HILCC: A Hierarchical Interface to Library of Congress Classification

A project report on the development of an operational prototype
LCC-based subject interface to electronic resources at Columbia University Libraries

Stephen Paul Davis, Columbia University Libraries
December 2001

1. Introduction

Those who follow the progress of library-based information access and retrieval technologies will, if pressed, be obliged to admit that libraries and the automated system vendors that serve them have done little in the last decade to improve subject access to our print and, now, online collections. Much has of course been written and proposed in the library and information science literature about possible new strategies for access and retrieval, but few new approaches have actually been developed, tested and implemented in recent generations of library OPACs. Some would attribute this variously to: the marginal economics of library automation's niche marketplace; the timid approach vendors have taken to their feature enhancement processes; the enormous technical infrastructure changes libraries and vendors have had to absorb over the last ten years in order to stay even minimally current with new technologies; the aging systems of classification and subject analysis that continue to serve as our cataloging standards; the difficulty of innovating in OPACs when developers are constrained by the heavy hand of Z39.50 and fear the loss of interoperability with consortia and other cooperative systems; and the rise of the Web and the seemingly universal appeal of know-nothing, shot-in-the-dark keyword-Booleanism.

Even the traditional library strategy of creating and displaying "syndetic structures" of cross-references has foundered in the online environment. While most vendor-based library systems do now at least support the loading and integrated display of subject cross references in the OPAC, keeping them current and adequately customized to the local collection has proven to be too expensive and time consuming for many libraries. In institutions that have actually made this investment, the dismaying truth is that retrieval set displays can sometimes be overwhelming for users because of clumsy OPAC design and functionality. (A subject search for *political science* in virtually any large university's OPAC illustrates the problem nicely.) Research library users who bravely attempt subject searching are apt to encounter screen after screen of see-also references listed in alphabetical (i.e., conceptually random) disorder, followed by repetitious subject heading entries trailing endless subdivisions. Library patrons may be forgiven if they turn to the OPAC's keyword search function instead, or worse, flee to Google.

In this context it seems telling that, as soon as the Web became generally available, librarians almost instantly began coding and uploading informal subject-oriented lists of resources, whether of online databases, ejournals, or "Internet pathfinders." Perhaps unsurprisingly, these initiatives tended to come not from cataloging departments but from reference staff, selectors and collection development officers. The popularity of this approach was such that, as librarywebs grew and flowered, they would often sprout a number of different subject menus -- overlapping, dissimilar -- that were developed (and sometimes even maintained!) by different staff members or departments.

With the stunning growth of both commercial and noncommercial electronic resources and the commensurate need to collect and make them accessible to library users, the impossibility of 'scaling up' the creation and maintenance of manually created electronic resource lists has become increasingly apparent. In response to this, some libraries have already come to the conclusion that library cataloging -- whether vendor-supplied, shared or original -- may in fact be the best resource for meeting the challenge of providing flexible access to our burgeoning virtual libraries, and in a way that preserves conceptual integration with our enormous print and other non-electronic collections. The role of cataloging may not be as different in the near term future as some have predicted.

Recognizing the critical role of cataloging departments in allowing libraries to come to grips with issues of scale (as well as those of consistency and quality control) does not, of course, solve the problem of providing better subject access to our electronic collections. However, it does at least clarify the ground rules.

We can hope that the new generation of OPACs and their companion digital library platforms, such as Endeavor's ENCompass, will be able to deliver more flexible and effective user interfaces while also providing broad integration of access to our electronic and print collections. It is heartening to see that some commercial information vendors and aggregators, including OCLC and RLG, have in fact made great strides in recent years in enhancing their proprietary interfaces, improving functionality and design and providing users better guidance during the search process. These companies' deep pockets and commitment to product improvement have clearly benefited our patrons; at the same time they make the disparity between their systems and our OPACs even more glaring.

In response to these trends, a few libraries have begun to experiment with exporting cataloging data from their LMS's into newer, non-library database platforms and toolsets as a way to create more innovative Web-based access to electronic resources outside the context of the OPAC. Since these experiments are unhindered by the constraints of traditional library automated systems, they may yield interesting and useful new approaches to exploiting the subject and other metadata available in catalog records -- approaches that may perhaps eventually find their way back into library OPACs.

This paper describes one such project, undertaken by Columbia University Libraries, to develop a system for the automatic generation of browsable, web-based, subject-oriented presentations of our electronic resources by mapping LC classification numbers present in catalog records to a hierarchical subject vocabulary derived from the Library of Congress classification schedules.

2. Background

In 1997, Columbia Libraries developed an SQL database and publishing system -- called the Master Metadata File (MMF) -- to support the automatic generation of Web pages listing reference databases, online reference texts and other electronic resources as they were added to our digital library collections. Each online reference resource was described in a brief "profile" record that was keyed into the MMF via a secure, web-based CGI form. Normally this process -- both the description and the keying -- was done directly by the individual reference librarian or selector who had assumed overall responsibility for the title. This brief "profile" record was in fact quasi-AACR2 MARC-compatible metadata; but since public services staff members were creating the records we decided it was more politic to call these descriptions profiles rather than metadata records. (Examples of the various components of this project may be found in Carol Mandel's "Manifestations of Cataloging in the Era of Metadata" [outline & slides], ALCTS/LITA Institute on Managing Metadata for the Digital Library, May 4-5, 1998, available online at: http://www.columbia.edu/cgi-bin/cul/resolve?cul.1BILCW)

As part of profile creation, the staff person was asked to assign one or more broad subject categories to the resource (see Table 1) as well as one or more standard resource type or genre designations such as "Abstracting & Indexing Service," "Ejournal" or "Full Text Resource." The profile also included additional description and annotation, including scope, keywords, search tips and related resources.

Subject Categories	
Art, Architecture & Music	Law & Legislation
Business & Economics	Science & Technology
General & Interdisciplinary	Social Sciences
Humanities & History	

Table 1: Columbia LibraryWeb subject categories, ca. 1997

From this information we batch-published HTML listings by title, by subject category and by resource type. All links in these browsable listings, when clicked, took the user to an individual "about" screen -- also batch

generated from the MMF -- that contained a description of the resource along with a stable, proxied (etc.) persistent URL link in the form of a "connect" button (see sample profile in Appendix A). HTML listings and about screens were regenerated on an ad hoc basis -- usually several times a week -- whenever any resource profile was added, changed or deleted.

This approach to producing browsable subject guides to our electronic reference resources worked well for a time, but inevitably began to show its weakness as the number of reference databases grew and as we began offering increasing numbers of electronic journals and texts to our users via the Web. We needed to move to a more scalable and less manual solution for access to electronic resources. We also needed a way to replace the informal subject categorization we had used so far with something more authoritative and comprehensive but also more specific and "granular."

In planning for ways to support an expanded Web-based presentation of our resources, we were aided by the Libraries' strategic commitment -- which had evolved over the same period -- to continue to provide standard cataloging for as many of our electronic resources as feasible, whether through shared, original or vendor-supplied records. What this meant in concrete terms was that we could anticipate relieving selectors and reference librarians of the need to create increasing numbers of "profile records," and instead extract MARC records for these same electronic resources from our LMS for loading into the Master Metadata File. Once loaded into the MMF, we would then be able to use standard programming and scripting tools (such as PERL, C++ and Java) to write batch or real-time interactive Web-based presentations for our users.

Among other benefits, this approach would give us access to the various subject-oriented elements present in these records -- not only subject headings but also geographic area codes, contents notes, and LC classification numbers. We would still continue to provide a way for selectors and others to create metadata records directly in the MMF, but this would be limited to resources not considered appropriate for full cataloging (e.g., ephemeral web sites or commercial services under evaluation) or those that had simply not yet gotten to the head of the cataloging queue.

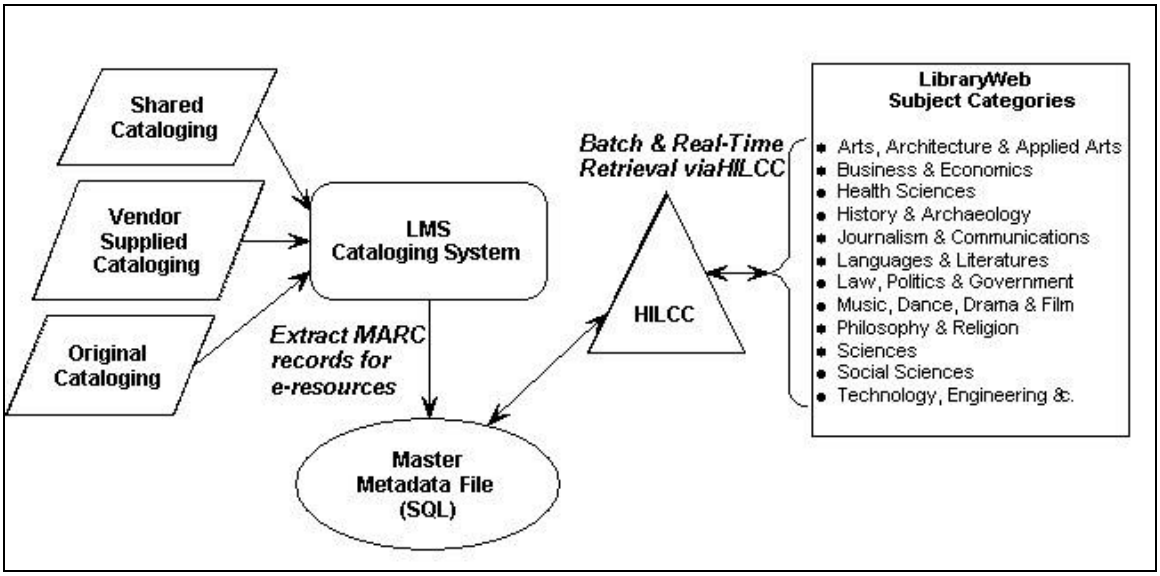

Chart 1: Workflow for extracting eresource records from the LMS for loading into Columbia's Master Metadata File (MMF) where they are accessible from subject category menus via HILCC

Our goal in expanding the MMF project in this way was not to develop anything like a fully functional end-user search and retrieval system that duplicated OPAC functionality. We did, however, see it as an opportunity to experiment more flexibly with the Web-based presentation of our electronic resources. One of our first priorities was to investigate ways we might exploit the subject content carried by LC Classification (LCC) numbers.

3. LC Classification as the Basis for a Hierarchical Subject Interface

[FOOTNOTE: For the uninitiated, LC Classification was developed at the Library of Congress beginning about 1899 and was not, contrary to rumor, based on the organization of Thomas Jefferson's library but rather on Charles Ammi Cutter's *Expansive Classification* (1891). A convenient outline of LCC is available online at: http://www.loc.gov/catdir/cpso/lcco/lcco.html. LC Classification numbers constitute the first portion of LC call numbers and are carried in MARC21 field 050, subfield a.]

We chose to focus on classification rather than subject headings because classification seemed better suited to our proximate goal, namely the creation of a browsable subject interface in which each electronic resource would appear under the one subject category that best reflected its content. We felt that this type of single-entry listing would provide a simple, easily navigable approach for many kinds of subject discovery, and would allow us to work toward replacing the various manually-created subject menus and listings in our libraryweb with an automatically generated subject interface. Moreover, we already knew from our experience with vendor-supplied OPACs the apparently insuperable difficulty of developing user-friendly search and retrieval interfaces for LC subject headings. Although we did in fact extract and load LC subject headings from catalog records into our MMF, our immediate purpose was to use them as enriched keywords for indexing by our Web-based search engines.

At about the same time that Columbia began working on this project, librarians at other institutions were evaluating the use of Dewey Decimal Classification (DDC) as a possible hierarchical system for access to Internet resources. (See, for example, Vizine-Goetz, D. "Using Library Classification Schemes for Internet Resources." In: *Proceedings of the OCLC Internet Cataloging Colloquium*, San Antonio, Texas, January 19, 1996. URL http://staff.oclc.org/~vizine/Intercat/vizine-goetz.htm). It appeared to some that DDC would lend itself to this type of subject browsing better than would LCC because of Dewey's intrinsically hierarchical design. Even if true, however, no one could possibly envision Columbia or other ARL libraries changing over (or back) to Dewey for all their cataloging, or even just the cataloging of electronic resources.

As we studied LCC, we saw -- as Vizine-Goetz and others had observed -- that underneath its enumerative top-level, there were potentially useful hierarchies at the second and subsequent levels as revealed by captions and the indispensable, conceptually significant indentations present in the schedules as a guide to classifiers. For example, the list of LC class number ranges shown in Table 2 actually reflects the useful hierarchy shown in Table 3.

B
BM
BM 495-532
BM 497-509
BM 497-497.8
BM 498-498.8
BM 499-504.7
BM 507-507.5
BM 508-508.5

Table 2: Selected class ranges from LCC's Philosophy, Psychology, Religion schedule

Philosophy, Psychology, Religion.				
	Judaism			
		Sources of Jewish Religion. Rabbinical Literature		
			Talmudic Literature	
				Mishnah

				Palestinian Talmud
				Babylonian Talmud
				Baraita
				Tosefta

Table 3: LCC captions corresponding to the class ranges in Table 2.

On the other hand, many LCC hierarchies do seem outdated or at least less useful in the sense that naive users might have trouble inferring lower level subject categories from the upper levels of the hierarchy. For example:

G	**Geography. Anthropology. Recreation.**		
GT		Manners and Customs (General)	
GT 485			Churches and church going
			etc.

Table 4: Excerpt from LCC's Geography, Anthropology, Recreation schedule

C	**Auxiliary Sciences of History**		
CB		History of Civilization	
CB 156			Terrestrial evidence of interplanetary voyages
CB 158-161			Forecasts of future progress
			etc.

Table 5: Excerpt from LCC's Geography, Anthropology, Recreation schedule

T	**Technology**		
TX		Home Economics	
TX 341-641			Nutrition. Food and Food Supply
TX 950-953			Taverns, Barrooms, Saloons
TX 1100-1105			Mobile Home Living
			etc.

Table 6: Excerpt from LCC's Technology schedule

Beyond these types of hierarchical infelicities, LCC has well-known problems of terminology, consistency, balance and what might be called "world view," many of which have been discussed in the literature, at conferences, and doubtless internally at the Library of Congress. Some examples:

- **Class D "History: General and Old World"**. A single class schedule, D, is allocated for the history of all regions of the world, apart from America; two entire schedules, E and F, are dedicated to America and the United States. This certainly places the cultural and conceptual framework of those who created in LCC in stark relief. At the very least the term "Old World" must now be seen as outdated and culturally biased toward the Western Europeans and Americans who invented the New World/Old World dichotomy.

- **Class E, "History: America"**. Toward the beginning of schedule E the following hierarchy appears:

E	**History: America**			
E 151-889		United States		
E 184-185.98			Elements in the population	
E 184.5-185.98				Afro-Americans
E 186-199			Colonial history (1607-1775)	
			etc.	

Table 7: Excerpt from LCC's History: America

This is probably not a hierarchy that one would want to present to users directly.

- **Class B, "Philosophy. Psychology. Religion"**. Under Class B no fewer than one-third of the second level terms are given over to Christianity, including one for "The Bible," which is not even identified as the Christian Bible. And while there may once have been excellent reasons for sandwiching Psychology between Speculative Philosophy and Aesthetics, the intercultural rudeness of including Theosophy -- an eccentric American cult dating from the late 19th century -- in the same caption as Islam might at the least bewilder our patrons.

B	**Philosophy. Psychology. Religion**	
BC		Logic
BD		Speculative Philosophy
BF		Psychology
BH		Aesthetics
BJ		Ethics
BL		Religions. Theology. Rationalism
BM		Judaism
BP		Islam. Bahaism. Theosophy, etc.
BQ		Buddhism
BR		Christianity
BS		The Bible
BT		Doctrinal Theology
BV		Practical Theology
BX		Christian Denominations

Table 8: Excerpt from LCC's Philosophy, Psychology, Religion schedule

So in short, our working group realized quickly that we would not be able to use LCC "out of the box," as it were, to create a browsable user interface. Rather than being discouraged, however, this recognition was in fact somewhat freeing. It meant that, while we would indeed be able to exploit the LC class numbers themselves, the notion of using LCC's language and structure more generally was so clearly out of the question that to proceed at all we would have to develop different entry vocabulary and create alternate hierarchies.

Since we planned to build this presentation hierarchy outside the cataloging system and then process catalog records with LC class numbers against this mapping schema, we would have the freedom to experiment with different presentation strategies -- adding to, changing and rearranging as we needed. We could easily, for example, reformulate the Philosophy and Religion presentation to:

HILCC Presentation Hierarchy:			*Corresponding LC class number ranges:*
Philosophy & Religion			
	Philosophy		
		Philosophy (General)	B 1-5802
		Aesthetics	BH 1-301
		Ethics	BJ 1-2195
		Logic	BC 1-199
		Speculative	BD 1-701
	Religion		
		Religion (General)	BL 1-290, BL 350-632.5
		African Religions	BL 2390-2490
		Ancient Near Eastern Religions	BL 1600-1695
		Bahaism	BP 300-395
		Buddhism	BQ 1-9800
		Christianity	BR 1-1725, BS 1-2970, BT 10-1480, BV 1-5099, BX 1-9999
		European, pre-Christian Religions	BL 689-980
		Hinduism	BL 1100-1299
		Islam	BP 1-253
		Jainism	BL 1300-1380
		Judaism	BM 1-990
		Mythology, Comparative	BL 300-325
		North & South American Religions	BL 2500-2592, E98.R3
		Oceania Religions	BL 2600-2630
		Rationalism, Atheism, Secularism	BL 2700-2790
		Other religions	BL 660-687, etc.

Table 9: Philosophy & Religion section of HILCC

Or we might take a more taxonomic approach, where for example Christianity and Islam branch from Judaism; or a highly enumerative, alphabetic approach giving equal time to each religion & sect. This realization helped inspire us to continue our planning.

---

4. Project Planning

A working group with members from our Bibliographic Control Department and Library Systems Office was convened in 1997 to begin working on a "Hierarchical Interface to LC Classification" (HILCC).

[FOOTNOTE: The working group currently includes the following Columbia Libraries staff members: Rick J. Block, Stephen P. Davis, Kate Harcourt, Sarah H. Witte, Robert A. Wolven; in addition, former staff member Jeffrey Sowder contributed to the development of HILCC.]

Our informal objectives for this Phase I pilot project were to:

a. Develop a preliminary classification map and subject category hierarchy for each LCC schedule;

b. Invite reference staff, selectors and other subject specialists to assist with reviewing and revising the various subject areas of this preliminary HILCC map;

c. Extract a test dataset of catalog records for ejournals from our LMS;

d. Convert and load test record sets into our Master Metadata File;

e. Specify and program one or more web presentation options;

f. Review results & revise;

g. Put HILCC into production as an operational prototype;

h. Continue to gather staff and user feedback; continue to make corrections, updates and enhancements;

i. Operationalize the weekly extraction and loading of catalog records for ejournals from the LMS into the MMF and the regeneration of the browsable HILCC interface to our eresources;

j. Expand the dataset to include other e-resources beyond ejournals;

Almost immediately, however, as we faced the daunting task of coming to grips with the full range of LC class schedules, we felt the need to place certain kinds of limits on our work in order to make the project manageable within a reasonable timeframe, especially since the work was being done by staff fully engaged with other primary responsibilities. Some helpful working assumptions that evolved were:

- In the current phase of the project, we would create no more than three hierarchical levels at any point in the schema, even if it might be desirable to go to a fourth or fifth;
- Our main goal was to provide an effective end-user interface to LC Classification numbers, not to spend time rearranging LCC itself;
- It was not our goal to create a universal subject interface, only one that would serve the Columbia community;
- We should not attempt to duplicate functionality already available in the current generation of OPACs;
- We should proceed by creating operational prototypes that could be assessed and improved over time;
- HILCC should be seen as only one of several possible ways for our users to search for resources by subject; it did not need to solve all or even most subject retrieval problems.

We more or less managed to keep these guidelines in view as we proceeded.

---

5. Working Design Principles and Considerations

Perhaps the best way to describe the process of developing HILCC is to review the design principles that evolved as we worked. Some of these ideas had been articulated at the outset of the project, but for the most part they grew out of questions and issues raised by the ongoing work itself. The discussion below is organized around individual design and development decisions made during the course of the project.

a. **The first level display should include no more than twelve (or so) categories.** We felt that in many respects, HILCC's first-level menu organization was the most important design challenge. People's ability to find relevant resources without repetitive backtracking depended directly on their ability to choose the correct top branch of the hierarchy. The first menu needed to be clear enough for undergraduates but not too simplistic for faculty and graduate students searching outside their usual domain. It needed to be balanced in terms of disciplines and subject areas, but also reflect Columbia's actual online collections and academic curriculum.

We began with an informal survey of other university and commercial sites that already provided general-purpose browsable listings by subject category to see what could be learned. With regard to the specific question of how extensive a top-level listing might be, we did come away with some impressionistic results. The list below shows the number of first-level subject category menu items we found in a typical selection of sites. (The numbers below reflect the sites' first level menus as of December 2001, but the counts are virtually the same as those we obtained in 1997. Shopping and other non-substantive links have been omitted.)

MSN Search - 10	BUBL (Strathclyde University) - 10
Hotbot - 14	NC State University - 12
Yahoo - 14	LC American Memory - 13
Excite - 15	WWW Virtual Library - 14
Netscape Search- 16	

Based on these results, we felt comfortable proceeding with a menu of between ten and fifteen top-level items, finally settling on twelve as our working target. The number and content of those categories did evolve considerably during the course of the pilot, however, as we received feedback from different groups of library staff. Still we did ultimately end up with a twelve-item main menu.

We also decided early on that we would attempt to follow several principles that were (and still are) considered to be good design practice for web page usability, namely a) that web pages should ideally take up less than one "screen," without scrolling, on a standard 600 x 800 pixel display; and b) that the user should ideally be able to reach any desired item in no more than three mouse clicks from the main page; and c) that generous use of white space as a design element is pleasing to the eye and helps avoids visual clutter. As many others have found before us, applying all three of these principles simultaneously is often virtually impossible, particularly when working in a complex and deeply hierarchical domain. We can say with a high degree of confidence that we were not entirely successful in achieving these design goals.

Since our starting point was LC Classification with its twenty-one separate schedules, we were immediately obliged to combine and merge HILCC categories in ways that we hoped would yield a list with straightforward terminology; also one that would allow users to easily discern the likely conceptual boundaries of each category. The results of this exercise in redividing the world of knowledge may be seen in Table 10.

Main Subject Categories	
Arts, Architecture & Applied Arts	Law, Politics & Government
Business & Economics	Music, Dance, Drama & Film
Health Sciences	Philosophy & Religion
History & Archaeology	Sciences
Journalism & Communications	Social Sciences
Languages & Literatures	Technology, Engineering & Applied Sciences

Table 10: HILCC Level 1 Categories (ca. 2001)

In most cases our selection of categories seemed reasonable and sensible for the Columbia environment; a few however were recognized from the outset to be less effective and were flagged for early review during Phase II in conjunction with user assessment studies. Terms that might have been useful as unifying top-level categories were sometimes rejected because of their perceived unfamiliarity to users and replaced with an enumeration of second level terms. For example, we initially selected the term "Applied Sciences" to cover technology, engineering and computer science; but informal student feedback made it clear that the term "applied sciences" is not only <u>not</u> generally used, it was considered utterly opaque by our users. So we fell back here as elsewhere on a caption that enumerated rather than summarized the major subareas, i.e., Technology, Engineering & Applied Sciences. Likewise, the more generally used rubric

"Performing Arts" that is often used to describe music, dance, drama and film was deemed confusing at Columbia, chiefly because these areas are for the most part not actually taught as performing arts here. (Columbia's major areas of music instruction and research, for example are ethnomusicology, historical musicology, music theory and composition.)

The process of settling on top-level categories proceeded iteratively over time and left those of us involved with a new appreciation of the inherent difficulties in trying to organize knowledge according to a generalized schema. The categories chosen for the top level of course reflect biases of individuals in the working group and other library staff who contributed to the effort. They also reflected explicit and implicit collection development policies, which in turn reflect the curriculum and research priorities of Columbia University.

For example, subject areas relating to many of Columbia's major professional schools, which are served by the central library system -- such as Architecture, Business, Engineering and Journalism--figure prominently in top-level categories. Columbia's Health Sciences Library and Law Library are relatively independent from the main library and did not participate during Phase I in the development of HILCC, with the consequence that those sections of HILCC are less well developed and adhere more closely to the LC Classification outline than they might have otherwise.

Some library staff have suggested that our top-level organization is weighted disproportionally toward the humanities; this is something we will revisit during Phase II. How we would in fact be able to determine what an equitable balance of top-level categories might be remains, at this point, mysterious.

Additional biases in HILCC can easily be recognized by a simple list of other institutions' top-level categories that do <u>not</u> appear in Columbia's, e.g., Agriculture, Bibliography / Library Science, Computer Science, Education, Generalities, Geography, Military Science, Naval Science, Recreation, Sports, Reference, Veterinary Medicine.

b. **HILCC's overall hierarchy should be no more than three levels deep.** It was clear from the beginning that some sections of LCC lent themselves, in theory, to four, five and six-level hierarchies. For pragmatic reasons and usability considerations we decided to limit ourselves to three hierarchical levels in Phase I of the project.

In some cases we did this by skipping over hierarchies implied in LCC. E.g., for class ranges BL1000 - BL1299 we used the simpler:

Philosophy & Religion -- Religion -- Hinduism
rather than LCC's:
Philosophy & Religion -- Religion -- History & Principles of Religion -- Asian, Oriental -- Hinduism

Some extensions of HILCC to a fourth or even fifth level do seem inevitable, however, and are already in the planning stages for Phase II. We will likely interpose levels of geographical hierarchy in the history portion of HILCC, and a few other places. For example, for the class ranges DD1 - 9999, we anticipate using:

History & Archaeology -- Regions & Countries -- Europe -- History of Germany
rather than LCC's:
History: General & Old World -- History of Germany

It is clear that we were not able to meet our ideal page length target of "one screen" when enumerating at the second and third levels of the HILCC hierarchy. In Phase 2 we will be aiming for more compact and flexible presentation techniques, such as JavaScript rollovers, expandable folder hierarchies, etc.

c. **The degree of granularity should be relative to the actual resources available.** In determining how enumerative and specific to be at any given hierarchical level, we decided to be guided by the current depth of our electronic holdings in a specific area. We recognized in doing this that we would need to revisit and revise portions of the hierarchy periodically as larger numbers of resources were added to our digital collections, but felt this would be an important and useful process in any event.

As a separate but related issue, if at a given level of HILCC -- at whatever the degree of granularity implemented -- there were in fact categories for which there were no electronic resources yet available, we decided to retain and display these unused categories as "grayed out" menu items on the assumption that this would aid users in recognizing that they had indeed navigated to the correct place in the hierarchy, but that there were no "hits" for their topic, e.g.,

Arts, Architecture & Applied Arts

Visual Arts (all)
General
Decorative Arts
Drawing, Design & Illustration
Painting
Photography
Print Media
Sculpture

d. **Subject categories should be built from LCC but not constrained by it.** It was apparent to us that in some cases the existing structure of the LCC tables would indeed provide the basis for a reasonably coherent user presentation. In others, however, we would either need to ignore the explicit or implicit hierarchies present in LCC schedules or else reorganize them in order to create a balanced and usable interface. For example, LCC's schedule G, "Geography, Anthropology and Recreation" did not seem a useful category grouping for end user presentation & navigation, and the G class ranges were variously reassigned to the sciences and social sciences.

e. **A specific LC class range should map to only a single location in the HILCC structure.** We initially attempted to map some classification ranges into more than one HILCC category, but realized that this would add substantially to the maintenance overhead and the programming requirements of the project. We also recognized that the issues raised by this kind of double mapping were really just the beginning of a larger analysis and planning effort that would be needed to address the challenge of presenting interdisciplinary resources to users (see also section g following).

An area in which we were especially tempted to map a single class range to more than one HILCC location was that of Psychology (BF). LCC of course embeds Psychology in the middle of Philosophy, quite near Religion and "The Occult Sciences." Although it is perhaps possible that this was once the ideal location for it, modern psychology has subdisciplines that fall variously into the social sciences, the sciences and the health sciences. In HILCC we finally decided to position Psychology under the Social Sciences; but because Columbia's Psychology Library has traditionally been designated a science library, we also manually created a duplicate link in our user interface from within the Science hierarchy.

f. **HILCC processing and output should accommodate multiple LC class numbers appearing in a single bibliographic record.** Although standard catalog records rarely include more than one classification number for the purposes of capturing different subject aspects of the same work, we anticipated this occurring on an ad hoc basis at Columbia as selectors and reference staff requested that important electronic resources be available under different subject trees.

In fact, we had previously established a precedent for multiple subject category assignments in our initial Master Metadata File implementation where, when selectors created profiles for individual reference databases, they were able to select up to three subject category areas under which the title would be listed. In part this was a reflection of the broad scope of some of the large

databases; ***Annual Reviews,*** for example, contains extensive resources in sciences, health sciences and social sciences.

g. **The categorization and presentation of interdisciplinary resources should be addressed separately from the main HILCC effort.** Reference staff and selectors consulting on the HILCC project emphasized to the working group the importance of also providing better and more comprehensive displays of interdisciplinary resources than could be derived from a generalized mapping of LCC numbers. This need had so far been addressed at Columbia by the manual compilation and maintenance of specialized guides or "Internet pathfinders" on interdisciplinary topics such as women's studies, African-American studies, Middle East studies, etc.

We sketched out a possible method of addressing this need through the creation of separate, customized versions of HILCC for the different interdisciplinary areas. These selective classification maps could be used -- perhaps in conjunction with other elements found in the catalog record -- to extract, filter and display more targeted presentations of subsets of our electronic resources.

It did seem, however, that this would depart somewhat from the project's primary objective of creating a generalized subject interface to all our electronic resources. It also seemed clear that, even more than with the generalized HILCC interface, these interdisciplinary "mini-HILCCs" would require a higher level of ongoing maintenance since they would inevitably need to reflect Columbia's evolving academic organization and curriculum, the specific collection development policies of our different collections and departmental libraries, not to speak of the particular strengths and interests of the library staff members responsible for particular subject areas. The working group agreed that this was an important area that we would nonetheless need to postpone to Phase II.

h. **The user interface must include composite, summary lists at the first and second levels of each hierarchy.** Once a user manages to locate the "correct" top-level category, it still may be unclear how far down a hierarchy to drill to find a specific narrower concept; or in some cases the user may just want a complete view of the subject at a second or third level. We thus assumed more or less from the beginning that we would need also to provide combined listings at the first and second levels. In this way, users who give up before finding their way down to the lowest level of the hierarchy still have a chance of finding relevant resources.

For example, if a user is trying to locate electronic resources dealing with European public policy, and has chosen the correct top-level menu category, she or he might navigate down the following hierarchy:

Law, Politics & Government -- Political Institutions & Public Administration -- Europe

But if the user gives up before drilling all the way down to "Europe," all relevant resources on that specific topic would still also be findable under: *Law, Politics & Government (all)* as well as *Political Institutions & Public Administration (all)*. These higher-level artificial composite listings might of course be lengthy, but the desired resources would at least be present and findable through determined browsing or judicious use of the browser-based find command.

i. **When feasible, terminology used at the lowest level of the hierarchy should be meaningful and unambiguous when displayed independently.** In order to have the capability of flexibly displaying and repurposing HILCC output in different contexts, we came to the conclusion that the terminology at the lowest level of the hierarchy, at least, should be constructed such that it could be displayed independently of the hierarchy. For example, we would use:

Social Sciences / Education / <u>History of Education</u>
rather than the elliptical:
Social Sciences / Education / <u>History</u>

The term "History" used alone in a list or display would obviously be ambiguous or misleading, whereas "History of Education" could stand on its own.

One of the immediate reasons for this decision was that we wanted to use these lowest-level HILCC terms to produce a separate topical subject index to be used in conjunction with HILCC, since it was clear that a topical list would be easier for some kinds of resource discovery -- particularly when the terminology and conceptual organization of the subject was straightforward. (See Table 11)

Topical list:	***From corresponding HILCC category:***
[...]	
Biochemistry	Sciences -- Chemistry -- Biochemistry
Bioengineering	Technology, Engineering & Applied Sciences -- Engineering -- Bioengineering
Biography	History & Archaeology - Biography
Biology	Sciences -- Biology (all)
Biomedical Engineering	Health Sciences -- Biomedical Engineering
Biophysics	Sciences -- Biology -- Biophysics
Book Studies & Arts	Social Sciences -- Education & Scholarship -- Book Studies & Arts
Botany	Sciences -- Botany
Business	Business & Economics -- Business (all)
[...]	

Table 11: HILCC-derived Topical List

Planning for and implementing a successful "inheritance" schema for displays of hierarchical data is complex and not something we felt we could necessarily take for granted in the design of HILCC. We recognized, though, that the decision to create stand-alone terminology would itself lead to more questions about how we should formulate these lowest-level terms when there was no obvious equivalent in common usage, or when phrases might perhaps be inverted to help cluster like content in an alphabetic listing, e.g.,

History of Education
[...]
Theory and Practice of Education

or:

Education, History of
Education, Theory and Practice of

In Phase 2 we will compare terminology at the lowest level against LCSH to see if corresponding terms can be borrowed for use in HILCC. We may find that we need to store two versions of the lowest-level term, one suitable for display in a hierarchy, and one capable of standing alone without ambiguity.

6. Implementation Issues -- Sciences

Many of the types of issues that arose during HILCC development can be seen in miniature in the Sciences hierarchy, which was one of the first to be prototyped. A partial summary of decisions made in constructing this section follows along with brief characterizations of the discussions that led to them. (Refer to Appendix B, the Science portion of HILCC with a listing of corresponding LCC ranges; and Appendix C showing selections LC Class Schedule Q with corresponding LCC captions.)

- **Terminology Changes.**
  - LCC's Level 2 caption "Astronomy" (QB) was expanded to "Astronomy & Astrophysics," essentially elevating a third level LCC category to the second level of HILCC. This change was made to reflect the combined orientation of Columbia's academic and research programs in these fields and the Libraries' corresponding collecting policies in these areas.

  - LCC's Level 2 caption "Neurophysiology and Neuropsychology" was broadened to "Neurosciences," which was felt to be a more useful summary of the scope of material collected and classed in LCC's QP 351-495 range.

- **New Intermediate Groupings.**
  - We created an intermediate hierarchy under Sciences for "Earth and Environmental Sciences," into which we moved Ecology (QH 540 - 549.5) Environmental Sciences (GE 1 - 350), Forestry (SD), Geology (QE), Human Ecology & Anthropogeography (GF), Meteorology & Climatology (QC 851 - 999), Natural History (QH 1 - 278.5), Oceanography (GC) and Physical Geography (GB).

    This collocation was undertaken in part because "Earth and Environmental Sciences" has become an important focus at Columbia both in terms of academic organization and curriculum focus. Since this is clearly a cross-disciplinary grouping, however, we may want to revisit this in Phase II when we have the capability of creating separate, customized interdisciplinary mini-HILCCs.

  - We combined LC's "Human Anatomy" and "Physiology" sections into a single HILCC category "Human Anatomy & Physiology" at the second level, moving Anatomy, Neuroscience and Physiology to separate subcategories at the third level.

  - A new Level 3 category "Mathematical Statistics" was created from the LCC range QA 273-280.999, which includes LCC captions "Probability" and "Mathematical Analysis". This was done so that different areas of the classification schedule relating to statistics of various kinds could be brought together with one another, e.g., in generating an overall guide to electronic resources in statistics and quantitative methods. Again, it seems likely that the interdisciplinary "mini-HILCCs" envisioned as part of Phase II might lead us to reconsider whether to retain this type of artificial category in the master version of HILCC.

- **Rearrangement of Hierarchies.**
  - Biochemistry and Radiation Chemistry are at the fourth level of LCC captions; in HILCC they were moved up to the third level directly under Chemistry for HILCC, so that they did not disappear from HILCC's three-level hierarchy.

  - Microbiology (QR) was moved from Level 2 in LCC to Level 3, under Biology, in HILCC.

- **Merging of Schedules.**
  - Much of LCC's Class S (Agriculture) was folded into HILCC's "Sciences" hierarchy under two new intermediate groupings "Animal Sciences" and "Plant Sciences". (Note, however, that Botany and Zoology were retained in this Phase as Level 2 categories, chiefly because of their traditional places in science collections and curricula. As ongoing review of HILCC proceeds, such decisions may well be reconsidered.)

    The decision to merge Class S into the HILCC "Sciences" hierarchy was made largely because materials relating to Agriculture and related fields are not collected in any significant way by Columbia Libraries nor has Agriculture figured significantly in Columbia's curriculum. The rationale for this decision illustrates concretely why different institutions might organize HILCC-like subject hierarchies quite differently. Still, Columbia

does in fact have printed and electronic resources relating to Agriculture, notably those that we receive as part of the Federal Depository Library Program as well as those that appear unbidden in ejournal aggregator packages such as ProQuest. Moreover, Columbia's newer curricular and research emphases on Earth and Environment Sciences include the effects of global climate change on agriculture, so again, HILCC may need to evolve further in this area as elsewhere.

7. Next Steps

Assessment. Although no decisions have been made about a Phase II for the HILCC project, some kind of user assessment will be needed, whether separately or as part of other planned LibraryWeb usability testing. We will shortly have some 5,000 records available in our Master Metadata File -- for databases, ejournals, e-texts, etc. -- that we will make accessible through the HILCC interface; this should provide the critical mass needed for a targeted user assessment. Even so, the effectiveness of HILCC may be difficult to test. The success of an individual user's interaction with HILCC may depend on a number of factors, including:

- the appropriateness and relevance of the classification numbers originally assigned to our electronic resources;
- the aptness and recognizablity of the specific HILCC terms Columbia selected for that specific LC class range;
- the transparency of the hierarchy in which Columbia positioned the specific HILCC subject term;
- the usability of the specific web design and navigation functionality which we build into and around HILCC;
- the placement and description of HILCC in Columbia's LibraryWeb relative to other types of subject searching, and the way the user's expectations are shaped in advance of using it;
- the length, organization and transparency of the list of resources retrieved after navigating a hierarchy;
- the actual presence or absence of specific electronic resources in our collection of interest to the user.

There are also the many user-based variables -- such as the user's purpose in searching, past online experience, other systems used, etc. -- which we already know how to capture and record if not always how to interpret.

But what may be impossible to design a test for is the overall value of browsing electronic resources by subject categories in the first place. In the brick and mortar world, the value of organizing print collections for browsing has always been difficult to quantify; but libraries have nonetheless spent many millions of dollars over the decades in order to create and maintain them. When questions have arisen about the cost-benefit of classified collections, the arguments in favor cited convenience, serendipity, benefit to the non-specialist and such. Do these values translate to the online environment?

There are some practical considerations bearing on the effectiveness of a HILCC-like approach that are difficult to anticipate, most notably the issue of scalability: what may seem useful and manageable against a list of 5,000 electronic titles may look quite different when the list has grown to 50,000 or more.

Another even more basic concern may be the actual availability in source records of the LC Class numbers themselves. As large libraries begin increasingly to purchase catalog records, for ejournals, etc., from aggregators or specialized service bureaus, they also become reliant on those companies to obtain or themselves assign relevant LC Classification numbers. At present this is not always the case, and the cost of adding class numbers to thousands of purchased catalog records would be prohibitive for most institutions.

Phase II. Assuming Columbia's HILCC project proceeds to a Phase II, the following improvements and extensions to HILCC would be in queue for consideration.

- **Subject Specialist Review.** Continued review of the various subject areas of HILCC by library subject specialists and/or faculty as we add more electronic resources to our digital collections.

- **Interdisciplinary mini-HILCCs**. Perhaps the most urgent finding from our Phase I operational test was the importance of extending the HILCC project to include customized presentations of interdisciplinary resources. In research universities especially, new interdisciplinary institutes, centers and projects are created (and sometimes disappear) quickly. Library staff need a simple but flexible way of creating targeted listings of resources by computer-assisted means, both on-demand and for public presentation of our digital library collections.

- **Review of HILCC Terms Against LC Subject Headings.** Work has already been done at the Library of Congress and elsewhere to begin to correlate LC subject headings with relevant portions of the LC Class Schedules. Improving HILCC terminology by use of LC subject headings may benefit users and provide the Library more options for using HILCC as a basis for newer kinds of subject-oriented access.

- **Resource Presentation Using Additional Metadata Elements.** We recognize the importance of using other metadata elements in combination with HILCC for browsing and searching for materials; these additional elements or aspects include: genre or format (e.g., ejournals, e-texts), geographic content (e.g., online resources from or about sub-Saharan Africa), "reference-ness" (e.g., key online databases in Public Affairs).

- **New Interactive Modes of User Discovery.** The availability of rich metadata extracted from catalog records and made available via a robust SQL/Web retrieval framework potentially provides a powerful toolset to experiment operationally with more "intelligent" forms of user interaction such as:

    - basic research dialogues with users, allowing search strategies to be refined, expanded, limited, etc.
    - content mapping and visual navigation, allowing users to see the depth of content in certain collection areas, the relationships between content clusters
    - interactive query optimization with user-assigned relevancy weighting
    - creation of a non-specialist cataloging interface for faculty or other researchers to prepare metadata and integrate their resources into Columbia's digital collections using HILCC categories

8. Conclusions

Hierarchical, Term-Mediated Classification Systems. Judging solely by the continuing presence and even expansion of hierarchical subject-oriented menus in the major commercial search portals, one would have to come to the conclusion that this type of interface is considered an essential access tool by firms that have a more-than-altruistic interest in helping people locate resources. Taking Yahoo as perhaps the most notable example: not only does its home page prominently feature its proprietary classification hierarchy, Yahoo also does something that we in Libraries should have been doing long since. In response to a keyword search, rather than presenting the user with a list of "raw" hits, there's a simple but far more intelligent response:

- **Category Matches**
    - **[hits]**
- **Sponsored Links**
    - **[hits]**
- **Web Site Matches**
    - **[hits]**

"Category Matches," promoted to the top of the result screen, are in fact links back into Yahoo's classified hierarchy. The OPAC search for "political science" proposed in the introduction of this paper is, by contrast, a relief and a pleasure to execute in Yahoo if one navigates via the category matches. Of course -- this being Yahoo -- once you retrieve the actual resources at the bottom of a selected hierarchy, the content may well be outdated, eccentric or wrong; but the functionality and interface design that gets you there puts library OPACs to shame.

How enormously helpful it might be for patrons if our OPAC keyword searches returned:

- **Subject Category Matches (browse by subject category)**
    - **[hits]**
- **Subject Heading Matches (narrow your search)**
    - **[hits]**
- **Reference Matches (key resources)**
    - **[hits]**
- **Individual Title Matches**
    - **[hits]**

Lest one think that Yahoo -- famously staffed by bona fide librarians -- is an outlier, Google too has a powerful, if somewhat less developed, classification system that, again, displays at the top of the raw results screen the relevant category hits from their classification system. Ditto MSN, Overture (formerly GoTo), Galaxy, Oingo/Applied Semantics (an excellent search system!), Teoma and Webcrawler. In many respects creating a classification system simplifies the work of portal system designers and gives them a powerful and manageable conceptual toolset. (If instead they had somehow been persuaded to wrestle with our tens of thousands of LC subject headings and their tangle of references, they too might have given up in despair.)

By contrast, keyword and subject heading searches by themselves more often than not yield ambiguous results. A keyword searcher is always, effectively, a beginner; one must proceed each time by trial and error, usually without knowing entirely whether one has succeeded or failed. The results are often mysterious; it is unclear whether one has selected the best keywords (or subjects or references). In the usual case, the searcher may indeed retrieve some results but never really know whether these are -- even objectively speaking -- all, some or any of the results most relevant to the topic at hand in the system being searched.

Hierarchical classification systems, if mediated by entry vocabulary, explanations and a well-designed navigation framework, can offer the benefit of more easily guiding users through a process of recognition, filtering and decision making than do other subject retrieval methods. Visible hierarchies can help show users how the information provider has actually organized the resources conceptually. Armed with this understanding, the user may learn more quickly and in real time how to find resources of interest than with other methods.

Hierarchical Interfaces to Library of Congress Classification. During the course of this project we were sometimes struck by the utter hubris that would be needed today to undertake from scratch the creation, implementation and maintenance of a generalized classification system that was: accurate and socially sensitive; learned but not esoteric; consistent, proportional and conceptually balanced; suitable for students and scholars (and citizens and members of Congress) ... and capable of remaining all these things for more than a few months or years. But while we wait for a new Charles Ammi Cutter -- and a new Herbert Putnam with the vision to push this forward -- what we in research libraries have to work with for now is LC Classification.

The HILCC project at Columbia is still at a relatively early stage, particularly in terms of the possible functional integration of HILCC with keyword or guided searching methods. The process of LCC mapping and terminology selection that is at the heart of HILCC has been time-consuming, complex and is still not complete, accurate or authoritative even on its own terms. Still, I believe we have succeeded admirably in our original objective, namely to provide within our LibraryWeb an integrated, browsable subject-oriented presentation of our electronic resources.

In a broader context, it may well be feasible for others to take advantage of what we have learned from the HILCC project in at least two areas. Those institutions or consortia with the resources to build services outside their OPACs might consider taking Columbia's LCC map and reworking it for their own environments and collections. [FOOTNOTE: Columbia's HILCC documentation and current version of the LCC map is available at: http://www.columbia.edu/cgi-bin/cul/resolve?cul.1BQN3R] If others do indeed take this on, Columbia would be grateful for feedback and suggestions and more than willing to answer questions or review proposals.

On the other hand, after having spent a great deal of time attempting to come to grips with the many small and large mysteries of LCC, such as how and why "Hyperborean, Indian, and Artificial Languages" are clustered together in PM 1-9021 and whether there's a more scholarly and/or socially sensitive way to present them to our users, we might be forgiven for wondering if the Library of Congress might after all deserve to take on the role of developing and managing an authoritative, general-purpose version of HILCC. In this imaginary scenario, LC would act as coordinator and co-developer as it does with various other standards, and ultimately act as distributor for the schema and its updates as it does with other cataloging-related products. Ideally HILCC would also be used directly within LC's own information systems as an interface to its own electronic resources. This would help ensure that HILCC was grounded in actual eresource collections and end-user interactions and, perhaps, given some priority for thoughtful development and refinement. Those of us on the outside who subscribed to the product service would be free to tweak it for their local environments and, of course, kibitz.

It might be, though, that LC has even more to gain from such an approach. It is conceivable that a tool such as HILCC could provide a platform with which to model revisions or perhaps even a complete overhaul of LCC. (Surely someday we deserve an LCC21 or at least an LCC22!) A key benefit would be that LC (and other) catalogers would be able to continue using "old LCC" indefinitely while schedules were being modeled, discussed, revised, discarded and reworked. HILCC itself provides the "switching system" or crosswalk between the old and new. One can imagine many different development and implementation possibilities worth exploring.

In conclusion, then, as libraries collectively hunker down and face the prospect of trying to manage and provide access to an exploding number of electronic resources with, as we are warned, fewer and fewer staff resources; and as we try to distill for ourselves and our parent institutions what it is exactly that we have to offer the information world apart from offsite storage facilities and eresource license management; let me suggest that classification, used alone and in combination with keyword searching, appears to be well suited to resource discovery within a broad heterogeneous information environment and, if developed intelligently, could be a key tool in our evolving knowledge-based environment.

[Note: The author would like to thank Bob Wolven and Patricia Renfro, both of Columbia Libraries, for useful and timely feedback on this report.]

Appendix A: Sample Electronic Resource Profile (1997)

The metadata record for "Art Index" included the information displayed in the About profile below, and was designated by the library selector to appear under two subject categories from an early prototype of HILCC, namely:

1. **Art, Architecture & Music**
2. **General & Interdisciplinary**

About: Art Index

Connect via WWW

Coverage

Subject Description: 2 part index to art periodicals in Western European languages: Part 1 (420 journal titles): from 1929-1984; Part 2(376 journal titles): from 1984 to the present. Beginning in 1994, abstracts are included. After the late 1990s, some full text is available (lacking illustrations).

Subject Keywords: archaeology, architecture, art history, city planning, computer graphics, crafts, film, folk art, graphic arts, industrial design, interior design, landscape architecture, museology, painting, photography, sculpture, television, textiles, video

Years: Part 1: 1929-1984; Part 2: September 1984 to the present with abstracts starting in Spring 1994, and some full text beginning in the late 1990s.

Content Type: Citations, abstracts, and some full text.

Updated: Part 2: updated monthly.

Regions: U.S. and international.

Sources: Part 1: 420 journals indexed; Part 2: 376 journals indexed

Publishing Information

Provider: *Content:* H.W. Wilson Company.
Electronic Presentation: H.W. Wilson Company.

Copyright: H.W. Wilson Company.

Availability: ***This resource is available only to current faculty, staff and students of Columbia University.*** More information about LibraryWeb access

Searching

Search Methods: Subject, author, title and other access points.

Related Resources

- After you have found citations, you will need to do a title search in CLIO to see whether Columbia owns this publication, for example, `t=education update` If you can't find it in CLIO, speak with a reference librarian.

Appendix B: Hierarchical Interface to LC Classification: Sciences

The Sciences section of HILCC (as of July 2000), constructed from portions of LC Schedule Q, S and G.

General Sciences		Q 0-390.999
Animal Sciences		SF 0-190.999, SF 191-275.999, SF 360.99-599.999, SF 600-1100.999, SH 0-400.999, SK 350-579.999
Astronomy & Astrophysics		QB 0-991.999
Biology (all)		*[composite of all Biology]*
	General	QH 301-504.999, QH 506-539, QH 573-705.999
	Biophysics	QH 505-505.999
	Microbiology	QR 0- 502.999
Botany		QK 0-989.999
Chemistry (all)		*[composite of all Chemistry]*
	General	QD 0-65.999
	Analytical Chemistry	QD 66-145
	Biochemistry	QD 415-449
	Crystallography	QD 732-999.999
	Inorganic Chemistry	QD 146-97.999
	Organic Chemistry	QD 198-414
	Photochemistry	QD 700-731.999
	Physical & Theoretical Chemistry	QD 450-624
	Radiation Chemistry	QD 625-699
Earth & Environmental Sciences (all)		*[composite of all Earth & Environmental Sciences]*
	Ecology	QH 540-572
	Environmental Sciences	GE 0-350.999
	Forestry	SD 0-669.999
	Geology	QE 0-996.599
	Human Ecology & Anthropogeography	GF 0-900.999
	Meteorology & Climatology	QC 851-999.999
	Natural History	QH 0-278.5
	Oceanography	GC 0-1581.999
	Physical Geography	GB 0-5030.999
Human Anatomy & Physiology (all)		*[composite of all Human Anatomy & Physiology]*
	Human Anatomy	QM 0-695.999
	Neurosciences	QP 351-495.999
	Physiology	QP 0-350.999
Mathematics (all)		*[composite of all Mathematics]*
	General	QA 0-74.999, QA 101-272.599, QA 299.5-939.999

	Mathematical Statistics	QA 273-299.499
Physics		QC 0-999.999
Plant Sciences (all)		*[composite of all Plant Sciences]*
	General	SB 0-450.899, SB 599-998.999
	Agriculture	S 0-972.999
Zoology		QL 0-991.999

Appendix C: Library of Congress Schedule Q (Selections)

A selection from LC Schedule Q showing detail corresponding to HILCC Science categories.

General Science			Q 1-390
Mathematics			QA 1-939
Astronomy			QB 1-991
Physics			QC 1-999
	Meteorology. Climatology		QC851-999
Chemistry			QD 1-999
	General (including alchemy)		QD 1-65
	Analytical Chemistry		QD 71-142
	Inorganic Chemistry		QD 146-197
	Organic Chemistry		QD 241-441
		Biochemistry	QD 415-436
	Physical and Theoretical Chemistry		QD 450-801
		Radiation chemistry	QD 625-655
		Photochemistry	QD 701-731
	Crystallography		QD 901-999
Geology			QE 1-996.5
Natural History. Biology			QH 1-705.5
	General Natural History		QH 1-278.5
	General Biology		QH 301-705
		Biophysics	QH 505
		Ecology	QH 540-549.5
Botany			QK 1-989
Zoology			QL 1-991
Human Anatomy			QM 1-695
Physiology			QP 1-981
	General		QP 1-345
	Neurophysiology and Neuropsychology		QP 351-495
Microbiology			QR 1-502