"On the Record" but Off the Track

**A Review of the Report of
The Library of Congress Working Group on
The Future of Bibliographic Control,
With a Further Examination of Library of Congress
Cataloging Tendencies**

Thomas Mann

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

March 14, 2008

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

Major points

The Working Group's *Report* is off the track in many of its major assumptions, assertions, and recommendations:

- In ignoring the very real and important differences between the research needs of scholars and those of "quick information" seekers

- In not understanding what the Library of Congress Subject Headings system (LSCH) continues to accomplish in providing crucial overviews of relevant literature across multiple languages, accomplishments that are neither equaled nor superseded by Web 2.0 search mechanisms

- In assuming that the capacity to search across multiple environments is more important than the capacity to search efficiently and comprehensively within any of them individually

- In its tacit endorsement (in spite of a few unintegrated paragraphs to the contrary) of uncontrolled keyword searching as being more important than controlled-vocabulary searching (since keywords are the only elements that *can* be searched across multiple environments)

- In disregarding the need for cross-references, browse-menus of LCSH subdivisions, and scope notes as integral elements of vocabulary control for book collections in research libraries

- In calling for the movement of cataloging data from online library catalogs, which can display both cross-references and browse-menus of subdivided LSCH terms, to a Web environment that can display neither

- In not even mentioning the importance to scholars of maintaining browsable, onsite book collections arranged in Library of Congress Classification (LCC) subject categorizations

- In calling for the "de-coupling" of Library of Congress Subject Headings strings into individual word "facets," thereby entirely eliminating the elaborate and detailed network of cross-references and browse-menus that have been professionally constructed and expanded for over a century—which network is crucial to providing systematic overviews of books relevant to a topic

- In not realizing that the "de-coupling" of LCSH that would also eliminate the scope-match level of indexing specificity achieved by precoordinated strings, which conceptual level of subject access continues to solve problems, both of overview-provision and of preventing information-overload at excessively granular retrieval levels

- In ignoring the elaborate existing network of integral linkages between LCSH precoordinated subject strings and LCC class numbers, which linkages would be yet another casualty of "de-coupling" the strings

- In failing to recognize that readers cannot combine, postcoordinately, individual facets whose existence they cannot think of in advance, and which are more hidden than revealed by their segregation into separate Topic, Time, Geographic, and Form facet "silos" rather than merged in a single, unified browse-menu

- In calling for an opening up of LCSH to "non library stakeholders" whose input (if allowed into library catalog environments, rather than Web pages

linked to library catalog records) would entirely disregard, and directly undercut, the necessary maintenance of cross-references and browse-menus, as well as undercut the principle of "uniform heading" itself, which is necessary for vocabulary control

- In failing to note the crucial distinction that, in providing overview perspectives, the relevance-ranking of keywords is not at all comparable to the conceptual categorization of resources under standardized retrieval terms

- In disparaging the functions of the Library of Congress as "alpha" library (in spite of its sole acquisition of the nation's Copyright deposits, its unmatched scale of foreign-book purchasing, and its unique and unrivaled financial support by every taxpayer in the country) in maintaining professional cataloging standards, while calling instead for a distributed, de-centralized, and open-to-all system in which, effectively, what is everyone's responsibility will quickly become no one's responsibility

- In ignoring the economic reality that tax-supported cataloging work done centrally at LC more than pays for itself in the savings that accrue to thousands of other individual libraries, in all Congressional districts

- In asserting that the digitization of special collections (especially textual rather than visual collections) of use to comparatively few scholars, is now to be regarded as a higher priority than maintaining the LCSH and LCC cataloging systems, which are of use to scholars and libraries in every Congressional district in the nation

- In ignoring the substantive reasons for having, and maintaining, different controlled vocabularies to begin with, suitable to the distinctive needs of different user groups

- In biting off more purposes for "bibliographic control" than cataloging operations, alone, can possibly accomplish even with a combination of both LCSH and LCC and Web 2.0 search mechanisms

- In ignoring the integral need for the education of researchers (provided via both point-of-use reference service and class instruction) in the total system of bibliographic control, in exchange for naïve beliefs in the capacities of both Web 2.0 collective indexing and under-the-hood programming across multiple search environments

- In conspicuously failing to provide any concrete examples of how the Group's call to put everything into a "unified" and "Web-based" environment would produce improved, rather than diminished, research results in comparison with an overall system that utilizes both

professionally-created LCSH cataloging (with cross-references and browse-menus) and algorithmic/democratic Web 2.0 mechanisms in separate (but linked) search environments—an overall system that *also* provides browsable physical book collections shelved in LCC subject categorizations in a third environment; and multiple other resources in additional environments with yet other distinctive search and retrieval capacities.

In spite of these difficulties, there does exist a sensible path forward that allows for all of the Working Group's calls for "outside" Web 2.0 inputs to be utilized as supplements to, rather than replacements for, proven LCSH and LCC mechanisms—a compromise that will allow systematic, comprehensive, and scholarly access to the nation's library collections to be maintained while also making full use of the new Web possibilities of relevance ranking, democratic tagging, folksonomy referrals, etc. But even this path will be severely compromised if LC's own cataloging managers succeed in carrying out their current reorganization plan, which entails rewriting the Position Descriptions of LC's cataloging staff to de-professionalize their distinctive work by minimizing or eliminating their need for subject expertise, and to burden them with new and time-consuming acquisitions responsibilities.

*　　*　　*

In responding to *On the Record* I can only repeat what Francis Jeffrey famously said in opening his review of Wordsworth's *The Prelude*: "This will never do." The recommendations of this *Report*, while well-intentioned, are unfortunately so naïve about the requirements of scholarly research that, if implemented in the particular way proposed by the Group, will seriously undercut the capacity of scholarly researchers everywhere to pursue their topics systematically and at in-depth levels, rather than haphazardly and superficially. (A more reasonable, and practical, solution to this problem will be noted at the end of this review.)

Please note up front: no one is suggesting that maintaining the Library of Congress Subject Headings (LCSH) and Library of Congress Classification (LCC) systems are the only things that we need to be doing. Keeping systems that work, however, and that solve real problems for researchers better than any proposed alternatives, while we also pursue additional enhancements, is not the "straw man" of "merely protecting the status quo" that it is often misrepresented to be.

Scholarship different from quick-information seeking

The *Report* does not recognize the fact that the needs of the Library of Congress (and other research libraries) are quite different from the needs of school libraries, or public libraries, or special libraries—and that libraries' needs are different precisely because the

communities they serve are so different. The Working Group makes no substantive distinctions among types of libraries or levels of users.

Specifically, it makes no substantive distinctions between scholars and those who may be called "quick information" seekers. The *Report*, in saying that "separation of communities of practice . . . is no longer desirable, sustainable, or functional," would seem to suggest that a "one size fits all" approach to research is the goal to aim for, as though researchers themselves formed a single, "unified" community. But they do not. Here are only some of the differences between scholarly research and "information seeking":

> **1**) Scholars seek, first and foremost, as clear and as extensive an *overview* of *all* relevant sources as they can achieve. They want to see "the shape of the elephant" of their topic (the reference being to the fable of the Six Blind Men of India). They want to see not just the full panoply of its different important parts but also how the parts fit together and are related to each other.

> **2**) Scholars seek to find relevant works in conceptual relationships to earlier (and out-of-print) books, and to those that may appear decades after the works' publication; therefore speed in cataloging is *not* the hallmark of quality service for scholarly purposes, because any book may be useful at any time during the course of its existence, not merely in the first six months after its publication. If the work cannot be found in relationship to other relevant works, its individual utility is substantially vitiated rather than enhanced. (Moreover, the book's utility is *eliminated* if it cannot be found to begin with, because its peculiar keywords are unknown to the researcher, and it fails to show up in standardized categories [i.e., under uniform headings] which, themselves, *can* be found systematically.)

> **3**) Scholars are especially concerned that they do not overlook sources that are unusually important, significant, or "standard" in their field of inquiry.

> **4**) Scholars do not wish to duplicate prior research unnecessarily or to have to "re-invent the wheel."

> **5**) Scholars wish to identify *whole books* on their topic, especially when they are trying to get an overview of its extent, in preference to excessively granular retrievals that dredge up every paragraph in any book that happens to mention their topic in passing.

> **6**) Scholars wish to be aware of cross-disciplinary and cross-format connections relevant to their work, and to the larger (and perhaps multiple) conceptual perspectives within which their topic falls. They want to see relationships of books (and other resources) not just to those on the same topic, but also to other relevant works "off to the side" that impinge on their subject in interesting ways.

7) Scholars need systematic access to resources in hundreds of languages, not just in English.

8) Scholars particularly appreciate mechanisms that enable them to *recognize* highly relevant sources—either within catalogs or on classified bookshelves—*whose keywords they cannot think up in advance*, to type into a blank search box. The larger any collection is, the more it requires access via recognition mechanisms (e.g., conceptual categorizations, cross-references, browse capabilities that show more, or other, relevant items than just those they know how to ask for).

9) Although they are more cognizant of the need for diligence and persistence in research, and of the requirement to check multiples sources, and of the need to look beyond the "first screen" display of any retrievals, scholars also wish to avoid having to sort through huge lists, displays or clouds—from *any* source—in which relevant materials are buried within inadequately-sorted mountains of chaff having the right keywords in the wrong conceptual contexts, or out of context entirely.

Two points need emphasis here:

- First, scholarly researchers do indeed *want* these things—even if they do not expressly *say* they want them in user surveys (which may not have asked all or any of the right questions). And we can discover this fact, predictably and repeatedly, in any situation in which a reference librarian shows a scholar how to *solve* any of these problems. (Researchers are routinely delighted to be informed about better, more efficient, and more comprehensive search options than they know how to ask for.)

- Second, most of these outcomes cannot be brought about by simultaneous searches of keywords (or any other data) *across* multiple "environments"; they require sequential searches that change depending on feedback provided by previous steps, and that also make use of the widely-variant search capacities *within* particular environments (e.g., capacity to collocate relevant works having different terminologies, capacity to provide browse-menus of unanticipated "side" options, capacity to do citation searching, capacity to limit by date/language/format, etc., etc.) Such capabilities—necessary to scholarship but not to "quick information" seeking—are lost in "seamless" searching in a unified Web environment.

The fact that there are entirely different professional associations for Public, Research, Special, and School libraries—each with different needs (otherwise there would be no

need for separate groups)—does not disturb the egalitarian aspirations of the Working Group, to open up LC cataloging to anyone who might want to join the existing cooks in the kitchen. The recommendation (4.3.3.1) to "Make vocabularies cross-searchable and interoperable" is incredibly naïve. It apparently does not occur to the Group that different controlled vocabularies aim at *different levels of subject specificity* to begin with—e.g., LCSH vs. Sears List—and for that reason alone cannot be made cross-searchable without destroying their utility in solving the very different problems of their very different user groups.

English language headings

Further, while it is undoubtedly politically correct to point out that "Emphasis on textual strings as identifiers binds entries to a single language and thus hampers efforts to internationalize both authority files and bibliographic files," (*Report*, p. 24), the fact remains that the provision of *uniform* subject headings *in English* solves problems for the American researchers whose tax and tuition moneys are *paying* for such solutions. For example, in American library catalogs one will get a better overview of Italian-language books on **Venice** by typing in that Anglicized form, and "limiting" to Italian, than by typing in the keyword "Venezia" to begin with. Even foreign scholars themselves will get the best overview of foreign language works (in German, French, and Greek) on "tribute payments in the Peloponnesian War" by typing in the English language uniform heading **Finance, Public—Greece—Athens**; and so on. If we wish to contribute to the internationalization of bibliographic control, the best thing we can do is to provide an LCSH system that does indeed solve rather than exacerbate vocabulary control problems, and that is available for translation by other countries' taxpayers, rather than to dumb down LCSH to the point that it no longer controls what it needs to control. English is already functionally established as the international language of scholarship to a much greater degree than the Working Group apparently wishes to recognize. It helps scholars *everywhere* if they have access to *one* system that makes books in *all* languages systematically discoverable at the same time and in the same ("uniform heading") groups. Making LCSH more "translatable" is a particularly bad idea if it entails, as the Group recommends, the "de-coupling" of its precoordinated strings that give the system its unmatched power in providing the systematic-overview perspectives on subjects that all scholars, internationally, require (see the **Afghanistan** example below). Without such power to begin with, the system is hardly worth translating into other languages.

Inappropriate business model

Although some users do in fact need deeper and more systematic access to information than others do, the Working Group is skeptical of this very point:

> Many libraries have chosen to produce all their metadata to satisfy the needs of their most sophisticated users, *despite the fact that such users are but a small percentage of their total user base* [Note: this remark is apparently based on

"business model" assumptions.] They do so on the unproven assumption that all users will benefit from the greatest detail in cataloging. [p. 31; emphasis added]

Perhaps, rather, "many libraries" produce their records on the assumption that satisfying the needs of their most sophisticated users is crucial because the scholarly research of *those* researchers is *unusually important* to the intellectual health of the entire nation, no matter what percentage of the totality of all researchers they may be. Perhaps the many academic and research libraries—unlike small public libraries—actually assume, not that "all" users will benefit from the greatest detail, but that *these* researchers, unlike "quick information" seekers, *need* more systematic and comprehensive access to books than the vast majority of those who want only "something." (The latter inquirers can always search records having "the greatest detail" at any level, either comprehensively or as superficially as they please; but the former group cannot search comprehensively without that detail being present.) Perhaps the only "unproven" assumptions are the Working Group's notions that a) scholarly research, since its practitioners form "but a small percentage" of "all" users, is evidently of correspondingly small importance (i.e., too small a market share to be catered to with systems expensive to maintain); and b) that the extra requirements of scholarly research are no longer to be regarded as important if they stand in the way of creating a utopian "one size fits all" retrieval system for the entire world.

The need for multiple environments with different capabilities

Although the *Report* occasionally does seem to recognize the existence of "a diverse community of users, and a multiplicity of venues where information is sought," the implications of this reality do not play out in its actual recommendations. Particularly telling are the assumptions articulated on page 10:

> Different communities of bibliographic practice have grown up around different resource types: library collections of books and journals; archives, journal articles; and museum objects and images. As these resources and others become increasingly accessible through the Web, separation of communities of practice that manage them is no longer desirable, sustainable, or functional. . . . Consistency of description *within any single environment, such as the library catalog*, is becoming *less significant than the ability to make connections between environments,* from Amazon to WorldCat to Google to PubMed to Wikipedia, with library holdings serving as but one node in this web of connectivity. In today's networked information environment, bibliographic control cannot continue to be seen as being limited to library catalogs. [Emphasis added]

The assumptions here are that:

> 1) "seamless" access *across* a variety of environments is much better than "seamed" access that segregates one search environment from another;

2) that such capacity to search *across* different environments is a more important consideration than the capacity to search efficiently and systematically *within* any one of them; and

3) that the "different communities" involved will be better served by eliminating peculiarities in their local environments that inhibit cross- or federated-searching.

I must directly challenge such naïvete, and for good reasons. Seams separating different research resources are in fact not only desirable for scholarly research; they are necessary.

Problems with seamless access

One reason that "seamless" searching is held up to be the goal is, apparently, that "users" say they *want* it. But are there no more factors to be considered than just that? Let me offer an analogy: users of health care services *want* miracle drugs that will target their problems, without any dangerous side effects. And they *want* such drugs to be freely available. And they *want* them without having to go to a particular place—their doctor's office—to undergo a physical exam, during which their doctor may well inform them that the side effects they don't know about can in fact be very injurious to them; and during which the same doctor may well prescribe a variety of much better treatments, from a whole range of options of whose existence and operations the patients had no idea at all. Further, patients *want* those miracle drugs without having to consult any pharmacist, either; and the pharmacist may well notice and point out additional undesirable effects of combining the desired drugs with other medications the patient is taking, unknown to the physician.

Why then does the health care profession not simply dumb down its complexity and give people what they "want," immediately and "remotely"? Perhaps it is because what they "want" regarding health care does not match reality, and may therefore do harm not only to those who want such immediate gratification themselves, but also to the many *others* who depend on the continued existence of a proven system that can indeed deliver much better results, even if entailing greater inconveniences in access, or delayed gratification.

The analogy to libraries is of course not perfect—but the point is still valid: What uninformed people *say* they "want" may not in fact be conducive to their best interests, and may in fact do them—and others whose needs they are ignoring—more harm than good if they are not apprised of either the shortcomings of their desires, or the benefits of alternatives they know nothing about. Perhaps we need to remind ourselves that the "free access to everything in the Web environment" desire cannot match reality until such time as the Copyright Law itself is entirely repealed, and intellectual property is eliminated on the altar of a socialist pipe dream.

One weakness of the analogy is that obviously librarians cannot require their own educative intervention in the research process, at least insofar as it is done on the open Internet. To say this, however, is not equivalent to saying that such intervention is

unimportant, useless, or not of major benefit to those who must pursue substantive research inquiries. Moreover, research libraries attached to universities probably *can* require students to take "research methods" classes, through various Departmental connections. Any librarian who cannot routinely show *most* researchers many more sources, and much better search options than the users think of on their own in pursuing scholarly-level inquiries—or that can be brought to their attention only by "under the hood" programming—is sadly lacking in professional ability. Showing readers more options than they can think of, however, entails much more than making multiple databases cross-searchable by uncontrolled keywords (cf. the extended example in my previous "Peloponnesian" paper at www.guild2910.org, a companion-piece to this one, and to which I will have further occasion to refer).

Seamless searching across multiple sites or databases is not the Holy Grail of our profession. Indeed, the only elements that can be searched *simultaneously* (or via links) across the environments mentioned above (Amazon, WorldCat, Google, etc.) are uncontrolled keywords. (It is significant that the Working Group's attempt to provide an actual example of what it is talking about ignores higher-quality databases such as *Historical Abstracts, Public Affairs Information Service, PsycInfo,* and *WilsonWeb*, all of which have very different controlled vocabularies.) You cannot search controlled elements across multiple environments when most of those environments do not use the same controlled terminologies to begin with, when the various controlled vocabularies aim at entirely different levels of specificity tailored to different audiences, and when most of the environments also lack other human-standardized data (e.g., geographic area codes, language codes, country of publication codes, capacities to limit results by format [review articles, bibliographies, obituaries, etc.]).

Further, if you *do* simply type uncontrolled keyword inquiries into a controlled database, you are usually assured of missing *most* of what that database actually holds, relevant to your interest—and beyond that problem, whatever you do find is likely to be buried in large retrievals having the right words in the wrong contexts. On the other hand, if you do in fact have to search the various environments *sequentially* (rather than simultaneously), then a disregard of the individual and peculiar strengths of each one, by limiting one's search to only the keywords-held-in-common-*across*-environments, becomes a prescription for utterly superficial research: it will routinely miss at least as much relevant material as it finds, and will probably miss all of the best material available on one's topic. (In the "Peloponnesian" paper I provide concrete examples of how much is missed by cross-searching different databases via only the keywords they have in common.)

The need for defined linkages and relationships, not merely individual words

Unfortunately, the Working Group seems to assume that if libraries merely assign cataloging "data"—meaning single-word index-terms—to individual records or Web pages then we have all done our job. Moreover, they apparently believe, further, that anyone in the entire "supply chain" can add this (or at least some) kind of "data" to records, so we ought not to be exclusive in limiting contributions from a wide variety of

sources. What the Group completely overlooks, however, are precisely the most important parts of library *catalogs*—the very parts that *do not reduce to "data" (or metadata) attached to individual records.* The Group not only has an excessively blinkered vision of what constitutes LCSH in its terminology; it also fails to see the need for a catalog "environment," with search capacities different from those in a Web environment, without which LCSH entirely loses its capacity to provide overview perspectives.

An obvious problem (obvious, at least, to librarians outside the Working Group) is that *no vocabulary control system with uniform or standardized headings can be maintained without cross-references.* (I will point out, below, that LCSH in particular also *requires* browse-menus of precoordinated strings of terms for its control.) And neither cross-references nor browse-menus can be searched within general Web platforms such as the Amazon, Google, WorldCat, etc., "environments" that the Working Group has in mind. The Group says, however, that "bibliographic control cannot continue to be seen as being limited to library catalogs." Since LCSH *vocabulary* control (as opposed to much more vague "bibliographic" control in general) cannot be brought about without cross-reference and browse-menu mechanisms *that are necessarily peculiar to library catalogs*, it follows, evidently, that the "bibliographic" control "across communities" sharing a general "Web platform" is such that uniform headings themselves (and the means to *find* them) are regarded as no longer necessary—i.e., as long as uncontrolled keywords (added by anyone in the entire "supply chain") can be searched *across* multiple environments, *that* (apparently) is now to be regarded *as* "bibliographic control." But this is nonsense—and nonsense that is very dangerous to scholarly research.

What is utterly lacking in the Working Group's understanding of the "data" that needs to be present on catalog records, then, is any awareness of the extraordinarily useful *formally-defined network of interconnections and relationships* among LC subject headings themselves, and, further, between the LCSH headings and LCC class numbers—a network that cannot be sustained by opening it up to uncontrolled and inconsistent data-contributions produced by anyone, anywhere.   These relationships are necessary to scholars who are trying to find, not just "something," but an overview perspective of "the shape of the elephant" of their topic.

Library cataloging itself—applied primarily to *book* collections—is only one component part of a much larger system constituting "bibliographic control" in general. And the goal of cataloging is not merely to provide researchers with "something quickly," no matter where the searchers are physically located, within or outside of library walls. Its purpose, first and foremost, is to show "what the library has"—i.e., in its own local collection, onsite. (Further, access to any local research-level collection, onsite, *also* includes the capacity of researchers to subject-browse book collections shelved by LCC class numbers—a major additional component of the larger "bibliographic control" system that is conspicuously overlooked by the Working Group.) As distasteful as this assertion may be to some Internet enthusiasts, we must examine the reasons for saying

that it is nonetheless true, even if unpleasant. It does little good to have OCLC provide access to the catalog records across 50,000 collections simultaneously if the individual catalogs that are being merged are themselves of shoddy quality, and fail to show, accurately, what each institution has individually, while also failing to integrate those local holdings into an overall system with at least a minimum of standardization to its retrieval terminology. Success in searching *across* collections is much more dependent on the quality of records *within* each individual collection than the Working Group wishes to recognize.

A concrete example of what would be lost in either a Web environment or a faceted catalog

An example from real experience is needed; what follows is only a partial presentation of the LCSH treatment of a particular country that people often ask about these days; the same kind of array of aspects and cross-references will *predictably* be found under any other country (or other substantive topic) as well:

Afghanistan
Afghanistan—Armed Forces—Officers—Biography
Afghanistan—Anniversaries, etc.
   NT Independence Day (Afghanistan)
Afghanistan—Antiquities
   [DS353]
   NT Atishkadah-I Surkh Kutal Site (Afghanistan) [formally linked elsewhere to
       DS375.A84]
       Delbarjin Site (Afghanistan)
       Kapisa (Extinct city)
       Shortughai Site (Afghanistan)
       Tilly Tepe (Afghanistan)
Afghanistan—Bibliography [most of which have Z3016 class numbers]
Afghanistan—Biography
Afghanistan—Biography–Dictionaries
Afghanistan—B oundaries
Afghanistan—Boundaries—Tajikistan—Maps
Afghanistan—Civilization
Afghanistan—Civilization—Bibliography
Afghanistan—Commerce
Afghanistan—Commerce—History
Afghanistan—Constitutional history
Afghanistan—Defenses—History—20th Century—Sources
Afghanistan—Description and travel
   [DS352]
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—India—Sources—Bibliography—Catalogs
Afghanistan—Foreign relations—Iran—Sources
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
     [DS355-DS371.43]
     NT   Ghaznevids
           --Anti-terrorist operations, 2001-
               USE Afghan War, 2001- [formally linked elsewhere to
                    DS371.412-DS371.415]
         --19th Century
            NT   Afghan Wars
Afghanistan—History—Bibliography [all of which have class number Z3016]
Afghanistan—History—Chronology
Afghanistan—History—Dictionaries
Afghanistan—History—20th century—Sources
Afghanistan—History—Soviet occupation, 1979-1989
     [DS371.2]
     This heading may be subdivided by the subdivisions used under individual wars
     BT   Soviet Union—History—1953-1985
         --1989-2001
            [DS371.3-DS371.33
         --2001
            [DS371.4-DS371.43]
            NT   Afghan War, 2001 [formally linked elsewhere to
                  DS371.412-DS371.415]
Afghanistan—History—Soviet occupation, 1979-1989—Bibliography
Afghanistan—History—20th Century—Chronology
Afghanistan—Imprints
Afghanistan—In art—Catalogs
Afghanistan—Juvenile literature
Afghanistan—Kings and rulers
Afghanistan—Languages

NT   Afshar dialect
           Bashgali language   [formally linked elsewhere to PK7055.G3]
           Brahui language   [formally linked elsewhere to PL4621-PL4624]
           Dardic languages   [formally linked elsewhere to PK7001-PK7070]
           Dari language   [formally linked elsewhere to PK6871-PK6879]
           Khowar language   [formally linked elsewhere to PK7070]
           Munji language   [formally linked elsewhere to PK6996.M8]
           Nuristani languages [formally linked elsewhere to PK7050-PK7055]
           Ormuri language
           Turkmen language
           Waigali language   [formally linked elsewhere to PK7055.W3]
           Wotapuri-Katarqalai language [formally linked elsewhere to PK7045.W6]
           Yazghulami language [formally linked elsewhere to PK6996.43]

Afghanistan—Literatures
NT   Afghan wit and humor
           Brahui literature   [linked elsewhere to NT Bahui poetry]
           Dari literature   [linked elsewhere to NT Dari fiction, Dari poetry, Dari prose literature]
           Khowar literature   [linked elsewhere to NT Khowar poetry]
           Pamir literature   [linked elsewhere to NT Folk literature, Pamir]
           Persian literature   [linked elsewhere to three BT and twenty-three NT headings including Children's literature, Persian; Ismaili literature; Shiite literature; Travelers' writings, Persian]
           Pushto literature   [linked elsewhere to three NT headings]
           Tajik literature   [linked elsewhere to PK6978, two BT and eight NT headings]
           Uzbek literature   [linked elsewhere to two BT and ten NT headings]

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

We need to look more closely at the implications of this **Afghanistan** example; and we need to note many shortcomings of the Working Group's assumptions and recommendations.

What the Working Group has overlooked

1) Most of the headings—as in this example—throughout the entirety of LCSH are multiple word strings. They are not individual words; they are precoordinated strings right from the start.

2) It is the browse-menu of strings that enables researchers to recognize aspects of their topic that they cannot specify in advance, and that are not "caught" by the cross-reference structure. The larger the collection being dealt with, the more researchers need exactly such recognition mechanisms—menus of options that enable them to pick out search alternatives that would never occur to them on their own. In this example, typing **Afghanistan** alone would not bring up any cross-references to the hundreds of "off to the side" subdivision-aspects listed in the menu above—i.e., the latter are revealed *only* by the *browse-menu* of the precoordinated strings appearing underneath each other, not by Broader Term, Related Term, or Narrower Term (BT, RT, or NT) links. There are 478 subdivision-strings of the topic **Afghanistan** listed in the full browse display in LC's catalog; and this roster is immediately followed by **Afghans**, with a further 51 subdivision-strings of its own. (Most researchers would not think of **Afghans** as a separate heading—until they see it presented as a recognizable option in a browse-menu.)

The display of these "side" relationships is, for retrieval purposes, a *structural* element of LCSH that is just as important as the BT, RT, and NT links themselves. It is only by these menus that tens of thousands of headings having free-floating subdivisions are controlled—i.e., while catalogers, at the input stage, may know the rules for assigning subdivisions, researchers at the output stage, who do not know the rules, can find the relevant subdivisions in a systematic manner only by recognition of their existence and contextual placement within menus such as the above.

Such an array of hundreds of aspects of **Afghanistan** cannot be comparably matched by a word cloud, even if type-size differences mirror frequency of term use. Word clouds having more than about twenty elements are very hard to take in—but readers do not have comparable problems with alphabetical subdivision displays in a single vertical list. (In a real OPAC browse menu, unlike the above example, the number of "hits" for each string is specified, so researchers can also gain, immediately, a precise knowledge of how many records will be brought up by each string.) Even though, in the history of library science, browse-menus pre-date word clouds and relevance-ranked displays, they are nonetheless far superior to either in conveying overview information and in allowing simple recognition of complex wholes.

Indeed, if a topical LCSH heading such as, say, **Monasteries**, is subdivided geographically, readers immediately and intuitively recognize from the early part of the alphabetical array (**--Albania**, **--Armenia**, etc.) that country subdivisions are present, and

that they will have to scroll down to **–United States** if that is what they really want. The recognition of the need for such specificity is immediately "forced" on their awareness, even if they were not looking for it. (In a facet-display, the first step in the recognition-awareness of options is lost—researchers must actively seek geographic subdivision information whose existence is not *immediately* apparent, and whose importance in producing the best level of specificity for their inquiries is also not immediately obvious.)

A major weakness of word clouds is that they cannot show cross-references, scope notes, or further subdivisions of their own terms. *This is not to deny the utility of such displays for other purposes*; but we must remain clear about the differences between catalog search environments and Web search environments. The former are much more conducive to scholarly "overview-provision," as above. I have had the experience of showing thousands of researchers the existence of such browse-menus connected to their research topics; and not only do they *not* complain that the displays are too long or "too hard," they usually thank me for "opening up" their topics so that they can see a full array of aspects whose vocabulary designations (form, extent, and specificity) they could never have thought up in advance.

3) Such an array as the above cannot be displayed nearly as efficiently by "faceting" the subdivisions into separate silos (geographical, topical, form, chronological) that have to be searched apart from each other.

In the first place, breaking up the strings into separate facets *causes all of the cross-references throughout the entirety of LCSH to vanish*. The Working Group has utterly overlooked this crucial fact. The very elaborate and detailed network of references in LCSH does not merely link individual words together—it links tens of thousands of formally-established *precoordinated strings* to each other. In this example, will the linguist who clicks on a separate facet for **Languages** realize *in the absence of cross-references that are dependent on the precoordinated string* **Afghanistan—Languages** that there are entirely separate categories (above) defined for a dozen specific Afghan languages and dialects? (Even without a statistical study, I will venture an answer: No. She will miss both the existence of the list and the overview of its range of inclusion.)

These particular NT terms (**Afshar dialect**, **Bashgali language**, etc.) could not be *positioned* to begin with under **–Languages** and **–Literatures** if those terms were not subdivisions themselves, *formally linked-by-precoordination* to **Afghanistan**. Without precoordination, such NT cross-references cannot be intelligibly located in the overall scheme—i.e., if the name of every individual language, worldwide, were simply cross-referenced alphabetically under the individual facet-term **Languages** in general (lacking precoordination to a specific country or group), no one would be able to see the language-names *in limited clusters* that effectively point out their geographical and cultural affinities, unencumbered by hundreds of irrelevant and indiscriminate juxtapositions to all other languages worldwide. The precise linkages brought about by the browse-menus and cross-references convey information that is very important to scholarship—*relational information that is not conveyed by the individual terms themselves*. This is precisely what would be lost, entirely, if the naïve recommendations of the Working Group were

followed, in opening up LCSH to "non-library stakeholders," and in destroying precoordination in favor of using only individual terms or facets rather than conceptually-defined strings.

The Working Group is apparently ignorant of this reality. Let me propose an (imperfect) analogy to a crossword puzzle: the Group proposes opening up LSCH to "non-library stakeholders" (Recommendation 4.3.1.2), re-enforcing its previous call to accept data indiscriminately from the entire "supply chain"—i.e., "data from others (e.g., publishers, foreign libraries) that do not conform precisely to U.S. library standards" (Rec. 1.1.1.1). Given *the network of interrelationships* of LCSH elements to each other (cf. above example), this would be much like the editors of the *New York Times* crossword puzzle accepting outside "contributions" to their "Across" word column, regardless of the horizontal length of the proposed terms, and equally regardless of whether or not the terms can be integrated in relationships to any of the vertical words. Such openness would indeed make the system more "democratic"—but it would also destroy the system itself, which is constituted by much more than just the individual words by themselves. *Without the relationships being formally defined and presented for inspection, there is no system.* Without contributions that do in fact "conform precisely to U.S. library standards" there is no LCSH—there is only a powder of disconnected fragments and facets that are related to each other only haphazardly, non-contextually, and non-*systematically*.

The Working Group seems also not to know that many previous proposals to "facetize" LCSH have been repeatedly discussed and rejected, by the Airlee House Conference (1990) and the Bicentennial Conference on Bibliographic Control for the New Millenium (2001)—the papers of the latter conference spelling out in great detail the many substantive *reasons* for not going in that direction. Yet another major study of subject cataloging *explicitly recommends the maintenance of left-anchored browse displays of LC subject strings*. It is "Recommendations for Providing Access to, Display of Navigation within and among, and Modifications of Existing Practice Regarding Subject Reference Structures in Automated Systems," (*LRTS* 49 [2005], 154-66), from The Association for Library Collections and Technical Services (ALCTS) Cataloging and Classification Section Subject Analysis Committee (SAC). This study is the product of nearly ten years' work by *three* SAC subcommittees, charged (in its own words) with "investigating the theoretical, pragmatic, and political dimensions of improving subject access through better use of reference structure data." And a fourth major consideration of precoordination, LC's own internal study on "Library of Congress Subject Headings: Pre- vs. Post-Coordination and Related Issues"—done in March of 2007—also strongly recommends against breaking up the LCSH strings into individual facets. Given the Group's close contacts to LC, it could easily have obtained the internal study. (It is now available at < <http://www.loc.gov/catdir/cpso/pre_vs_post.pdf>.)

The Working Group is oblivious to any of these studies; none of them are referenced in their text or even listed in the *Report*'s bibliography. Is there any wonder that one senses an agenda being pushed, rather than even-handed inquiry into what cataloging methods actually work best? (It is remarkable, by the way, that much of the research on

facetization has been funded all along by OCLC—whose own WorldCat *cannot display either cross-references or browse-menus of precoordinated terms*. Why, however, should the rest of us naïvely accept OCLC's oversimplified software to begin with, in our own OPACs, especially when it directly undermines the ability of scholars to perceive *overviews* of relevant book literature?)

Perhaps this needs to be said explicitly: the Library of Congress is not a subsidiary of OCLC. Nor is LC a subsidiary of Google, whose own software also cannot display even simple alphabetical lists of subject headings, let alone cross-references or browse-menus of strings. LC's primary purpose is not to be a "feeder" mechanism for either external organization. The *first* responsibility of LC is to catalog *its own*—and the nation's—*unique* copyright-deposit collection, a collection *also* unique in the breadth of its overseas acquisitions, and *thereby to make the necessary webs of interconnections* that are needed for efficient retrieval, and that no other library is in a position to see.

4) Breaking up the strings also causes all of the tens of thousands of links between LCSH terms and LCC class numbers to vanish because the class designations, too, are linked not simply to individual words but to precoordinated *strings* of terms. Note, even in the one very small roster above, how many LCC class numbers are formally linked to various different LCSH strings or phrases. It would seem obvious that you cannot give *one* classification number to everything written on **Afghanistan**; the topic has too many aspects that need to be distinguished from each other. But the separation, and definition, of these aspects is brought about precisely by the precoordinated combination of Afghanistan with a second term—e.g.:

> **Afghanistan—Antiquities** [DS353]
> **Afghanistan—Description and travel** [DS352]
> **Afghanistan—History** [DS355-DS371.43]
> **Afghanistan—History—Soviet occupation** [DS371.2].

Without the combination being formally defined *as* a combination, you could not create a separate class number, in each case, different from the number given to most general works on **Afghanistan**. A postcoordinate combination of heading + facet cannot generate a class number without human (cataloger) intervention.

The need for such linkages points out one of the major differences between LCSH and conventional thesauri—the latter do not have to link their various verbal headings to any classification numbers. They can provide single-word terms in part because they, unlike LCSH, do not need *also* to define classification numbers that embody a relationship of ideas rather than a single concept. LCSH does not "reduce" to a conventional, single-discipline thesaurus; it, unlike other controlled lists, has to define relationships among different subjects (both verbal LCSH and alpha-numerical LCC) *across* all areas of knowledge, not merely within narrow slices of the subject-universe. And without precoordination of linked concepts this cannot be done even nearly as efficiently as it can be *with* precoordination.

Although LSCH is permeated from beginning to end with tens of thousands of such links to LCC, the Working Group seems to be blithely ignorant of this fact. (Indeed the range of linkages between LCSH and LCC is even more extensive than is formally written down: it has long been cataloging practice to link the first LCSH term assigned to any book to the LCC class for it; and thus an enormous body of "past practice" is readily available for inspection within the catalog itself, even apart from formally-authorized links.) The Group simultaneously recommends, however, "Provide LCSH openly for use by library and non-library stakeholders" (4.3.1.2) *and* "Increase explicit correlation and referencing between LCSH and LCC and Dewey Decimal Classification (DDC) numbers" (4.3.1.4). The first recommendation, if it is interpreted as opening the system to "contributions" from non-library contributors, *would destroy the strings and cross-references*, because non-library users—coincidentally, much like the Working Group itself—do not perceive to begin with the overview-networks and webs of relationships into which their "data" [read: individual words] must be integrated. Non-library contributors don't see the "crossword puzzle" relational requirements. And the second recommendation cannot be accomplished at all *without* the strings and cross-references.

5) Breaking up an easily-browsable *single* roster of 478 topics under **Afghanistan** into subcategories of Topic, Time period, and Form "facets" would require much more pointing and clicking to *see in its entirety* (i.e., to gain an overview of the shape of "the whole elephant") than simply skimming a single alphabetical/vertical list, large sections of which can be passed over very quickly. Such a (literally) disjointed facet-display violates the Principle of Least Effort in a way that will hide the overall presentation of relationships by requiring readers to do much more work (in clicking and backtracking) to reconstruct conceptual connections that are crucial to retrieval at the best level of specificity (whole books rather than granular pages) for their topic, and that should not have been severed to begin with.

Specifically, breaking up the strings into facets not only confuses the important relationships of the topic-period-form facets to **Afghanistan** *but also severs their relationships to each other*. The extent of "relational" data within LCSH, in other words, is not confined to cross-reference and browse-menus that show connections among different strings; there is additional important relational data embodied *within the individual strings themselves*—information that is lost without the precoodination.

For example, consider the string **Afghanistan—Defenses—History—20th Century—Sources**. A reader who can see only the facet **Sources** in a separate "Form" silo (adjacent to **Afghanistan**), disconnected from its *context in the string* may click on it because she is looking for primary sources on "kings in Afghanistan"—but only when she goes through further clicking will she see that **Sources**, in its relationship *here* with **Afghanistan**, has nothing to do with kings, at which point she will have to backtrack and try again. Indeed, she would have to click on *every* Form and Time Period facet, in turn, to see if the result is connected to the sub-topic **Kings and rulers** because the "faceted" Form and Time elements are connected only to **Afghanistan** by itself, not *contextually* to **Afghanistan—Kings and rulers**.

In contrast, inspecting a single browse display would enable her to immediately recognize which Form and Time-period designations are attached to (*or absent from*) which Topical facets—without endless clicking back and forth. *Faceting undercuts rather than enhances simple recognition capabilities when compared with browse-menus of precoordinated strings.* While faceting would indeed simplify the assignment of individual words at the cataloger/input stage, it destroys the simple recognition-capabilities that are necessary at the user/output stage. Contrary to the widely-touted mantra, facetization does not "make the data work harder"; it makes the *user* work harder, forcing her to reconstruct post-coordinately relationships which she otherwise could simply have recognized immediately, in a *single* initial "pass" through the browse-menu system. Again, it is a stunning violation of the Principle of Least Effort in information-seeking behavior. "Least effort" is supposed to refer to the level of work done by the users, *not the catalogers*. "Least effort" on the users' part *means* "most effort" on *our* part. The better the browse-menus we create, providing the most contextual information and conceptual linkages, the less work the researcher has to do to gain an overview of "the shape of the elephant." The less work *we* do, however, the more pointing and clicking and backtracking among separate "silos" the *readers* have to do.

(By the way, it is nonsense to assert that users can achieve through post-coordinate Boolean combinations the same results that they can achieve through recognition of precoordinate strings, especially those having multiple subdivisions. It takes some actual experience with real readers, outside academic ivory towers, to know why this is so: the reality is that *it will never occur* to users to think of anything even close to the range of 470+ terms to put *into* any Boolean combinations with **Afghanistan**. For example, will the historian who combines **Afghanistan** AND **History** post-coordinately realize that she is missing scores of other elements such as **Antiquities** (with numerous cross-references of its own), **Bibliography**, **Biography**, **Chronology**, **Commerce**, **Civilization**, **Description and Travel**, **Encyclopedias**, **Ethnic relations**, **Foreign relations**, **Military relations**, etc., etc., etc.—*all* of which may well be of interest to someone studying the history of the country? Will the same historian who is indeed interested in such particular (although unexpected) aspects of the subject also be able to think of them in all of their contextual relationships with each other, as in **Afghanistan—Foreign relations—India—Sources—Bibliography—Catalogs**? The answer again is "No"—even if there is no statistical study to verify the obvious fact.

6) This kind of **Afghanistan** overview menu, once broken up, cannot be reconstituted by "under the hood" query expansion. *Nor should it be*—who on earth would want every BT, RT, and NT cross-reference (and the further cross-references generated by those terms themselves), *and* every "off to the side" aspect of Afghanistan (as shown in the browse-display subdivisions) to be included automatically by a query-expansion of the heading **Afghanistan** by itself? The *explicit seams* among both cross-references and browse-displayed subdivision-aspects are of much more use to researchers in enabling them to recognize what they really want—*segregated from what they do not want*—than any "black box" operations working "under the hood," which cannot be scrutinized to see what, exactly, they are including, or *failing to exclude*, in their query expansion.

7) Neither the cross-reference links nor the browse-display menus can be reconstituted by adding "democratic tags" to individual bibliographic records. Why not? Because readers attach their tags *only to the item they have in hand*—they do not add cross-references (e.g. **Afghanistan—Antiquities** linked to NT **Atishkadah-I Surkh Kutal Site**) or browse-menus of "off to the side" terms (**--Bibliography**, **--Encyclopedias**, **--Historical geography**, **--History—Chronology**, etc, etc.). *Relationship data such as exemplified above cannot come from people who are tagging only individual items they have in front of them*—especially when they are not in a position to see even a tiny fraction of the works available to LC catalogers, in 450+ languages from all over the world. Yes, *some* readers will indeed add comments or notes mentioning "outside" connections—but scholarly researchers will not be able to rely on such connections being made *systematically and predictably* within a Web environment.

8) The works retrieved under each of the above subdivisions will usually be *whole books* on their topics—or at least books having a substantive portion of their texts devoted to the subject. In other words, such "scope match" headings *solve* the growing problem of full-text retrievals that are much too granular. LCSH strings do not turn up hundreds or thousands of books that simply have relevant keywords somewhere near each other on the same page, while the rest of the book's content is irrelevant. There are indeed times when a researcher wants to know what is in the fifth paragraph on page 237; but *when he is trying initially to get an overview* of the most relevant works on his topic, he does not want his retrieval cluttered with every possible mention of the specified words, no matter how small the reference, in irrelevant texts whose presence conceals the most important *whole books* by burying them within huge retrieval sets. Anyone who has ever done a Google search knows that Google's search mechanisms exacerbate rather than solve this problem. LCSH—even though it was created before computers—*solves problems of information overload that are now created and aggravated by computer and Web-environment retrievals.*

Why *control* continues to be necessary

Let's be clear on why we strive, in the first place, for the control provided by uniform headings: it is because the scholarly world has known for centuries that authors do not use the same terms to refer to the same subjects, even within the English language, let alone across the 450 other languages LC also must collect for Congress and the American people. The problem of keyword variations (across multiple languages) is best solved by collocating all of the variants under a single *uniform* (i.e., standardized) *heading*, which is *added to* each work, rather than transcribed from it. This cataloger-added uniform heading thus becomes a *retrievable term held in common by all of the works*, so that any researcher who finds this term alone can thereby retrieve all of the works to which it is attached, whose own vocabularies display widely variant keywords for the same concept. "Vocabulary control" is a rock-bottom principle of library science, and nothing in the last generation of computer developments has invalidated it—no matter how counter-fashionable it may be to say so in "blogland." A sample of the latter commentary is provided by one recent blog response to the Working Group' *Report*, which says:

"Every time I hear someone talking about "controlling" bibliographic data, I chuckle, a low throaty laugh intended to convey my disbelief that anyone thinks we will still be controlling anything in fifty years. . . . Many of us in LibraryLand worry that we're just one black swan away from "game over," but not the muckety-mucks of cataloging. They [are] needily [sic] grounded in beliefs and practices the rest of us see as not only foolish and outdated, but pernicious." [cf. Google Blogs]

What is present in this remark is a superficial, shot-from-the hip, emotional expression of personal distaste; what is conspicuously lacking is any argumentation, evidence, examples, or concrete experience to back it up. "Preaching to the choir" may be a common practice in blogland; but such intellectual vacuity is no substitute for an actual understanding of what LCSH and LCC accomplish that tags, folksonomies, and relevance-ranking do not. (Indeed, it is likely that the present review itself will be vacuously dismissed as a "rant," an "amusing" paper, or a mere call to "maintain the status quo," by those in LibraryLand who are incapable of writing a substantive response to it. [It will be evident from this paper's concluding sections, however, that I am endorsing a plan that takes us quite a way beyond the status quo.])

When we have the LCSH system *including its cross-references and browse-menus of precoordinated subdivisions* to provide entry into book literature, we effectively have at a minimum the means whereby anyone who uses the system can find, systematically and efficiently, the full range of books on any subject, in any language, from any time period—in or out of print—in the largest library that has ever existed anywhere on earth in all of human history. (This is no small feat by itself, even apart from any use of the same system by hundreds of other research libraries to provide similar access to their own holdings, and to network catalogs together.)

The genius of LCSH's *control* is that it gives us systematic pathways to gain a reasonably comprehensive overview of the full range of book literature on any topic, even though we may not have any prior subject expertise in the subject to be researched, may know nothing in advance of its vocabulary (in multiple languages), its component parts, or its relationships to other topics—narrower, related, broader, or tangential:

- Using LCSH, we do not have to think up all relevant keyword synonyms or variant phrasings of the same idea, even in English (let alone across all other languages simultaneously).

- We do not have to specify in advance, precisely, all keywords relevant to our topic because we can predictably rely on the system to show us more than we know how to ask for. The system enables us to recognize, systematically and predictably, what we cannot specify as we enter any new subject territory.

- We will be able to find, immediately, *whole books* on our topic, rather than excessively-granular retrievals of thousands of full-texts that simply have some of the words we want near each other at the page- or paragraph-level, in

irrelevant contexts. We will not be overwhelmed by "information overload"—let alone routinely overwhelmed.

- We will be able to view records for the book literature *within conceptual boundaries* that assure us that any keywords we want will appear within contexts that are truly relevant to our interest.

- We will have multiple menus to select terms from—menus of cross-references, scope notes, and browse-displays with the relationships of subject categories spelled out for our inspection and easy recognition, not hidden in "black box" operations under the hood.

- We will have roadmaps of the subject, and its interconnections to other subjects "off to the side," surrounding it and diverging from it, *displayed in a single roster*, that will bring to our attention related areas of concern that we didn't realize we could ask for. We will thereby be enabled to see not just individual "toenails" or "eyelashes" of "the elephant"—we will be able to get a good sense of the overall shape of the animal as a whole, with all of its conceptual aspects *connected and related to each other.* We will not be burdened with a jumble of disconnected individual parts (again, at excessively granular levels) that give us no indication of what important aspects or offshoots of the subject we may have entirely overlooked in our initial query.

Further, the genius of the *control* of the Library of Congress *Classification* (i.e., LCC as opposed to LCSH) is that it enables whole books relevant to a particular topic to be browsed systematically, within limited conceptual boundaries defined by the classes, such that any discovery of the right words will probably appear within the desired context; and further, researchers browsing down to the page and paragraph levels of classified books will be able to *recognize* relevant words in a variety of languages within works conveniently shelved proximately, as well as illustrations, maps, charts, tables, running heads, sidebars, typographical variations for emphasis, bulleted or numbered lists, footnotes, bibliographies, book-thicknesses, and binding conditions—any of which data may provide the key to researchers' discovery of useful information that they did not know how to ask for via a blank search box. (See the University of Chicago study, cited below.)

The difference between encouraging use of LCSH by non-librarians in Web environments vs. allowing them to undermine its structures of relationship in catalog environments

Some members of the Working Group might immediately respond that "Obviously the system outlined by Dr. Mann (the **Afghanistan** example above) is much too complex to be maintained by non-librarians, or in Web contexts generally." Further, they (and others) might sensibly point out that neither LCSH nor LCC can be "scaled up" to deal with billions of Internet sites. And I fully agree with both observations. But they would apparently conclude, "Therefore, abandon the complexity." I would conclude, in contrast,

"Therefore, do not allow access to its creation or maintenance by people ('non-library stakeholders') who don't know what they are doing, and who are utterly oblivious to the utility of the tens of thousands of cross-references, browse-menu displays, and LCC linkages that are integral to the system's operation within library-catalog environments. And also do not even try to have professional catalogers apply LCSH systematically to all Web sites. By all means, however, do allow non-librarians to take and apply ready-made LCSH headings *wherever they want in Web environments*; but do not allow them to 'contribute' uncontrolled terms of their own that won't fit into the 'crossword puzzle' relationships defined by professional catalogers, and that are necessary for searching *within the catalog environments* that, alone, can exploit the full retrieval power of *all* of the data."

Indeed, work is proceeding already along the line of allowing greater export and use of LCSH *outside* library catalogs (as in a Simple Knowledge Organization Schema SKOS/RDF application in Word); but encouraging this desirable outcome is quite different from 'going in the other direction' and allowing non-librarians to add their own terms to LCSH *within* library catalog environments. The distinction is that *tagging* with LCSH is not the same thing as *cataloging* with LCSH—i.e., adding LCSH headings to sites or pages *within Web environments* will necessarily sever the linkages of the headings to each other via cross-references and browse-menus, which the Web environments cannot display. Such application of disconnected individual LCSH terms to Web sites is therefore not the equivalent of employing the same terms within online library catalogs, which do display the interconnections and relationships of the headings to each other.

A direct comparison of LCSH cataloging results to Web 2.0 tagging results

It is undeniably true that the LCSH system is complex—*but so is the literature of the entire world, on all subjects and in all languages, and from all time periods, that it has to categorize, standardize, and inter-relate.* You cannot "reduce" an overview of the complex structure of the literature on **Afghanistan,** or on any other substantive topic, to an Internet display of "relevance ranked" keywords or word clouds. One need only try the simple experiment of typing "Afghanistan" into the "Tags" search box of LibraryThing (http://www.librarything.com/search) to see the paucity of its Web-based overview-provision mechanisms in comparison to the browse-menu, above, from LC's online catalog. (Note that LibraryThing's home page says that it covers "Over twenty-three million books on members' bookshelves"; LC's most recent *Annual Report* [2006] lists its Classified Collections as having 20,492,006 volumes.) The Web 2.0 access mechanisms of the commercial site are wonderful supplements to LC cataloging—but would be utterly inadequate replacements for it. I can only ask readers to do what the Working Group has failed to do: *Compare the results.* Which system provides a better *overview* of the literature? Which one maps out the entirety of "the elephant," with all of its parts and their interconnections, more understandably?

The extreme—and *growing*—complexity of the world's book literature is a rock-bottom reality that will not vanish simply because neither the Working Group nor LC

management wishes to pay for professional catalogers with the subject expertise to manage it. A complex situation "on the ground," in the real world of scholarship, does not lend itself to any simplistically-elegant theoretical solution. The complexity of the material to be managed requires high-level professional *thinking*, not just computer algorithms and non-standardized tags (*lacking relational ties*) contributed by anyone with access to the Internet.

While many today would facilely open up access-mechanisms to "the wisdom of the collective mind," the same theorists seem not to notice that LCSH itself *already embodies a truly amazing collective wisdom*, created by many hundreds of professional catalogers over more than a century—professionals *with both subject and language expertise*, many of whom have been uniquely in a position to see much of the literature of the entire world (not just the volumes on their own bookshelves at home). The fact that their knowledge is embedded in an easily-teachable system (cf. http://www.loc.gov/rr/main/research/) means that the rest of us don't have to have those levels of knowledge—we can simply *recognize systematically* the full range of options and relationships that the experts have already mapped out for us.

I invite anyone who believes that democratic tagging, folksonomies, faceting, or Web 2.0 indexing mechanisms will provide a comparably adequate *overview* of the book literature on Afghanistan to provide an actual example of what they are talking about, such that a *comparison of results* is possible. I especially invite the Working Group itself to provide such an example. Word clouds cannot provide such systematic overviews; Endeca's facet-system (which *cannot display any cross-references*, and which also effectively buries browse-menus) cannot do it; EBSCO's "grokking" displays cannot do it; Amazon cannot do it; Google Books cannot do it. But *this systematic overview-mapping still needs to be done* if American scholars are to gain systematic subject access to the books of the entire world.

I submit, further, that scholars still need the capacity to browse books (in multiple languages) shelved in the same subject categorizations within library bookstacks; and they continue to insist on the importance of this shelf-browsing capability even though both the Working Group itself and many library administrators do not wish to hear them. See especially, the excellent planning study done in May, 2006, by Andrew Abbott for the University of Chicago's Regenstein Library at < http://www.lib.uchicago.edu/staffweb/groups/space/abbott-report.html >, especially Secton VI.A.

The necessity of distinctive leadership by the Library of Congress

The Working Group goes off the tracks, not only in its misunderstandings of cataloging vs. Web 2.0 inputs, but yet again in its denigration of the unique role of the Library of Congress in the nation's overall system of bibliographic control. It is an unavoidable reality that the taxpayers of this country, and their Congressional representatives, have made LC indisputably the "alpha" library of the whole world—no matter how jealous or

resentful of that situation some observers may be. This reality is indicated by three abiding facts:

> 1) No other library receives this nation's Copyright submissions (even the medical and agricultural works that go to the NLM and the NAL come *through* LC).

> 2) No other library has anything close to LC's resources for overseas acquisitions across the board, from all countries and in all languages.

> 3) No other library is regularly and predictably funded by every taxpayer in the entire country at any comparable level.

The fact that we can integrate all of our huge General Collection book material in to a single retrieval system, crafted in such a way that researchers can *systematically* (not haphazardly) find resources within it *that they cannot specify in advance* is one of the glories of the entire history of world scholarship, even if members of the Working Group are unfamiliar with the operation of the system. In his testimony before Congress of May 7, 1996, Librarian of Congress James Billington said the following:

> Knowledge and information services to the U.S. Congress and central government cannot be made to depend on the acquisitions, deaccessioning, and access policies of other less publicly accountable repositories whose basic commitment must necessarily be to the shifting priorities of their own more limited constituencies. A substantial, universal collection in one location is more cost-effective for the American people than the expensive and labor-intensive process of a decentralized, coordinated collection. Other libraries would be unable to sustain the high cost of accepting permanent national-level responsibility for housing and servicing collections of the caliber and scope of those at the Library of Congress—without far greater total federal subsidies than the national library currently receives.

> The American library community depends increasingly on the Library of Congress to maintain national scholarly resources for the study of other nations. The 1996 report of the Association of Research Libraries affirms the key role of the Library of Congress "in building comprehensive collections of global resources," and notes that in most libraries "cutbacks in foreign acquisitions are driven by local demands, with little consideration of the effects on the entire North American system for [acquiring] highly specialized global resources." Other libraries are freed up to develop their own collections more selectively and economically by the assurance that they can utilize the Library of Congress as the "library of last resort."

> The assumption that electronic networks will include, or have easy access to, all the material that Congress and the nation will need in the future is almost certainly wrong. Most past knowledge and much future knowledge—particularly

in countries where crises are likely to arise—will continue to be available only in hard copy for many years to come. . . . The Library of Congress has long exerted strong and unique leadership in library classification primarily and precisely because the professionals here have intimate contact with the collections.

Is this testimony from a decade ago now irrelevant? Is the work that our catalogers do with our own book collections less important now than it was then? I think not: the 1996 *Annual Report* of the Librarian records that 221,991 new volumes were added to LC's General Collections in that year; in 2006, our *Annual Report* records the addition of 445,545 new physical volumes—a figure more than twice as large. More than a thousand new books, from all over the world, are being added to LC's General Collections every working day. The world's *books* have not stopped coming in—there are more of them now than ever before. And only one Library in the world is in a position to see so many of them *in relation to each other*; only one Library has—in numbers that are staggering every year—this range of direct and "intimate contact" with such an amazing variety of resources from, literally, the entire globe. We still need first and foremost to be able to *find* all of *these* books in *this* collection, especially when researchers cannot specify in advance which particular ones are most relevant to their topics. We need, *more than ever before*, the *system* that enables us to *recognize* works that are *conceptually* relevant to topics, works whose keywords cannot be specified in advance, even in English let alone simultaneously in hundreds more languages.

Dr. Billington also had this to say in his Congressional testimony of March 20, 1997:

> Jefferson's ideals of a "universal" collection, and of sharing knowledge as widely as possible still guide the Library. With Congressional blessing, it has grown to serve Congress and the nation—largely as a result of four milestone laws: (1) the copyright law of 1870, which stipulates that two copies of every book, pamphlet, map, print, photograph, and piece of music registered for copyright in the United States be deposited in the Library [N.B: not at OCLC in Ohio]; … (3) the 1902 law which authorized the Library to sell its cataloging records inexpensively to the nation's libraries and thus massively subsidize the entire American library system; ….
>
> We already provide … a little-known subsidy of some $268 million worth of annual cataloging services to the nation's entire library system.

It is worth noting that in 2007 LC's current Director for Acquisitions and Bibliographic Control, Beacher Wiggins, "assessed the costs of cataloging at $44 million per year" (*Library Journal*, 15 August 2007, vol. 132, no. 13). It would seem that maintaining LC as the "alpha library" that it necessarily must be—given it *unique* acquisitions level based on widespread taxpayer support—saves those same taxpayers over $200 million dollars a year in what would otherwise have to be assumed as local expenditures. (Indeed, the savings to the nation today are likely to be greater, rather than lesser, than the 1997 figures indicate, since so many local libraries have had to absorb so many cuts to their

own budgets, and so many rely increasingly on cataloging copy supplied by other institutions.)

The Working Group's emphasis on the lack of explicit legislation designating LC as the "national library" coincidentally mirrors the (new) view of LC management itself, that such lack of statutory labeling is now to be regarded as a justification for "service shedding" and "divestment" of LC's cataloging leadership responsibilities. And yet it can reasonably be argued—by both the American Library Association and its separate Washington Office, one fervently hopes—that LC's longstanding and unquestioned assumption of these duties for more than a century, and the Library's frequently-expressed pride in discharging them, gives these functions a kind of "common law" legitimacy—i.e., that such a long-settled arrangement creates a *reasonable expectation* on the part of LC's "common law marriage" partners in all Congressional Districts that the relationship will not now suddenly be abrogated simply because no formal "double ring ceremony" was performed in 1902. Indeed, the arrangement has survived various periods (two World Wars and a Great Depression) that were much more stressful than today's economy is to the nation's library system.

The matter of priorities at the Library of Congress: maintaining LCSH and LCC vs. digitizing special collections

The leadership that LC has assumed in creating and maintaining the constantly expanding LCSH and LCC systems has resulted in an astonishing contribution to scholarly research, especially in contrast to the palpable inadequacies of Web 2.0 mechanisms for sorting, filtering, and presenting systematic overview perspectives on huge collections of material. Should these book-cataloging operations, so demonstrably useful to libraries in every Congressional District, now—and suddenly—be given a lower priority in funding at LC than the digitization of its special collections? Answer: No. Why not?

The continued provision (and expansion) of this system is *much more important to scholarship overall*, in all subjects, all time periods, and all languages, than is the digitization of any locally-held and narrowly-focused special collections. *Anyone* doing scholarly research, worldwide, can profit from using the LCSH system—even if he or she has access only to LC's online catalog—in identifying the range of books relevant to any topic. In contrast, only a few highly specialized scholars will profit from increased access to *most* of our specialized collections in non-book formats. Obviously Prints & Photos is an exception: there is an endless demand for visual images in all subject areas, and what we might call "special" collections in this realm do not have the subject or time-period limitations that attach to most manuscript collections. I realize that putting *any* collection at all on the Internet will likely generate thousands of "hits" on it; but a mere count of hits will not indicate whether researchers actually profited from it, were disappointed by it, or regarded it as a cluttering presence getting in the way of what they really wanted. For example, LC has many *hundreds* of manuscript collections such as these:

Clarke, Frank Wigglesworth – diaries – geologist chemist

Elkisch, Paula – collection – 1924//1949 – psychiatrist, consultant
Harth-Terre, Emilio – collection – engineer
Keller, Louis – soldier
McCormick, Lunde Dupuy – 1895-1956 –naval officer
Peattie, Donald Culross, 1896-1964 – naturalist-writer
Shead Family – papers -1863/1872
Swidler, Joseph Charles, 1907- lawyer
Wells, David Ames, 1828-1898 – economist – public official

Undoubtedly every such collection is of interest to someone, no matter how apparently obscure it may appear to a general observer—I have been a reference librarian too long to believe that obscure resources will always remain unused. And while I think all such collections should be processed, with finding aids created and put online, I do not think that, *as a rule*, the *digitization* of this kind of voluminous "special" material can be justified as *more impor*tant than LC's maintenance and continued application of its unmatched cataloging and classification system, which has a utility much broader and deeper in its effect on the entire scholarly world.

The Working Group is off-track in changing these priorities (digitization of special collections over maintenance of LCSH/LCC) because it evidently does not grasp, to begin with, what LC cataloging can do, that cannot be matched by folksonomies, democratic tags, and algorithmic rankings of keywords. All the king's horses and all the king's men cannot re-assemble the conceptual wholes, or the relationships among them, that are created by vocabulary control mechanisms once they are lost. Algorithmic "relevance ranking" cannot do it because ranking is simply not the same thing as conceptual categorization.

Let us not be misled by the term "relevance" here. What is going on in such "relevance" determinations is merely keyword-weighting—and if the wrong keywords are typed in to begin with, then all of the massaging of their display-order by computer algorithms will not re-create the lost conceptual groupings, and their interconnections, that need to be there instead if overviews are to be achieved. Algorithms will not find in any systematic manner the full range of words (in hundreds of languages) that the researcher fails to key in to begin with; nor will they exclude the appearance of thousands of records having the right words in the wrong contexts.

Vocabulary control brought about by professional catalogers knowledgeable about both LCSH and particular subject areas *will* bring together disparate phrasings for the same subject, and *will* exclude mountains of irrelevancies at the same time. The serious and persistent problems created by a *lack* of vocabulary control continue to need just the solution that this control provides—regardless of the fact that the principles of control were arrived at during a time before computer technologies. And solving *these* problem continues to be much more important to scholarship in general than the digitizing of its special collections by LC itself. Indeed, when so many commercial companies are eager to digitize such a variety of collections, why should taxpayers, during a time of increasing federal deficits, have to pay for any digitization projects when private companies are

already so heavily involved in such projects? But no private money will maintain the LCSH and LCC cataloging systems.

LC's managerial attitude toward its own cataloging system

I strongly recommend the very opposite of what the Working Group proposes in this area. It is much more important to the overall scholarly community in this country for LC to divert funds *away* from digitizing special collections and *into* the expansion of its cataloging operations, restoring our traditional requirements for both language and subject expertise. LC's current thinking—demonstrably incorrect, as exemplified by the **Afghanistan** example—is that any cataloger should be able to provide subject cataloging in any area. This is the assumption behind LC's pending reorganization of its entire cataloging operation. The apparent managerial belief at high levels within LC (already present, even before the *Report* of the Working Group) is that "de-coupling" LCSH strings into individual words will eliminate the need for subject expertise, because *anybody* can assign *individual words* to catalog records. The loss of cross-references, browse-menus, and links between LCSH and LCC are not regarded as important—indeed, they are apparently *not even noticed* by administrators who do not use the system themselves. Indeed, over the last few years, LC's professional catalogers have been assaulted by a string of outside "experts," called in by management, whose talks have maintained all of the following:

   a) that professional input can be minimized, if not eliminated, by under-the-hood programming;

   b) that catalogs which simply display LCSH words in "faceted" form, without being capable of showing any cross-references, and which bury browse-menus, are the new models to emulate;

   c) that cataloging input from "anywhere"—from vendors, from unreviewed copy derived from any source in OCLC, from democratic tagging—is just as good as professional cataloging (even though it is oblivious of the "crossword puzzle"/syndetic interconnections);

   d) that catalogers themselves should not "agonize" over trying to maintain either the principle of uniform heading, or worry about where (or even if) terms fit in long-established networks and webs of interrelationship;

   e) that speed of processing, rather than accuracy, is now to be regarded as "the gold standard" of quality; and

   f) that since "the perfect is the enemy of the good," catalogers should not even bother to *try* to do their best work.

This is utter nonsense—and nonsense that will seriously undermine scholarly research. Indeed, the blunt fact is that advice not to "agonize" over the quality of one's work flies

directly in the face of both the Code of Ethics of the American Library Association to "provide the highest level of service" and to "strive for excellence," as well as the Eight Values of the Library of Congress, to provide "Service: Best possible service," "Quality: Highest quality in *every* aspect of our activities," and "Excellence: Encouragement and support of staff excellence."

And yet the above assumptions evidently constitute the understanding of many within LC's own cataloging management regarding the workings of its own system. As confirmation, it is more than a little noteworthy that the Library's new *Strategic Plan* for 2008-2013 bends over backwards to avoid even mentioning the word "cataloging" in describing our basic operations and responsibilities—as though LC's proud past is now considered to be an embarrassment by its current administrators (cf. < http://www.loc.gov/about/mission/ >).

It's time someone said this out loud: The real enemy of "the good" is not the perfect, but rather the slipshod, the partial, the unsystematic, the haphazard, the superficial, and the shoddy. No one maintains the "straw man" position that "the perfect" is attainable to begin with. Is it not desirable, however, to have professionals *striving to do their best* rather than *striving to achieve mediocrity*? Is it not better to have professional catalogers striving to provide the necessary *systematic-overview* mechanisms required by scholarship rather than simply to provide "something quickly"—especially if that "something" can just as readily be provided without professional input at all, via algorithms and democratic (as opposed to *professional*) input of any terms at all from "anywhere in the supply chain"?

In line with the above assumptions, LC management has *already* attempted to re-write its cataloger Position Descriptions in a "hybrid" way that both minimizes their need for subject expertise and also burdens them with new and time-consuming acquisitions responsibilities that have never been theirs in the past. This drastic reorganizational change, if it is implemented as planned, will directly undercut cataloging quality throughout the nation's shared networks. Although the Library has technically "announced" the reorganization, it has not spelled out the new "philosophy" behind the move, which, not surprisingly, is basically in accord with points a) through f) above, as well as with the views of the infamous Calhoun Report, which has been highly praised by LC's cataloging management. (It is unclear whether this proposed change is linked to attempts this past year by the Librarian's Office to stifle the voice of the Library of Congress Professional Guild, using tactics that have outraged many in the labor and library communities nationwide.)

Many—perhaps most—of the English-language catalogers will soon be expected to do subject cataloging in most subject areas, i.e., without their having the expertise needed to maintain the complex and topically-specific "crossword"/syndetic relationships within elaborate and extensive subject fields such as **Afghanistan**—or **Art, Business, Education, Mathematics, Music, Shakespeare, Women, United States**, etc. LC management apparently believes that cataloging can also be speeded up, as well as simplified, by dumbing down not just its standards for catalogers' subject expertise, but

also by eviscerating the LCSH system itself to make it more hospitable to the creation of single-word headings rather than precoordinated strings—with all of the attendant (but disregarded) losses to contextual meanings of terms, to cross-references, to browse-menus, and to linkages to LCC. The unarticulated philosophy of the new scheme is that providing "something" quickly, outside of relational and contextual structures, is more important than providing a systematic overview of what the Library has.

This dumbing down of LC's *subject* cataloging operations *across the board* will have even greater impacts on every other library in the country than its previous Series Authority mistake. Again: the work in cataloging is not, as some administrators seem to assume (or wish), simply the assignment of individual-word headings; it is the creation of scope-match strings and the integration of them into multiple, and crucial, networks of relationship defined by cross-reference, browse-menus of precoordinated strings, and links to LCC class numbers—relationships that must be maintained and systematically expanded by subject experts, and which cannot be magically duplicated by the "collective wisdom" of Web 2.0 inputs, or by non-standardized terms "contributed" by anyone anywhere in the entire "supply chain."

The financial difficulties of libraries that are expected to assume more of LC's responsibilities

One paragraph from the Working Group's *Report* stands out, as it is so sensibly inconsistent with the recommendations, elsewhere, that more libraries other than LC itself assume more and greater cataloging responsibilities:

> … [O]ver the past century [many] libraries have not only reduced the number of staff in their cataloging operations, but also have reduced the proportion of staff who are professionally educated to catalog. Cataloging personnel in most libraries are predominantly paraprofessionals whose training often does not include the creation of authoritative name forms, subject analysis, or in-depth description. Thus, when LC makes decisions that have a substantive impact on the flow of authority work or bibliographic records, these libraries are unable to compensate for the loss without the addition or reallocation of resources. The libraries that are most dependent on LC for bibliographic data are often the smallest and least well funded, and are therefore the most vulnerable to any LC cutbacks . . . ." [p. 16]

It is not mentioned by the Working Group—although it is mentioned in the Librarian's testimony, above—that libraries everywhere are subject to unpredictable funding curtailments; and so mere access to OCLC's large pool of resources is not a solution to problems created by LC's new cutbacks and attempts at "divestment" and "service shedding." Indeed, the same "market forces" that are causing libraries other than LC—LC, which has *unique* taxpayer support from the entire nation that shields it from fluctuating market forces—to divest their own operations are the very same forces that will prevent them from assuming greater cataloging responsibilities within the

"distributed," "non-centralized," and "cooperative"—add: "inadequately funded"—network of volunteers envisioned by the Working Group.

The need for LC to expand rather than contract its cataloging operations

In order to make up for the economically-forced divestment of responsibilities taking place in other libraries throughout the country, LC's own internal priorities should therefore be to channel *more* funding into doubling or tripling the size of *its* cataloging staff, especially in hiring professionals (rather than technicians) with subject and language expertise—and in maintaining their Position Descriptions as *catalogers*, rather than as hybrid cataloger-acquisitions workers, without the requirement of having appropriate subject knowledge:

- *That* is where the most good will be done for every Congressional district in the country.
- *That* is the course of action that would bring the greatest approval from LC's oversight committees.
- *That* is the course of action that would bring us the greatest approval from the American Library Association and its Washington Office.
- *That* is the course of action that would be most beneficial to scholarship in this country.

Digitizing LC's special collections simply does not translate into tangible or cost-saving benefits for any Congressional Districts; but providing more high-quality LC cataloging does provide exactly the direct benefits that are most needed by every public or research library in every District. As Dr. Billington's testimony (above) indicates, the cost of centralized work at LC, funded as a public good via nationwide taxpayer support, more than pays for itself in enormous savings to local libraries everywhere. No such "paying for itself" benefit accrues to the digitization of special collections.

Perhaps the Library of Congress should expend less effort in worrying about its corporate "Brand" in the marketplace and more effort in discharging its duties to scholarly researchers whose importance cannot be judged by "market share" calculations; less effort in raising private funds for projects that Congress will not pay for, and more effort in utilizing its public funds effectively for the greatest public good; less effort in providing unexpected extras (e.g., digitizing special collections) and more effort in continuing to provide the essential services that are reasonably expected from it by libraries in every Congressional District.

If LC claims that it does not have enough funding for its cataloging operations, the problem lies not so much in the level of its Congressional support, but in the misguided internal allocation of its funds to operations of lesser import rather than greater import. The Library of Congress needs to spend its limited funds on doing precisely the kind of thing that Google, Amazon, LibraryThing and other Web-based services categorically refuse to do: creating high-quality cataloging and classification metadata, cross-

references, and browse-menus of precoordinated, controlled headings—the mechanisms necessary for providing *systematic overviews* of the book literature of the entire world.

The need for user education in the total system of bibliographic control

There is yet another area where the Working Group goes "off the track," this time in failing to notice the need for an additional necessary element in the overall system of bibliographic control: education of users. This must be provided by reference librarians, working both in direct contact with individual researchers, and in class presentations to larger groups. It is incredibly naïve to think that most remote users—particularly scholars rather than "quick information" seekers—will be able to achieve any systematic overviews of the existing literature of their topics without feedback or prior instruction on the capacities, limits, and scopes of the hundreds of online sources they turn to. Indeed, such feedback may often alert them to vast ranges of valuable sources that are not online to begin with—reference collections, classified bookstacks, published bibliographies, people-contact sources, etc., among them. Let's also not overlook the fact that the Copyright law is never going to be repealed; and nothing short of outright repeal could possibly allow "everything" to be freely available online to remote searchers everywhere. Nor can we overlook the fact, demonstrated by studies of OPAC user logs as well as experienced by reference librarians every day, that most researchers, when left to their own devices, are quite unsophisticated in doing computer searches.

The latter point is relevant to the importance of both cataloging and user education. A specific objection frequently raised in connection with cataloging is the common observation that LCSH "just isn't used by researchers—especially by remote users out of contact with reference librarians—and that they prefer keywords instead." In a sense this is true—but only in a sense. My experience in helping thousands of researchers over three decades is that most of them, in typing uncontrolled keywords, *think* that they are actually asking for a *categorical concept they have in mind*—e.g.., they believe that typing in the keywords "wisdom literature" (without the quotation marks) *includes* what they really want ("Egyptian ethics"); or that entering the keyword phrase "Low countries" *includes* everything on Netherlands, Belgium, and Hainaut (within Belgium) specifically. They think that typing in "Cockney" will bring up all the catalog records on relevant linguistic studies even though they have titles such as *Ideolects in Dickens, Bernard Shaw's Phonetics, The Muvver Tongue,* and *Die Londoner vulgarsprache.* What they "prefer," in other words, is based on a serious misunderstanding of what their "preferred" search technique is actually capable of delivering.

Uninstructed researchers think along these lines because no librarians have ever taught them about the differences between subject headings and keywords; most people fuzzily-assume that keywords by themselves *are* subject-*category* terms because they *want* them to be and they don't know the technical differences between "controlled" vs. "uncontrolled" vocabularies.   The upshot is that the very same researchers "who don't use LCSH" *equally do not know how to do efficient keyword searches.*

The problem that scholars have in gaining an *overview* of the entire "shape of the elephant" of their topic is one of the most difficult tasks they encounter. As professionals, we librarians need to aim at solving the hardest problems that confront researchers, not the child's-play straw-man difficulties of merely providing "something" quickly and remotely, and only in English. It just won't do to leave researchers in the situation of the Six Blind Men of India, who immediately found "something" about the animal, and simply accepted whatever they found quickly to be all that there is to know on the subject.

There is much more involved in overview-provision than can be brought about by library catalogers alone—or by Web 2.0 contributions alone, or by "under the hood programming" combined with federated searching. For a concrete example, I must refer again to my previous "Peloponnesian War" paper (www.guild2910.org)–the task of providing such an overview of relevant literature on "tribute payments" in that war cannot possibly be done by algorithms or under-the-hood-programming—nor can it be done solely in a Web environment to begin with.

Two kinds of user education are required for the proper operation of an overall system of bibliographic control. One is classroom instruction—and such classes need to cover much more than just the question of "how to think critically about Web sites" (cf. "Peloponnesian" paper, pp. 34-38 for a suggested minimum outline of topics to be covered).

The other is point-of-use instruction. The need for reference interviews has not magically vanished just because it doesn't work as well in a Web environment as it does in person. The task is rather complex: we need not only to get people efficiently to the best sources for their topics but also to steer them away from sources that appear attractive, but that will waste their time. We also need to help them formulate their search vocabularies— often not simply in verbal terms (controlled or uncontrolled, and specific to the peculiarities of particular databases) but in coded terms as well (e.g., geographic area codes, industrial classification codes, biosystematic codes, chemical ring structures, etc.). We need, further, to alert them to *format* considerations (encyclopedia articles, literature review articles, bibliographies, chronologies, personal narratives) that they never think of on their own, but that will make their retrievals much more efficient. And, still further, we need to point out the utility of multiple powerful search techniques, beyond the simple typing of uncontrolled keywords into a blank search box—e.g., using browse-menus in OPACs, doing citation searches or related record searches, browsing in the library's classified bookstacks, etc. Under-the-hood programming is no substitute; it fails to alert readers to all of the best options; it clutters their retrievals with out-of-context irrelevancies; and it conceals the most powerful retrieval features of many of the best databases because it reduces all searches to keyword inquiries.

Classroom instruction is needed to convey basic overview information on the diversity not only of sources available, but of search techniques themselves. No matter how good the lectures, however, the fact remains that there is a skill element in doing good research—a kind of "savvy" in mediating between questions (often poorly phrased to start with) and resources. This is neither conveyable in talks nor machine-replicable. But

such skill in "piloting" cannot be regarded as a mere afterthought in the overall system of bibliographic control if we wish to promote real scholarship and to maximize use of our library collections.

And yet the need for user education and guidance (as well as access to libraries-with-walls) is entirely ignored by the Working Group, whose purview of "bibliographic control" cannot see anything that does not appear on a "remote" user's computer screen.

Again, one wonders what *experience* the Group members have in actually using research libraries, especially to find information in subject areas in which they have no prior expertise. Do they really believe that a capacity to find "something" quickly and remotely, and only in English, is sufficient for scholarly research? Do they really think providing *overviews* of the relevant works on a subject—i.e., mapping out "the whole elephant"—can be accomplished solely in a Web environment, with term-weighting of keywords and uncontrolled tagging as adequate replacements for (rather than additions to) cross-referencing and browse-menus? Even on this narrow point itself, it won't do for the Group to simply assert (as they do on page 19 of their *Report*) that vocabulary-controlled subject headings continue to be necessary, on the one hand, while on the other hand in the rest of their document they effectively undercut or eliminate the mechanisms that are necessary for *finding* the controlled terms.

A prudent solution overlooked by the Working Group

A prudent way to solve most of the problems discussed above exists: the Library of Congress and other libraries should indeed be open to accepting cataloging or tagging data from anywhere in the supply chain—but such contributions should be added only to Web sites that are *linked to* the records in library catalogs, which must be maintained in separate environments with necessarily different search and display softwares.

The Library of Congress already has an excellent prototype in place, one that points the way to a more comprehensive system of opening Library records to Web audiences. It is the Flickr project, in which thousands of photographs from LC's collections have been put into the Web-based Flickr site (www.flickr.com/photos/library_of_congress/) and opened within that environment to all of the advantages of Web 2.0 democratic tagging and commentary. The Flickr records are linked to the controlled records within the Library's separate Prints and Photographs Online Catalog (PPOC), a catalog which the Prints & Photographs Division insists on maintaining separately because of its control features that are lacking in Flickr (http://lcweb2.loc.gov/pp/pphome.html).

A second semi-prototype also now exists at LC: its Digital Table of Contents (D-TOC) project, a product of its Bibliographic Enrichment Activities Team (BEAT). A recent description is as follows:

> The team's best-known project is the creation of digital tables of contents data (D-TOC), either as part of bibliographic records *or as separate files linked to them*. During the Library of Congress fiscal year 2007 (October 1, 2006-September 30,

2007), software developed by BEAT enabled the inclusion of tables of contents directly in 18,023 records for ECIP galleys and the creation of 20,389 additional D-TOC for published books. The cumulative number of "hits" on the D-TOC server since 1995 surpassed twenty million over the weekend of November 23-25, 2007. Other BEAT projects this fiscal year *linked* the Library's online catalog to more than 5,200 sample texts, brief biographies of 58,862 authors, 1,239 book reviews, and publishers' descriptions of 63,821 new publications.

[www.loc..gov/ala/mw-2008-update.html; emphasis added]

In other words, here are proto-examples of uncontrolled elements from a variety of places in the "supply chain"—elements that exist and can be searched in the Web environment—being *linked to* library cataloging records in a way that does not take the records themselves out of the necessary catalog environment which displays the cross-references, browse-menus, and scope notes that the Web environment cannot show. The D-TOC project is not currently open to democratic tagging from any users, as the Flickr site is; but the important point is that D-TOC and Flickr both demonstrate ways in which the best of both worlds can be made available, in combination, to researchers anywhere.

I believe this is the kind of solution that can actually please just about everybody. It does require a compromise on the part of the Web enthusiasts, to recognize that separate (although linked) environments must be maintained. But doing only this would prevent the "baby from being thrown out with the bath water"—the concern of those who, like myself, see grave losses to scholarly access if all "bibliographic control" is forced exclusively onto the Procrustean bed of the Web environment. We need to think outside that box alone. If we maintain library catalogs in separate "environments," we can—through links—allow (and welcome!) all of the inputs recommended by the Working Group: those that "do not conform precisely to U.S. library standards"; "democratic tags" that do no conform to any standards whatsoever; vendor- and publisher-supplied data; or data from anywhere at all in the "supply chain." We can indeed "have it all" as long as we recognize, to begin with, that more than one environment is necessary for different "bibliographic control" mechanisms to function optimally. Just as the transition from card catalogs to OPACs produced major new search capabilities—not just keyword searching, word truncation, and Boolean capabilities, but also (and equally important) the display of browse-menus that had been effectively hidden in card catalogs—the addition of Web 2.0 capabilities in search sites linked to OPAC records will mark a major advance in research and retrieval capabilities.

Although I have labeled this a "compromise" position, it is nonetheless one that reflects reality over theory. Moreover, any position demonstrating compromise among various voices (especially one that will not harm the interests of thousands of libraries in local Congressional districts) is something that can readily be justified to the Library of Congress's oversight committees.

The need for the American Library Association to act

I would add, finally, one additional point; and obviously I am speaking here only as an individual citizen and librarian: the American Library Association and its Washington Office need seriously to mount a lobbying effort targeted *specifically* on:

> a) insisting, contrary to the Working Group's recommendation, that maintenance of LC's cataloging operations must be regarded as a much higher priority for all of the nation's libraries than is the digitization of LC's special collections, and

> b) reversing LC's proposed plan to re-write the Position Descriptions of its professional catalogers, and to reorganize their entire department, in such a way as to minimize (or even eliminate) their need for subject expertise, as well as to burden them with acquisition responsibilities that properly belong to other professionals. More, rather than less, subject (and language) expertise is required across the board at the Library of Congress. The drain of professionalism from the Cataloging department, caused by increasing retirements that management does not see fit to remedy through more hiring, has already become very serious.

The latter problem is separate from any recommendations proposed by the Working Group, although the Group's *Report* will undoubtedly be appealed to as a justification for the plan, which is going forward without consultation of outside stakeholders whose own cataloging operations depends on LC's output. If the Library of Congress succeeds in dumbing down its own subject cataloging operations through this reorganization, there will be serious negative consequences for all American scholars who wish to pursue their topics comprehensively and at in-depth research levels, and for libraries in every Congressional District whose financial constraints make them more dependent than ever on the continued supply of quality subject cataloging from the Library of Congress.