age." Gorman, Michael. Co-published simultaneously in *Cataloging & Classification Quarterly* (The Haworth Information Press, an imprint of The Haworth Press, Inc.) Vol. 36, No. 3/4, 2003, pp. 5-17; and: *Electronic Cataloging: AACR2 and Metadata for Serials and Monographs* (ed: Sheila S. Intner, Sally C. Tseng, and Mary Lynette Larsgaard) The Haworth Information Press, an imprint of The Haworth Press, Inc., 2003, pp. 5-17. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

http://www.haworthpress.com/store/product.asp?sku=J104

10.1300/J104v36n03_02

INTRODUCTION

There is an old story about a fisherman fishing in a private river who, on being told "You can't fish here!" replied "But I am fishing here." This tale is not only illustrative of the difference between "may" and "can" that you should have learned at your mother's knee, but also of human beings' propensity not to realize that they are doing impossible things until they are told so. The impossible thing that librarians have been doing for a long time is classification–the reduction of the almost infinite dimensions of knowledge to a straight line from 000-999 or A to Z. Who but someone with the innate hubris of a cataloguer would dare to catch a flying complex subject and pin it to its exact place on that straight line to remain there forever? Now we have another impossible thing to do and we have to do it knowing that it is impossible–a thought that would never have occurred to, say, Melvil Dewey! This new impossible thing is to bring order to chaos, to trap lightning in a bottle, to take an electronic document with the life-span of a May bug (and, most likely, the cosmic significance of a May bug) and make it part of an arranged and harmonious world–in short, to apply some kind of bibliographic control to the disorder of the Internet and the Web. If there are, as politicians and futurists claim, untold intellectual riches out there in cyberspace, what use are they if they cannot be found, and found, and found again? Until we have a Networld and a Webworld with the attributes of a well-organized major library, how will we ever know if those worlds have the potential to rival Libraryworld?

The great irony of our present situation is that we have reached near-perfection in bibliographic control of "traditional" library materials at the same time as the advent of electronic resources is being seen by some as threatening the very existence of library services–including bibliographic control. Before considering the question of "cataloguing the Web and the Internet," it is salutary to review the great achievements of the past thirty years. In considering where we are going, it is necessary to know where we have been.

UNIVERSAL BIBLIOGRAPHIC CONTROL

When the ideal of Universal Bibliographic Control was first advanced[1] thirty years ago, the international library community was only beginning to discern dimly the possibilities of the interconnection of international standardization and library automation. International standardization was

at a very early stage (far closer to an ideal than a reality) and the ideal of each item being catalogued once in its country of origin–the resulting record being made available to the world community–seemed far from a practical realization. Records were exchanged between countries (mostly between national libraries), but in the most inefficient manner possible–print on paper–and, since they resulted from different cataloguing codes and practices, were integrated into catalogues with great difficulty. The choice was between incorporating international records without alteration–something that degraded the catalogue very quickly–or doing such extensive revision (and retyping) that it would have been cheaper and quicker to catalogue the item oneself *ab initio*.

MAchine-Readable Cataloging (MARC) was in its infancy when Universal Bibliographic Control was proclaimed as an ideal,[2] the International Standard Bibliographic Description was still being drafted,[3] and, despite the Paris Principles,[4] cataloguing rules in different countries lacked a common basis for the assignment and form of access points ("headings"), and adhered to different descriptive practices. It was, I believe, the confluence of a need (national and research libraries throughout the world needing less expensive and more current cataloguing) and a means (automation and, more specifically, the MARC format) that has brought us nearer to Universal Bibliographic Control than anyone would have dreamed possible thirty years ago.

The idea of a universal bibliography is nearly as old as bibliography itself.[5] The idea of economies in bibliographic control by means of sharing catalogue records between libraries (cooperative cataloguing) or purchasing catalogue records for other (usually national) libraries goes back to, at least, the middle of the 19th century. In fact, the American librarian Charles Coffin Jewett drew up his cataloguing rules[6] specifically for a proposed scheme by which the Smithsonian Institution would produce "separate, stereotyped titles" to be used in the catalogues of American libraries. In these, and in the hugely successful Library of Congress catalogue card service and the *National Union Catalog* to which it gave rise, we can see bibliographic needs and desires that lacked only an appropriate technology to be met. In hindsight, it is easy to see a trajectory of inevitability that made MARC, the International Standard Bibliographic Descriptions, AACR2, and other vehicles of international bibliographic standardization seem more the result of historical forces than the often faltering and separate steps they were in truth.

BACKGROUND AND HISTORY OF THE STANDARDS

Each of the three standards I mention had original purposes that were quite different from their eventual impact on international standardization. MARC was brought into being originally to facilitate the creation of Library of Congress catalogue cards on demand. The International Standard Bibliographic Description evolved from the Standard Bibliographic Description drawn up by a committee appointed as a consequence of the International Federation of Library Associations and Institutions' International Conference of Cataloguing Experts.[7] The Standard Bibliographic Description was seen, among other things, as a means of standardizing the presentation of descriptive data so that it could be machine-translated into MARC (hence the stylized and individual punctuation). AACR2 was the culmination of decades of effort to bring uniformity to cataloguing practice in the English-speaking world, and, particularly, to reconcile British and North American descriptive cataloguing practices. Each of these three standards metamorphosed and had an impact far beyond the anticipation of all but the most far-sighted. It is instructive to recall how and why each developed and expanded, because we need to understand that the bibliographic world (just like the real world) is full of unintended consequences and the ripples from a stone thrown in one part of the bibliographic pond may eventually cover it all.

The MARC format is, by any standards, an historic achievement. It has been the main force in international standardization from a practical point of view. It is, literally, the engine that has made Universal Bibliographic Control possible. The journey from the caterpillar of the automation of card production to the beautiful butterfly of today has been long and largely successful. It is worth pointing out, however, that its origins and original purposes (including being a carrier format rather than the way in which bibliographic information is stored and manipulated) have created drawbacks that should be hardly surprising when one considers we are dealing with a thirty-year-old standard.

The structure of MARC is that of the catalogue card, when computer systems call for a different approach. Be that as it may, the fact is that there are tens of millions of MARC records in the world; MARC is accepted and used throughout the world; MARC is the basis for almost all automated bibliographic systems (including commercially produced systems); and, no practically feasible or demonstrably better system has been advocated. It should be unnecessary to point out that MARC is merely a framework standard–that is, it is a way of storing and making

data capable of manipulation that has been formulated in accordance with content standards (cataloguing codes and the like). I would not trouble to point that out were it not for the frequent references to "MARC cataloguing" in writings about metadata and "simplified" cataloguing. There is, of course, no such thing as "MARC cataloguing"–MARC is the way in which we encode the results of the cataloguing process and has little or no influence on that process.

One of the two documents studied at the International Conference of Cataloguing Experts was a comparison of descriptions from cataloguing agencies throughout the world. The document revealed a great commonality of the information found in such descriptions and the order in which that information was presented. It found differences in the abbreviations used and other stylistic matters (mainly due to language differences) but was able to propose a conflation of the descriptions that formed the basis of what became the Standard Bibliographic Description and later the International Standard Bibliographic Description. Originally, the idea was to create a basis for agreement across cataloguing codes on the relatively non-contentious matter of descriptive data. Soon, however, this was supplemented by the idea that universally used distinctive punctuation, clearly identifying the areas and elements of the Standard Bibliographic Description, would not only aid in the understanding of bibliographic data in unfamiliar languages but could also be used in automatic translation of that data into MARC records. It is no coincidence that the areas and elements of the International Standard Bibliographic Description correspond exactly to the relevant fields and subfields of the MARC format. In accordance with the theme of stumbling toward standardization, it should be noted that both MARC and the International Standard Bibliographic Description were developed initially for books and only later generalized into standards for all types of library material.

The second edition of the *Anglo-American Cataloguing Rules* (AACR2) is, in fact, nothing of the sort. It was politically expedient at the time to identify this new code as a revision of the previous *Anglo-American Catalog[u]ing Rules* (1968), but AACR2 is completely different from its predecessors in several important ways. One need only cite, for instance, the facts that AACR2 is a single text (unlike its predecessors, which came in North American and British versions), is the most complete working out of the International Standard Bibliographic Description for materials of all kinds, and represents the triumph of Lubetzkian principles, which the first AACR signally did not. AACR2 quickly transcended even the historic achievement of being a

unitary English-language cataloguing code to become the nearest approach to a world code we have. In the words of the introduction to the Italian translation of AACR2:[8]

> *Le Regole di catalogazione*, nella loro seconda edizione, sono il codice più diffuso nel mondo (sono state pubblicate in gran numero di lingue diverse) e l'unico che-di fatto-svolga le funzioni di codice catalografico internazionale. [The Cataloguing rules, in their second edition, are the world's most widely used (they have been translated into numerous different languages) and the only rules that are, de facto, an international cataloguing code.]

This state of affairs is partly due, of course, to the dominance of the English language (in its various manifestations) in the modern world. It is also due, in part, to the fact that AACR2 represents the most detailed working out of the principles of author/title cataloguing set forth in the Paris principles and based on the analysis and pioneering work of Seymour Lubetzky;[9] and of the application of the International Standard Bibliographic Description family of standards to all library materials.

Here we stand then, on the brink of Universal Bibliographic Control for all "traditional" (i.e., non-electronic) materials with a universally accepted format for exchanging bibliographic data, a universally accepted standard for recording descriptive data, and a quasi-universal cataloguing code that is either in use in, or influencing the codes of, most of the countries in the world. Is there any reason in principle why we should not bring electronic documents and resources into this architecture of bibliographic control? The answer is "no." Are there practical reasons why this task is formidable? The answer is "yes."

INTEGRATING ELECTRONIC DOCUMENTS

The attributes of a well-regulated library are well known to us all. They are organization, retrievability, authenticity, and stability. There are those who claim that electronic documents and sites (assemblages of electronic documents) are different in kind and not just degree from all of the other formats that human beings have used to communicate and preserve knowledge across the centuries. This is, essentially, an implausible notion–after all, at the end of the day we are still dealing with texts, images, sounds, and symbols–but its strongest support comes from the evanescence and mutability of electronic documents. Those characteristics, which any true librarian deplores, are really the logical

outcome of the history of human communication–each format produces more documents than its predecessor, and each is less durable than its predecessor. It takes a long time to make many copies of stones bearing carved messages, but those messages can be read centuries later. You can send a message from Chicago to Addis Ababa in a twinkling of an eye, but that message may be expunged in a second twinkling. Many electronic documents are like those minute particles of matter that are only known because scientists can see where they have been during their micro-milliseconds of existence. Does an e-mail message exist if it is deleted unopened?

It seems to me that we should know, more precisely than we do now, exactly what it is that we are dealing with when we talk about organizing documents and sites in cyberspace. These can be divided into five types:

1. *Ephemera.* Libraries have always been far more selective than is generally acknowledged when it comes to their collections. I am not talking now of selection within formats (books, records, videos, etc.) but of ruling out, consciously or unconsciously, vast areas of recorded information. Much of the stuff that we used to ignore now shows up on the Internet and the Web. To demonstrate this, just do a search on any subject and review the few thousand "hits" with a view to imagining their tangible analogues. Personal Web pages are the electronic versions of scrapbooks and diaries–of keen interest to their compilers but to few others. Restaurant reviews? Press releases in digital form? Association newsletters? Weather forecasts? Lists of Australian university faculties? Syllabi? Advertisements? On and on it goes–acres of the cyberworld full of ephemera. What else is out there?

2. *Print-derived resources.* One useful sector of the Internet is composed of many documents and sites that are derived from the print industry and are dependent on the success of that industry for their very existence. These do not, by and large, present much of a technical problem. We know, in principle, how to catalogue different format manifestations of texts and graphic publications; thus, extending that knowledge into cyberspace is not a massive intellectual challenge. Further, print-derived electronic resources are far less transient than their purely electronic counterparts.

3. *Commercial sites and pornography.* People anxious to sell you something populate much of the electronic frontier. From e-tailers to business-to-business sites to pornographers, they are all pursuing the Great American Capitalist Dream in the sure and certain knowledge that not only is there a sucker born every minute but also that he or she is likely to spend a lot of time online.

4. *Electronic journals.* Most electronic journals are, of course, based on the products of a flourishing print industry. Forecasts over the last decade predicted that electronic journals will supplant print, but no one has, as yet, produced an economic model for such a major change and there are, at this time, a microscopic number of commercially viable true electronic journals. The problem is, of course, that the whole concept of a journal (serial assemblages of articles which are paid for in advance–whether they are ever read or not) seems inapplicable to the electronic age. Many problems in adapting to technology are caused by simply automating procedures or resources and not rethinking the whole issue. Why not, in an age of electronic communication, provide services that deliver desired articles on demand and charge the users only for the articles that they read? In such a world, the "journal" would no longer exist and libraries would be cataloguing at the level of what S. R. Ranganathan called "micro-thought"–a level that we have always left to indexing and abstracting services.

5. *Digitized archives (textual, sound, and visual).* One of the most important and valuable achievements of the electronic age is the way in which large archives have been made available to global audiences. Those archives (which are unique, by definition) have, hitherto, been accessible only to researchers with the means and time to travel to the location of the archives. To take a well-known example, the Library of Congress's *American Memory Project*[10] is a vast assemblage of pamphlets and other texts, graphic items, films, sound recordings, maps, etc., that is taking advantage of digitization and the Web to give the world access to the untold riches of the Library's archival collections. Other institutions have created Web archives of coins, stamps, posters, manuscripts, prints and drawings, early films, sound recordings, photographs, and every other conceivable means of communication, including artefacts. There has long been a great divide between library cataloguing and archival cataloguing. The former concentrates on individual manifestations of works and the latter has been largely concerned with creating finding aids for assemblages of documents. In the twenty-plus years since the appearance of AACR2, there has been some movement on this matter to bring the two cataloguing traditions closer together.[11] Although the two may always operate at different levels, there is no reason why their cataloguing practices cannot be harmonized and the results of such harmonization applied to the various parts of the *American Memory Project* and other such digital archives.

Here are the fundamental problems we encounter in trying to organize electronic documents and sites (other than those that are by-products of the print industries):

- there are too many of them
- a lot of documents and sites have never been, and never will be, of interest to libraries and library users
- the vast majority of such electronic documents are of temporary use, local use, or no use at all
- we have little or no guarantee that any given electronic document is what it says it is
- we have little or no assurance that any given document or site will be the same when next located, or that it will even exist
- there is nothing like the level of standardization of denotation and presentation that we find in books and other tangible library materials.

HISTORICAL ANTECEDENTS AND OMENS

Far from being unique, these problems are uncannily like those of manuscripts and early printed texts. The manuscript swamp from which the early printed text emerged, taking fumbling steps at first, resembles nothing so much as the electronic swamp that we now confront. We are far from the exciting world of innovation and creativity that is presented by those who hope to make money from the "information age." What we are seeing is a cultural reversion, not cultural progress. The problem with the manuscript culture was that many texts were lost, many were altered in copying, many lacked such things as titles and discernible authors, and all lacked publishers and distributors. A library burning to the ground today is a local tragedy; a library of manuscripts burning to the ground was a cultural catastrophe from which there was no recovery. Anyone who has tried to catalogue electronic documents and sites will tell you that they are elusive and shape shifting, they often lack titles and discernible authors, they may or may not exist tomorrow, they are subject to unpredictable change, and, once lost, they are lost forever. Sound familiar? There is, of course, one huge difference between the manuscript age and the looming electronic age. Pre-Gutenberg manuscripts were, by definition, created by an educated elite. Anyone doing a search using a search engine like Alta Vista is soon made painfully

aware that cyberspace is littered with the productions of ignorant, semi-literate, and/or crazed individuals.

What shall we do about this reversion, this "back to the future" impending catastrophe? Those who throw their hands up in despair will surely be forgiven, but we librarians love action and will seek an answer no matter what. We need, first, to decide what it is we seek to organize. We can recognize pornography when we see it as well as any Supreme Court justice. We can recognize commercial enterprises that need no help from us in bringing themselves to the attention of potential customers. We can probably recognize the ephemeral (though one librarian's ephemeron is another's invaluable cultural resource). That still leaves us with three large classes of material–print-derived resources, digital archives, and truly electronic resources of, at least, potential value to library users. As I have said, there is no doubt that we could relatively easily bring the print-derived resources into the world system of bibliographic control using links from existing records, multi-layered records, and full records using existing standards.

Having eliminated the other slices of the electronic salami, this still leaves us with a sizeable chunk. That chunk consists of the potentially worthwhile scholarly and information resources that exist only in cyberspace and may or may not be retrievable at any given time using search engines that use free text keyword searching–well known to be the very worst information retrieval system conceived by human minds. There is something seductive about the "surfing" metaphor (especially when one remembers that surfing can be exhilarating but you end up more or less where you started, only flat on your face in the sand), but, as a Californian, I think another Golden State metaphor–panning for gold–is more apt. Are we doomed to stand in cold streams for the rest of our lives, engaged in the stoop labor of panning through the dross in ever-hopeful search of the glint of the one nugget among all the grit and stones? The answer is . . . maybe.

METADATA–THE THIRD WAY

I have gotten this far in the paper without mentioning the word "metadata," but will break that silence now. The idea behind metadata is that it is a Third Way, approximately half way between cataloguing (expensive and effective) and keyword searching (cheap and ineffective). Some believe that the future belongs to metadata–this is hard to believe given that its best-known example, the Dublin Core is an ill-formulated

subset of the MARC record. Let us go back to the question of what it is that we are going to do about the worthwhile purely electronic resources that we wish to separate from the rest of the Internet and the Web. In my view, there are four possibilities. We can:

- identify and catalogue them according to standards we use for other materials
- identify them and take a subset of MARC (a framework standard, not a content standard) and call it "metadata," if that makes us feel better, to be filled with content according to bibliographic standards (either fully applied or dumbed down) by cataloguers and paraprofessionals
- identify them and take a subset of MARC and allow that framework to be filled with any content by anybody
- leave them in the murky waters of the Internet to be discovered or not discovered as determined by the karma of the searcher on the day in question.

These possibilities, obviously, range from the expensive and the effective to the inexpensive and the ineffective. There are also permutations and gradations, but those are essentially the choices before us. My belief is that "metadata," as presently conceived, will evolve toward standardization of elements and content and will be indistinguishable from real cataloguing in a relatively short time. That applies, of course, to those resources that are deemed "worthwhile." The rest will go their merry way to use, neglect, or oblivion with few tears shed.

Before we get to any kind of control, there is the question of identification of "worthwhile" materials. Again, we have choices. They are, first, a Grand Plan for cyber-collection development and, second, a grass roots movement in which individual libraries and librarians, and groups of libraries choose and catalogue the documents, resources, and sites they deem worthwhile. If you liked the drive for a national "information policy," you will love the years of striving for a national cyber-collection policy. It is not my cup of tea. The second approach will be a reprise of the history of libraries. Individuals and individual libraries built collections, one choice at a time, over many years. It was not until much later that union catalogues and library collectives brought those individual collections into a national system. The difference this time is that the benefits of the work of individual libraries and groups can be made available to all contemporaneously. Let a thousand INFOMINEs bloom, and record by record, collection by collection, "worthwhile" Internet resources will be organized and made available in what will come to be a national system following nationally-accepted standards.

CONCLUSION

Last, and most important, what is the point of all of this if the resources identified and catalogued are not preserved? Those more optimistic than I look to gigantic national electronic archives maintained by governments and private companies that will ensure the indefinite survival of the electronic records of humankind. The cost and practicalities of such schemes boggle the mind and defy credulity. We can, of course, ignore the problem and hope it all turns out right in the end–after all, that is what we are doing now. Alternatively, we could turn to the only known way of preserving massive numbers of texts and images–print them on acid-free paper. If you are inclined to laugh at that suggestion, I would recommend that you explore the financial costs and the cultural costs of the alternatives, and keep an open mind.

Metadata is a buzzword that is losing its buzziness, but real problems and real issues lurk behind all the pomposity and techno-babble. What are we going to do about identifying and making accessible the valuable records of humanity that are only available in electronic form? How are we going to deal with the mutability and evanescence of those records? How are we going to preserve those resources and transmit them to posterity? We will only answer these questions if we employ wisdom and insight, are cognizant of the lessons of history, and work with the interests of all our users, present and future, in mind.

AUTHOR NOTE

Michael Gorman is Dean of Library Services at the Henry Madden Library, California State University, Fresno. From 1977 to 1988 he worked at the University of Illinois, Urbana, Library as, successively, Director of Technical Services, Director of General Services, and Acting University Librarian. From 1966 to 1977 he was, successively, Head of Cataloguing at the *British National Bibliography*, a member of the British Library Planning Secretariat, and Head of the Office of Bibliographic Standards in the British Library. He has taught at library schools in his native Britain and in the United States–most recently as Visiting Professor at the University of California, Berkeley, School of Library and Information Science (summer sessions).

He is the first editor of the *Anglo-American Cataloguing Rules, Second Edition* (1978) and of the revision of that work (1988). He is the author of *The Concise AACR2,* 3rd edition (1999); editor of, and contributor to, *Technical Services Today and Tomorrow,* 2nd edition (1998); and editor of *Convergence* (proceedings of 2nd National LITA Conference), and *Californien*, both published in 1991. *Future Libraries: Dreams, Madness, and Reality*, co-written with Walt Crawford, was honored with the 1997 Blackwell's Scholarship Award. He published *Our Singular Strengths* in 1997. His most recent book, *Our Enduring*

Values, published by ALA in 2000, was the winner of ALA's 2001 Highsmith award for the best book on librarianship. Mr. Gorman is the author of hundreds of articles in professional and scholarly journals. He has contributed chapters to a number of books and is the author or editor of other books and monographs. He has given numerous presentations at international, national, and state conferences.

Michael Gorman is a fellow of the [British] Library Association, the 1979 recipient of the Margaret Mann Citation, the 1992 recipient of the Melvil Dewey Medal, the 1997 recipient of Blackwell's Scholarship Award, and the 1999 recipient of the California Library Association/Access, Collections, and Technical Services Section Award of Achievement.

REFERENCES

1. Kaltwasser, Franz Georg, "Universal Bibliographic Control," *Unesco Library Bulletin* 25 (September 1971): 252-259.

2. Avram, Henriette, "The Evolving MARC System: The Concept of a Data Utility," In *Clinic on Library Applications of Data Processing*, *1970* (Urbana, Ill.: University of Illinois, 1971), pp. 1-26.

3. Gorman, Michael, "Standard Bibliographic Description," *Catalogue & Index* 22 (Summer 1971): 3-5.

4. Chaplin, A.H., "Cataloguing Principles: Five Years after the Paris Conference," *Unesco Library Bulletin* 21 (May 1967): 140-145.

5. Gesner, Konrad, 1516-1565. *Biblioteca universalis, sive, Catalogus omnium scriptorum locupletissimus, in tribus linguis, latina, graeca, & hebraica* (Zurich: Apud Christoph Froschauer, 1545).

6. Jewett, Charles Coffin [1816-1868] *On the Construction of Catalogues of Libraries, and their Publication by Means of Separate, Stereotyped Titles*, 2nd ed. (Washington: Smithsonian Institution, 1853).

7. "Report of the International Meeting of Cataloguing Experts, Copenhagen," *Libri* 20, nos. 1-2 (1970): 105-137.

8. *Regole di catalogazione Angloamericane*, 2. ed., revisione del 1988. (Editrice bibliografica, 1997), p. vii.

9. *The Future of Cataloging: Insights from the Lubetzky Symposium,* (Chicago: ALA, 2000).

10. Available at <http://memory.loc.gov/>.

11. See, for example, Hensen, Steven. "NISTFII and EAD: The Evolution of Archival Description," *American Archivist* 60, no. 3 (1998): 284-296.