What is Distinctive About the Library of Congress in both its Collections and its Means of Access to Them,

and

The Reasons LC Needs to Maintain Classified Shelving of Books Onsite,

and

A Way to Deal Effectively with the Problem of "Books on the Floor"

by Thomas Mann

The judgments made in this paper do not represent official views of the Library of Congress. Further, this paper does not claim to represent the view of everyone in AFSCME 2910. It seeks primarily to assure that considerations of great importance to the mission of the Library of Congress, as viewed by its scholarly and governmental clientele, will be given proper weight in both managerial and Congressional decision-making. I am grateful to the Guild for providing a forum in which these issues can be raised and discussed.

Summary

The Library cannot solve its space problems by adoption of a "digital strategy"without seriously damaging our larger mission to promote scholarship of unusual scope and depth. If the Library's own access to its own general book collection were to be dumbed down to only the levels of subject access provided by Google, Amazon, or Internet search mechanisms, we would effectively be endorsing, and institutionalizing, the level of ignorance exemplified by the Six Blind Men of India. We would no longer be able to satisfy the information needs of Congress or other federal agencies, or those of advanced scholars, in a timely or efficient manner. It is not enough simply to have comprehensive collections; we must equally have *efficient* access to them. While the subject access-mechanisms (LCSH cataloging in our OPAC and LCC shelving of books in the stacks) needed to provide unusual comprehensiveness and unusual depth of scholarship are themselves expensive to maintain, they are the very things that enable us to realize our distinctive mission, to provide our distinctive services, and to discharge our distinctive responsibility to provide maximal access

to the Nation's unique copyright-deposit collection, while simultaneously enabling other research libraries in hundreds of Congressional districts, which copy our systems, to achieve substantial cost-savings themselves.

To deal with the problem of books-on-the-floor, what we need is a systematic, professional weeding of our Capitol Hill collections to determine which volumes will next go to remote storage in Module 5 at Fort Meade; and Recommending Officers must be given the necessary time on an ongoing, weekly basis to undertake this huge task. Because of the delayed funding of that next module, however, we now have just the time we need in order to undertake this systematic examination and weeding of the general collection. This option is much preferable to that of relying on wholesale digitization of our books, especially since the latter process (as in Google or Amazon models) could never extend to our entire collection (for copyright reasons alone), and because it would simultaneously create more problems of access than it would solve, in ways that would directly and severely undercut our distinctive mission.

* * *

In her Memorandum (6/29/09) to selected Division Chiefs and other staffers, on the subject "Meeting on Digital Strategy – June 29, 2009," Associate Librarian for Library Services Deanna Marcum proposes several points for serious discussion, among them (quoting her words):

- "Given the very small number of staff and researchers who utilize the classified collection for browsing purposes, can we justify the costs and limitations that shelving scheme imposes?"
- "Should we be devoting precious space on Capitol Hill to house materials which are (or will soon be) easily accessible online through Google and others?"
- "Are libraries at risk of survival? Are they going as quickly and surely down the same road as newspapers? Is the Library of Congress distinctly different from other libraries?"
- "Is there an economically justifiable argument to be made to Congress that we should continue to acquire and maintain physical books? Amazon, a company less than fifteen years old, became the world's largest retailer of physical books. They now report over 35% of their sales are of digital books, and they expect this to be their core business going forward. . . . Should the Library seriously consider giving priority to digital books?"
- "For any book now published digitally, should we procure (demand through Copyright) the Kindle version, and make it the preferred service copy? There are problems to be solved here, but is this the future?"

It is a truism in the field of polling that the way in which questions are framed to begin with has a noticeable effect on nudging respondents' answers in certain directions; the options contained within the question

are given more prominence than those left unspoken. It is thus noteworthy that several of these questions are immediately "framed" by comparisons of the Library:

(a) to Google, which is described simply, without any question or further analysis, as making materials (primarily books, in this context) "easily accessible";
(b) to newspapers, which are "quickly and surely" going down a road that puts them "at risk of survival." The unspoken implication is that since newspapers, operating on a business model (i.e., to make profits from providing current information), cannot compete with the Internet, then the Library of Congress itself (which is here being compared to newspapers) is in similar danger (because LC is to be regarded through a similar "business model" lens?); and
(c) to Amazon, whose business model (again, oriented toward the goal of making a profit rather than promoting scholarship) in "going forward" with digital book sales (Kindle) is here held up, without any notice or consideration of alternative models, as the implied touchstone of what is "economically justifiable."

The comparison of LC to newspapers

These recommendations all imply, without explicit articulation, a mission for the Library of Congress that does not notice to begin with how distinctive the Library's actual mission is, in sharp contrast to the missions of Google, newspapers, or Amazon. I will consider Google and Amazon below; as for the likening of LC to newspapers, a whole range of concealed propositions in the comparison need to be brought out from under cover:

- Newspapers are dependent on making profit; LC, in contrast, is not a business, but is rather a federal agency supported by national taxes to fulfill responsibilities other than generating cost-recovery income. Every newspaper in the country could fail without affecting the continued existence of LC; we are not in the same boat.
- Newspapers fill a niche, in the overall information universe, of providing current information; LC fills a very different niche, that of promoting scholarship in all subjects, all time periods, and all languages, not only within current sources but especially within its vast retrospective holdings unmatched anywhere else.
- Newspapers do not have long-term preservation responsibilities; LC does. And LC must fulfill these responsibilities as a public service even though it does not make any profit by doing so.
- Newspapers do not have the responsibility to provide maximal access—or any access at all—to the country's unique collection of copyright-deposited books; LC does.
- Newspapers' niche puts them in competition, in the marketplace, with the Internet; LC's mission and unique responsibilities do not put our distinctive functions in such competition—we do

> important things other than just providing current information, and we also provide scholars with important and distinct *alternatives* to the *inadequacies* of the Internet.

One wonders why LC's management is comparing the Library to newspapers in the first place—is it simply (and naïvely) because of an unarticulated assumption that, because both "provide information" of some sort, they must *therefore* both be regarded as in competition with the Internet for "market share" of information-seekers? Such an assumption disregards, right from the start, any understanding of the distinctive niche (in both mission and mechanisms) of the Library of Congress. (Note that the Calhoun Report, which was both commissioned and highly praised by Dr. Marcum, does indeed regard libraries as in competition with the Internet for "market share," and it also rejects any important "niche" function for research libraries [see pages 3-5 and 7 of http://www.guild2910.org/AFSCMECalhounReviewREV.pdf].)

Distinctiveness in preservation responsibilities and in not being a commercial business

In this paper I will leave aside considerations of long-term preservation, a concern for the Library of Congress that does not touch newspaper publishers at all. Nor is it a concern of either Google or Amazon. Both of the latter are businesses—and business can fail whereas the U.S. government cannot, short of defeat in all-out war or failure to accommodate internal revolutionary sects. With legally-enforced taxpayer support from the entire country, and with legally-mandated copyright deposits being centralized in its operations, the Library of Congress is not a business; it does not have to buy that huge copyright-component of its holdings, and it is protected from market forces of "having to show a profit" or to "increase market share" in ways that no other library in the country is. And it will remain protected unless Congress or the Library's own management changes its mission in a way that voluntarily undermines those cost-savings and protections.

It must be noted especially that along with the *unique privilege* of receiving millions of free books comes the corresponding *responsibility* to provide maximal access *to that particular* unrivalled collection. This is a responsibility that, apart from the National Library of Medicine and the National Agricultural Library, no other library has.

The business-model assumption of economic "profit" applies to LC only insofar as its collections (*and* its means of gaining access to those collections, which systems we maintain for copycat use by hundreds of other research libraries) contribute to the well-being ("profit" in more than just an economic sense) of the Congress, the Supreme Court, other federal agencies, and scholars generally. Any "profit" we generate accrues *not to LC itself* but to the better functioning of the Nation that continues to see value in the maintenance of a unique, centralized copyright-deposit collection. Of course LC must manage its operations in a fiscally prudent

manner, and no one disputes that; but prudent management means one thing when the goal is to make a profit or to increase market share, and quite another when the goal is to promote scholarship for the good of the whole country, in ways that no other institution can manage (due to both a lack of comparable collections and a lack of comparable access to them).

Library Services' persistent assumption of a business modelfor the Library

If LC's mission had been, before the present, based on a business model, the Library would have vanished during the Great Depression, if not sooner, along with so many other libraries that did not produce cost-recovery profits for their own operations. It is noteworthy that even the biggest corporations—unlike the federal government—that are unshielded from market forces may not be here even a few decades from now. Who in 1959 would have thought that in 2009 General Motors would seek bankruptcy protection—and then, further, require a huge government bail-out to protect it from market forces? The lesson here is that the long-term existence of either Google or Amazon *is itself* by no means guaranteed—but LC's existence *is* guaranteed. Even that, however, is now questionable when Library Services continues, consistently, to formulate its thinking in terms of an inappropriate business model. Note, again, how closely the assumptions within Dr. Marcum's frame of reference, in her decidedly slanted questions (above), mirror the assumptions of the Calhoun Report, which she has highly praised, in calling for the adoption of a business model for the Library's operations. (It is meaningless to pay lip-service to rejecting such a model while at the same time consistently framing one's thinking exclusively within its terms and assumptions.)

LC's distinctive mission to promote scholarship of unusual comprehensiveness and depth

I take it as given that it is LC's core mission to *promote scholarship*—and particularly scholarship of *unusual comprehensiveness* and of *unusual depth*. This *raison d'etre* tacitly underlies our serving since 1870 as the Nation's centralized depository for copyright acquisitions; it underlies our unique maintenance of multiple overseas offices for purchases from all countries and in all languages; and it similarly underlies our maintenance of multiple blanket order arrangements and exchange programs throughout the developed world—in almost every case at a level unmatched by any other library. It also directly underlies the outreach efforts we make in bringing world-class scholars to offices physically inside the Library, via the Kluge program and other residencies administered by the Office of Scholarly Programs. This understanding of our mission constitutes the very thing that attracts these scholars to the Library of Congress in the first place. And it certainly underlies our commitment to serving the information needs of Congress; as Mr. Jefferson himself said, "There is in fact no subject to which a Member of Congress may not have occasion to

refer." (My twenty-eight years as a reference librarian here provide daily confirmation that, indeed, *no* subject is outside the purview of what our researchers—Congressional or other—ask for. See the examples below.)

This goal of promoting scholarship of *unusual* comprehensiveness and unusual depth forms the obvious beginning of an answer to the question "Is the Library of Congress *distinctly* different from other libraries?" It is truly astonishing, and equally dismaying, to staff throughout the institution that the question of whether LC is "distinctively different" can be seriously asked at all, let alone that its answer must explained to the Library's own administration.

Distinctiveness not simply in the size of collections,
but in the technical mechanisms needed to find what is in them

Dr. Marcum's list of questions (above), with their many concealed propositions and blurred distinctions, shows little awareness of the *distinctive technical requirements* of LC's operations that are necessary to fulfill its own mission. By "technical requirements" I mean specifically the systems of *cataloging* (in our online public-access catalog [OPAC]) and *classification* (in our shelving of books), that provide the *means of access* to our collections that are themselves substantially different from the access mechanisms of relevance-ranked keywords provided by Google and the "more like this" customer linkages of Amazon. The latter are useful supplements to our own systems, but by themselves are wholly inadequate for the promotion of scholarship (rather than quick information-seeking).

The latter much-less-expensive-Internet search mechanisms would suffice to support the *very different* goal of providing researchers merely with "something quickly" on their topics of interest. This goal could indeed be met, in a majority of "information seeking" cases, by Google or Amazon (or Wikipedia). But the acceptance—even advocacy—of the Google/Amazon frame of reference for *all* research effectively creates a Procrustean bed that re-defines *access* to books in a way that positively precludes most of the substantive scholarship that it is the Library's distinctive mission to promote.

And yet Dr. Marcum's appeal to Google and Amazon as touchstones for LC's operations contains the implication that their search mechanisms are indeed adequate for LC's own mission. One is reminded, from the peculiar way in which her questions are articulated here, of her remarks of March 24, 2004, that with the digitization of full-texts by Google, "Library of Congress cataloging would not be needed in these circumstances" ("What Is Going on at the Library of Congress," http://www.guild2910.org/AFSCMEWhatIsGoingOn.pdf p. 5). One is further reminded of Dr. Marcum's favorable views of warehouse storage mechanisms even for onsite library collections (i.e., not just those in remote storage facilities)—mechanisms that do not require subject-classified shelving for books (*ibid*., pp. 9-13; also http://www.guild2910.org/AFSCMEMoreOnWhatIsGoing.

pdf, pp. 11-12).
In her current list of questions, Dr. Marcum explicitly says that "Google and others" (evidently Amazon/Kindle) make such collections "easily accessible"—as though simply digitizing book collections and putting them "online," all by itself, *constitutes* the provision of "easy access" to them. The clear implication is that the search mechanisms of Google and Amazon—relevance-ranking of keywords and "more like this" customer ratings/ranking—are all that is needed for the provision of *scholarly* access.

The alternative search mechanism provided by Library of Congress—Library of Congress Subject Heading (LCSH) cataloging in our OPAC, and classified shelving by Library of Congress Classification (LCC) in our bookstacks—are *not even noticed* in her very skewed portrayal of what makes book collections "easily accessible."

The questions remain for scholars, however (those who actually use LC's collections): Are books in fact "easily accessible" if they are digitized but not *findable* by Internet search mechanisms? Are not **alternatives** to Internet search mechanisms *also* necessary, to *solve* the problems of access that are in fact created and exacerbated by inadequate Internet search mechanisms?

An alternative frame of reference

Before giving concrete examples of the inadequacy of Google and Amazon search mechanisms to provide the kinds of access to books needed to fulfill LC's distinctive mission, let me first try to be as explicit as I can about the frame within which I view the Library's responsibilities to the American people in administering our unique, taxpayer-supported and legally-mandated-deposit book collection—"unique" in its immense and localized aggregation, not in its particular titles—with a view (again) to a mission of promoting scholarship of unusual scope and depth, rather than just providing "something quickly." There are different parts to this alternative frame of reference:

1. Consideration of the scope of the book collection itself
2. Consideration of the scope of the *means of access* to the collection
   a. Access mechanisms showing, systematically rather than haphazardly, the *range* of books relevant to any topic
   b. Access mechanisms providing *depth* of access into the contents of the individual relevant books.

Of course there are more resources in LC's collection than just book-formats: maps, manuscripts, motion pictures, sound recordings, prints and photographs, microforms, sheet music, subscription databases, et al. What is at issue in the present discussion, however (in the comparison to Google/

Amazon), is our general book collection and the means of access that we provide to scholars for discovering what is in it. It is in this area, especially, that LC is most distinctive from other libraries.

The first concern has to do with recognizing the importance of, and maintaining, the immense and unrivalled *scope* of LC's book collections, covering all subjects, all time periods, and all languages (with major qualifications only for Medicine and Agriculture). Although most of the same books exist individually in other libraries, and can be located via WorldCat and other mechanisms, it is their *physical contiguity* at LC, especially on Capitol Hill, that gives them an aggregate, synergistic power unrivalled anywhere else. When the same books can be perceived *in relationships to each other*— relationships of conceptual categorization defined by both LCSH cataloging and LCC shelving—and can also *be retrieved quickly*, without the hassles and delays of inter-library lending or remote storage transportation, then their functional utility is immensely increased. The Principle of Least Effort in information-seeking has been verified in dozens of empirical studies: what is easily and quickly findable gets used more readily than that which is not—even by senior scholars. "Findable" means both noticeable and retrievable. It is a factually false claim that Google Books provides comparable access— Google cannot display the texts of most books published in the 20th century (the vast bulk of LC's holdings) due to copyright restrictions; nor can it show the conceptual relationships created by LCSH or LCC (examples below) due to the serious inadequacies of its keyword-only search mechanism. (Relevance ranking of specified keywords is not at all the same thing as conceptual categorization, which groups similar works together no matter what keywords, or even what language, their authors may use.) Nor can Amazon, with its focus on current in-print, English-language books provide anything resembling comparable access to LC's general book collection.

This first concern, however, has already been adequately addressed by the Librarian of Congress in his letter of February 14, 2007, to Inspector General Karl W. Schornagle, responding to the IG's recommendation that, in Dr. Billington's words, "the Library should not even attempt to build comprehensive collections." The Librarian points out how LC's collection is distinctively different from those of university libraries:

> I am compelled to offer a more nuanced case for the role and responsibilities of the Library of Congress, an institution that is unique in the world. The report [from the Inspector General, *Library Services: The Library's Collections Acquisitions Strategy*, Dec., 2006] assumes that the Library of Congress is just one of many libraries collecting materials that can be used by everyone. This is simply not the case.
>
> [Question: Why is Library Services still asking, in 2009, if it *is* the case?]
>
> The Library of Congress is the authoritative, responsible knowledge institution for the Congress and the United States

government. University libraries, by and large, collect materials to support their institutions' curricula: consequently, there are great similarities among their collections. The Library of Congress's comprehensive collections policies allow us to provide very-difficult-to-obtain, obscure materials that provide information to our lawmakers that is available nowhere else. Quite often, it is precisely the marginal material of today that proves to be most useful at a later date. . . .

Since it is the responsibility of the Library of Congress to provide authoritative answers to Members of Congress and their staffs, it is essential that the Library have the most comprehensive collection possible of peer-reviewed, validated information that is found in books and scholarly journals. . . . The enormous legacy print collections that will never be digitized must remain available for answering questions and providing support for in-depth research...

[Question: Why does Library Services, in contrast, regard these as "materials which are (or will soon be) easily accessible online through Google and others?"]

Recommendations to develop collaborative arrangements for acquisition with other institutions are simply not practical. Our responsibility is to provide answers to Congress and staff quickly and authoritatively. . . .

[Question: Is the digitization of special collections now to be regarded as a higher priority than maintaining a full range of access mechanisms for answering such inquiries?]

Even in those cases when other libraries hold the material we need, the interlibrary loan process used among libraries across the country is labor intensive and time consuming. We cannot wait two to three weeks while requests are issued and materials are sent to us. . . .

Acquisitions is an inescapable top priority if we are to continue fulfilling our basic mission of serving Congress efficiently and "preserving a universal collection for future generations."Determining what must be acquired must be governed basically by experienced curators weighing quality judgments not by regulations about quantities—let alone by a study that suggests a presumption of overall reductions at a time when knowledge is more important than ever for our economy and security. . . . We cannot state emphatically or often enough that no other organization has the range of responsibility that the Library of Congress has.

[Question: Why is Library Services still asking, in 2009, "Is the Library of Congress distinctly different from other libraries?"]

Significant damage to the Library's core mission could resultfrom studies that seem to advocate cutting back on the national collection.

In terms, then, of the question of LC's *scope of acquisitions*, in collecting a comprehensive range of books (and other materials), the Librarian has provided a definitive answer, that LC is indeed distinctively different from all other libraries, and must remain so in order to discharge its unique responsibilities. One must wonder, however, why this answer is not reflected in the "frame" of questions asked by Dr. Marcum.

"Scope" entails more than just size of collections: a full scope of different access-mechanisms *is* also *required*

In considering scope of access to LC's collections it is useful to think in terms of the fable of the Six Blind Men of India who were asked to describe an elephant. One grasped the elephant's leg and asserted, "The elephant is like a tree"; one touched the animal's side and said, "The elephant is like a wall"; one grabbed the tail and said, "The elephant is like a rope"; and so on with the tusk ("like a spear"), the trunk ("like a hose"), and the ear ("like a fan"). Each of the sages latched on to some information quickly, and each concluded, on the basis of what he found "easily accessible," that that was all there is. None perceived the number or the variety of the elephant's other parts, nor did any perceive the importance of the relationships and interconnections of the parts. The problem of understanding was not solved simply by the entirety (the "full scope") of the elephant being actually present, because a very important additional factor was lacking: none of the Blind Men had *the mechanism needed to perceive* that scope.

Researchers who have only Google and Amazon search methods available to them are left permanently in the situation of the Six Blind Men: they have no means to see, systematically (rather than haphazardly), the full range of resources that lie well beyond the purview of relevance-ranked keyword retrievals and customer recommendations. While the latter are indeed desirable (and sometimes indispensable), they are by no means sufficient by themselves to promote *scholarship* of *unusual scope*.

I spoke above of different component parts in the "alternative frame" of reference that is needed in assessing LC's distinctive mission and operations. Providing a full *scope of means of access to our collections* is just as important in this frame as providing a *full scope of collections themselves*.

Included within the full scope of other search mechanisms (beyond those of Google and Amazon) that are necessary—especially to provide *overview perspectives* rather than just "something"—are those providing access:

- by LCSH subject cataloging in our OPAC,
- by controlled, uniform name and title headings in our OPAC,
- by uncontrolled keywords within hundreds of commercial databases,
- by published bibliographies and professionally-created finding aids,

- by specific types of literature such as encyclopedias and literature-review articles that can be efficiently identified and cleanly segregated from all other types,
- by citation searches (showing where any source has been subsequently referred to),
- by related-record searches (retrieving articles having footnotes in common, regardless of keyword disparities),
- by people-contacts and subject experts (including reference librarians and curatorial staff), and
- by special reference collections carefully selected by professionals and segregated from the general collections.

None of these avenues of entry into our collections is duplicated or matched by Google or Amazon. (I have omitted, here, mentioning the additional necessary mechanism of access "by subject-classified shelving" because browsing in bookstacks is usually a very poor way to gain an overview of the literature of a large topic. Focused browsing within limited ranges of shelves, however, is nevertheless important, but for a very different purpose: its strength lies in providing *depth* of access to *particular* volumes within a *limited* topical grouping [see below] rather than in providing overviews of the *range* of *all* book-groupings relevant to the topic—the function handled by LCSH cataloging.)

Only the first of these alternative avenues of access—i.e., subject cataloging in our OPAC—need concern us at the moment: In providing professional *subject cataloging*, via the Library of Congress Subject Headings (LCSH) system, LC creates and maintains a crucial means of access to its book collection that is itself *distinctively different* from the access provided by Google or Amazon—and, in what it accomplishes in systematic overview-provision, it is *far superior*.

Tens of thousands of examples are possible here; for the present we will have to let one suffice: the *subject cataloging access* to books on "Afghanistan" that is infinitely more efficient in providing an overview of the whole scope of our relevant collections than could be provided by either Google or Amazon search mechanisms. And I need not emphasize how important it is to our national interest that Congress, and scholars generally, have access to as much knowledge on this subject as we can possibly provide. A researcher using LC's online catalog can easily call up a browse-display such as the following:

Afghanistan
Afghanistan—Antiquities
Afghanistan—Bibliography
Afghanistan—Biography
Afghanistan—Biography—Dictionaries
Afghanistan—Boundaries
Afghanistan—Civilization
Afghanistan—Civilization—Bibliography

Afghanistan—Commerce
Afghanistan—Commerce—History
Afghanistan—Constitutional history
Afghanistan—Defenses—History—20th Century—Sources
Afghanistan—Description and travel
Afghanistan—Economic conditions
Afghanistan—Economic Policy
Afghanistan—Emigration and immigration
Afghanistan—Encyclopedias
Afghanistan—Environmental conditions
Afghanistan—Ethnic relations
Afghanistan—Fiction
Afghanistan—Foreign economic relations
Afghanistan—Foreign public opinion
Afghanistan—Foreign relations [numerous subdivisions]
Afghanistan—Foreign relations—Great Britain
Afghanistan—Foreign relations—Sources
Afghanistan—Foreign relations—United States—Sources
Afghanistan—Gazetteers
Afghanistan—Genealogy
Afghanistan—Geography—Bibliography
Afghanistan—Guidebooks
Afghanistan—Historical geography
Afghanistan—Historiography
Afghanistan—History
Afghanistan—History—Bibliography
Afghanistan—History—Chronology
Afghanistan—History—Dictionaries
Afghanistan—History—20th century—Sources
Afghanistan—History—Soviet occupation, 1979-1989
Afghanistan—History—Kings and rulers—Biography
Afghanistan—Imprints
Afghanistan—In art—Catalogs
Afghanistan—Juvenile literature
Afghanistan—Kings and rulers
Afghanistan—Languages
Afghanistan—Maps
Afghanistan—Officials and employees
Afghanistan—Periodicals
Afghanistan—Pictorial works
Afghanistan—Poetry
Afghanistan—Politics and government
Afghanistan—Populations
Afghanistan—Relations—India
Afghanistan—Rural conditions
Afghanistan—Social conditions
Afghanistan—Social life and customs
Afghanistan—Social policy
Afghanistan—Statistics

Afghanistan—Strategic aspects
Afghanistan—Study and teaching
Afghanistan—Yearbooks

Such "road map" arrays in our OPAC enable scholars who are entering a new subject area to *recognize what they cannot specify in advance*. They enable scholars to see "the shape of the elephant" of the book literature on their topic *early* in their research.

Neither Google nor Amazon makes such *systematic overviews* of subjects accessible at all, let alone "easily accessible."

Subject cataloging in our OPAC accomplishes the goal of extending the scope of scholars' inquiries by showing them more of the full range of what is available than they know how to ask for before they are exposed to it. LCSH cataloging enables them both to recognize a much broader range of topical options within their subjects that would not occur to them otherwise; and it *also* enables them to pick those aspects of interest in a way that separates them from other aspects that would only be in the way, as clutter, without this roster of conceptual distinctions to choose from.

The serious problem of overwhelming clutter—retrieval of too much 'junk' having the right keywords in the wrong conceptual contexts—is one that is *created* by Internet book-search mechanisms but *solved* by LC cataloging. Google and Amazon searchers cannot see "the shape of the elephant" of their topics; LC catalog users can, much more efficiently.

Our LCSH cataloging is a truly amazing service that we provide routinely—but it is one that would be lost entirely if we dumb down our access-provision to only what a Google or Amazon "model" allows. As a reference librarian who is called upon every day to help scholars get efficiently into subject areas that are new to them, I am repeatedly astounded by what such overview-provision mechanisms accomplish that neither Google nor Amazon (nor OCLC's WorldCat) can even begin to approximate. (Note, however, that none of these services have the responsibility to Congress which LC has.)

The actual browse-display in LC's online catalog shows over 480 subdivisions of Afghanistan, so the above is only a sampling. Further, I have not listed any of the cross-references to still other relevant headings that are provided, nor any examples of the range of keyword-variant titles—in scores of languages—that may be included within the coverage of any *one* of these headings. (For an example of the scope of completely unpredictable title keywords retrieved, all in one place, by the single LCSH precoordinated heading-string **Afghanistan—History**, see the brief discussion at http://www.guild2910.org/searching.htm.)

Further still, LC now has a new way (the Subject Keyword option in our OPAC's Basic Search menu) to bring up, systematically, a browse-menu of

all other headings in which **Afghanistan** is itself a subdivision of another topic, for example:

Abandoned children—Afghanistan
Actors—Afghanistan
Administrative law—Afghanistan
* * *
Buddhist antiquities—Afghanistan
Cabinet officers—Afghanistan—Biography
* * *
Muslim women—Afghanistan
Muslim women—Education—Afghanistan—Bibliography
Rural women—Afghanistan—Social conditions
Sex discrimination against women—Afghanistan
Single women—Legal status, laws, etc.--Afghanistan
Women—Afghanistan—Interviews
Women—Afghanistan—Social conditions
Women and war—Afghanistan
Women in development—Afghanistan—Case studies
Women in Islam—Afghanistan
Women in politics—Afghanistan
Women journalists—Afghanistan
Women—Legal status, laws, etc.—Afghanistan
Women refugees—Afghanistan—Social conditions
Women—Services for—Afghanistan—Directories
Young women—Afghanistan—Biography

Neither Google Books nor Amazon—*nor WorldCat*—has a search mechanism that is nearly as efficient in laying out the scope of options available to researchers, enabling them to recognize aspects of their topic, right at the beginning of their research, in ways that they could never specify in advance when confronted by only a blank Internet-type search box. While the latter does indeed offer an apparently "simple" means of searching, its limitations condemn scholars to remain perpetually in the situation of the Six Blind Men of India. The literature of the world—which LC alone among libraries attempts to collect in a systematic fashion—is itself very complicated, very disparate, and very tangled and interconnected in ways that cannot be revealed by Web search mechanisms. The amazing scope and diversity of the world's literature itself is a rock-bottom fact that will not vanish simply because we naïvely want to believe it can be made "easily accessible" through a single search box, backed by "under the hood" computer algorithms working on keywords and democratic tags.

In short, the LCSH system of cataloging in our OPAC provides a means of *seeing* the full scope of LC's book holdings, in a systematic manner, in a way that is *distinctively different* from what can be shown by Google, Amazon, or any Web 2.0 mechanisms. We provide menus of options and road-maps of the "subject territory" of any topic in a way that enables researchers to see, systematically (rather than haphazardly), more than

they know how to ask for, and in ways that clearly segregate relevant from irrelevant sources on whatever the topic may be. Our search mechanism is *distinctive* in its capacity to show "the shape of the elephant" in a way that the several alternative mechanisms cannot—even though the others are *also* necessary for their own distinctive capabilities.

Depth of access *entails problems (and solutions) different from* scope of access

If our mission is indeed to promote scholarship of unusual comprehensiveness (scope) and of unusual depth, then we must also look at the *distinctive means of access* that allows LC to provide *depth* of penetration into its collections. And this consideration is an additional structural element of the alternative frame of reference (beyond that of a "digital strategy" modeled on Google and Amazon) needed to understand LC's overall distinctiveness.

In his letter to the Inspector General, Dr. Billington addressed part of this issue, without fully articulating all of the implications, by pointing out that "The Library of Congress's comprehensive collections policies allow us to provide very-difficult-to-obtain, obscure materials that provide information to our lawmakers that is available nowhere else. Quite often, it is precisely the marginal material of today that proves to be most useful at a later date." I fully agree; but I must point out again that we need not just the range of obscure materials that prove to be so useful, but also *the means of access to* their *full-texts* that cannot be provided by any digitization project.

The Library of Congress Classification (LCC) scheme *applied to the shelving* of our book collections is the alternative means of access that gives us distinctively deep penetration to full-texts within the Nation's unique copyright-deposit collection in ways that cannot be duplicated by either Google or Amazon.

I realize that such a statement is counter-intuitive to people who do not *use* the collections—i.e., their questions are: How can classified shelving of books provide any important access to their contents that is *not also* provided by full-text digitization? If the books' contents are fully digitized, doesn't that in itself make them "fully accessible"?— i.e., how can it be maintained that something *else* is needed?

The first problem with such questions is their underlying—and false—assumption that the Copyright law will be successfully circumvented to allow the full text digitization of the vast bulk of 20th century books (which also comprise the bulk of LC's own general collections). The full-text display of most of these books will certainly be precluded; but even their full-text indexing (by Google) continues to be very much a matter of questionable legality.

The second problem is that of confusing scope (or size) of collections with the scope of the *means of access* to the collections. These are not the same thing. Let me therefore *demonstrate* the substantive differences between access provided by text-digitization vs. that provided by classified shelving of full-texts, with concrete examples that zero in on LC's distinctive responsibility to serve the national government:

A Question from a Supreme Court Justice

In March of last year (2008) I received a rush request from a librarian at the Supreme Court library, that one of the Justices needed to confirm the statement that "the United States occupation zone in Germany after World War II encompassed 5,700 square miles and a population of over 18 million people." I haven't the faintest idea why any judge would want such information; but if the question comes from the Supreme Court then it cannot be cavalierly dismissed as unimportant. It must be answered, and answered quickly. Period.

I first tried the subscription databases *America: History and Life and Historical Abstracts* in hopes that someone had written an article on the occupation zone; but the results were much more diffuse than I wanted. So I tried our online book catalog. Just as an initial stab I did a keyword search of "occupation" and "zone" and "Germany," with a limit on the search, that I wanted only records published between 1945 and 1947. Within the 145 records that came up, I spotted one pretty quickly that had a formally established corporate name on it: **Germany (Territory under Allied occupation, 1945-1955: U. S. Zone). Office of Military Government**. I offered a silent "thank you" to our catalogers for their authority work in grouping relevant works together under this standardized name form, for when I searched on this standardized term I found a very focused pool of records. There were 18 hits; one of them had the word "population" in its title.

Since this was a rush request, I immediately went back to the bookstacks to look at this one pamphlet. This initial item did indeed have population figures for the American zone in 1947, but no square mileage figure. Right next to it, however, was another report that had a 1946 population figure—17,174,367—close to the "18 million" in the original inquiry, but obviously being a very different keyword character string. And it also had an area figure for the American zone—but in square kilometers, not in square miles. That was no problem, as the figure could easily be converted. The significant point, here, is that the chart providing the area figure did not say "square kilometers" written out—it said simply "sq. km."

The equally significant point is that this particular pamphlet is indeed digitized in Google Books; but, even so, I could not *find* it there. If you search Google Books for the three words with which I started my own search in LC's online catalog, namely "occupation" and "zone" and "Germany," and limit to publications between 1945 and 1947, you get 653

hits; the exact pamphlet I found in the stacks showed up as the *307th item in the Google "relevance ranked" display*. (I could find it in the list only because, at that point, I already knew the precise source I was looking for. The number and the ordinal position are those as of March, 2008; Google displays, however, change not only from one day to the next but also, frequently, from one minute to the next.)

I cannot emphasize the following point enough, because it is so strongly counter-intuitive to administrators who do not actually have to do such searches, or find such information, themselves; and yet it is nonetheless true: You cannot "progressively refine" such a set of 653 items down to the right pamphlet by simply typing in extra keywords. Why not? Because the terms "18 million" or "square miles"—the words contained in the judge's question—are not the words that actually appear in the table between paragraphs 2 and 3, on page 6; nor do they appear anywhere else. In order to do "progressive refinement" you have to know in advance which exact words will produce the refinement you seek; and *it is precisely that knowledge which we lack when we are moving around in unfamiliar subject areas*. In fact, I could not get the relevant table to show up at all, even in snippet form, even after I had discovered the right keywords (via stacks-browsing), in spite of the fact that I could view other snippets from the same pamphlet. The Google software is such that it won't show you *every* snippet containing the words you type in; and the company is playing it safe, legally, in not providing full-text views of post-1923 works, such as this occupation zone pamphlet.

The point is this: even if the Google keyword search software would display every instance of every word asked for, I still would never have known in advance the precise keywords (like "sq. km.") that I needed to type in—I would havetyped the phrase "square miles" written out, because it would not have occurred to me to think in terms of kilometers, let alone in abbreviations of terms.

The fact that the pamphlet is digitized does not mean that it is therefore "easily accessible"—quite the contrary: it is not *findable* because Google's keyword *search mechanism* does not provide adequate *access* to it.

By using the classified bookstacks, however, I employed a different and, in this case, a much more efficient search technique for approaching the same (or at least a comparable) body of literature—a technique that enabled me *to recognize what I could not specify in advance*, in a blank search box. I could find the source I needed because it was physically right next to the one that I started off looking for—and the one I was looking for was itself one of only 18 records, not one of 653. And I could skim both that initial full text—down to the level of its individual tables—and the one right next to it *quickly*, precisely *because* they were physically shelved right next to each other, within a limited class. I did not have to browse all 535 miles of our Capitol Hill bookshelves—I had only a very limited range of materials (less than one shelf) to inspect. In contrast, any search of Google

Books does indeed simultaneously search all of its tens of millions of texts. The relevant material in the classified shelving-array was thus effectively *segregated from, rather than merged into*, hundreds or thousands of irrelevant sources having the same keywords in undesired contexts.

Classified bookstacks allow researchers to find *through recognition within full texts* what they don't know how to ask for: we can look not just at tables of contents (which can be included in OPAC records), but also maps, charts, tables, illustrations, diagrams, running heads, highlighted sidebars, binding condition, typographical or color variations for emphasis, bulleted or numbered lists, prefaces, footnotes, bibliographies and back-of-the-book indexes—most of which cannot be included in OPAC records at all, or added retrospectively to millions of existing records—and all within *limited* physical areas. Such quick and focused browsing provides *deep access via recognition* in ways that digitized libraries of the very same texts do not.

Moreover, authority control in *library catalog records* enables researchers to zero in *on which very limited areas of the stacks* they should go to, in the first place: the established corporate name form **Germany (Territory under Allied occupation, 1945-1955: U. S. Zone). Office of Military Government** enabled me to segregate relevant records from this body from all of the others having the same keywords ("occupation" and "zone" and "Germany") in irrelevant contexts.

Let me step back a second and put this in perspective: the 535 miles of bookshelves that LC has on Capitol Hill is greater than the distance between the District of Columbia and Charleston, South Carolina. And we have about 18 million books in over 470 languages on those shelves. (This does not count our offsite Fort Meade holdings.) Nevertheless, because of the subject-classified shelving arrangement, I was phoning the population and square mileage figures, and then FAXing them, to the Supreme Court librarian within about a half hour of my searching the catalog—and part of that time was taken up just in running back and forth through the tunnel between two of our buildings, and waiting for the stacks elevators. That anybody could find this requested information at all within such a huge and diverse collection, without knowing in advance which words to ask for—and find it quickly—is pretty amazing, even to me (and I'm the one who did it). But it was indeed possible *because of the prior work done by our professional catalogers* in creating the classified shelving arrangement. Search results coming from that avenue of access, enabling a researcher to simply recognize what he cannot specify in advance, within full texts (most of them both copyrighted and out-of-print) in a definably-limited subject-class area, cannot be matched by either Google or Amazon retrievals. The Library of Congress thus provides a *distinctive alternative* to Internet search mechanisms that in *some* cases is far *superior*.

Note the combination of factors needed to answer this Supreme Court question:

1) LC had to have acquired the obscure 1946 pamphlet to begin with—i.e, the scope of our collections had to be unusually comprehensive right from the start. (WorldCat lists only a dozen other libraries that have the same work.)

2) LC had to provide a *means of finding* it:

- that provided *fast* access (onsite rather than in a remote warehouse);
- that provided *deep* access, at page-and paragraph-levels, and even to the level of individual words (which level is not provided by OPAC records);
- that provided *systematic* access, via authority work in the OPAC enabling me to target which specific stacks area to focus on;
- that allowed *recognition* of the pamphlet's terminology ("17,174,367" and "sq.km.") that could not be specified in advance;
- that enabled the correct item to be found *without excessive clutter*—i.e., within a manageably small conceptual grouping defined by the classified shelving, thereby preventing the relevant pamphlet from being buried within mountains of irrelevancies having similar words ("occupation", "zone") in the wrong conceptual contexts.

Here is a major additional point: *at LC alone* we have similar on-shelf, full-text-level, subject-access-by-proximity to *every* item in our *unique aggregation* of 18 million books onsite—no matter how old or in what languages the books may be—rather than only to much smaller collections (as in university libraries), or only to copyright-free full texts (as in Google Books), or only/primarily to English works currently in print (as in Amazon). No other library can match the range of subject-arrayed full texts that we have here. We have both a very distinctive collection to begin with *and very distinctive access to* its contents that would be lost without classified shelving.

This is the point that was not fully articulated in Dr. Billington's reponse to the Inspector General. It is the synergy of a uniquely comprehensive collection *coupled with the way it is shelved* that provides LC's unmatched ability to respond immediately to the information needs of the U.S.government. If all of the obscure, seldom-used items in our collection—such as this occupation zone pamphlet—were sent to remote storage, then we *could not find them ourselves once they are removed from classified shelving*. (Books in remote storage are not shelved by subject-classification numbers; they are housed randomly in tubs, but are still retrievable by bar codeslinking the volumes to the tubs in which they reside.)

In our present configuration, *every item in the collection effectively serves as an entry point to the full-texts of every other item shelved nearby*—i.e.,

the discovery of only one potentially relevant pamphlet *also* gives *us* the capacity to examine, quickly and at full-text level, all of the items it is shelved next to. If we take any of those initial items off the shelf and send them to storage, then we effectively have much less full-text-level access to all of the other items, nearby, that it would have brought immediately to our attention. (Here is another truth that is counter-intuitive to those who do not themselves use the collection: searches of catalog records by class number in the OPAC are simply no substitute in such cases—such OPAC searches do not provide comparably-**quick** access, or full-text-**level** of access, or simple *recognition* access of unpredictable *text* words [e.g., "sq. km."]—especially if the catalog records refer to items stored offsite [thereby further tilting the slope against "least effort" retrieval]. Class-number searches of records in the OPAC, as substitutes for searching the actual books in classified order, may sound appealing to cost-cutters who themselves don't have to do such searches at all, let alone under deadlines; but they just do not *work* for those of us who do.)

And yet Dr. Marcum, referring to subject-classified shelving, speaks only of the "limitations that shelving scheme imposes"—with no mention at all of the *immense expansion of subject access* that it *creates* to *precisely* the "very-difficult-to-obtain, obscure materials that provide information to our lawmakers that is available nowhere else" (in Dr. Billington's words).

No catalog record contains the depth of information contained in the actual book it points to; and at LC, our pre-1968 catalog records (PreMARC) are skeletal to begin with—they lack even subtitles, contents notes, and series statements. We especially need classified shelving for maximal access to the bulk of our "very-difficult-to-obtain and obscure" older books—such as the 1946 pamphlet on the occupation zone. For just such cases of providing quick access of unusual depth for important governmental purposes, policy "must be governed basically by experienced curators weighing quality judgments not by regulations about quantities"—e.g., "the very small number of staff and researchers who utilize the classified collection for browsing purposes."

A Question from a Member of Congress

Last September ('08) I helped a Member of Congress who was actually in Main Reading Room, wanting "a list of books or other sources from the Revolutionary period on the subject of economics that would have been known to the Founders." To make a long story short, among the many sources I turned up, one in particular was appreciatively described by the Member as "the mother lode" when I brought it out of the stacks. It is a 1967 article entitled "A Note on Jefferson's Knowledge of Economics" from *Virginia Magazine of History and Biography*. I discovered its existence—while the Member was still in the room—not through any Google or Amazon search (how many individuals' names or different verbal designations are needed to define "the Founders"?), nor through any search of our subscription databases—which I also tried—but through

a published bibliography shelved in the Main Reading Room's reference collection. And I did not know in advance the title of this bibliography or name of its compiler. All I knew was that there is a *classification area* of the bookshelves where bibliographies on Adams, Jefferson, and the other Founders are *likely* to be found, in the Z8000s (i.e., book-length bibliographies on individuals, alphabetically sub-arranged by the person's name who is the subject of the bibliography).

Within that very limited area of the collection I could quickly recognize the relevant works that I could not specify in advance (one of them refers to itself as a "reference guide" rather than a "bibliography"); this area of the shelves immediately showed me three Jefferson bibliographies, all annotated, right next to each other.

If these three sources had been shelved not by subject but by dates of accession (1982, 1983, 1993) they would have been widely separated by thousands of intervening volumes in the reference collection, or by millions in the regular stacks. Moreover, one of the three bibliographies does not list this article at all; but I would not have known in advance which of them does have it, if I had had to request the volumes from a storage warehouse—a very time-consuming process that would not have served the Member's interest.

The fact that the relevant bibliographies were shelved together, by subject, enabled me to skim through them very quickly, and to find not just the general topic of the above article (listed simply under "economics" in two of the books' indexes), but a precise annotation within the full-text of the annotated bibliography that reads "List of books in TJ's library which were also cited in Smith's *Wealth of Nations*; TJ had 94 of the 149 authors cited." This indeed is "the mother lode."

Making our collections "maximally accessible"

In his transmittal letter covering the *Library's Strategic Plan: Fiscal Years 2008-2013*, Dr. Billington refers to LC's mission to make its collection "maximally accessible" (http://www.loc.gov/about/strategicplan/2008-2013/StrategicPlan07-Contents_1.pdf). Are our collections in fact "maximally accessible" if the very access to them in question is constrained through the narrow channels of Google and Amazon keyword search softwares?—and if it is further constrained through their display-limitations that cannot show the full-texts of most copyrighted books? This is not a rhetorical question; it is a serious matter and it has a definite, objective answer: the collections are *absolutely not* "maximally accessible" under such conditions. Nor are they even "easily accessible," as Dr. Marcum claims (above). These Internet-type access channels *create* as manyproblems as they solve for scholary research: in this case, none of these copyrighted bibliographies is full-text searchable in either Google Books or Amazon. And even if they were, it would never have occurred to me to think of searching a keyword string such as "cited in Smith's *Wealth*

of Nations." I could recognize the importance of the words when I saw them within the full-text of a lengthy annotated bibliography—a source that, itself, I could find quickly, within a very limited class area—but I could
never have specified those exact words in advance. (Those who claim miracles for "progressive limitation" of large keyword sets demonstrate remarkably limited experience in having to find works in subject areas outside their own expertise. But LC reference staff have to identify and retrieve, quickly, materials on any subjects that any Member of Congress can think of, even when we ourselves do not know what the best
search terms may be. But because of our distinctive access mechanism of classified shelving we can "make" the shelves show us what we don't know how to specify.)

It would therefore be destructive of LC's distinctive mission to promote scholarship of unusual depth if the Library decided to stop shelving its unique aggregation of books in the distinctive way that provides exactly the deep access required by a Supreme Court Justice and a Member of Congress.

Of course, *most* of the questions we get do not require that kind of access—but then *most* of them *also* do not require citation searching or related-record searching or referrals to subject experts or use of the full range of encyclopedias that LC alone can offer [example below]. But *this* Library needs to have the fullest possible set of access tools immediately available for the comparatively few times when *all* of our access mechanism are indeed necessary, and necessary under deadline pressure. We have professional responsibilities here that Google, Amazon, and OCLC do not have— responsibilities that, in fact, are shared by no other research library (including all Association of Research Libraries institutions, either individually or collectively).

A question on globalization

Let me provide another example of the unique advantages that scholars have at LC from having the fullest possible set of access tools readily available, on a moment's notice This was a question from an academic researcher, not a government official. He said he was interested in writing on "globalization," but wasn't sure yet how he was going to approach the topic. Since encyclopedia articles are often wonderful starting points for people who are moving into new or unfamiliar subject areas, I could show him very quickly (via one of our subscription databases) that overview-articles on different aspects of "globalization" appear in all of the following titles:

Encyclopedia of Religious Revivals in America
Encyclopedia of Community
Blackwell Encyclopedia of Sociology
Encyclopedia of Communication and Information

Berkshire Encyclopedia of World History
Encyclopedia of Evolution
Encyclopedia of Clothing and Fashion
Encyclopedia of Management
Encyclopedia of Information Science and Technology
Encyclopedia of American Foreign Policy
Encyclopedia of Democracy
Globalization: Encyclopedia of Trade, Labor, and Politics
International Encyclopedia of Environmental Politics
Encyclopedia of African History
Routledge Encyclopedia of Philosophy
Routledge International Encyclopedia of Women
Tobacco in History and Culture: An Encyclopedia
International Encyclopedia of the Sociology of Education
Encyclopedia of the Arctic
Encyclopedia of Government and Politics
Social Science Encyclopedia

Since only a "very small number of staff and researchers" would ever "utilize" such a range of specialized resources, does that mean, then, that the Library of Congress need not collect that full range? or that LC can send these works to remote storage, from which each would then have to be requested for overnight (or lengthier) delivery? or that we can rely on Kindle versions (which do not exist)? The fact that these encyclopedias are easily accessible—most of them within open reference collections, and browsable in relation to other full-text reference sources—gives them a cumulative power that in turn gives any researcher at LC a *distinctively different* advantage in approaching such a topic, which he would not have in any other library in the world. The very same sources that would indeed be "little used" in other libraries *become* "more used" *here* because of the synergies brought about by their physical aggregation in our unmatched reference collections.

In other words, the bean-counting objections to the maintenance of classified shelving are remarkably similar to the Inspector General's objections to assembling a "universal collection" of texts, in one physical location on Capitol Hill, in the first place. The fact is, the vast bulk of our book holdings, because they are copyrighted, cannot be read at full-text level by anyone who is physically outside our reading rooms. As much as our publicity emphasizes how much material we are digitizing for free distribution on the Internet, virtually none of that material comes from our general book collections. Since *only* a *very small number of staff and researchers* can actually *use* most of the 21.5 million volumes we have collected—i.e., those who are physically in our reading rooms, whose number is indeed "very small" in comparison to the number of those producing hits on our Web site—do low "body counts" here mean, then, that we should no longer collect a universal scope of materials to begin with, because the number of onsite scholars is so small in comparison to

Internet users? (Should we eliminate the expense of the Kluge Scholars program, or of the NEH-funded scholars in residence, because they form "only a very small number of staff and researchers" who can make full use of the scope of our onsite collections?)

The proposal to cease collecting a universal scope of material, physically centralized at LC, is exactly what the Inspector General suggested—and it is exactly what the Librarian of Congress decisively rejected; in his words, "The enormous legacy print collections *that will never be digitized must remain available for answering questions and providing support for in-depth research*." That centralized physical collection, no matter how many people use in on Capitol Hill, enables *anyone* who does come in the door to freely pursue scholarship of both unusual scope and unusual depth [see next example], *which cannot be comparably achieved anywhere else*. And that is our mission. Any "digital strategy" that overlooks this rock-bottom fact cannot work; we must think "outside the box" of digitization alone, and proceed on the assumption that LC must maintain both onsite collections, *maximally accessible only to a comparatively small number of onsite researchers*, as well as electronic resources freely available on the Web.

A Question on the Qu'ran

Just one more example: In March of this year I helped a Muslim scholar who was interested in how a particular Arabic word for a kind of palm tree had been translated in various English versions of the *Qu'ran*. I went back into the stacks four times for him, each time bringing out armloads of all I could carry of our various English translations, *all shelved together* in BP109. He was *very* thankful—he had only one day to visit the Library, and he told me at the end of the day that he would never have seen so many different English versions in one place anywhere else in the world. The important point is this: even if the same books had been collected elsewhere, he would not have had anything like comparable *access* to them if they had not been *physically shelved next to each other* to facilitate easy retrieval and easy comparisons of full-texts. Nor could he have typed in, within any digitized collection of texts, the English translations of the particular Arabic word that interested him—because he did not know to begin with which English terms were being used. That was exactly what he needed to find out—not what he could specify in advance. When Muslim scholars themselves have unrivalled access to Qu'ranic materials here (as nowhere else), should we eliminate the very means of access that richly satisfied this gentleman because he was a member of only a "very small number of staff and researchers" who benefited by our classified shelving?

Or, rather, isn't it *precisely because* this Library can handle that "very small number" of such inquiries—from Supreme Court Justices, Members of Congress, academics, and foreign visitors from all over the world—in ways that no other library can, that give scholars such *distinctive* results in coming here? Isn't *this* the *unusual* kind of service the Librarian of

Congress appeals to in his reply to Mr. Schornagel, in rejecting the Inspector General's bean-counting assumptions? We have that kind of access available at a moment's notice, in all subject areas, across all languages—no matter when such questions come and no matter what their source, either governmental or private. And we have that unmatched capacity precisely because the Library of Congress is *not* like Google or Amazon.

If we no longer have classified shelving as one (albeit not the sole) important means of access to our immense aggregation, then we are not making the books either "maximally accessible" or even "easily accessible" when the most difficult questions arise. In maintaining this *distinctive means of access* to our *unique aggregation* of books we are providing a vital *synergy* for scholars that is simply not profitable for the private sector to furnish. This is precisely the role that needs to be played by the *one* government agency, tax-supported by the entire country, that is *responsible* for providing maximal access to the Nation's sole collection of copyright-deposit books, in ways that cannot be duplicated by Google, by Amazon, by OCLC, by the *entire* Internet, or by *any* other library.

Providing just this is the kind of maximal access is something we used to be proud of around the Library of Congress. But now, apparently, we're supposed to ignore our history, radically dumb down our search capabilities, eviscerate our own understanding of our mission, and just provide "something quickly" to Congress and the American people, rather than scholarship of unusual scope and unusual depth—and why? Because Library Services consistently tries to force the Google and Amazon paradigm onto LC's operations, as models for the service we ought to be providing, and no longer sees LC as distinctively different in either its mission or in the technicalities of its operations needed to carry out that mission.

How to deal with very real space shortages

What should we do, then, to deal with our very real space shortages, if we cannot rely on a digital strategy to solve the very real problem? Several things:

1) Don't "slant" the portrayal of the problem to Congress as being worse than it is; and further, don't "frame" the problem to begin with as though digitization now represents the *only* solution. The Library has dealt with the same problem of books "piled up on the floor" repeatedly for over 130 years. When we became the centralized locale for copyright deposits, before the Jefferson Building was opened, then-Librarian Ainsworth Rand Spofford's legendary memory could still find things within the piles of books that then existed. (Granted, there's no single person like him anymore; but there is still a huge distributed range of staff knowledge of the collections—a resource that can readily be appealed to as part of the solution to current problems.) When the Jefferson Building itself was

opened, the planners at the time thought they had left enough room for all of the copyright deposits for the whole 20th century; but the stacks were filled up by the 1920s. We then filled in two open courtyard areas, and lobbied for the Adams Building, which opened in 1939. That, too, rapidly filled up by the 1960s, with books on the floor; and relief was not provided until the Madison Building opened in 1980. Of course, that too filled up, and now we have remote storage facilities at Fort Meade and Landover (and separate collections at Culpepper). At every stage of our history, books became "piled up on the floor" and yet the Library survived without throwing the baby out with the bathwater—specifically, without abandoning the classified shelving needed to fulfill its mission (to promote scholarship of unusual depth) in favor of either height-shelving or accession/barcode storage for Capitol Hill books. There is no need to abandon classified shelving now, either. Doing so would cripple our ability to provide just the kind of *unusual* access for Congress, the Supreme Court, and advanced scholars, that Google and Amazon cannot provide at all.

2) Given that, in Dr. Marcum's words, "We know from experience that funding for Fort Meade modules is slow in coming, when it comes at all," then it is increasingly important for LC management to convey to Congress, with much greater emphasis, the message that LC is indeed distinctive, that Google and Amazon do *not* provide workable models for us to follow, and that therefore funding for another module is imperative if we are to maintain our distinctive capacity for promoting the kinds and levels of scholarship that cannot be done anywhere else. Given, further, that "the proposal to begin design work for module 5 was cut from the 2010 budget," then *management needs to do its own job*, in lobbying all the harder for its restoration in a future budget. And will that really be asking Congress for anything unusually expensive?—i.e., since the various modules are largely similar anyway, how big of a problem is it to design another one that will basically be a carbon copy of what's out there already? Similarity of design for add-ons is why we chose modular construction to begin with.

And, yes, we can live—as we have so often before—with *some* books "piled up on the floor" in the meantime. This is not a disaster; it never has been in the past. It is a manageable problem.

3) One very important component of the solution to books-on-the-floor is not even mentioned in Dr. Marcum's memo, quite possibly because it does not fit within the "all digital frame" within which she limits her inquiry—i.e., the very title of her memo is "Meeting on Digital Strategy"—and yet it is exactly what Dr. Billington appeals to in his letter to the Inspector General: "Equally important are the librarians and curators who know this material intimately. There is *no satisfactory substitute* for knowledgeable staff members who have built collections over long years" (emphasis added). In the present situation, there is no satisfactory substitute for those very same knowledgeable staff members who alone are capable of *weeding* the onsite collections, to determine which works (and copies) shall go to storage when funding for new modules does become available. This point

needs emphasis: wholesale digitization cannot take the place of what the Library's professional staff can do in adjusting the collection's size to its space on Capitol Hill.

Recommending Officers (ROs) at the Library are currently weeding our general collections, on a rush basis, to get a half-million books off the floor by November of 2009. (Apparently the floor-clearing must be finished prior to the Library's testimony before Congress early next year.) ROs are the professional staff in charge of collection development in various specific subject areas; our responsibilities are divvied up, in part, by the different class areas in the LC Classification system—and the bookstack areas corresponding to those classes. We are currently inserting yellow slips in the volumes we pick to go to our remote storage warehouse in Ft. Meade, Md. But much of this present work, due to the deadline, is quick and dirty—we are looking, right now, especially for long runs of serials or comparable runs of annual volumes (or annual sets, such as encyclopedias). Serials that are shelved in classified order do not reveal the contents of their individual articles as readily as books that are shelved within the same class areas (as with the Occupation Zone pamphlet, above). The reason is that any long run of a serial may contain articles on many thousands of individual subjects; but the necessary placement of the entire run of the journal in only one class area does not show those individual articles in any noticeable relationship to the corresponding books on the same wide variety of topics. Access to journal articles has always been provided most efficiently by commercial indexes and databases (and footnotes within scholarly sources themselves) rather than through either LCSH headings in library catalogs or LCC shelving in the stacks. Such runs of serials are therefore prime candidates for weeding to remote storage (especially if their contents are duplicated in JSTOR or other databases). Their absence allows more room for retention of obscure monographs such as the Occupation Zone pamphlet, the revelation of whose content is much more dependent on classified shelving.

But a focus on weeding serials cannot solve all space shortage problems in the long run. It is unavoidable that we must also select books for transfer to Ft. Meade. I, for one, would love to have the time to go through our deck areas covering Philosophy, Psychology, and Library Science, to slip individually all of the books (and copies) that I think can reasonably be sent to remote storage in a way that still preserves access to "very-difficult-to-obtain, obscure materials that provide information to our lawmakers that is available nowhere else." Having the books in every such class slipped by the professionals in charge of those areas would be the best way to clear the floors when the next "crunch" comes. If management does its own job, it is inevitable that we will indeed get a Module 5 for additional remote storage—even if it takes a few more years than we would like.

But shelf-by-shelf examination of individual monographs, throughout LC's 535 miles of bookshelves, is not something that can be done quickly or under a short deadline. Assuming I had as much as three hours per week

to weed all of our B and Z classes, I suspect it would take me a couple years to do a good job of it.

We have just the time we need

In this connection, however, the current delay in securing funding for a new module at Ft. Meade can reasonably be regarded as very much a blessing in disguise: if we cannot have a new module for a few more years, then we can use just that intervening time to intelligently "slip" which books should go there when the space does become available. *We now have exactly the time we need to pre-select which books should go to Module 5: let's make the best use of it*. This is a much more practical way to deal with space shortages than any reliance on wholesale digitization of texts by Google or Amazon. (Of course the weeding process would have to be an ongoing practice rather than a one-shot project; but that is exactly what I am recommending. Time in the stacks should be regularly scheduled for all ROs, as a permanent feature of their jobs.)

An added benefit of such a practice would be an increase in Recommending Officers' own knowledge of their own areas of the collections, derived from direct and systematic reviews of their entire stacks areas. I can foresee one spin-off benefit already, since I've experienced it myself in the current round of weeding. One problem that we have with the new reorganization of cataloging is that when new books are received, they first go to a central area in the Madison Building where they are examined by librarians from the different reading rooms, to spot those titles or sets that should be flagged immediately for assignment to a particular reference collection. (It is most efficient to do the cataloging and labeling of the volumes *as* reference books in their initial pass through the system, rather than to have to add the additional cataloging and labeling data as a second operation.) In past years, before the reorganization, the librarians who examined these new titles could also slip them "for examination"—i.e., to be sent for review by the various subject-specialist Recommending Officers. That way the ROs could physically examine interesting new volumes within their subject areas—even those that were not routed for formal assignment to the reference collection in their Division. In the new workflow, however, the slips "for examination" are now being discarded, and only those "for reference assignment" are attended to. What this means is that many new books of interest are now *not* brought to the attention of the Library's subject experts; and it also means that many volumes which ROs would assign to reference collections (if they inspected them personally) wind up in the general bookstacks rather than in the reading rooms. This is a problem that would be largely solved if ROs spent time in the bookstacks themselves, directly examining everything there—they could decide not only which books should be yellow-slipped to go to remote storage, but also which ones should be re-assigned to reference collections. Better reference collections would be a benefit to reference service throughout the entire Library.

Let's say that every RO would be required to spend, say, three hours per week in the stack areas for which he or she is responsible over the next three years. (The number of hours cannot be firmly determined at this point.) If each RO had the time to decide, at the level of individual volumes (i.e. not just long runs of serials), what can be weeded to storage, with a professional judgment of how easily each could be found in storage, *we could still maintain the classified shelving needed to answer those questions of unusual depth (that cannot be answered by digital sources) while also eliminating the next round of the "books on the floor"* problem. Once it is slipped, this material could be removed very quickly to Fort Meade when the Module 5 becomes available. There are undoubtedly "problems to be solved" here, too; but, given the *access of unusual depth* that is at stake, such a project is worth the Library's effort.

One of the "problems to be solved" is of course limited staff time. There are only so many more things that can be done by people who already have full-time jobs. What activities need to be sacrificed when push comes to shove? In this regard, Dr. Marcum says the following in a recent memo to LC staff ("Friday's News – September 18, 2009"):

> We expect to know relatively soon exactly what [budget figures] we have to work with, but I shall be asking all of the divisions to trim their budget requests to the amount they had in 2009. *Doing new things inevitably means cutting back on existing activities*. [Emphasis added]

We need to step back a moment and really examine that last sentence. When any group of people, from a family to a large corporation, is faced with the need for belt-tightening, what should determine its priorities? Is it at all normal, or prudent, to cut back on the essentials of their operations in order to "do new things"?

I submit that organizations should first seek to protect those things that they are *responsible* for, and which most directly effect the *mission* (or even the life) of the group. A parent who is faced with less income but undiminished mortgage payments and food bills will not say, "Let's stop eating three times a day so that we can get a *new* video game system." A soft-drink company that jettisons its tried-and-true product for something *un*tried but *new* (think of the debacle of "New Coke") will find that it has acted foolishly—not only against the interests of its clients but contrary to its own best interests. In the tough economic times that all institutions are facing, the fiscally prudent response of any organization is to *first* protect the functions that define its core mission, and to provide the services is *responsible* for, and which its clients *rely* on it to provide. What that means, then, is "protecting the functions that serve our core mission is in fact more important than 'doing new things,' especially if the new things that are being pursued undermine that very core mission." Again, I submit that LC's mission is to *promote scholarship of unusual scope and of unusual depth*. We cannot continue that "existing activity" if we abandon

either the system of LCSH cataloging in our OPAC or the system of LCC classified shelving in our bookstacks. These are the very mechanisms that enable the Library of Congress to provide that *unusual access* to its unique, publicly-funded, legal-deposit-mandated collection—access that cannot be matched or even approximated by Google or Amazon. And keep in mind, further, that neither Google nor Amazon (nor OCLC nor the ARL) has LC's responsibility to serve the information needs of Congress, the Supreme Court, or any federal agency. LC is indeed very distinctive; in Dr. Billington's words: "The Library of Congress is the authoritative, responsible knowledge institution for the Congress and the United States government." The access mechanisms provided by Google, Amazon, and WorldCat are not adequate for the kind of work we are required to do here. This is not to say the latter are useless—the reverse is true—but it is to say they are *not nearly sufficient* for what the Library of Congress needs to be able to do, especially for its local clientele, and especially on short notice. It is quite true that this local clientele itself, including Members of Congress, Supreme Court justices, Kluge and NEH scholars, and Muslim researchers among others, forms a "very small number" of people compared to the universe of those relying on Google and Amazon. But then the latter group of researchers will never find all that we can discover at the Library of Congress. They do not have our means of access.

Priorities in regard to cooperative agreements with other libraries

In regard to providing services to outside libraries, I believe LC needs to respectfully disagree with two statements in a recent report, *Transformational Times*, from the Association of Research Libraries (http://www.arl.org/bm~doc/transformational-times.pdf). One is that "If libraries turn inward and focus on protecting local resources, they could pull back from essential cooperative work" (p. 6). The fact is, if local collections are not adequately cataloged and classified to begin with, then they cannot be *found* within any merged pool of collective resources, such as WorldCat. (Indeed, WorldCat's search software is almost as bad as Google's or Amazon's, in that it, too, effectively precludes the possibility of researchers' gaining the overview-perspectives, as with **Afghanistan** [above]. The OCLC system cannot display either browse-menus of precoordinated headings *or* cross-references among broader, related, or narrower subject terms.)

Further, LC's cataloging of its own, unique copyright-deposit collection is becoming even more crucial to the operation of other research libraries, which are facing draconian State-level budget cuts that LC is protected from. The unique extent of the material that we get *here*—the scope of our book acquisitions that Dr. Billington rightfully insists on maintaining—is becoming more, rather than less, important to other libraries because they can no longer spend as much for cataloging the same titles first, when their own budgets are being cut in ways that LC's is not. They need, because of increasingly stringent cuts to their own cataloging operations, more than ever to utilize the cataloging data for the same books provided by the Nation's copyright-deposit library. The new (October, 2009) *Library of*

Congress: Study of the North American MARC Records Marketplace from R2 Consulting (http://www.loc.gov/bibliographic-future/news/MARC_Record_Marketplace_2009-10.pdf) states clearly:

> **Library of Congress cataloging continues to be widely valued:**
> Libraries, vendors, and cooperatives speak with their actions. There is heavy reliance on LC's output throughout all segments of the profession and industry. [p. 4]
>
> LC records are the most highly sought, period. [p. 12]

The best thing that LC's management can do for the cooperative benefit of other libraries is to increase the level of book-cataloging that we do here, precisely so that hard-pressed libraries in all other Congressional districts can rely *more* on what we do locally.

The second misleading statement in the ARL report is that "Special collections distinguish a major research library from all others, and provide incomparable value to graduate and faculty research and teaching" (*ibid*. p. 16). While there is some truth to this regarding most ARL libraries, the fact remains that LC has both privileges and responsibilities quite unlike those of any other ARL facilities; and as Dr. Billington pointed out, it is the unparalleled scope of LC's *general* collection that *most* distinguishes us from other research libraries. Indeed, it is our general collection itself that forms the single most "special" collection we have, thanks to our unique copyright-deposit privilege. Further, the mechanisms of access that we provide to *this* collection—i.e., the professional cataloging and classification (with full Cuttering of class numbers) that we create here—have the greatest effect in generating cost-savings for other libraries. That fact needs to be kept clearly in mind in evaluating statements such as "Doing new things inevitably means cutting back on existing activities."

Conclusion

The Library cannot solve its space problems by adoption of a "digital strategy" without seriously damaging our larger mission to promote scholarship of unusual scope and depth. If the Library's own access to its own general book collection were to be dumbed down to only the levels of subject access provided by Google, Amazon, or Internet search mechanisms, we would effectively be endorsing, and institutionalizing, the level of ignorance exemplified by the Six Blind Men of India. We would no longer be able to satisfy the information needs of Congress or other federal agencies, or those of advanced scholars, in a timely or efficient manner. It is not enough simply to have comprehensive collections; we must equally have *efficient access* to them. While the subject access-mechanisms (LCSH cataloging in our OPAC and LCC shelving of books in the stacks) needed to provide unusual comprehensiveness and unusual depth of scholarship, here, are themselves expensive to maintain, they are the very things that

enable us to realize our distinctive mission, to provide our distinctive services, and to discharge our distinctive responsibility to provide maximal access to the Nation's unique copyright-deposit collection, while simultaneously enabling other research libraries in hundreds of Congressional districts, which copy our systems, to achieve substantial cost-savings themselves.

To deal with the problem of books-on-the-floor, what we need is a systematic, professional weeding of our Capitol Hill collections to determine which volumes will next go to remote storage in Module 5 at Fort Meade; and Recommending Officers must be given the necessary time on an ongoing, weekly basis to undertake this huge task. Because of the delayed funding of that next module, however, we now have just the time we need in order to undertake this systematic examination and weeding of the general collection. This option is much preferable to that of relying on wholesale digitization of our books, especially since the latter process (as in Google or Amazon models) could never extend to our entire collection (for copyright reasons alone), and because it would simultaneously create more problems of access than it would solve, in ways that would directly and severely undercut our distinctive mission.