Knowledge life cycles: renewal and obsolescence

By David F. Kohl

Resumen: La educación, el trabajo e incluso el sentido de la identidad de los profesionales de la información, así como los servicios prestados por las bibliotecas, han sido conformados por la tradicional necesidad de recoger, cuidar y poner a disposición del público una enorme colección de objetos físicos. Como la responsabilidad de esta infraestructura física desaparece, tenemos que repensar la forma en que vamos a realizar nuestra misión para recoger, organizar, conservar y poner a disposición los recursos de información de la humanidad. Se tratan diversos temas que están transformando el entorno de los profesionales de la información, como la digitalización, las colecciones digitales, el acceso abierto, la abundancia de información...

Palabras clave: Replanteamiento profesional, Reinventar la profesión, Cambios tecnológicos, Profesional de la información, Bibliotecarios, Bibliotecas, Digitalización, Ambiente profesional.

Título: Ciclos de vida del conocimiento: renovación y obsolescencia

Abstract: The traditional need to collect, care for and make available a huge collection of physical items has shaped the education, work and even the sense of identity of information professionals, as well as the services provided by libraries. As the infrastructure related to physical items increasingly disappears from our responsibilities, we need to fundamentally rethink how we go about our mission to collect, organize, preserve and make available the information resources of the human enterprise. Various topics that are transforming the information professional’s environment are discussed, including digitization, digital collections, open access, information abundance, etc.

Keywords: Professional rethinking, Reinventing the profession, Technological changes, Information professionals, Librarians, Libraries, Digitization, Professional environment.


DOI: 10.3145/epi.2009.jul.03

I AM PLEASED AND HONORED to be invited here today to talk with you at this Ninth ISKO Conference – and festive time in Valencia¹. I am particularly pleased because the topic of knowledge obsolescence and renewal has been a fascinating one for a librarian, such as myself, to reflect upon.

Unlike most of the things in our life knowledge doesn’t become obsolete because it rusts, or wears out, or gets diluted or is given to someone else. Almost uniquely among all the elements of our everyday life, knowledge becomes obsolete only when it is no longer a helpful guide to reality.

History provides a multitude of examples of fundamental shifts in human thinking, of examples of knowledge becoming obsolete, with profound consequences for human history and progress: the Ptolemaic view of the universe giving way to the Copernican revolution, or centuries later Newton’s worldview being supplanted by Einstein. But the example which most readily comes to mind here in Valencia is a man working for Spain, namely Christopher Columbus. As you know, in the late Middle Ages Europeans had discovered the riches of the Orient, of India and China, and their spices, silks, and gold. Sailing eastward to trade with these cultures made huge fortunes for both governments and individuals. But there was a problem—a lengthy detour around Africa, months spent sailing first south and then back north to reach the markets of the Orient. A direct route could potentially provide trading fortunes beyond the dreams of avarice.

The key stumbling block² to finding such a direct route was not the ships or sailors’ skills or even navigation technology. It was the need for a changed mind set, a new way of looking at the world. An earlier lingering view³ of the world as flat meant that the only way to go east was to sail east—and that meant a huge detour around Africa. Columbus, however, had developed a different vision of the world. He came to see the world
as round. That seemingly simple change in a way of thinking had huge implications. It meant that you could go east by sailing west, and so could arrive in the Orient without the huge African detour. As it turned out, the world was a more complicated place than Columbus expected and although he didn’t find a new route to the Orient, he did discover a “New World” and changed the course of history.

The lesson here is the power which our view of the world has over us and the degree to which it determines what we consider possible. Today, for librarians and information science professionals generally, I would argue that a “Columbus moment” has come. I have increasingly come to feel that the digital revolution represents not just a more efficient way to sail to the Orient around Africa, but an opportunity to “go east by sailing west” and thereby truly discover a New World. More prosaically put, we need to see the digital revolution as not just a better way of doing business, but a fundamentally different way of doing business.

**Nothing more than sailing a little faster in old ships**

Let me illustrate the difference between these two approaches by very briefly reviewing the use of digital technology in libraries in the last 40 years. By and large it has been a history of using digital technology to solve the old problems of a print world rather than to see and exploit the opportunities of the digital world.

The first serious applications of digital technologies in libraries was to enhance library procedures for dealing with print collections, specifically circulation systems and catalogs. Based on the *Quadraplanar structure* developed at the University of Chicago in the early 70s the goal of this and similar automation was use of digital technology for inventory control and record display of print (and other physical items) in library collections. While we talked about it in revolutionary terms, it simply represented a more efficient and accurate way of doing traditional library tasks. We were, so to speak, simply sailing slightly more well run ships around Africa.

The next key development was a better way to provide intellectual access to the print collection. In the 1980s we saw the development of online searching, i.e. digital databases which began to increasingly replace the library’s print indexes and abstracting reference tools. Costly and complicated to use (at least for the untrained) such resources were generally mediated by specially trained librarians to here too provide better access to print library resources. Although once again hailed as revolutionary this development too simply represented a more efficient and effective way of delivering traditional library services.

By the early 1990s a more fundamental transformation was beginning to creep into the library world as digital content began to seriously enter the collection. Joining such pioneers as *Iepsr (Inter-University Consortium for Political and Social Research)* longitudinal databases maintained on stacks of computer tapes, were CD formatted databases such as Sir Chadwick-Healey’s *English Poetry* (700 - 1900), compact discs for music collections and even, in the US, the *Government Printing Office* was increasingly distributing its massive information avalanche in a digital CD format.

By this time libraries had progressed to the point that they saw themselves providing not just print resources but resources in a variety of formats, e.g. microforms (3 main types with subtypes), audio tapes (two types), slides (two types), video tapes (two types), etc. So digital discs and tapes were seen as just one more addition to the United Nations of collection formats. More profoundly, although these new digital materials were again seen as a revolutionary development they were still packaged in a physical format, even at times deliberately made to look like books so that they could be shelved with the “real” collection. We were still just sailing better ships around Africa.

**Discovering the digital New World**

In the late 1990s the ground began to truly shake as the first of two even more radical digital tsunamis hit. Developments such as the Big Deal (mass journal purchases) and *Jstor* (retrospective digitizing of core journals) created a world-wide explosion of access to digital journal literature while print access to this same literature began to plummet. Both libraries and their users began a slow but steady withdrawal from print journal publications while simultaneously journal publishers not only provided digital versions of their journals going forward, but also began comprehensive digital retrospective conversion. While I am aware of no full and systematic census reporting the degree of academic journal digitization for the 20,000+ academic journals, patron use and library purchase patterns clearly indicate that digital journals are now the major and increasingly exclusive format for academic journals. Almost universally these digital journal collections are not held in libraries.

By the middle of the first decade of the new millennium the second radical digital tsunami struck. Google, *Microsoft* (for a while), the EU and *Open Content Alliance* had begun massive book digitization projects. In a complementary manner access to government documents also shifted. In the US, for example, rather than continuing to provide the digital information on physical discs, the government has made a substantial shift to information access via online websites. Meanwhile
we saw the development of high density storage facilities and a growing migration of library physical collections to them. Taken together these developments in conjunction with those of the journal literature earlier, reveal the same profound two part message for librarians.

“Digital information is no longer simply an additional format; it is increasingly the only format”

The first part of the message is that digital information is no longer simply an additional format for 21st century libraries; it is increasingly the only format. Not just the central elements of traditional library collections, journals, books and government documents, but music, the visual arts, everything can be and, in fact is, now represented digitally. The library increasingly is no longer a United Nations of formats, but a single world government which is the digital format.

“Digital information is increasingly independent of physical storage devices”

The second part of the message is that information is no longer, in an important sense, physically based. Not only is all information is increasingly digital, but that digital information is increasingly independent of physical storage devices such as CDs or computer tapes. Software, whether tools such as Microsoft Office, content such as Netflix movies, or research data such as Hubble astronomical data, is increasingly distributed not in physical packages but as data streams. It seems inescapably clear that while we have yet a way to go, the tipping point has been passed and the destination is obvious. Except for maintaining a relatively small archive of important print materials, libraries will no longer be able to find their identity as a storage place for information artifacts whether print on paper books or digital CDs. In short, our image of the library as a building filled with books needs to change. I believe it is not inaccurate to say that information is rapidly moving into the “cloud” and is, as far as libraries are concerned, no longer physically based. This represents a profound change in how libraries must now learn to think about themselves and their mission.

So what kind of road do we see ahead of us as we attempt to look ahead with a new vision? Although I think it is probably too early to have definitive answers, the shape of at least some of the key issues is already emerging. Let me briefly comment on a few of these which may prove helpful in starting the creative juices flowing as you prepare for the coming panels, discussions and presentations of this conference. These issues arise in three broad areas: technological challenge, sustainable funding and unfinished business. Remember, the goal is to help us see with new eyes, rather than through the limiting notions of our past.

I. Technological challenge: the information environment has changed from scarcity to glut

I can still remember the days when finding any information on a topic was a triumph. Now, a few keystrokes in a Google search can frequently turn up thousands, sometimes millions, of potential information sources. Like food, information has become available in quantities unimaginable even a century ago. And just as our bodies and cultural habits have a difficult time dealing with this new abundance (even the Chinese are reporting an obesity problem), so also our libraries and other information structures find traditional tools and procedures inappropriate to the new tasks demanded of them.

It seems to me that there are two key issues emerging here:

1) how we prepare digital information so that it can be retrieved efficiently and effectively, and
2) where do we find the expertise for identifying useful and reliable information?

We are in a completely different world today from that time when a 19th century William Poole sat on the cross town bus each morning with a stack of journals and a notebook and indexed journals articles on his way to work at the Boston Athenaeum. One of the biggest differences between the print and the digital world is the amount of information available. The quantity of digital information is today so great that it can only be organized for retrieval by machines. Although humans continue to be the end users of information (so far) the interface which makes that information retrievable is by necessity increasingly machine driven. This raises fundamental questions about the role of traditional cataloging and classification schemes. What role, if any, do such traditional systems? Do they play a useful role in our understanding and creation of the developing semantic web and information ontologies? In short, do we still need Dewey, LC and Cutter? This question not only has implications for professional practice but for library and information science education (and accreditation) as well.
But even accepting the idea of machine indexing is insufficient if we have in mind today’s tools.

**Automated search**

The emerging challenge is automated access and retrieval from the “deep web.” Even Google, who this past summer announced that its crawlers had identified and indexed its one trillionth website, only provides access to the surface of the web. The full content of the databases which underlie many of the web sites remains opaque and unavailable to automated indexing. This is content such as individual book records in library collections, airline flight schedules, underlying statistical charts and tables, and the like which presently can only be accessed by a searcher typing in a query. For example, present search programs can identify hotels in Valencia, and I can then go to the individual sites to check room prices for a given date. But I cannot simply ask Google to find me the best room price in Valencia on a given date. Preliminary work on deep web access is underway by such groups as Kosmix and DeepPeep but we have barely begun the journey to the web’s information content.

**Spread knowledge**

A second aspect of the ongoing technological challenge is how do we most effectively use the human component in identifying useful and reliable information. Traditionally, the main strategy has been to rely on a small body of experts to provide guidebooks, encyclopedias, book reviews and even catalog records. Given the relatively high costs and physical limitations of communication in the print era this worked relatively well. The 1911 Encyclopedia Britannica is a perfect example with articles on physics by Ernest Rutherford, on poetry by Charles Swinburne, on philosophy by Bertrand Russell, and many, many other expert contributors. But as the barriers of cost and physical limitation have fallen in the period of streaming digital information, there is a serious question as to whether the community at large might be a better and more reliable provider of and guide to good information. Such experimentation, e.g. wikis, libraries allowing patron tagging and Amazon.com encouraging community provided book reviews, has mushroomed. This fundamental shift in thinking was nowhere more clearly illustrated than in the announcement late last year that the EB would be reconstructed as a wiki. This is a long ways from 1911 and the stellar cast of experts. Like all major transformations, however, we need to proceed with care. Both the community generated book reviews on Amazon.com and the entries of the Wikipedia have turned out to require serious adjustments and safeguards. And Nature’s year long experiment with open peer reviewing of submitted articles was recently discontinued as a failure. There is a great opportunity here, but not a simple one to grasp.

"The community generated content is a great opportunity but not a simple one to grasp"

**II. Sustainable funding: how do we reliably pay for digital information?**

Folk wisdom tells us that money is the root of all evil. These days that saying seems to reflect more than just a little insight. But as mature adults, we know that money is more complex than just causing serious mis-

http://www.kosmix.com/

http://www.deeppeep.org/
chief in the marketplace. It is also the basis, indeed a necessity, for opportunity. In the maturing digital environment issues of funding and support are both critical matters and still far from being resolved.

To talk reasonably about funding, I think it is necessary to clarify a number of issues so that we are talking to each other rather than past each other.

The first issue is that information costs money – always has, always will. Information is not free. The whole debate about Open Access for example, is not whether information should be free, but who pays for it. The argument is whether a central government, or a local institution, or an author, or an individual user, or a not-for-profit granting agency should pay for it. While our users often make the mistake of thinking that information is free, we as librarians and information professionals need to be wiser. What has changed is what we pay for. In the print world creation of the physical item and its distribution involved significant costs. This is not the case in the digital world where the significant costs have been transferred to quality issues (editorial work) and certification of academic value (peer review).

“Open Access is not new: every community served by a library is an OA environment”

The second issue is a misunderstanding of the question of Open Access. Typically the question is framed as: is OA desirable or is OA possible. In fact, in a very real sense we already have widespread OA and it’s not new. Every community served by a library is an OA environment. Students, professors and staff do not pay a direct charge to use the library nor do different amounts of library use cost a patron different amounts of money. For a member of a library community all the library’s information is open access. A key part of a library’s identity is that it is a mechanism for breaking the link between accessing information and paying the costs of that information. In essence, therefore, we have thousands of islands of open access communities throughout the world. What we are really discussing when we talk about Open Access is the degree to which we expand and join those separate OA islands into a national, regional or even world continent. A kind of information Pangea.

The third issue it is important to be clear about is the role of traditional faculty in the world of digital information costs. There seems to be a myth in the library and information science community that faculty solidly support OA. And this is unfortunately not true. While there are very loud individual faculty voices arguing for OA, traditional faculty as a group, particularly in their societies, are extremely divided about the funding of digital information. The evidence for this state of affairs comes only partially from almost universal reports of faculty reluctance to deposit their papers and articles in local institutional repositories (forcing university administrators to require such actions). Primarily faculty ambivalence is revealed through the actions of the academic societies:

– Not long ago the American Psychological Association announced that authors in its journals would not be allowed to provide their articles to publically available institutional repositories;

– It is an open secret that many academic societies who contract with either not-for-profit or commercial publishers for the publication of their journals actively encourage bidding wars in order to maximize income from their journals with resulting major increases in journal prices. For example, Project Muse at Johns Hopkins which publishes a large number of academic society journals announced a 15% increase in its journal subscriptions for 2009, even though Icolc and ARL have issued a call for a worldwide moratorium on journal price increases during the present financial emergency;

“Generally speaking, the group of traditional faculty doesn’t support OA”

– IEEE (originally Institute of Electronics and Electrical Engineers) and ACS (American Chemical Society) have long opposed OA and kept subscriptions high;

– Even Harvard, in a perfect example of academic schizophrenia on this topic, last year reported a vote by its college faculty to publish only in OA journals while almost simultaneously its Business School announced both a major price hike on its prestigious Harvard Business Review and a policy of not offering discounted consortial pricing. Needless to say the HBR is not an OA journal.

The basic issue here is that journal pricing and OA cannot be dealt with only by library and information professionals but must involve the larger academic and research community who have complex and even conflicting agendas.

Adding to the confusion and almost completely unaddressed so far is the issue of research informa-
tion generated outside of the university environment—though often by faculty through special arrangements with government or commercial institutions—in areas such as pharmacology, medical engineering, or defense. This proprietary or classified scholarly information is completely unavailable at present.

“Much university research generated through special arrangements with government or commercial institutions is classified and completely unavailable”

It is in this context then that we need to explore the issue of sustainable funding for academic research and dialog. It is my sense that we are still very early in this process, but that the ultimate solution will be complex and evolutionary, not simple and ideological. For example, PubMed Central appears to show promise as does the University of California system’s recent agreement with Springer to allow UC faculty publications in those journals to be OA. On the other hand last year Yale Library announced that it would not participate in BioMed Central because the cost was prohibitive and the model considered unfair; Brown University Library reported a similar assessment of the Public Library of Science and last year the Swedish university group which maintains the DOAJ (Directory of open access journals) felt it necessary to ask the library community for donations to support the site. On a more personal level a striking example of faculty ambivalence came to my attention at this year’s midwinter ALA meeting when a faculty member angrily told me that she was returning to publishing only in commercial journals given what she felt were the outrageous page costs of OA journals on the one hand and the attractiveness of the increasingly liberal author retained rights of the commercial journals on the other. In short, I strongly suspect that the most productive way to view such turmoil is to see it as a creative fermentation, not as an ideological struggle of perfect good against perfect evil; and with an outcome that is likely to be multifaceted and practical rather than unitary and simple.

III. Unfinished business: crossing the last major river...

The major elements of traditional library collections have always been books, journals and government documents. While journals and government documents have reached digital maturity, the path of books has been more difficult and slow. In the last two years, however, we have begun to make serious progress in tackling this monumental task and if we have not yet crossed the river we are certainly in the boat and rowing hard. There are two key tasks here:

1) digitizing these massive collections and,

2) providing access to this digital collection.

Digitizing the monographic collections involves three further sub-issues largely settled but whose effectiveness and implications we are still sorting through. These are: funding source, technology, and copyright. Although various beginnings have been made in recent years to digitize small, specialized monographic collections in Europe and North America using local or EU grant funding, serious digitization did not begin until Google, followed quickly by Microsoft and then large scale grant and government funding, developed serious programs of monographic digitization. Microsoft has since dropped out, but three distinct modes of funding continue. Google represents the main commercial thrust partnering so far with major North American and European libraries (7 million titles); the Internet Archive, a project of the Open Content Alliance represents a committed approach to grant money (1 million titles), and Europeana and European Digital Library are probably the best examples of government funding. While there has been a certain amount of argument regarding the pros and cons of each approach, I’m not sure any of it really matters. Most libraries, at least in the US, seem to see the digitization projects as essentially practical problems requiring practical solutions. Many libraries, in fact, use multiple funding models, even at times with the same book being scanned repeatedly in order to fit the conditions of different projects. Still, entrusting a major part of our cultural heritage to a commercial entity, Google, the dominant player by far...
at this point, requires caution and thoughtful reflection. The considerable financial advantages which Google provides libraries in this area needs to be carefully balanced against possible negative tradeoffs.

The question of developing the necessary technological tools may also be largely over. Some years ago when I was at Gottingen in Germany I was shown the complex scanner and procedures for digitizing their copy of the Gutenberg Bible. It was not exactly custom made, but close to it. This January at ALA Midwinter there were something like six vendors actively selling machines to digitize monographs. While the decision of which specific scanner to use remains, the projects and their tools are now mainstream, easily available, and apparently affordable for institutions.

For a while it appeared that copyright would be a serious stumbling block. But the recent settlement between Google and copyright holders seems to have removed this as a serious issue in the US and possibly, given the harmonization of copyright laws between the US and the EU, may provide guidance for copyright issues connected with EU digitization. Here too, although we are not completely out of the woods, there does appear to be daylight ahead.

Digitizing the monographic collection, however, is only half the story. Users need a way to access this material. The earlier idea of expecting library and information users to access monographic information by means of institutional or home computer screens has clearly not worked. What we have done is turn such users into mini-publishers who use computers to find information and then print out hard copy. Recently, however, the increasing popularity of the Sony and Kindle readers and their growing number of “me-too” devices suggest that the tipping point for a truly paperless solution may be at last approaching. Underlining this new direction is the striking and rapid decline of print technology throughout society generally—newspapers, books, and magazines facing readership loss and even bankruptcy.

To be sure issues remain for ebook readers. There is yet to be developed a software standard which would allow Sony e-books to be read on Kindle and vice-versa. And while impressive advances have been made in reader technology in terms of battery life, back-lit screens, and liquid paper, they still can’t do color and we still don’t know what the optimal or acceptable screen size is. It is intriguing that while some in the West still wonder if the screens are large enough, Japanese teenagers—some of the most sophisticated users of digital technology in the world—apparently are downloading and reading books on their cell phones. As often the case in the past this may be a foretaste of things to come as both businesses and libraries in-
creasingly report an interest in making their websites mobile-friendly.

And there remains the nagging possibility that the best use of digital technology may be still to produce print products. Several months ago the University of Michigan announced an experiment in print on demand where digital books (out of copyright) could be reproduced as a printed, bound book for $10.00 (slightly less than 8 euros) each. To put this in context, keep in mind that the cost of just loaning this same book via ILL costs $31.00. With a digital collection it is 3 times cheaper to print out and give the book away than provide it through ILL from another library. The idea of a fully digital collection which is printed out on demand probably shouldn’t be ruled out just yet.

A final issue which has not been much addressed at all but is a significant one is what to do with the printed books after digitizing. Clearly there will be many heirloom items (e.g. Gutenberg Bibles, works of art and historically significant items) which libraries will want to retain in active on-site archives. But there will also be thousands upon thousands of items whose only value is their intellectual content and so largely superfluous once that content is digitally transformed (it is important to note here that digitization of the print monograph collection means not just ascii text but also a digital image of the page as well if we are to have true preservation). My guess is that it would make sense to consolidate such materials into economically efficient high density storage, remove unnecessary duplication, and view them as interim emergency backups to a digital system which is still under development. Catalonia has already taken an important step in this direction with the establishment of a high density storage facility in Barcelona. Ongoing funding, determining appropriate levels of duplication, and serious political issues surrounding the discarding of books remain concerns to be worked out.

Conclusion

In summary, the message today is that the digital transformation of information simply doesn’t fit into the traditional library and information structures of the past. Nowhere is this more clearly seen than in the change in how information is stored and accessed. In the print world where information was stored in physical artifacts, it was necessary not only to build huge buildings but to place them physically at the heart of the university or in the departments themselves so that they could be accessed conveniently by students, faculty and staff. Digital information is not constrained in this manner. While it too requires an infrastructure it is a communications infrastructure, not a storage infrastructure. It can easily be available any time and any place.

The reason this fundamental change is so important to libraries and information professionals, is that so much of our education, our work, the services we provide and even our sense of identity has been shaped by the traditional need to collect, care for and make available a huge collection of physical items. As the responsibility for this physical infrastructure disappears, we need to fundamentally rethink how we go about our mission to collect, organize, preserve and make available the information resources of the human enterprise. We need to see our historical mission with truly new eyes. Like Columbus we must adopt a new view of the world and begin to understand the possibilities of this new vision, how we too can go East by sailing West. Today and in following days we will be exploring the details of just how to make this momentous trip. It’s a big task, but an exciting one, and an undertaking that will help define our common future for decades if not centuries to come.

Thank you.

Notas de la Redacción

1. En Valencia era la Semana de Fallas.
2. The key stumbling block = El escollo clave
3. lingering view = persistente visión
4. Quadraplanar data structure
Sistema informático para poner en un catálogo común la información bibliográfica de varias bibliotecas, conservando su operativa individual. http://www.jstor.org/pss/4306754
5. Inter-University Consortium for Political and Social Research (Icprs) http://www.icpsr.umich.edu/
7. William Frederick Poole publicó en 1848 Index to Subjects Treated in the Reviews and Other Periodicals. En 1853 lo mejoró con el título Index to Periodical Literature, obra considerada precursor del The New York Times Index.
8. 1 trillion (EUA) = 1 billón (España) = 10³³
9. there is remains the nagging possibility = sigue la persistente posibilidad

David F. Kohl, Dean and University Librarian, Emeritus.
2929 Courtropes Lane, Cincinnati, OH 45244, USA (Dean and University Librarian, Emeritus, University of Cincinnati) Tel.: +1-513 624 0503 david.kohl@uc.edu