ELECTRONIC RESOURCES:
WHICH ARE WORTH PRESERVING & WHAT IS THEIR ROLE IN LIBRARY COLLECTIONS?

by

Michael Gorman
Dean of Library Services
California State University, Fresno

Libraries across the world are spending increasing amounts of money on the acquisition of, and giving access to, electronic resources of all kinds. In addition, those libraries are devoting increasing amounts of human resources to advise and teach library users how to use electronic resources. These expenditures of money, time, and effort are, whether we like it or not, at the expense of more traditional library collections and services. To take just one example, an annual subscription to an expensive electronic resource is money that would otherwise have been spent on journal subscriptions. The difference is, of course, that the former expenditure is for a transient resource. It gives the library's users access to the current manifestation of the resource, with no guarantee that they have access to past manifestations or that users in the future will have access to "back files." If a library purchases a print journal, the act of purchase guarantees access to issues of that journal in perpetuity. Even if the journal ceases publication or the publisher goes out of business, the issues will continue to exist and be available in bound volumes or microform for as long as the library wishes. If, as they are wont to do, electronic resource publishers go out of business, the content of their electronic publications may well be lost forever. In the print world, preservation is a largely passive and routine activity. If we are to preserve electronic resources, libraries will have to be far more aggressive and active. Surely, if we are to justify the increasing amounts of money and human resources we are devoting to electronic documents and sites, we must ensure that money and time is not spent on impermanence and instant gratification.

There is no doubt in my mind that *the* major issue facing libraries today is that of the preservation and onward transmission of the human record. This task has been accepted, usually tacitly, by many generations of librarians and archivists. The fact is that we librarians and archivists, and we alone, are responsible for something that everyone now takes for granted—that each generation will know more than preceding generations. That cumulation of the human record is made possible because useful knowledge has been recorded and preserved and can be the basis for the creation of new knowledge, which, in turn, is preserved and made available by librarians and archivists. This cycle may seem almost commonplace but it should be remembered that the Age of Print may turn out to be an aberration in human history—a scant few centuries in which very few texts and images were lost and we used a commonplace technology (print on paper) to ensure that the people of the future know all that we know. This was not true of the Manuscript Age that preceded the Print Age and may not be true of the Digital Age, unless we, as stewards of the human record, take appropriate steps.

What have we done to ensure preservation of the human record in the Print Age and are there lessons we can draw from that experience? The first thing was a stable technology used within a powerful economic model. We humans printed texts and images on paper, made many copies of each document and distributed them widely throughout the world. That process was run, in most instances, on the reasonable desire of authors, publishers, and booksellers to be recompensed for their labors. In addition, we have, in recent decades, developed, if not yet perfected, a system of global bibliographic control so that the human record was not only preserved but also could be retrieved and shared globally. It was an age of fixity, authenticity, and stability—a time in which preservation and bibliographic control were two sides of the same coin. Now, in contemplating digital documents

and resources, both preservation and cataloguing are in doubt and each is dealt with often as if it were completely separate from the other. When it comes to preservation, our first problem is posed by the fact that, in the Print Age, we have relied on publishers to decide what was and was not worthwhile. Once published by a reputable publisher, a book or other document was deemed, *ipso facto*, to be worthy of preserving. This is by no means so in the digital arena. Electronic "publication" no longer implies any editorial judgment or any imprimatur. The orderly world of the creation, editing, publishing, selection, and preservation of books is in danger of being replaced by the electronic version of a vast, global town square with millions of people shouting on top of their voices. How are we to make any sense of this cacophony, recognize the worthwhile voices, and preserve their productions for future generations?

I believe that the answer lies in some innovative and strong-minded research—in particular, we need an enumeration and taxonomy of the Web and the Internet. Huge, inhuman numbers for electronic sites and documents are thrown around in discussing preservation and cataloguing, and they tend to depress rather than stimulate thought and discussion. What is one to do when confronted by the news that there are billions of Web sites or that popular sites register millions of "hits" in a week? There is nothing that one can do with such "information" and, in truth, it is usually designed to impress rather than to inform—to be a reinforcement of the idea of the inevitability of the dominance of electronic communication. In addition, phrases like "cataloguing the Web" have been used in a form of false egalitarianism that proposes that all electronic documents and resources are equal. The result of gigantic numbers and semantic vagueness has been a sort of despair that says that we will never be able to bring the products of the Digital Age preserve any substantial portion of the digital swamp or bring that portion under any recognizable form of bibliographic control. I propose that we combat this electronic triumphalism by embarking on a serious enumeration and taxonomy of the Web and the Net that is aimed at identifying and isolating those documents and resources that are worth cataloguing and preserving. In creating a taxonomy, I would propose that we consider a number of variables including:

> ➢ is the resource commercial or not (i.e., is its main purpose to sell a product or service);
> ➢ is it derived from print, archival, and other tangible documents or not;
> ➢ what is the nature of the entity issuing the resource—e.g., academic, learned society, scientific/technological, individual, commercial publisher, etc.:
> ➢ and is the document(s) or resource static, cumulative, or constantly changing.

There are other variables, of course, but that is the point of research. We can start with the variables I propose and add others as examination of facts warrant.

We tend to think of electronic resources as all belonging to one species. I think this is a great mistake. In particular, we should look at two crucial criteria. The first is the distinction between electronic documents and collections of documents on the one hand and sites on the other. The second is the relative mutability of the resource. We also tend to think that the most important demarcation is that between electronic and non-electronic documents. I believe that, when it comes to the preservation and cataloguing of electronic documents we should make other distinctions. Many such documents and sites are by-products of print and other publishing industries and, hence have far more in common with books, journals, sound CDs, films, etc. In addition, many sites are digitized archives that are either complete (i.e., static) or cumulative—that is, though they change, the change is in the form of additions and not deletions. Such sites are, again, analogous to "traditional" resources. A static (i.e., complete) electronic resource is the same thing, for our purposes, as a book. A cumulating electronic resource is the same thing as a serial (especially given the proposed new AACR2 definition of "integrating resource"). Web sites with constantly changing content, on the other hand, have no parallel in the world of print or publishing of other tangible objects. They are more like those curious structures one sees in the streets of Paris—erected to have posters plastered on them. The posters change, become torn and overlaid with newer posters, are removed and vandalized with graffiti, so that the content and visual effect differ from week to week. Only the site endures. Since the content is so unstable and shape shifting and since the sites themselves, unlike the Parisian structures,

can vanish overnight, they are difficult, if not impossible, to preserve and bring under bibliographic control. A possibly heretical question occurs to me—does that matter? Is the content of such sites of any enduring value? It is certain that the authors of such content place very little value on it, since they make no attempt to preserve it themselves. An enumeration of sites that gives an idea about their subject matter, their creators, and their life span is very different from the cataloguing and preservation of content to which we are accustomed, but it may be all that we can accomplish. Further, it may be all that such changing content deserves. I put this forward merely as a hypothesis, and leave it to researchers to show us whether that hypothesis is correct or whether there is some enduring part of the human record on these changing and vanishing Web sites that we should strive to preserve and, more practically, *can* preserve.

Think of axes in which the north of the vertical axis is labeled "non-commercial" and the south "commercial" and the west of the horizontal axis is labeled "static", the center "cumulative," and the east "ever-changing/impermanent." One way of approaching our central question would be to map all resources to the intersections of these criteria. As a working hypothesis, those resources that fall into the north-west quadrant—that defined as those non-commercial resources and sites that are static or cumulative—are likely to be the more valuable. Conversely, those that fall into the southeast quadrant—those that are more commercial and are changing or ever changing—are likely to be the least valuable.

Creating and maintaining the bibliographic control Web. I have just spoken about the interaction of preservation and cataloguing. I have also called for research aimed at resolving many issues centering on the definition of those electronic materials that need to be preserved and, therefore, should be catalogued. Further we need to create and maintain a structure of bibliographic control that will ensure the preservation of the records we create as well as the documents and resources they represent. The starting point should be the grand idea of Universal Bibliographic Control, first put forward more than a quarter of century ago, in which individual libraries, regions, and countries cooperate to produce and share records without redundancy. Then there is the question of cataloguing and metadata. My view of the latter is that it is an ill-considered attempt to find some kind of Third Way between the wilderness of search engines and free text searching and the grand architecture of bibliographic control that librarians have developed over the last 150 years. I think that metadata is the product those who with no knowledge of, or regard for, cataloguing—they are bibliographic alchemists seeking the philosopher's stone that will offer us effective cataloguing without expense and effective access without controlled vocabularies. There is no such thing and the sooner that notion is disposed of the better. Instead of the sterile discussions and failed schemes of metadata, we need enquiry and undisputed facts—in short, national and international agreements based on experience and reality. Those discussions should be devoted to developing an internationally agreed data set, a set of agreements on international controlled vocabulary data bases, interfaces between the artificial language of classification and the "natural language" of subject headings, and a developed international MARC format. We need a combination of research into framework formats, content formats, international exchange structures, and database management—and, above all, to clarify the distinctions between the various elements of the international bibliographic control architecture. Just as good research could lead to an international Grand Plan for preservation, it can lead to a complementary Grand Plan for bibliographic control of all documents, irrespective of format.

These are interesting times and we must do our best to rise to its challenges—in particular in ensuring that we play our historic role in preserving, cataloguing, and transmitting the human record.