

## Whither Academic Information Services in the Perfect Storm of the Early 21st-century?

© Michael A. Keller 2006

What constitutes the perfect storm in the information age? Here are a few indicators:

1. Nearly ubiquitous network access; low-cost computers and PDAs; plentiful, affordable magnetic memory;
2. a commercial culture that is increasingly “just in time” with ordering, delivery, and payment services predicated on constant network access; a popular Internet culture seemingly devoted to openness; an investment culture focused on quarterly profit reports that drive stock prices – a short-term view of the viability of a company;
3. Google; Yahoo; MSN; millions of information providers offering free or fee services; millions of non-commercial information providers in the blog-o-sphere, including Wikis and Rss feeds;
4. Course management systems; collaboration environments and applications drawing globally widespread communities more closely together; institutional information environments lacking branding of the agencies responsible for them, namely the libraries;

The combination of these vectors gives rise to many of the papers on this program that predict new models and new paradigms for various intermediary agencies: libraries and other information services; publishers and other information providers to name but two. These forces give rise to social movements, like the open access movement, that may accelerate disintermediation beyond the effects of the various indicators just named. In the opening keynote address, Lynne Brindley spoke not

just with the authority of her office, but on the basis of an extensive study, about the directions she is taking the British Library to sustain its role as a great national cultural entity. Jerry Campbell in a recent article entitled “Changing a Cultural Icon: The Academic Library as a Virtual Destination” in the Educause Review wrote: “Today, however, the library is relinquishing its place as the top source of inquiry.”<sup>i</sup> While I agree with Lynne’s approach, especially for a nationally chartered cultural institution, I think Jerry Campbell’s theme and philosophy regarding the future of libraries are too centered on the profession of librarianship itself and not enough on our clients. My own position is that we in the libraries, the merged organizations of libraries and information technology support agencies we might call “cybraries”, and academic computing organizations derive our value and validate our missions by serving and supporting the researchers, the readers, the students, and the faculty making use of our services, our collections, and our facilities, even while they exploit information and information services provided by others. If our professions are to survive, they will do so, because we provide services that our clientele want and need. In short, I support the client centric view of information services, both traditional and innovative ones.

Some cases that prove my point also suggest possibilities for cybraries to adapt already extent services and to more flexibly allocate their inherent capacities to their clients. Several large research libraries, those at the University of Michigan, Stanford, Harvard, the New York Public Library, and Oxford, have agreed to let Google digitize large portions of their book collections in order to improve intellectual access by word indexing to the contents of the millions of books potentially to be treated by the project . Of course, those who would use the Google Book Search service have to be able to get to the web at all. Suits and counterclaims, caviling critics and hungry lawyers to the contrary, the motivations

for participation were and are both intellectual access and digital preservation of information assets. Other similar projects have arisen, principal among them the Open Content Alliance and the British Library project involving Microsoft. All of these projects pre-suppose adherence to copyright laws, though the Google project depends strongly upon the U.S. fair use defense provisions for the intermediate digitization in order to create digital indexes to works not in the public domain. This defense was specifically strengthened in a very recent U.S. district court decision, *Field versus Google*, filed on 19 January 2006.<sup>ii</sup> A federal district court in Nevada has ruled that the Google cache feature does not infringe U.S. copyright law. The Google cache holds works for indexing, but does not necessarily display whole pages of the works in the cache. The ruling clarifies the legal status of several common search engine practices and could influence future court cases, including the lawsuits brought by book publishers against the Google Library Project. The Google Book Search, Open Content Alliance, and British Library book digitization have been preceded and accompanied by numerous other book digitization projects, of course. Each and all of these seek to improve the nature and extent of intellectual access to books as well as to begin major digital book preservation programs, admitting that digital preservation programs are all more or less experimental at this time. We might characterize in very broad fashion the information services arising or enhanced by these sorts of projects as discover and locate functions as well, perhaps, as the request and deliver functions.

Some of you may be aware of Sakai, the project involving several U.S. institutions that led to an open source collaboration application that is also a course or learning management system. Sakai and other course management systems provide a range of modules that support teaching and learning, making more efficient most course administration functions, like registration, section assignment, testing, and grade

reporting, while making possible or enhancing certain pedagogical functions, such as delivering lecture notes and videos, collaborative document creation, and 7x24 discussion sessions on course topics. Among the modules of these course management systems is that of the e-reserve, now re-constituted to become the set of references to required and recommended texts, videos, and other resources, some of them possibly physical along with the web authorization function that makes access to networked information resources compliant with copyright, again based in the U.S. on the fair use provisions. We might characterize some of these functions provided by the course management systems as locate, gather, deliver, create, and sharing functions. We also contemplate the provision of e-portfolio services so that student and faculty “authors” might easily store their best “papers” for long-term access in our digital repository, a service that our alumni affairs officers relish. This service has been underway for some time at several institutions, among them the National University of Singapore through the libraries, and has proven to be a very popular one.

So, between the book digitization projects and the course management systems we can see several services provided via the web for individuals without the direct intervention of cybrarians or i.t. professionals. The ones mentioned are discovering, locating, requesting, delivering, gathering, creating, and sharing. The methods for providing these functions arise from these sorts of systems: online catalogs of library holdings; search engines; course management systems; institutional information topographies; web browser environments; and the web itself. The users will benefit further when other systems are developed for true federated searching.

The fears evident in the suits against Google by the U.S. Authors Guild and five major trade publishers certainly inform Jerry Campbell’s paper and many

librarians who agree with his underlying premise that libraries are being superseded by the likes of Google and Yahoo. Are we being dis-intermediated by such web services? Or are we gaining valuable new tools for our clients for many of the discovery and retrieval tasks for small and large bites of information?

One response to that fear is contemplating some of the functions beyond those presently offered by Google. It is important to qualify in advance these functions to say that while we all hope to offer our campus communities on-line reading functions, we will do so under the provisions of the appropriate laws, particularly in the U.S. that of section 108 that is presently under study now for possible amendments appropriate to the digital age.<sup>iii</sup> The recently reported study with recommendations for amendments concerning so-called orphan works is relevant as well.<sup>iv</sup> With these legalities in mind, the functions beyond reading on the web and vastly expanding the digital resources one might find in an e-reserve, some of the services we at Stanford contemplate, based on digitized books whose copies are owned or licensed by Stanford, are:

1. taxonomic indexing, providing access to ideas more or less independent of the exact expressions of ideas across many texts;<sup>v</sup> this has been accomplished in the HighWire environment;
2. associative searching, providing access to texts by statistically associating terms within individual texts and providing a relevance ranked list of similar texts based on the degrees of difference in the association;<sup>vi</sup> this has been accomplished in the WebCat Plus environment in Japan by Prof. Akihiko Takano at the National Informatics Institute;
3. hyperlinking of explicit citations in footnotes, endnotes, or bibliographies to the digital versions of the cited references;<sup>vii</sup> this has been accomplished

- first by HighWire and subsequently through the CrossRef service;  
hyperlinking of textual references might come next;
4. Novel graphical interfaces to ease navigation and discovery in complex information topographies; <sup>viii</sup> this has been accomplished by the Groxis company through their service called Grokker;
  5. More subtle searching, allowing readers to direct their queries to metadata, titles, full texts, or abstracts; Bjorn Olstad's paper, coming later in this conference, promises to offer some new insights in this service and on graphical interfaces; this has been accomplished by HighWire;
  6. alerting services for readers so that they will know what new information resources have been added each week to the available collections; these alerts will be driven by the express interests of each patron, based either on specific keywords or entire texts, or by taxonomic or associative "fingerprints"; <sup>ix</sup> this has been accomplished by HighWire;
  7. recommendation services, perhaps categorized by the nature of the user, whether faculty member or researcher or student; <sup>x</sup> this has been accomplished in a fashion by Amazon;
  8. application of the "info-tools" service provided now by Ebrary that makes getting dictionary definitions, map locations, and biographical sketches convenient while reading a text; <sup>xi</sup>

Each of these services would improve intellectual access, speed research, improve teaching and learning, and generally make our readers more self-sufficient in their discovery and retrieval behaviors. Each will require focused work by librarians and programmers to create the necessary algorithms, programs, and applications.

Some will require systems integration. The hyperlinking service will need a lot of hand work, because of the numerous styles of citation employed in the texts in our

collections over the centuries. In short, there is no end in sight for information professionals to craft, operate, support, and improve systems that will provide the services just enumerated.

Beyond the empowering of our readers and users, there may be some “high touch” services will continue to engage subject specialists, meta-data specialists, and other technical professionals. For specific communities of scholars and students, the creation of special web environments for synthetic or even specific collections will be a powerful new set of services. These web environments exist now in the form of “knowledge environments” or subject gateways providing numerous information, communication, and publishing services to narrowly defined communities of scholars or on very specific topics. An example of this is the Harvard College Library site entitled “Women Working 1800-1930” that features a carefully selected collection of manuscripts, publications, images, and other documents from the Harvard collection as well as a set of teachers’ resources and related links.<sup>xii</sup> Another example is the Matthew Parker Online Library of Corpus Christi College, Cambridge University; the site is still in development.<sup>xiii</sup> For each of the 537 Anglo Saxon manuscripts, an extensive bibliography of secondary resources, modern editions, reference works, and searching functions will be supplied. The site will eventually support courses in English Medieval studies, paleography, and history as well as collaborative research for scholars anywhere. These two projects illustrate two clusters of services cybrarians might provide in the near and distant future; one is the assembling of resources to make the research and pedagogical functions more penetrating and just plain easier. The other is the function of bibliographers in assembling information and keeping it up to date to guide novices just entering a new domain of interest.

Turning now to the traditional functions of librarians, consider the findings of an internal study Stanford librarians performed to assess how many books that we acquired in physical format were available in digital form. The study involved a random sample of 1373 English language titles drawn from 22,892 titles with imprint dates between 2001 and 2005 acquired in FY2005. The question we were asking was “how many of the physical books we acquired were also available in digital form, whether by leased access or purchase for local retention?”. Of that random sample of 1373 titles, 13.2% or 181 titles were available in digital form, of which further analysis showed that only 3% or 41 of the titles or about were available for purchase for local retention and use. In addition to the results of these studies, admittedly limited ones, there is notable reluctance by book publishers to engage digital publishing and network distribution, much less experimenting with new, mixed media genres that could only be presented on the web. That reluctance is based on the desire to maintain their current levels of sales and thus of profitability, to be sure, but is as well a reflection of the difficulty of reading using existing technologies. Laptops are too cumbersome. Desktop computers enforce a limited range of postures if one is to read. PDAs have too small screens for reading extended works. The new reading appliances, such as the new Sony e-book reader, marketed as the Librié in Japan and arriving in the US this month, are not yet proven, though the experience in the Japanese marketplace suggests a small consumer base for e-books and e-reading appliances. This suggests that physical books, and particularly ones with extended texts and larger formats will be with us for some time. We look to the recently announced intentions of Harper Collins to create their own library of e-books for new insights. Google, by the way, will index that collection and from that index direct readers to the Harper Collins site, copying a relationship developed between Google and HighWire some years ago.

With lots of physical books in our immediate future, some of us can breathe more easily perhaps.

Earlier, I glossed over the idea of federated searching, but want to return to it now. Google indexes about 9 billion pages. Google Scholar is providing lots of specialized capacities, but for an as yet limited range of materials of scholarly interest. We believe that there are about 100 billion pages on the Web, about 90% of them behind some sort of access control. Google, Yahoo, MSN, and others cannot reach those web sites and thus cannot index them. Numerous publishers and governments control those 90 billion pages and provide access and indexing to them under contractual terms. Librarians have been the selectors and aggregators of access to that very large majority of the web's contents. That there are too numerous indexes and too few innovative analysis, discovery and retrieval functions for that majority is a serious limitation on the easy and effective use of that material for study and teaching. Librarians, principally through their library management systems vendors and other third parties, are attempting to devise and operate reliable and consistent meta-search engines. An example of well-functioning meta-search across a significant, but limited set of disparate information resources is the Flashpoint service provided by the Research Library at the Los Alamos National Laboratory. Flashpoint provides a single search interface to BIOSIS, the Web of Science databases, the Engineering Index, INSPEC, MathSciNet, and PubMed. This is a good, wide target, but the service works because LANL can re-format the various databases into a single format; it is thus not the same thing as a true federated indexing service. Much work is needed in this area.

Another topic of considerable interest should be seen as another service or even as a collection of services. That topic is, of course, digital repositories of various

designs for long-term persistence of bits and bytes. The phrase digital archiving only loosely defines the function, but I mean both that one and institutional repositories. There are several papers on these related subjects, so I will not comment further other than to say that the services provided by large dark archives, whether on a central basis such as the KB or on a distributed basis such as LOCKSS and its derivative CLOCKSS, and the services provided by institutional repositories to function efficiently need to be included in the same architecture as all the other services mentioned heretofore.

The Aquifer program of the Digital Library Federation is developing a series of middle-ware services, standards, and collection policies that, when joined with other user services like searching and gathering, will enable users to see collections in many institutions as one, to retrieve and make use of digital objects from those many silos of content. The idea of Aquifer is to generate or adapt the machine-to-machine tools, the metadata standards, the data formats, authorization mechanisms, and other services so that a certain level of reliable and consistent end-user services will make distant collections interoperate with local ones.<sup>xiv</sup> Essentially Aquifer is about building or adapting tools as well as building or adapting standards.

However, tools and standards like the ones on the Aquifer agenda need to be seen in the context of an architecture as well.

The notion of a digital architecture that is based not primarily upon systems, but upon a service framework is very much in the minds of the members of the Digital Library Federation. Thanks to the initiative of Lorcan Dempsey and a group of technical leaders from DLF institutions, DLF has begun to look to a Service Framework to organize the work of providing user services, web services, and library management functions in a rapidly changing environment. The Services Framework is not an architecture per se, but an organizational tool helping us to

think about the integration of systems, applications, and standards while keeping the perspective of the user, the customer very much in view. To provide a bit more detail on this concept, let me quote from the Services Framework report to the DLF Board in May 2005:

What is a service? A service is any functional component which it is useful to talk about as a unit. At one level, of course, the library itself is a service. The reference desk is a service. The catalog is a service. We recognize that a growing proportion of library services will be encapsulated in automated systems.

A service may be directly delivered to a user interface. However, not all services will be directly used by humans. Some services will be used by other applications, or 'intermediate consumers', which build applications from one or more services. In this case, a service may provide a machine interface to the consuming application. Examples of intermediate consumers might be a campus portal, a course management system, a meta-search engine. Increasingly, library applications themselves will be consumers of services, assembling their functionality from different components.

This modularization will become more common as libraries and their digital environments become more complex and interconnected. They need to be able to build systems flexibly and responsively. They need to reduce the cost of making changes. All of this points to the need for a modular approach, building up applications from simpler parts. This in turn raises several questions about the nature of those components and how they communicate:

- At what levels of granularity and aggregation should services be designed?

- How do we move existing services into a digital networked environment?
- How to we design services to be flexible and responsive?
- What machine and human interfaces should services support?
- How should services be designed so that they are reusable and interoperable?
- What functionality should be split out into a shared service infrastructure? rather than being created again and again (authentication, for example)?

A Service Framework helps us address these questions. Specifically it provides:

- A shared, consistent vocabulary for discussing library processes and services.
- Common ground for talking about processes and services within and across libraries.
- Common ground for talking to other domains – e.g., e-learning or e-research – about how they might take advantage of library services;
- A basis for identifying reusable components and interoperability needs.
- A roadmap for identifying priorities for shared attention or development, and for identifying gaps and ‘pain-points’. <sup>xv</sup>

We need now to consider the means by which information professionals will constantly alter their tools and methods while applying knowledge and experience to the on-going development of the services mentioned in this paper and dealt with in considerably more detail in most of the other papers of this conference.

Speaking as one who has led and demanded re-engineering at Stanford for some years, I affirm that once re-engineering has been accepted and shown useful results, most staff find the continuous re-engineering of the methods, the work, the information environment stimulating and satisfying. A simple, but profound change of perspective makes re-engineering at least palatable, both

organizationally and personally. That change of perspective is from the appreciation and valuation of work done every ten minutes at every person's desk to the comprehension of how that same work contributes to the highest mission and goals of the institution employing one. At Stanford, getting our technical services people to see that their work was not simply increasing the through-put of the acquisitions, serials, and cataloging departments, but the fueling of the great engines of education and scholarship with new sources. Once one has lifted one's view from the immediate to the strategic, it is easier to conceive of new methods and tools as well as to apply them. That same principle applies equally to the constant flow of new possibilities for all sorts of information professionals in our settings. It may very well be that the new systems and services implemented by folks like us or folks like our colleagues from Google will make some activities obsolete. That obsolescence should be seen in the larger ebb and flow of specialties needed for all sorts of other services, as opportunities for organizational and personal development. If fewer of my staff are needed for capturing or creating new bibliographic records for new books or e-books because our booksellers are providing us with better and cheaper ones, then more effort will probably be needed to operate the Stanford Digital Repository, tune up our acquisitions, and work directly with our clientele to understand both the structured information topography we provide as well as the means to navigate and assess the information chaos out there on the web.

From my perspective, visiting and making use of the information and information services of our virtual library should be as satisfying as visiting and using our physical collections and facilities. Certainly our facilities should be welcoming and our in-person services both easy to engage and very deeply comprehensive for the high touch, the personal interactions that cannot be replicated easily in the

virtual world. We imagine that some of our libraries in their next incarnation, Stanford's Engineering Library now in planning for instance, will ultimately not have any books at all. Yet Stanford's new Art Library will resemble in many ways the traditional art library with lots of big books as well as plenty of on-line services. One size will not fit all for the next several decades, we are betting. Our subject specialists and our reference librarians will still teach in about 100 courses each year the basic bibliographic literacy skills and maybe even an information heuristic, often using computer assisted instruction sets and our Sakai-based course management system, CourseWork. We need our clientele to become more self-sufficient for the vast majority of their information research activities so that when they really need help from deeply knowledgeable people, there is time enough for those high touch services.

All of the services and all of the functions performed in, by, and for libraries suggest the numerous roles librarians, cybrarians, and information technologists will be playing for the foreseeable future. The systems that support and enable:

- the selection and gathering of library materials, both analog and digital;

- the creation or adaptation of services providing intellectual access to content of the widest varieties;

- the distribution of access to the content;

- the interpretation of the content, from the level of individual facts and ideas to concatenations of information and opinion, to knowledge for students and faculty;

- the preservation of the avatars of the content, physical ones like books and virtual ones like digital images for art history research and teaching; and

- the analysis, manipulation, and presentation of the content for research, teaching, and learning

are ones needing the attention of real, honest-to-goodness information professionals and needing that attention for a good long while. The entry of Google and others into the gathering of materials in order to provide intellectual access and to distribute that sort of access broadly, as well as to provide content access where the laws permit it, should be considered yet another set of services that will enrich the lives of our clients, even if complicating our work. The expanding spectrum of possibilities should be embraced, not resented. If in the long run, such service providers as Google and Yahoo predominate and the value of libraries and cybraries are considerably diminished in the collective cultural consciousness, in the marketplace of ideas, so be it. Our roles and methods now, in the past, and for the future should be devoted to improving the lives of our clients, our readers, our users, more and more including disadvantaged people at long distances from the great edifices of collections and services we and our predecessors have built. Let us work so that the rising tide of access to information, opinion, and ideas lifts all boats everywhere.

---

<sup>i</sup> Jerry D. Campbell, "Changing a Cultural Icon: The Academic Library as a Virtual Destination" in *Educause Review*, v. 41 (Jan-Feb 2006), <http://www.educause.edu/apps/er/erm06/erm0610.asp>

<sup>ii</sup> United States District Court, District of Nevada, Case CV-S-04-0413-RCJ-LRL, Judge Robert C. Jones, dated 12 January 2006, filed 19 January 2006; [http://www.eff.org/IP/blake\\_v\\_google/google\\_nevada\\_order.pdf](http://www.eff.org/IP/blake_v_google/google_nevada_order.pdf)

<sup>iii</sup> See the report at <http://www.copyright.gov/orphan/>

<sup>iv</sup> Information about the study at <http://www.loc.gov/section108/>

<sup>v</sup> For an example of both taxonomic indexing and a graphical user interface based on it, see article browsing at <http://highwire.stanford.edu/lists/artbytopic.dtl> and the topic map linked on that page;

---

<sup>vi</sup> For further information see <http://geta.ex.nii.ac.jp/e/> ; to engage an associative search, see WebCat Plus at <http://webcatplus-international.nii.ac.jp/en/>

<sup>vii</sup> For example, see the article by Giedrius Vilkaitis<sup>#1</sup>, Isao Suetake<sup>#</sup>, Saulius Klimauskas<sup>#1</sup>, and Shoji Tajima<sup>#</sup> "Processive Methylation of Hemimethylated CpG Sites by Mouse Dnmt1 DNA Methyltransferase<sup>#1</sup>", in the Journal of Biological Chemistry, 280, p. 64-72 (January 2005) at <http://www.jbc.org/cgi/content/full/280/1/64> for both retrospective and prospective hyperlinks; other examples abound in the e-journal literature and in MedLine; CrossRef (<http://www.crossref.org/>) is a service linking the works of numerous e-journal publishers;

<sup>viii</sup> In addition to the HighWire topic map, see the Grokker search visualization service at <http://www.grokker.com/>;

<sup>ix</sup> For an example of this service, see <http://highwire.stanford.edu/cgi/alerts>

<sup>x</sup> For an example of this service, see the Amazon.com recommendation system, an explanation of which is provided at [http://www.amazon.com/exec/obidos/tg/browse/-/13316081/ref=br\\_bx\\_c\\_2\\_0/102-7234960-1076906](http://www.amazon.com/exec/obidos/tg/browse/-/13316081/ref=br_bx_c_2_0/102-7234960-1076906)

<sup>xi</sup> See the info-tools service at <http://www.ebrary.com/>;

<sup>xii</sup> See the site at <http://ocp.hul.harvard.edu/ww/>

<sup>xiii</sup> The prototype site is at <http://parkerweb.stanford.edu>

<sup>xiv</sup> See <http://www.diglib.org/aquifer>

<sup>xv</sup> See the DLF site for the Service Framework: <http://www.diglib.org/architectures/serviceframe/index.htm>