What is Going on at the Library of Congress?

By

Thomas Mann

Prepared for AFSCME 2910
The Library of Congress Professional Guild
representing over 1,500 professional employees
www.guild2910.org

June 19, 2006

Thomas Mann, Ph.D., a member of AFSCME 2910, is the author of *The Oxford Guide to Library Research*, third edition (Oxford and New York: Oxford University Press, 2005) and *Library Research Models* (Oxford U. Press, 1993).

The judgements made in this paper do not represent official views of the Library of Congress.

What is going on at the Library of Congress? Several recent decisions by the current LC administration have produced firestorms of protest, both inside and outside the Library, that LC is abdicating its professional responsibilities to the national system of shared cataloging, as well as undermining its core mission to acquire, catalog, make accessible, and preserve its own unparalleled holdings–especially its book collections. Among these recent decisions are:

1. The commissioning of "The Calhoun Report" to provide an ostensibly objective cover to justify abandoning the system of Library of Congress Subject Headings (among other recommendations to downgrade LC's cataloging and classification operations). [See < www.guild2910.org/AFSCMECalhounReview.pdf >.]

2. The unilateral decision to stop creating Series Authority Records, in violation of the standards the Library previously agreed to in the national Program for Cooperative Cataloging. [For an overview of the damage this will do to researchers, see the national petition, with 3,495 signatures, to protest this move at <

http://www.petitiononline.com/MARC830/petition.html >

3. The decision to accept digital formats for preservation purposes in place of paper copies or microfiche for traditional materials that are not "born digital." (In May of 2006 Beacher Wiggins, the Director for Acquisitions & Bibliographic Access, and Deanna Marcum, Associate Librarian for Library Services, agreed to cancel print titles from the Emerald group of publications in exchange for access to digital versions; in the previous month they agreed to stop collecting American doctoral dissertations on microfiche from University Microfilms in exchange for electronic access to digital versions. (LC, according to its national library responsibilities, had up to then maintained the only full set of dissertations in preservable form in any American library.) It was decided that "access" to electronic copies was less expensive. Recent papers on the astronomical economic costs of maintaining digital formats, as contrasted to microfilm–even apart from the still unsolved technical problems of emulation and migration–were simply disregarded; they may be found at < http://jodi.tamu.edu/Articles/v04/i02/Chapman/ > and < http://hurstassociates.blogspot.com/2006/04/article-digital-black-hole.html >.)

4. The decision by the Library's Copyright Office to dumb down the cataloging of copyright receipts by recording only the information on the registrant's application form, without any inspection of the actual deposited items. (An essay on the problems this will create for future researchers is at < www.guild2910.org/CopyrightCatalog.pdf >.)

5. Perhaps most disturbing of all, the continual starvation and dilution of the Library's book-cataloging operations over a period of several years, with the claim that "inelastic funds" necessitate a massive retrenchment in this area, when in fact it is the administration's change in "vision" that now ranks the digitization of copyright-free special collections as a higher priority for the Library's funding than maintenance of traditional book cataloging operations.

The cumulation of these decisions marks a tipping point in imposing an agenda that has long been the goal of the current administration–a goal that will have profound negative implications for research libraries, as well as for individual scholars, throughout the nation. This goal can be characterized as "moving the Library into the digital age"–the "digital age," however, as it is envisioned from a decidedly blinkered perspective.

Characteristics of the blinkered "vision"

Several characteristics of LC management's vision for the future of LC, and of research libraries dependent on LC systems, are matters of great concern:

- It puts much greater emphasis on digitizing LC's "special collections"

(manuscripts, maps, photos, etc.–in general, our non-book formats) at the expense of cataloging and classifying our "general collections" (i.e., the books).

- It effectively subordinates LC's book collections as being less important to scholarship than its non-book formats.
- Its new concentration on special formats entails a de-valuing of cataloging and classification operations for book formats–effectively a turning away from maintenance of the cataloging systems that research libraries everywhere depend on.
- It puts inordinate faith in Internet search mechanisms (Google-type relevance ranked keyword searching, Amazon-type user preference ["folksonomy"] tracking) for providing access to books, at the expense of entry through vocabulary-controlled headings and authority-controlled names, titles, and series, the loss of which will be profound for systematic (rather than haphazard) searches.
- Its advocates' confidence in the adequate operation of Internet-type search mechanisms that "relevance rank" keywords from digitized full texts is itself predicated on the wishful-thinking assumption that copyright restrictions on book digitization will be greatly relaxed in Google's favor.
- Its proponents are acting as though those copyright restrictions are *already* legally relaxed, by attempting to dismantle the access systems (cataloging and classification) that remain necessary for efficient access to printed and non-digitized books–as well as to any digital counterparts they may have.
- It asserts that substantial onsite collections of printed books are themselves no longer necessary in research libraries, and that books ought to be housed in remote storage warehouses to avoid "costly duplication" among research library collections; consequently neither LC nor any library following our lead needs to concern itself any more with the shelving of large collections of physical books in subject-categorized groupings for onsite subject browsing.
- Its belief that merging book records (either catalog records or digitized full texts) into a single huge database of "everything" searchable simultaneously ("seamlessly") through a single Google-type search box would effectively bury books amid mountains of irrelevancies that would make substantive scholarship much more difficult to accomplish.
- Its understanding of what "the user" wants is severely skewed by misrepresentations of what user studies actually report.
- Its proponents are relying on "feedback" from only a small coterie of like-minded "digital library" advocates, to the neglect of most users of LC cataloging and classification systems.
- It is essentially writing off efficient access to all books that are not in the English language–sources that would be systematically retrieved in the conceptual categories created by cataloging and classification but which will not be retrieved by Google-type keyword searches or Amazon-type reader referrals.
- It is effectively changing the very mission not only of LC itself, but of all research libraries, from that of promoting systematic scholarship, especially (though not exclusively) within book literature, to that of simply providing "something"

(usually not books) delivered quickly and remotely, and discoverable only through haphazard and non-systematic keyword or user-referral mechanisms.

It is perhaps hard for the outside world to believe that such massive changes in the national library system–and such prospective losses of access to *books*–have already progressed as far as they have. Any further debate on these issues, by either AFSCME 2910 members or the outside scholarly community, needs to be informed by a look at the pattern of LC management's own statements and actions.

In LC management's own words

In a May 5th, 2006 letter to the directors of the Association of Research Libraries, Deanna Marcum, the Associate Librarian for Library Services, says the following:

> When I spoke to the Association of Research Libraries shortly after I was appointed Associate Librarian for Library Services, I spoke of the necessity of rethinking our bibliographic infrastructure. I charged all of the directors with redesigning our services and products *with the needs of the end-user–the individual researcher–in mind*. I also charged the Acquisitions and Bibliographic Access Directorate . . . with streamlining their processes *to make information held by the Library of Congress more conveniently and more quickly accessible to our users. . . .*
>
> Despite the call from petitioners for restoration of funds to replace retiring staff, our decision to stop creating series authority records was not a cost consideration. We, like all of you, are looking for ways to invest more of our inelastic funds in *services that add value for researchers, students and the general public.* We are using the opportunity the comes with retirements to *reconsider areas in which to invest our staff resources.* We have millions of items in our ***special collections*** that are not available to the public because they do not have even cursory bibliographic description. *We know that increasingly our users go first to Google and other Internet search engines to find information they are seeking.* We have made it a priority to ***increase access to content rather than to continue bibliographic practices*** *that, though helpful to other libraries, do not add* ***immediate*** *value* ***for the user***.". . .
>
> *Big changes are on the way. The series authority records are but the first step in refocusing the Library of Congress* to take advantage of the promises of technology, *to focus on the actual needs of information seekers*, and to build a 21st century library that is as effective *in the digital age* as the traditional library *has been* in the *world of print*.
>
> [Emphasis added, here as below]

Dr. Marcum's vision for the "digital age," as expressed here, consistently reiterates her comments elsewhere, as reported (for one example) in the distributed minutes of LC's March 24, 2004, Cataloging Management Team (CMT) Meeting:

Deanna reminded the CMT of the "Out-Sell study" commissioned by the Council on Library and Information Resources and issued in October 2002, *Dimensions and Use of the Scholarly Information Environment: Introduction to a Data Set Assembled by the Digital Library Federation and Outsell, Inc.*, (available at http://www.clir.org/pubs/abstract/pub110abst.html). This study asked more than 3,000 faculty and students at nearly 400 academic institutions, ranging from community colleges to the Ivy League, where they went to fill their information needs. ***Deanna said the study showed that faculty and students have enormous respect and trust for libraries, but they don't use them***. *Instead, they turn first to Google.* Most members of the CMT nodded in agreement. Deanna said this finding raised the question of how much to invest in the ways libraries have been providing information. . . . She asked the CMT to think about cataloging ***from the perspective of users, not necessarily of other libraries***. She also predicted that the future of libraries would be closely tied to relationships with Google and other similar organizations.

[Arts and Sciences Division Cataloging Chief] Judy [Mansfield] mentioned research by Prof. Karen Markey Drabenstott (University of Michigan School of Information) showing that ***end users "always" construct keyword searches as one or two words***. . . .

Deanna said that the Library of Congress needed to undertake a massive effort to digitize its collections, notwithstanding obstacles posed by copyright law. The Library would need to negotiate settlements with intellectual property rightsholders . . .

Dennis [McGovern, of the Social Sciences Cataloging Division] pointed out that *Library of Congress Subject Headings (LCSH)* inform the *Dewey Decimal Classification* and the Faceted Application of Subject Terminology (FAST), and therefore have a life of their own even if students don't search directly using LCSH. LCSH is also important to researchers outside their own fields since the headings and reference structure present a view of the organization of knowledge in all fields. Numerous other databases use LCSH. In reply to a question from Dennis, *Deanna said that there is a good deal of research showing that subject analysis is less important now than in the past*. She repeated that *helping people get to information* is the work we have to do. . . .

Deanna said that Google was seeking agreements with libraries to digitize the content they owned. If Google succeeds in digitizing a library of eight to ten million volumes and making the content available on the Web, what happens to every other library? ***She said Library of Congress cataloging would not be needed in these circumstances. The LC role might be to augment the digital core with its special collections.***

. . . Jeff [Heynen, Chief of the History and Literature Cataloging Division] said that "our biggest problem" was that the Library of Congress is committed to applying consensual standards which take years to establish and change.

Deanna closed her remarks by saying she was thrilled that CMT members were thinking as broadly and creatively as they could. [emphasis added]

Of the many points touched on here, let us concern ourselves first with Dr. Marcum's expressed concern for "the user."

A skewed vision of the Library's users

The passages just quoted places emphasis, commendably, on serving "the needs of the end user." What is problematic is Dr. Marcum's vision of who that end user is. In the letter to ARL directors she says "We know that increasingly our users go first to Google and other Internet search engines to find information they are seeking." And in the CMT minutes she claims that the *Dimensions* study "showed that faculty and students have enormous respect and trust for libraries, but they don't use them." This assertion, however, is directly contradicted by the study itself, which found, from over 3,200 interviews, that 55.4% of all respondents (and 59.7% of undergraduates) still regard browsing as "an important way" to get information. Two thirds of faculty and grad students use print resources for research all or most of the time. 52% of undergrads use print for coursework all or most of the time. (For additional evidence of students' continued library use, from multiple other surveys, see "Survey of User Studies" at < http://www.guild2910.org/google.htm >.)

Second, she says–accurately–that students do turn "first" to Google. But her implication is that the typical student today uses *only* Google–i.e., the user does *not* use the library *in addition to* Google. This is clear from the way she has repeatedly represented that user in several of her speeches. Indeed, AFSCME 2910 members need to be clearly aware that Dr. Marcum has portrayed "the user" that research libraries need to be concerned with not as a serious scholar but rather as undergraduate with a certain mind set:

> And let's suppose that I am one of your students with a term paper coming due. And let's suppose further that I've been assigned to write about the foreign policy of President Fillmore.
>
> Now, in the old days, I might have walked to the library . . . . But today, let's say it's cold outside on campus, and I don't want to go to the library. *I want to stay in my cozy dorm room*, where I have a computer . . . .
>
> I ... have the option, *sitting there in my cozy computer-equipped dorm room*, of *ignoring the library*, and going online to, say, the Google commercial search service. *With Google, all I have to do is type my subject*–"President Fillmore Foreign Policy"–into a search box and click on "Go." If I have used "Advanced Search" to get only references containing all four words, up will come what Google calls the first ten references out of "about 14,200." I don't have to go through multiple organizational layers, guessing and clicking, to get *something* on my subject.
>
> [Speech on "The Challenges of Managing Information" to the 2005 Presidents Institute of the Council of Independent Colleges, 6 January 2005 [emphasis added]; same example used in "The Future of Cataloging" January 16, 2004 <

http://www.guild2910.org/marcum.htm >]

In another speech, to the Great Lakes Colleges Association, 13 April 2006, Dr. Marcum uses the example of a now college-aged "Calvin" from the comic strip *Calvin and Hobbes*, who also wishes to avoid the burden of "making the physical trek from his dorm all the way to the campus library." In past times he "still had to go to the library building to get [information]. *He didn't like that*," she says. She continues:

> Eventually, you [college administrators and head librarians] *made it easier* for him by setting up an e-mail address, through which he could consult with a reference librarian *without going to the library*. And you installed a computer program that enabled him to do key-word searching of the online catalog. But before long, Calvin looked at Hobbes, and Hobbes looked at Calvin, and they both nodded–yes, *there's an easier way. Go to the Web, get the Google search box, type in a keyword*, and get back ***not catalog information****, and* ***not advice****, but* ***information itself***–the stuff you really need *to write a term paper*. On the Web he could do that *without having to go to the library or anywhere else outside his dorm room, except that candy-bar machine*. Calvin *really* liked that.
>
> [emphasis added]

The "user" who is portrayed repeatedly in Dr. Marcum's speeches as someone who avoids the library entirely is not the user who is described in a wide variety of user studies (*supra*). Nevertheless, it is just *this* dorm-bound undergraduate whom Marcum consistently portrays as "the user" that research libraries should be striving to accommodate.

It is noteworthy that there are a variety ways, not one, to deal with the student who does not "like" to have to go "all the way" from a "cozy dorm" room into a physical library. One is to *educate* that person to an awareness of the range of resources and search techniques that he or she is missing by avoiding the library and confining search inquiries to the Google alone–as opposed to endorsing the rank superficiality of scholarship on "Millard Fillmore Foreign Policy" that Google, at its best, enables. (See "Research at Risk," *Library Journal* [July, 2005] regarding the results that Google actually turns up on Fillmore.) This is not Marcum's recommendation, however; she sees the Google Book Search project as essentially *eliminating the need for education*. In the same Great Lakes speech she says:

> The project, Google said, would greatly advance its stated mission, which is nothing less than, quote, "to organize the world's information and make it universally accessible and useful." Additionally, Google is working toward providing, quote, *"the perfect search engine,"* which, quote, *"would understand exactly what you mean" (in your requests as a searcher) and "give back exactly what you want."*
>
> That set of intentions is breathtaking.
>
> Google, it would appear, will greatly *accelerate* digitization of huge libraries . . . . Also, Google keeps working on ways to enable users of masses of

> digitized material to find what they want ***with ease and precision***.
>
> Hearing this, Calvin suddenly ***lost interest in literacy training***. A new time-machine-like technology was in sight to simplify his homework again.
>
> [emphasis added]

Marcum's solution to the problem of how to serve "the user," then, essentially recommends that libraries (a) put more digitized full-text content into Google, to increase the quality of what is available, and (b) to rely entirely on Google's search software to provide efficient access to it because (c) Google makes educational "advice" from librarians and "literacy training" on how to do research unnecessary.

Digitizing special collections as a priority higher than cataloging and classifying books

Further, Marcum recommends specifically that the Library of Congress concentrate on digitizing its special collections (mostly copyright-free non-book material) rather than expend labor on its general collections–its books. Note, again, the consistency of her 2004 and 2006 views:

> [2004] If Google succeeds in digitizing a library of eight to ten million volumes and making the content available on the Web, what happens to every other library? ***She said Library of Congress cataloging would not be needed in these circumstances. The LC role might be to augment the digital core with its special collections.*** [emphasis added]

> [2006] We have millions of items in our ***special collections*** that are not available to the public . . . . *We know that increasingly our users go first to Google and other Internet search engines to find information they are seeking.* ***We have made it a priority to increase access to content rather than to continue bibliographic practices*** *that, though helpful to other libraries, do not add* ***immediate*** *value* ***for the user***. [emphasis added]

Service to "the user" in this world does not value libraries' creation of mere catalog records ("bibliographic practices") for *books*. Nor does it value the intellectual labor of creating *authority controlled* names or titles, or *standardized* subject headings–why should it when Google's keyword search software is envisioned as providing access "with ease and precision"? Apparently "the perfect search engine" would "understand exactly what you mean" and "give back exactly what you want" simply on the basis of relevance-ranked keywords and user referral folksonomies ("people who ordered this also ordered these"). Is it any wonder, then, that LC is already moving to curtail authority work, or that it has commissioned an "outside" study–the Calhoun Report–to recommend that LC Subject Headings be entirely abandoned?

Endorsing ignorance

Further, "the user" in this vision does not just use Google *first*–he or she does not use the

"content" of any physical library *at all*. "The user" wants *only* "something" that can be found online, and won't bother to trudge "all the way" to the library to get any other "content." The highest levels of LC administration now believe that such behavior is to be simply accepted and endorsed as inevitable. In the digital-library mind set, it is an unshakeable article of dogma that "advice" from librarians and "literacy training" will always be ignored by the Calvins of the world, so why even bother to provide such education? This is tantamount to an abandonment of the ideal of education for an endorsement of ignorance–we shouldn't even bother to provide instruction or advice on information literacy, because "Calvin" wouldn't *like* that. The wilfully ignorant undergraduate *is* "the user" whose immediate wants we need to dumb down our operations to serve.

In Marcum's vision, providing whatever "content" we can digitize, as long as it is *immediately* and *remotely* available *through Google's search software*, trumps all other considerations of what LC in particular, and research libraries in general, ought to be doing.

Shared remote warehouses in place of substantive onsite book collections

In the new vision being promoted by LC's administration, it is not even necessary for research libraries to maintain substantial onsite book collections! Such collections, of course, can be browsed for subject access because they are shelved according to the subject classification system created by the Library of Congress; and browsing enables researchers to recognize relevant books whose keywords, in a variety of languages, cannot be specified in advance.

LC tried a decade ago to abandon the practice of shelving books in subject categorizations; the idea then was to shelve them by height, and within height groupings by acquisition order. Of course, abandoning the shelving of books by subject would also entail saving a lot of money in classifying them to begin with–the goal then (as now) was to abandon the maintenance of as much of the LC Classification system as possible (particularly Cuttering). That attempt, however, was stymied by a strong report from the Task Group on Shelving Arrangement that rejected the idea:

> Based on surveys, comments, and discussion with Library of Congress reference librarians and subject specialists, both in CRS and public reading rooms, learned opinion favors retaining a classified collection on Capitol Hill not only to enhance in-depth research for the most complex questions, but to enable continuous and efficient evaluation of collections for collections development and management purposes, and also to facilitate risk management assessment. . . .
>
> The consensus of all survey participants, as expressed in their written comments, is that the use of the stacks in browsable classified arrangement, allowing additional subject access to information in related books, is essential to meeting the mission of the Library of Congress, "...to make its resources available and useful to the Congress and the American people..."
>
> [*Working Group on Reference and Research Report to the Task Group on Shelving Arrangement*. October 30, 1997]

Even the Congressional Research Service weighed in with its own report, pointing out the adverse affects that height shelving would have on its ability to serve Congress:

> According to a recent CRS survey, 52% of CRS staff responding to the questionnaire indicated they regularly go to the stacks to browse or select appropriate books for 5% or more of requests. A majority favors continuing to shelve books needed for congressional research in classified order in the stacks. The survey identified several reasons for this opinion. In particular, the current system allows additional subject access that browsing like works in a single location provides, to meet the needs of Congress, and especially to meet RUSH deadlines. Short deadline requests handled by CRS comprised 79.8% of FY96 requests (388,500 requests out of 487,000).
>
> Of particularly serious concern to the Congressional Research Service is that under the proposed fixed location shelving scheme, materials most immediately affected would be current incoming books, those most frequently used for congressional research.
>
> Despite the gain of perhaps 33% shelf space for books that are arranged by size and fixed location (reportedly the experience of the New York Public Library), 87% of CRS staff who responded to the survey still preferred retaining subject arrangement of books on the Hill, even though this might mean more library materials stored off-site to continue classified arrangement of books needed for congressional research in the stacks. [*CRS Response to Shelving Alternatives in Library of Congress Stacks.* June 4, 1997]

Scholars throughout the country continue to regard browsing library book collections arranged in subject classified order as essential to their research; this fact is confirmed repeatedly in a variety of user studies (see, again, the Survey of User studies appended to the Calhoun Report review at < http://www.guild2910.org/ >). One survey elicited the finding that "The importance of serendipitous browsing in library collections *cannot be overemphasized* by the majority of faculty space holders."

That LC is once again primed to stop shelving books by subject was hinted at in a memo circulated to staff by Dr. Marcum on 11/7/03, in which she recommends the reading of an article by Richard Atkinson, "A New World of Scholarly Communication," that appeared in *The Chronicle of Higher Education* (November 7, 2003). Atkinson regards the duplication of printed books in multiple research libraries–the maintenance, in his words, of "many parallel, redundant research collections"–as a practice that is "outmoded and no longer affordable." He specifically labels as "self defeating" the maintenance of "the Association of Research Libraries' membership index--which ranks the association's more than 120 member libraries largely according to the number of volumes they hold on their shelves"; in his view, this traditional measure of library quality "provides no incentive for consortium members to forgo acquiring holdings that are otherwise available to the system as a whole. Even though the membership index rewards inefficiency and waste, we continue to treat it as a meaningful measure. . . .in a networked digital age, *excessive attention to the local management and ownership of physical materials impedes the responsible stewardship of the scholarly and cultural record*" [emphasis added]. His recommendation is that research libraries rely on shared remote warehouses for the storage of their book collections.

Remote storage warehouses, of course, do not shelve books in classified subject groupings; they are housed randomly in tubs (but retrievable through bar code correlations). Advocacy of such a scheme effectively eliminates the very possibility of browsing books on the same subject shelved contiguously–the books would neither be onsite to begin with, nor arranged by subject in the remote warehouses.

The Calhoun Report, commissioned by LC, strongly endorses Atkinson's proposal. Both Calhoun and Atkinson take Association of Research Libraries (ARL) libraries to task for maintaining onsite collections whose books duplicate each other (and which could be locally browsed).

That is cause enough for concern, but other statements made by Dr. Marcum point in the same direction. In her "Challenges of Managing Information" speech" (*ibid.*, 2005), she says:

> How libraries did it last night–indeed, in the last few decades–just isn't going to work anymore. It isn't going to work because what we were doing was loading physical libraries with such an increasing quantity of books and journals that we were pushing out the students and learning activities that libraries should serve. My guess is that the need for more shelves is what your campus librarians have complained about most.
>
> What you should do depends on your situation, but here's what some others are doing about this nearly universal problem. Some are freeing library space by acquiring or building relatively inexpensive repositories off-campus for print resources that are the least used.

No one has any problem with a recommendation to send the "least used" material to offsite storage. But Marcum continues:

> Some are containing costs further by collaborating with nearby colleges to build such off-campus repositories for their combined usage. Some colleges are collaborating further by making their collections accessible to students at all schools in a consortium, whose librarians then work together on *non-duplicative book and journal purchasing*.
>
> [emphasis added]

Warehouses, here, are to be used not merely for the "least used" books, but also to house copies of *current* books to avoid duplication among many local libraries. The concealed proposition, again, is that scholars no longer have any need to browse in substantial onsite book collections.

In a more recent speech, "The Future of Libraries in a Digital Age" (an "Address to Faculty and Librarians of Kenyon College, Gambier, Ohio, 17-18 April 2006"), Dr. Marcum cites with approval a number of ideas advanced by a colleague from California:

> I hold in high esteem a friend of mine in the library community named Jerry D. Campbell, who is chief information officer and dean of University Libraries at the University of Southern California, in Los Angeles. In a recent issue of *EDUCAUSE Review*, a journal

that deals with information technology in higher education, he has published an article entitled, "Changing a Cultural Icon: the Academic Library as a Virtual Destination." There he argues that, quote, "the library is relinquishing its place as the top source of inquiry." [< http://www.educause.edu/apps/er/erm06/erm0610asp >] Let's think about what he has to say.

The Web, he says, has become the world's largest, easiest-to-use store of information. Libraries might once have argued for their survival on grounds that much Web-accessible information is unreliable, and many publishers have resisted digitizing monographs. However, we're discovering that works not digitally available are getting ignored. And so libraries, themselves, are digitizing books. In addition, major efforts to digitize more books–millions, in fact–have recently been announced by search-engine companies such as Google, which plans to digitize large parts if not all of five major research libraries. *Now, almost everybody believes that eventually the Web "will have it all," says Campbell, who adds: "For most people, including academicians, the library–in its most basic function as a source of information–has become overwhelmingly a virtual destination."*

Dr. Cambell does not believe that the Library of Congress, his libraries at USC, or your own Kenyon should *immediately* close their doors. Academic libraries must continue current operations through a *transition* period *until everything is on the Web*. ... what I take Campbell to mean is that *eventually all of us will go to the open Web for scholarly material,* ***not to the campus library or even to its particular Web site****. We will go "Googling" or "Yahooing" via the Internet to find what we need from wherever.*

Do libraries then disappear? Not exactly, Campbell says. Instead, some services, though "derivative" and not collectively "a fundamental purpose for the library," in his view, may "hold the key to its future." What are these services?

[emphasis added]

The "services" in question are, according to Campbell and Marcum, providing "quality learning spaces," "creating metadata," offering "virtual reference services" (although "Campbell thinks the future of library reference service is uncertain"), "teaching information literacy" (although, in line with her comments above, Marcum adds, "Will making everything Web-accessible reduce the need for literacy training ... ? Perhaps ..." It is noteworthy that no reason occurs to Dr. Marcum to justify the continuance of instruction). The other "services" are choosing digital resources and "maintaining resource licenses," "collecting and digitizing archival materials," and "maintaining digital repositories."

It is noteworthy that the maintenance of substantial onsite book collections, arranged in subject-classified arrangements to facilitate browsing and direct access, does not even appear on the radar screen. It isn't even mentioned as *one* of the multiple "services" that future research libraries need to provide. It is simply not a concern of the digital age vision shared by Marcum, Campbell, Calhoun, and Atkinson. As Marcum mentions at the end of this speech, "Perhaps Campbell is right that the book-based library may eventually disappear and the Web may become the library of the future." That, in any event, is the clear goal: to get research libraries out of the business of having to deal with onsite printed book collections.

What then happens to libraries as places if they don't have substantial book collections on the premises? Dr. Marcum sees future library buildings as "providing spaces in which learners can use ... resources, individually and collaboratively, in multiple ways, including ways of their own invention. Computer banks, electronic classrooms, distance learning labs, special program facilities, collaborative study lounges, copy centers, even cafes and canteens and corners for contemplation–these are increasingly valued features of libraries. *Physical* libraries." (This is from her speech "Libraries: Physical Places or Virtual Spaces in the Digital World?" at the National Library of Medicine, November 5-6, 2003). As she said to the college presidents audience of her "Challenges of Managing Information" talk (*supra*), "If you aren't figuring out how to get a little café in your library like the big book stores have, well, just keep right on using it as a place that supports books, not students." Libraries, of course, should not be devoting their resources to maintaining onsite book collections; Marcum, Campbell, and Calhoun believe that, instead–not *in addition to*, but *instead*–they should be digitizing "everything" to provide "direct access," which is envisioned *only* as *remote* access to *digitized* texts, not as onsite-browsing access to printed books shelved in subject categories. The printed books, again, should be shelved in remote warehouses, to prevent "costly duplication."

Recapitulation

In sum, because the Google Book Search project is envisioned as providing access "with ease and precision" to "the digital core" of millions of *books*, it is no longer necessary for the Library of Congress to expend its efforts and resources in this area–LC should, *instead*, concentrate on digitizing its special collections. This radical change in priorities ("Big changes are on the way") has been a consistent feature of Marcum's vision for years. The reason is obvious: since "the user" can rely on Google digitizing and keyword searching to provide adequate access ("with ease and precision") to *books*, what's left for LC to do, then, is simply to digitize the unique, *non-book* collections in its custody that Google will not find at Michigan, Stanford, Harvard, Oxford, and NYPL.

Providing leadership in the cataloging and classification of book collections is no longer regarded as necessary at the top levels of LC administration. Cataloging and classification are themselves regarded as replaceable by keyword searching, relevance ranking, and user-referral mechanisms.

Objections

Are objections to this "vision" merely instances of "resistance to change"? Are they really just attempts to preserve "the status quo" when a wholesale "transition" to digital libraries is offered as the sole alternative to the maintenance of traditional libraries? Do objections from professional librarians reflect only short-sighted and selfish concerns to maintain the "comfort" of old-fashioned but outdated practices? Are so many professionals' objections, both from AFSCME 2910 and the larger research library community, simply crass attempts to keep unnecessary library jobs from being eliminated?

Or is it possible that the digital age proponents and consultants have only a very partial grasp of the truth, and that their attempt to force the whole of our responsibilities onto the Procrustean bed of a Google search box may actually be counter-productive, indeed *destructive*, to scholarship?

Is it possible that they in fact have no actual contacts with "the user" whose needs they profess to understand and represent?

Is it possible that their high administrative positions have isolated and insulated them from the needs of real researchers (not made-up cartoon characters), and from feedback on how badly Google actually works when the goal is scholarship rather than mere quick information seeking?

Is it possible that they do not grasp the importance of maintaining traditional book resources and access mechanisms within real libraries as *alternatives* to Google, when users' first-choice Google searches do not provide what they ultimately need?

Is it possible that they also lack feedback from any substantive first-hand experience in using their own libraries' book collections–the ones they are in charge of–for scholarly research?

Is it possible that they are misrepresenting the user studies to which they appeal as the ground for their beliefs?

Is it possible that they lack first-hand experience in struggling to read the e-book formats they would impose on everyone else?

Is it possible, when they confidently assert that *almost everybody* believes that eventually the "Web will have it all," that their circle of acquaintances is much too limited to be representative of either the library profession or the much larger scholarly world it must serve?

Is it possible that they do not perceive any real differences between quick information seeking, on the one hand, and substantive scholarship, on the other?

Is it possible that they do not understand that "hit and miss" retrieval mechanisms which provide "something" quickly, while appropriate (and necessary) in some situations, are no substitute for alternative mechanisms that provide systematic and comprehensive retrieval when much more than "something" is required?

Is it possible that, on the basis of misread statistics, highly selective feedback from a too-limited "choir," and lack of personal experience in using cataloging and classification systems, that they cannot discern if a real baby is being thrown out with the bath water?

One would hope that none of these possibilities is in fact a reality. If even a few of them were, then ALA President Michael Gorman's call for renewed emphasis on library education

would be fully justified.

The continuing need for Library of Congress subject headings in online catalogs

Since I have quoted so extensively from opposing viewpoints, I will give my own critics the opportunity to challenge some passages from my own book *The Oxford Guide to Library Research* (3rd edition; Oxford U. Press, 2005). It is a text that I wrote in the hope that it might play an important part in just the kind of educational effort Gorman has called for. In the current debates within the library profession, I hope the book as a whole will demonstrate that there is much more substantive value to traditional library practices than has been recognized by many proponents of the digital "vision." The first passage, from Chapter 2, provides an example of the continuing need for Library of Congress Subject Headings, and for online catalogs with mechanisms (unlike Google's) that can *display* such headings:

> One researcher interested in the history of Yugoslavia asked for help at the reference desk because, on his own, he'd simply done a Boolean combination of the keywords "Yugoslavia" and "history," and had been overwhelmed with too many irrelevant records. The solution to this problem was the use of the online catalog's *browse displays*. When doing a subject (not keyword) search under **Yugoslavia**, a browse display of many screens' length was automatically generated; it included headings such as these:
>
> **Yugoslavia–Antiquities**
> **Yugoslavia–Antiquities–Bibliography**
> **Yugoslavia–Antiquities–Maps**
> **Yugoslavia–Armed Forces–History**
> **Yugoslavia–Bibliography**
> **Yugoslavia–Biography**
> **Yugoslavia–Biography–Dictionaries**
> **Yugoslavia–Boundaries**
> **Yugoslavia–Civilization**
> **Yugoslavia–Civilization–Bibliography**
> **Yugoslavia–Commerce–History**
> **Yugoslavia–Commerce–Pakistan**
> **Yugoslavia–Commercial treaties**
> **Yugoslavia–Constitutional history**
> **Yugoslavia–Description and travel**
> **Yugoslavia–Economic conditions**
> **Yugoslavia–Encyclopedias**
> **Yugoslavia–Ethnic relations**
> **Yugoslavia–Foreign economic relations**
> **Yugoslavia–Foreign relations–Great Britain**
> **Yugoslavia–Foreign relations–Soviet Union**
> **Yugoslavia–Foreign relations–United States**
> **Yugoslavia–Geography–Bibliography**

Yugoslavia–History–1992-2003
[NT cross-reference to **Yugoslav War, 1991-1995**]
Yugoslavia–History–Bibliography
Yugoslavia–History–Chronology
Yugoslavia–History–Dictionaries
Yugoslavia–History, Military
Yugoslavia–History–Soviet occupation, 1979-1989
Yugoslavia–Kings and rulers–Biography
Yugoslavia–Maps
Yugoslavia–Pictorial works
Yugoslavia–Politic and government [with period subdivisions]
Yugoslavia–Relations–India–Bibliography
Yugoslavia–Relations–Pakistan–Chronology
Yugoslavia–Road maps
Yugoslavia–Social conditions
Yugoslavia–Social life and customs
Yugoslavia–Statistics
Yugoslavia–Strategic aspects
Yugoslavia–Yearbooks

These are only a sample of the full list. The researcher, in this case, was delighted: he could immediately see that he had many more options for his topic than he had realized. He was particularly excited by the **–Antiquities** subdivisions, which his keyword search under "history" had missed entirely.

Note also that all of this material would have been missed if the searcher had simply typed "Yugoslavia" and "history" into a blank search box in a massive full text database such as the newly-proposed Google Print project [now Google Book Search], which plans to digitize millions of books. The Google software cannot display browse menus of subjects-with-subdivisions and cross-references, allowing researchers to simply recognize options that they cannot specify in advance. Library catalogs provide much more efficient and systematic overviews of the *range* of books relevant to any topic. Searching for all relevant book texts via a simple Internet-type search box would be like trying to get an overview of a whole country while looking at it only through a bombsight. While the Google project may enhance *information seeking*, it will greatly curtail *scholarship*–which requires connections, linkages, and overviews–if it is regarded as a replacement for real libraries and traditional cataloging. (See Chapter 3 for more on Google Print [Book Search].)

I cannot recommend this too strongly: use your library catalog's browse displays. When there are multiple screens of subdivisions, *take the time to look through all of them*. You will *usually* be able to spot important aspects of your topic that you would never have otherwise noticed. This technique is almost tailor-made to solve the frequent problem of getting too much junk via keyword searches.

The larger a library's catalog, the more researchers must rely on *menu listings* that enable them to simply *recognize* relevant options that they could not specify in advance.

There are three such menus you need to look for: the cross-reference lists of NT and RT terms, the alphabetically adjacent narrower terms in the red books, and–especially–the rosters of subdivisions that automatically appear in online-catalog subject searches. A great deal of intellectual time and effort by catalogers goes into the creation of these menus; without them, you simply have to guess which terms to use; and, as in this **Yugoslavia** example, no one will be able to think up beforehand all of the relevant topics that could readily be of use in researching the country's history. Online browse displays of subject subdivisions are the kinds of things real users would kill to have in Internet searches–but Net search engines simply cannot produce them. This radical advantage is available to researchers only in library catalogs.

The continuing need for substantial onsited book collections shelved in subject classfied order

The second passage from *The Oxford Guide to Library Research*, from Chapter 3, is an example the continuing need for onsite book collections arranged in subject classified arrangements for in-depth browsing:

> Let me offer one more example of the need for focused browsing of contiguous full texts. A scholar from France, working on a study of the writer Paul Valery and his times, needed to pin down an important bit of information regarding Valery's connection to the famous Dreyfus case, in which a French military officer of Jewish descent was convicted of treason, and, only years later, acquitted. The woman had hearsay information from Valery's children and daughter-in-law that he had signed his name to a "petition" or "liste" on the issue at the time, but had no specifics of place or date. The large online *Tresor de la langue francaise* database of full text sources did not solve the problem, nor did biographies of Valery, nor did the *Historical Abstracts* or *Francis* databases (the latter having an emphasis on French studies), nor did two massive published bibliographies on Valery (each over 600 pages). I finally had to go back into the bookstacks, where, at the Library of Congress, we have 186 volumes on 6 shelves in the classes DC354-354.9 ("Dreyfus case"). As a shortcut, I was particularly looking for a volume that a browse display in the computer catalog had alerted me to, with the subject heading **Dreyfus, Alfred, 1859-1935–Trials, litigation, etc.–Sources**. (The **–Sources** subdivision indicates a published compilation of primary sources concerning the actual event.) This volume, shelved at DC354.8.Z65 1994, however, did not reprint or identify the particular newspaper petition with Valery's signature. On the shelf above it, however, I noticed another book which, it turned out, did indeed have the necessary information. As an extra serendipitous bonus, the same volume turned out to contain additional information about one of Valery's close friends–information that solved another problem for the researcher, that she hadn't specifically asked about.
>
> Once again, a search of the computer catalog–even by call number–could not identify *which one* of the 186 volumes had the exact information that was needed. If all of these volumes had been scattered by acquisition number, or shelved according to their many different heights, it would not have been possible to find the right book without

making separate requests and waiting for 186 individual deliveries. (The volumes could not even be delivered *en masse* if they were not shelved next to each other in the first place.) In the real world, the prospect of that degree of "hassle" effectively precludes the necessary scholarship from being accomplished. Libraries that do not take into account the Principle of Least Effort in information-seeking behavior are simply not functional, no matter how much money they may save; the purpose of a research library is to facilitate scholarship, not to save funds.

I emphasize the point because many library administrators these days do indeed think that "remote storage" techniques can be used within central library buildings themselves. This is something academics need to watch out for: if your library committees do not take active steps to prevent this erosion of shelving by subject, you will wind up with book collections that cannot be browsed *or* focus-examined down to the level of individual pages or paragraphs. In that case, you will no longer have systematic access to the "depth" parts of the books not contained on their catalog records: not just tables of contents and indexes, but maps, charts, tables, illustrations, diagrams, running heads, highlighted sidebars, binding information, typographical variations for emphasis, bulleted or numbered lists, footnotes, and bibliographies. All such material is readily searchable by focused browsing of subject-classified book collections. Further, the browsing search mechanism that presently allows you to recognize relevant books, or individual pages and paragraphs within them, will be replaced by blank search boxes on computer screens that require you to specify in advance, in detail, every word or phrase that may possibly be related to your topic.

The recently announced Google Print [Book Search] project, aims to digitize the full texts of fifteen million books, from a variety of research libraries, and make them freely available for keyword searching on the Internet. It is not yet clear exactly how the project will segregate works still under copyright protection (life of author plus 70 years) from those in the public domain; but in any event the announcement of the operation has caused some observers to assert that local, onsite book collections will no longer be necessary if every text is keyword searchable on the Internet.

There are, however, real problems with such a naive assumption. Those who hold it are apparently innocent of experience in the ways real scholars must actually work. Let me return to the above examples. . . .

Similarly, with the Paul Valery example, the researcher told me it is highly unlikely that she could have found the necessary information even in a huge full-text database like Google Print [Book Search]. One problem is that Google may not be able to mount copyrighted texts, which would include the French book that provided the information in this case. The other, more serious problem is that the researcher did not know in advance the right keywords to type in. Again, the Valery family members simply said the writer had, at some point, signed a "petition" or a "liste." It turns out that it was actually a subscription fund to provide money for the widow of one of the individuals involved in the scandal; and the text in question, *L'Affaire Dreyfus et la Presse* (Armond Colin, 1960), reports that the names were published in the journal *La Libre Parole* in 1898. The researcher, however, did not know the name of this journal in advance; nor

does the French text use the words "petition" or "liste" to describe the roster–it uses the words "souscription" and "souscripteurs" instead. In other words, the scholar would not have been able to type in the right keywords to find the information even if the copyrighted text were fully searchable online.

It is noteworthy that a search in the Google Web engine on "Valery" and "Dreyfus" already produces over 3,500 hits, and one on "Paul Valery" and "Dreyfus" produces 344. Keyword searching in a Google Print [Book Search] file is likely to produce similar mounds of chaff–especially since the single instance of Valery's name in the entirety of the *L'Affaire Dreyfus et la presse* (272 pages) would not have ranked this text at the top of any retrieval set derived from all of the words on 4.5 billion pages, or from the frequency of "hits" on this one very obscure book. Focused browsing in classified bookstacks, enabling scholars to simply recognize what they cannot specify in advance, remains crucial to advanced scholarship.

It is especially noteworthy that any proposed use of Google Print to replace classified bookstacks would entirely segregate foreign language materials into multiple electronic "zones" that could not be searched simultaneously by the specification of English keywords. With classified bookstacks, on the other hand, books in all languages are grouped together by subject in the same locales; and oftentimes researchers can simply notice relevant foreign books on a topic simply by their illustrations or other visual cues. Google Print enthusiasts would unwittingly re-create in reality the disastrous consequences mythologized in the Tower of Babel story.

Google Print [Book Search] will be a wonderful supplement to classified bookstacks in real research libraries, but a terrible substitute. The overriding reason is that mere relevance-ranking algorithms cannot solve the massive problem of out-of-context keyword retrievals in full text databases. Any large digitization project without the filtering, structuring, segregating, and channeling elements provided by traditional library categorizations would do much more actual harm than good–assuming, as the digital paradigm does, that digitized book collections would replace rather than supplement onsite print collections–because the efficient *categorizing of books by subject* is not a problem that technology can solve through any *ranking* algorithms of keywords. *Information seeking* at the level of finding discrete data would improve, but *scholarship* (being dependent on contexts, connections, and webs of relationships) would be made much more difficult under the "replacement" scenario. Any attempt at a structured overview of resources would be precluded right from the start by inadequate filtering, segregating, linking, and display mechanisms.

Faculty Library Committees need to be aware that most library administrators fail to note the distinction between (a) general browsing to see what's available, versus (b) focused searching for definite, and very specific, information likely to be found within a limited range of full texts, *recognizable* within that range even when its keywords cannot be specified in advance. If this difference is blurred, then all of the (valid) objections against general browsing as the primary way to do systematic research come into play, and none of the arguments for recognition-access to the depth contents of contiguous full texts are noticed. Although all historians, anthropologists, linguists, and others have experienced the advantages of direct access to classified bookshelves, almost no one

> bothers to write down the numbers of contiguous volumes examined, as in the above examples. It just takes too much time, and most academics have never perceived a need to do so. Nor do they articulate clearly the crucial need, in many research situations, for recognition access rather than prior-specification search techniques. (It is the dismissal by library administrators of any concern for recognition mechanisms that is especially galling to working academics, as it is usually done with the patronizing air that advocacy of anything other than computerized keyword retrieval is merely "sentimental" rather than rational.) Until recently, scholars could simply assume that no research library administrator would even think of undermining the practice of shelving books by subject. Unfortunately, that assumption is no longer a safe one–the abandonment is being actively promoted by bean counters who overlook the operation of the overall system in which the beans are situated.

The book provides numerous other concrete examples of the practical utility of LC cataloging and classification in solving the growing problem of too much junk being retrieved through Internet search mechanisms. The point is that Internet search mechanisms do not eliminate all difficulties–they in fact create as many problems as they solve. And those problems require *other* mechanisms for their solutions than Google and Amazon type access.

What is going on at the Library of Congress?

So what is the bottom line here? The question with which this paper began is, What is going on at the Library of Congress?

There is substantive evidence, provided by *patterns* of statements both from LC management and from the sources it relies on, that the Library of Congress is striving mightily to get out of the business of providing systematic access to a large collections of printed books through the provision of LC Subject Headings (in an online catalog that is not merged with Google) and through the provision of subject-categorized shelving of actual volumes arranged according to the LC Classification system. It sees the digitization of book collections being essentially accomplished completely by Google's Book Search project (and some others), in spite of copyright restrictions. It also envisions keyword searching of these digitized book texts, with computer-algorithm "relevance ranking" of the results (and Amazon-type reader-preference tracking) as being adequate to meet the new goal of research libraries, which is simply to provide *something* delivered *quickly* and *remotely* to "the user." Questions regarding the *quality* of resources made available on the Internet are all to be answered simply by digitizing *everything*–in spite of copyright restrictions and in spite of the fact that Internet search mechanisms cannot *find* the quality material, or adequately segregate it from the mountains of chaff, through keyword and user-tracking softwares.

As with research libraries in general, the goal to which LC in particular should now devote itself no longer includes the component of providing complete and systematic subject access to printed books (i.e., to show "what the library has" in its book collections); rather, in furtherance of the goal to provide merely "something" remotely, the new emphasis is on

revealing the content of non-book special collections over the contents of books. (Google will be trusted to provide the latter.) "The user" whom the Library is striving to satisfy is not the scholar who requires complete and systematic access to relevant books, nor does that user any longer require mechanisms enabling him or her to recognize important works (in a wide variety of languages) whose keywords cannot be specified in advance in a blank Google search box. Rather, "the user" whose satisfaction is paramount is the uninstructed undergraduate who has a term paper due quickly, who does not want to leave her cozy dorm room and trudge all the way to the library, who does not have time to read entire books in any event, who wants only English language material, and who cannot be bothered with anything that does not come up immediately in a Google search. Moreover, LC management's goal of focusing on the wants of *this* user includes the assumption that such students will not simply use Google *first*, and then go to the library, but that they will avoid going inside any library *at all*; this "user" wants *only* full texts that can be tapped into remotely. Digitizing those texts for remote access is now seen as *constituting* all that is needed to provide adequate "access" to them; no matter how bad, or how overwhelming, keyword retrievals may be, the proven alternative filtering mechanisms of conceptual cataloging and classification are now regarded as outmoded. And professional librarians who raise objections to the abandonment of cataloging and classification be dismissed as dinosaurs whose "resistance to change" springs not from their concern for the maintenance of high professional standards, but from a selfish fear of losing job security.

And so the Library's management in recent years has consistently disparaged the work of its own professional catalogers, has invited a string of outside speakers to tell them how irrelevant and outmoded their work is "in a digital age," has changed Library priorities to avoid hiring new professional catalogers as old ones retire, has restricted catalogers' ability to extend the subject headings system, and has generally sought to promote the belief that authority control, standardization of headings, and the shelving of actual books in subject categories–that all such practices are now no longer necessary because keyword searching in Google Book Search is, or soon will be, an adequate substitute. For some time now, LC management has been looking for every excuse it can find to dismantle as much as it can of its own cataloging and classification operations.

The imminent loss of the LC systems will have profound implications for research libraries throughout the world, and for all of the scholars in all academic disciplines who depend on them. The dismantling of these access systems, already taking place, will make scholarly research in large *book* collections much more difficult to do at all, and impossible to do in any systematic manner. Retrieval of books through Google and Amazon type search mechanisms, in place of systematic cataloging and classification, will produce results that are superficial, incomplete, haphazard, indiscriminate, biased toward recent works, and largely confined to English language sources. LC's abdication of leadership in cataloging and classification is on the brink of dragging down the capacities of research libraries all over the nation to promote substantive scholarship over "quick information seeking." In the new "vision," the Internet, and not large, onsite book collections, is now regarded as central to substantive research; and research libraries themselves are viewed mainly as feeder-streams to provide "content" to Google, rather than as providing *alternatives* to Google in both content and search method capabilities. This re-

centering of focus constitutes an abandonment of the mechanisms that provide systematic subject access to printed books. The national library of the United States is giving away the the birthright of American scholars in exchange for a mess of Internet pottage.

That is what is going on at the Library of Congress. If scholars in this country, in all subject areas, want to maintain efficient, deep, extensive, and systematic access to book collections in research libraries, they had better speak up now.