



BYTES (Books You Teach Every Semester)

**Final Report to the Andrew W. Mellon Foundation
Submitted on behalf of NERL
(The NorthEast Research Libraries consortium)
Ann S. Okerson and Paul Conway
Yale University Library
6 July 2001**



BYTES (Books You Teach Every Semester)

Final Report to the Andrew W. Mellon Foundation

Ann S. Okerson and Paul Conway

Yale University Library¹

INTRODUCTION

Common sense suggests that faculty at American colleges and universities will frequently be found to be teaching courses that draw upon some or many of the same or very nearly the same readings. It should be possible to define a hierarchy of texts -- "Books You Teach Every Semester" -- that turn up over and over again on collegiate reading lists. We have had the benefit of a planning grant to investigate the validity of this common sense assumption. Our goal, at least in the first instance, was not to explore the content of American academic culture, but rather to ask a pragmatic question: Does the coincidence of titles taught carry any implications for library management of that frequently taught information and new partnerships with publishers, faculty, and students?

Our hypothesis was that there would emerge from the BYTES study a list of titles so regularly taught that it would be valuable to find improved or alternative sources of supply for students -- sources that might transform the way in which a growing number of libraries provide electronic reserves services and readings. Possibilities for such transformation might include cooperative digitization by libraries of frequently assigned readings, collaboration with publishers (for material under copyright), or coordination of a larger joint project designed to identify, track, and manage a growing collection of such titles. Much would, of course, depend on the particular identity of the texts that emerge out of such pilot project to study frequently assigned readings.

BACKGROUND

The NorthEast Research Libraries consortium (NERL)² is a group of libraries created to collaborate on the acquisition of and access to digital information. On 8 October 1999, representatives of nine institutions that are members of NERL met with a representative

¹ Our thanks also to Joan Emmet, NERL Program Support Librarian, and to Paul Seeman, Research Assistant, for their considerable efforts on the BYTES project.

² See: <<http://www.library.yale.edu/NERLpublic>>

of the Andrew W. Mellon Foundation to discuss what research libraries need to understand about the rapidly emerging e-book phenomenon.³ Because many NERL libraries are interested in offering electronic books to their users, our interest was in the growing electronic availability of current imprints of more or less standard, scholarly, academic books. Our initial premise was that the way to understand the development of e-books was to license the output of netLibrary.com⁴ (currently the largest purveyor to libraries of electronic books from many academic and trade publishers). A consortial license, we thought, would allow us to learn more about the conditions under which readers will use electronic books. As the meeting proceeded, the group quickly agreed that the chief limitation of a usage-study approach is the dependence on an existing collection of e-books that falls far short of critical mass and that does not necessarily map to our readers' needs.

The group chose instead to take a pro-active approach, agreeing that research libraries must identify their users' needs, rather than react to the existence of randomly (though increasingly numerous) digitized books. NERL members all believe that more and more books will be converted into electronic form in the near future -- as the increasing level of grant support and venture capital for such projects signals. The October 1999 working party saw an opportunity to affect the marketplace as it emerges, rather than simply to react to what publishers and vendors may choose conservatively to do. The response from the Mellon Foundation to this type of approach was initially encouraging, and so the group developed a more concerted proposal.

Our proposal asserted that it is critically important for academic libraries to begin to influence publishers and vendors to digitize the books that students and teachers most need, rather than simply leave it to the marketplace to dictate commercial offerings. If the purveyors of scholarly books cannot be encouraged to create digital versions of needed texts, moreover, there may be a digitizing role for the libraries themselves. The key stakeholders in an endeavor that carefully digitizes scholarly books -- and those with the most to gain from its success -- will be the students and faculty who are frequent users of the holdings of academic libraries. Librarians are well positioned to bring desired information to their campus readership. Accordingly, NERL requested from the Andrew W. Mellon Foundation support for a pilot study with the expectation that a successful project would help ensure that academic libraries retain a central role in the selection and eventual digitization of critical monographic literature.

METHODOLOGY

A small steering committee from the nine institutions guided and oversaw the BYTES Project through its various stages. Members of this steering committee were also responsible early on for championing this project on their respective campuses, including

³ The institutions represented were Columbia University, Cornell University, Dartmouth College, Harvard University, New York University, Syracuse University, University of Connecticut, University of Massachusetts (Amherst), and Yale University. This group is a subset of the eighteen libraries that comprise NERL.

⁴ See: <www.netlibrary.com> for a fuller description of this company's products and content.

conversations with, wherever possible, faculty, staff, and any other key players. Members' particular hope was that the project would benefit from some faculty engagement.

BYTES proposed to conduct a prototype pilot project in which it would compile the participating libraries' reserve lists from the two full academic terms in calendar year 2000. BYTES did not intend to be an e-reserves project *per se* (i.e., it did not seek to evaluate electronic reserves services). Rather, the reserves metaphor is useful in signaling potentially heavily used materials. Furthermore, the fact that libraries maintain in small, sequestered collections such reading lists (and books) made reserves a good starting place.

Ideally the study would be conducted for two to three years, in the same semester of each year to catch all courses offered repeatedly as well as those courses only offered once every few years. However, a two to three academic year period of research would have prolonged the effort outside the range of timely recommendations. To keep the project to a manageable size, the group limited its investigations to humanities courses in two broad fields commonly taught to undergraduates: history and English-language literature. The assumption was that these disciplines would deliver a high yield of titles for study. English and history are subjects in which books, portions of books, and standard periodicals comprise the recommended course readings and in which aggregate enrollments are typically among the highest on any campus. Based on an initial sampling of Cornell, Syracuse, and Yale libraries' e-reserves lists in history and in English, we estimated that the total number of titles for the nine NERL libraries might be as high as 4,800 (history) and 4,000 (English) per semester. In the end, our sampling proved to be high, and the total number of books identified over the two semesters was closer to 13,000 than to 17,000.

Participants in the project realized that a small planning grant and relatively brief planning period could not test numerous sources and possibilities for works taught in the classroom. They also expected to include, at a future date, materials beyond those placed on library reserve. If the first test phases proved successful, later phases could introduce additional content possibilities, such as titles in syllabi, titles that faculty would like to include in required readings but could not easily locate, or perhaps bookstore purchases and books checked out of the respective libraries. The steering committee speculated that faculty may assign readings partly out of pure academic interest, but as well out of a sense of what is available for purchase and in libraries. Any experiment of this kind that began to affect the availability of titles would very likely have an impact on the selection and assignment of readings.

The BYTES steering committee began by establishing the format and the fields into which bibliographic records for the materials would be entered. Each of the nine reserves librarians (or other lead persons as assigned within their institutions) would enter their e-reserves selection into the specified database format, identifying their institution, each specific course, the type of reading (whole book, chapter(s) of a book, journal, etc.), number of copies on reserve, and other key data. At the close of the first semester, two

representatives from each of the nine institutions met for a day to review the methodology, discuss early findings, make a preliminary analysis, and assign additional tasks (such as a small "market" survey) to be carried out during the second semester. To support the data gathering effort for the Fall 2000 term, the analysis team prepared detailed instructions for gathering data about reserve items and formatting properly in the data template. **Appendix 1** presents the data gathering instructions.

The nine lists in each of two semesters were unified in the NERL office, under the supervision of the NERL and Yale Library staff, after which staff undertook the pertinent authority and query work to identify similarities, differences, and overlaps among courses and among specific readings in the nine institutions. Availability in electronic form was also researched for journals identified in the study. Although some important book titles have already been retrospectively digitized, identifying these becomes nearly impossible because there is no single "Electronic Books in Print."

After the close of the second semester, the group met again to review, analyze, and begin to develop the final report.

PROJECT PLAN

The BYTES project aimed to (1) identify a substantial core of scholarly books of high use and pedagogic importance for eventual digital access and (2) develop an effective methodology for identifying such books in the future. Therefore, certain steps and key questions had to be identified in the pilot project:

1. *Compiling the list(s).* A key outcome of the BYTES planning project included lists of books compiled over two academic semesters. As noted above, the lists were combined into one database and standards/authorities work was performed to advance the subsequent analysis.
2. *Identifying similarities, differences, and overlaps.* The Group asked such questions as, "What can be learned about the extent – or non-extent – of overlap in courses and readings, especially with regard to potential digitization of book materials? If there is little to no overlap in readings across institutions, are there patterns in the content data that might indicate potential for such overlap if libraries and suppliers made certain types of content available electronically?" BYTES hoped to consider whether synergy existed between already-digitized e-journals and book readings. Do e-journals help to suggest e-book candidates so that the combination will begin to create some kind of critical mass in given disciplinary subfields? This proved to be too sophisticated a question, however, for a project limited in time and scope.
3. *Testing and reviewing the methodology.* The group asked whether analysis of e-reserves lists is the most expeditious way to identify materials for digitization? Are there better or more efficient ways to generate this type of information? Is the information worth generating, after all? If this type of analysis were to

- become a routine and ongoing effort, have we identified a good way in which to carry it out?
4. *Market survey.* This involved the possibility of testing the notion of digitizing key books with selected faculty and students; of seeking to gain some insights into the extent to which users would find the digitization of the books useful; of trying to define or quantify project information in a meaningful way. How might faculty be induced to become stakeholders in a larger project project? What is the rationale for placing whole books on reserve lists? Are the whole books required reading? Sections of them (but it's more straightforward to put the entire book on reserve)? Might faculty assign books differently if they were online? Only a limited number of such interviews were conducted within the timespan of this project.
 5. *Digitizing options.* The project hoped to evaluate and prioritize the options for converting frequently assigned books into digital form. What are the economic, legal, and bibliographic requirements for doing so? Can the prospects for such conversion be improved in some way by the nine institutions cooperating operationally or financially? Ought libraries to lobby or contract with vendors to digitize the books? When, if ever, should the libraries undertake this work themselves? Above all, can such a digitizing project based on BYTES be put into production? Answers to some of these questions are, at best, suggested in our report.
 6. *Prepare a report* for the Andrew W. Mellon Foundation, describing and analyzing the BYTES pilot project findings, with the expectation of further distribution among all 18 NERL members and beyond, as appropriate.

RESULTS

Data Cleanup and Standardization

The analysis team received bibliographic information from the nine participating institutions in the form of discrete Microsoft Access databases. The methodology for creating the individual databases varied. Some institutions extracted bibliographic information directly from the library's online catalog. Other institutions keyed information from reserve lists directly into the database template. Still others used a combination of extraction and keying to create a full database for each semester. The analysis team created a single combined database for each separate semester by merging the information from the nine participant databases.

Cleanup activities focused initially on completing missing information for individual records, standardizing the content of data cells, and correcting typing or formatting errors. In order to prepare the database for querying, the analysis team added several entirely new fields and edited the content supplied by participants for uniformity of syntax. This work was broken down into seven phases.

1. *Additional Data Fields*

In order to assess various levels of similarity among items in the BYTES database, it was necessary to add several new fields to each records:

- a. **Standard Title:** The variety of methods used to generate each school's records introduced syntax heterogeneity that makes comparison across schools virtually impossible. For example, the same title might have been entered by three different schools in three markedly different ways; "Nature's Metropolis: Chicago and the Great West"; "Nature's Metropolis: Chicago and"; "Nature's Metropolis: Chicago and the Great West/William Cronon". Although these three items are identical titles, to a database, they are entirely different. In addition, variant editions of similar works often carry widely differing titles. Creating a standard title field in each record allowed cross-school title searching while preserving in the title field what might prove to be the useful quirks of each school's original submission.
- b. **Additional Authorship Information:** While the author field needed to be standardized in order to allow easy searching, part of that process involved excising information specifying secondary authors, translators, and the like. Since this information may prove to be valuable, it is included in the Additional Authorship Information field.
- c. **Publisher:** The need for this field emerged from the BYTES participants meeting in July 2000. The fall 2000 data gathering template included a separate field for publisher information. Data cleanup for the spring 2000 data involved searching each of the thousands of ISBN numbers in one or more bibliographic databases and adding publisher information to a new field for each record.
- d. **Editor:** Early discussions among the BYTES analysis team included the suggestion to distinguish edited works from those written by individual authors or teams. A field was added to the combined database and partially populated with available data.

2. *Standard Numbers*

Standard numbers distinguish among variant editions of similar works; without including them in the database, the analysis team could not calculate the number of editions per work and could not perform the most conservative type of cross-school duplication. The spring 2000 data gathering template specified the need for an ISBN for each title. Since books published before 1970 rarely have an assigned ISBN, the analysis team decided that OCLC accession numbers best filled the need for standard numbers. Over 3,500 standard numbers were added to the combined databases. When more than one OCLC record matched the information available for a given book, the search was limited to books held by the school in question. In a handful of cases, it was necessary to add the OCLC number for the book reported to be held by the greatest number of participants.

When standard numbers were available for both the cloth and paperback versions of a book, we chose the standard number for the cloth version.

3. *Separation of Concatenated Fields*

Several MARC fields – the title and publisher fields, for example – combine elements that must be separated prior to analysis. Data extracted directly from integrated library systems often included title fields that ended with author statements and publisher fields that combined the place of publication and publisher name. Cleanup trimmed authors from title fields and separated publisher names from location statements.

4. *Completion of Publisher Field*

While the publisher field was partially populated as a result of phases two and three, many records lacked publisher information (local reserves systems do not necessarily maintain full bibliographic information for reserves titles). Much of this data was added by consulting union catalogs and adding publisher statements on an item-by-item basis. Once enough information had been added it was possible to establish which major publishers were represented by which ISBN prefixes. Taking advantage of these ranges sped the process up considerably.

5. *General Cleanup*

Once the standard number and publisher fields were shored up, each field was examined for syntactical quirks such as trailing punctuation that might make comparison difficult. For example, a search for duplicate values would consider the entries “Penguin” and “Penguin,” to be completely different because of the placement of the comma. Most of this work was done by importing fields into Word, searching for punctuation marks followed by paragraph marks, trimming ancillary punctuation, visually checking each record, and pasting the finished product back into the database.

6. *Standardization of Author Fields*

As with the title and publisher fields, data in the Author fields reflects variations in local cataloging practices, thereby making comparisons difficult. The Author field was standardized into a new, parallel field by using the same techniques that were applied to the Standard Title field.

7. *Addition of Standard Title Field*

The standard number field allows edition-by-edition comparison of books in the database but does not support an assessment of title overlap across participating institutions. The addition of a standard title field allowed such comparison to proceed with a fairly high level of trustworthiness.

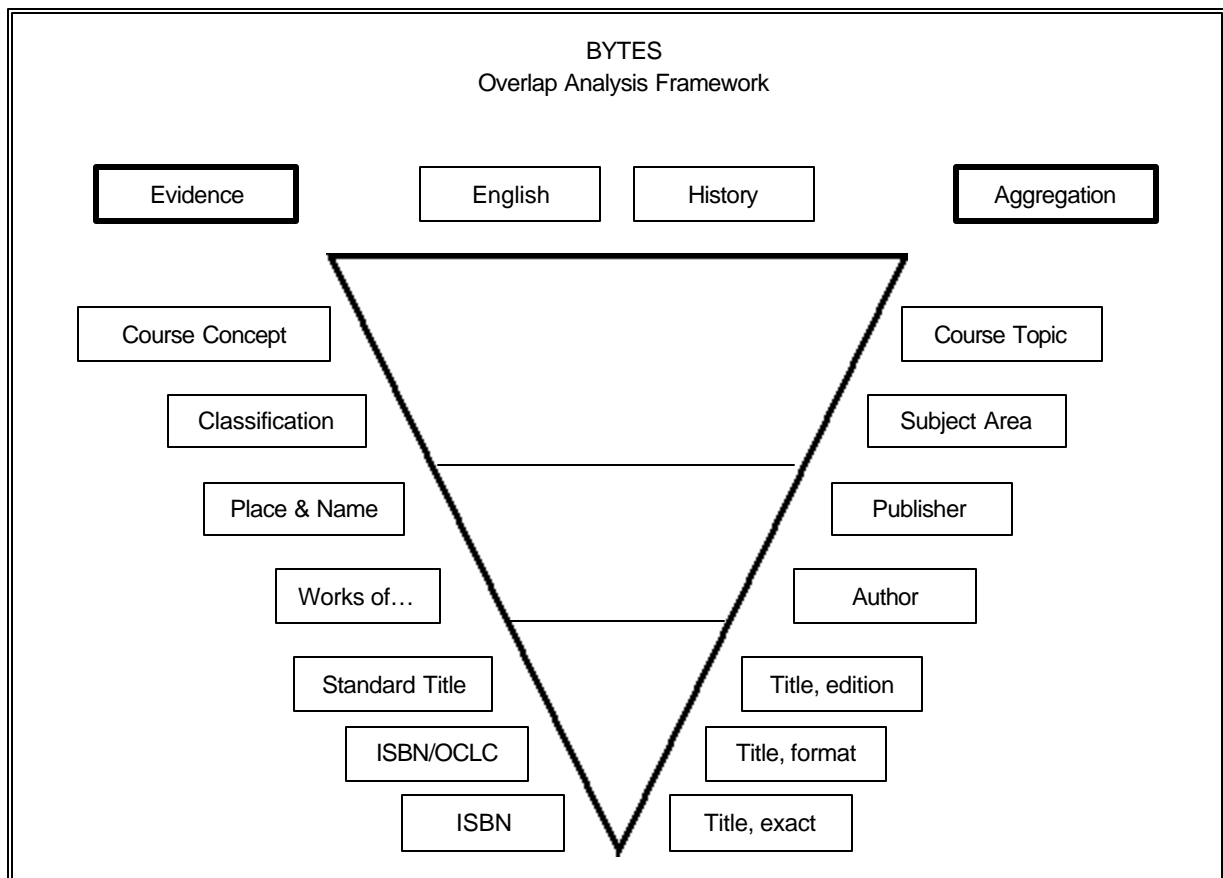
Logically, standard number duplication ought to be a subset of standard title duplication; the first step toward establishing a standard title field was limited to those records that shared standard numbers. When an existing title statement sufficed for our purposes it was used as the standard title. OCLC WorldCat was

consulted for guidance in creating standard titles of the new cases where no existing title statements were complete and error-free.

The next steps involved searching for matches in the Title fields and entering standard titles appropriately. The entire database was then sorted on the title field and visually scanned for duplication. Whenever the slightest discrepancy existed between records, union catalogs and the OPACs of individual schools were consulted to determine whether such discrepancies reflected the presence of entirely different books or variant editions of otherwise similar works. After # seven (above) was completed, a search-and-sort process was performed on the author field in a effort to uncover widely variant iterations of similar titles. Finally searches were conducted across several combinations of author, title, publisher, and published date fields to insure that no similar titles were excluded from the process.

Data Analysis Framework

The proposal to the Mellon Foundation specified in general terms the research questions that would be posed to the data gathered in the Spring and Fall terms of 2000. The research areas were not defined specifically in the proposal because the design of the database was determined only after the project had begun. The principal foci of the analysis of reserve lists were on the overlap of titles among participating schools and on the clustering of titles by author, publisher, call number, and date of publication. The analysis of overlap turns on two related issues: the definition of what constitutes a unique item and the definition of overlap patterns based on the concept of aggregation. The impact of both of these issues is illustrated in **Figure 1**.



The figure outlines an analysis concept that increasingly aggregates information from the database. In the figure, an inverted triangle represents an expanding concept of overlap. The right side of the figure defines the increasing aggregation. The left side describes the data required to undertake the overlap analysis. The framework assumes that the concept of overlap can be defined as a continuum from exact match of a discrete item to differential clustering at the level of intellectual disciplines. As one moves along an aggregation continuum, one would expect that the amount of overlap would increase while the precision of overlap would decrease.

The analysis team tested the intellectual viability of the overlap model with the BYTES steering committee. The group suggested that the following four aggregation types promised to generate the most interesting and useful findings.

- 1. Standard Title**

At the point of the triangle are those titles in the combined database for which ISBN numbers match exactly. At the next level of aggregation are titles that may exist in different formats (hardback or paperback, for example) or the same item published in two different countries (US or UK, for instance). Varying ISBN or OCLC standard numbers for a given title is the evidence for overlap. At a third level of aggregation are the same titles in variant editions, including reprints. The cleanup process clustered items that vary in ways other than author and title under the idea of the “standard title.” For overlap to exist, a given title must appear on the reserve lists of at least two participants, either as an exact match, a match that varies by format, or a match that varies by edition or characteristics other than author and title.

- 2. Author**

Beyond title matching, it is possible to conceive of clustering titles by author. At this level of aggregation, overlap would exist, for example, if two or more institutions held on reserve any of the works of William Shakespeare, even if overlap at the title level did not exist. The evidence for overlap is provided through a sort of the author field in the database, the content of which has been edited to provide for uniform spellings of first and last names.

- 3. Publisher**

At the next level of aggregation is the cluster of works by publisher. In this case, clustering occurs when two or more institutions have placed on reserve the works of a given publisher, even if there is no overlap at the title or author level. The evidence for clustering is contained in the publisher field of the bibliographic database. The field may be sorted without regard to country of publication.

- 4. Subject Matter**

Increasing aggregation at the level of subject classification of a given work yields higher levels of clustering among institutions. For the Fall 2000, the combined database included call number information for a substantial subset

of the titles submitted by the nine schools. Evidence of clustering is provided by truncating the call number to the primary Library of Congress class code (first letter). For overlap to occur at the level of subject matter, two or more institutions must hold titles on reserve in the Fall 2000 terms that share primary LC classification numbers.

Description of the Analysis Database

Each of the nine schools that participated in the study contributed data for nearly 13,000 book titles and an additional 205 journal titles to the combined database for both the Spring and the Fall 2000 terms. The size of the database as a combined entity as well as the size of the contributions of the individual schools makes it possible to draw inferences from the database about the shape of reserve activities in the participating schools. Because the nine participating institutions constitute a self-selected group, it may not be possible to project trends to the broader community of research libraries in the North East region. **Table 1** provides summary statistics that describe the scope of the collected data.

Number of Schools Represented:	9
Number of Courses, combined:	972
History:	590 (60.7% of total)
Spring 2000:	304
Fall 2000:	286
English Literature:	382 (39.3% of total)
Spring 2000:	194
Fall 2000:	188
Number of Book Titles:	12,933
Spring 2000:	6,682 (51.6% of total)
Fall 2000:	6,251 (48.4% of total)
Imprint Dates:	
Books published since 1926	12,778 (98.8% of total)
Spring 2000	6,589
Fall 2000	6,189
Books published before 1925	155 (1.2% of total)
Spring 2000	93
Fall 2000	62
Number of Authors:	7,162
Spring 2000:	4,333
Fall 2000:	4,104
Number of Frequent Publishers (>50):	51
Number of Book Titles Represented:	8,468 (65.5% of total)
SPRING 2000 DATA ONLY:	
Number of Journal Articles:	402
Number of Journal Titles:	205
Titles Available Electronically:	129

Table 1: Description of the Analytical Database

The table shows strong parity in the number of books on reserve during the two semesters and within the two disciplines under study. Just over 60 percent of the total number of titles are listed in courses of the disciplines of history, while the balance are titles assigned to courses in English language and literature. Most of the items on reserve during 2000 remain under the protection of the United States copyright law. Only 1.2 percent of the items included in the study were published before 1925. The limited scope of the study did not provide for the assessment of published content that may be in the public domain or where copyrights may have been retained by the authors and not ceded to the publishers.

Journal articles represent proportionately few of the items on reserve in the combined nine schools of the study. In part, the absence of journal material in the analysis database is an artifact of the way reserve materials are managed; few journal articles are cataloged in the online catalogs that serve as the source files for most of the items in the database. Additionally, journal materials included in some 40 course packs and other aggregated collections of readings had come to our attention but, for consistency's sake, these were not included in the database.

Duplication of Books Among Institutions

A key research question in the BYTES Project turns on the definition of the potentially elusive concept of “duplication.” How frequently two or more schools hold a particular book title on reserve in a given semester or a given year in part determines the extent to which it is possible to identify a corpus of books that may be needed in digital form. Books exist in several forms and formats. Books are recorded in bibliographic databases to varying degrees of uniformity. A given published book may exist through multiple printings and multiple editions, which may or may not be the result of substantive editorial work. Finally, a given book may be published by competing firms with slight variation in title. Variations across title, format, and editorial life could, on the one hand, result in many versions of essentially the same work. Alternatively, it may be desirable obscure slightly subtle (or potentially meaningful) variations across published works in order to identify duplication that is conceptually, if not rigorously true.

The project chose to mitigate variation across format and cataloging practices by defining the concept of “standardized title” as the unit of measure of title duplication across and among schools. A “standardized title” is a bibliographic entity that may vary in how it is recorded in the reserve lists of various schools in terms of spelling or format of title; in the presence or absence of leading articles (e.g., The, A, An); in variations in the ISBN reflecting different formats (paperback/hardback); and in multiple editions of essentially the same work. Titles entered by contributing schools in these different ways were collapsed during data cleanup phase into a new field in the database called “Standard Title.” No original data were deleted.

The most narrowly focused incidence of overlap occurs when two schools have the same title on reserve during a single semester. **Table 2** shows the number of titles shared by a given pair of schools. For example, Harvard and the University of Connecticut share 49

titles on reserve in the Spring 2000 semester. The lowest overlap is between Syracuse and University of Massachusetts, Amherst (4 titles). The greatest overlap, is between Harvard and Columbia (79 titles).

For the combined Spring and Fall terms, the pattern of duplication across pairs of schools remains virtually the same. In general, duplication across schools in terms of specific titles is very small. It is possible to speculate on the reasons for such low numbers. First, materials on reserve parallel the course offerings of a given school in a given semester. High levels of duplication by pairs of schools would require a similarity of course offerings in a given semester – an unlikely event, given the specialized focus of many teaching faculty. Second, higher duplication may be most usefully obtained by building a database of reserve materials over a two or three year period. A longer analysis period would capture information on most of the courses offered periodically on a multi-year rotation. Third, nearly exact duplication at the title level sets a fairly high bar for an analysis that is essentially a large sample of the possible materials available to students on a controlled or limited circulation basis.

Table 2. Title Duplication Between Pairs of Schools

Cornell	Dartmouth	Harvard	NYU	Syracuse	UConn	UMass	Yale	
33	35	79	17	31	41	13	52	Columbia
	13	42	9	16	28	14	32	Cornell
		23	15	6	15	9	15	Dartmouth
			25	31	49	24	48	Harvard
				7	12	9	14	NYU
					11	4	9	Syracuse
						14	16	UConn
							19	UMass

By shifting the analysis away from duplication of titles across schools to duplication at the title level independent of specific schools, another pattern emerges from the combined Spring/Fall 2000 database. **Table 3** shows the extent to which various schools share standardized titles on reserve. For example, 198 discrete titles exist on reserve at three of nine schools reporting, for a total of 594 volumes (4.6%) of the 12, 933 included in the combined database. No titles are on reserve at seven, eight, or nine schools simultaneously. Some 10,275 titles (79.4%) are on reserve at only a single school during the Spring or Fall 2000 terms combined. The fact that 2,658 volumes (20.6%) are shared on reserve at some level through calendar year 2000 is significant. This number represents the maximum title overlap found in the study at the title level.

Table 3: Overlap of Titles Across Schools

Number of Schools	Distinct Titles	Total Volumes
6	11	66
5	16	80
4	65	260
3	198	594
2	829	1658
1	10275	10275
		12933

It is important to note that a 20 percent overlap figure includes duplicate copies of a given title on reserve in a single school. For example, of the 16 discrete titles shared by five schools, five titles may be five copies of the same work, while the remaining eleven titles are single titles on reserve simultaneously at five schools. The analysis team decided to retain duplicate copies in a single school in this portion of the analysis. The rationale for this decision turns on the view that multiple copies in a single school is a clear statement of need for a particular title.

Appendix 2 is a list of the separate titles that are on reserve in two or more research libraries in the study during the 2000 school year. The table is a multi-faceted display. Organized alphabetically by standard title name, the list shows the author, imprint date, publisher, and standard number assigned to the title. The standard number is either the ISBN or the OCLC number for an identical title located in WorldCat. The list's most important value is to display how variant manifestations of the same work are held on reserve by various schools that participated in the study. For example, the 1995 edition of Philip Curtin's *African History* (Longman Press) is on reserve at Harvard and NYU, while the 1978 edition of the same work, published by Little, Brown, is on reserve at Dartmouth and Harvard. The list has many examples of this phenomenon, including two commercial editions of Joseph Conrad's *Heart of Darkness* held on reserve by four separate libraries.

Conceptual Aggregation Within the Analysis Database

Authors

Data on the authors of 12,933 discrete titles from nine academic library reserve lists provide the basis for an assessment of the usefulness of viewing books in terms of clusters of writers rather than as separate entities. Table 4 shows the extent to which the nine schools have multiple works by a given author on reserve in the Spring and Fall 2000 semesters. Only 231 books in the database have no specific authors. As might be expected, these items tend to be collected works or anthologies with no clear author or editor in the record. The remaining 12,702 titles in the database are distributed across some 7,162 separate authors. Of this author group, 1,275 (17.8%) have books on reserve in both the Spring and Fall 2000 semesters. Books written by the remainder are confined to one or another term.

The cluster distribution of authors across the database is striking in two dimensions. **Table 4** illustrates frequency of presence on reserve lists. The 80/20 rule does not quite apply to this table. Some 31 percent of the authors in the database account for just over 62 percent of the titles. The overall distribution is varied, however. For example, 11 authors have 12 works apiece for a total of 132 items on reserve. A total of 41 works of a single author are on reserve. Some 4,921 authors are represented in reserve systems by a single title in both semesters combined. This table does not factor out duplication of works by a given author within a single school. Harold Bloom is the author with the most works on reserve in the year 2000. The Yale Professor has 86 instances of 31

Table 4: Clustering of Titles by Author

Distinct Authors	Number of Titles	Total Volumes
No Author Listed	231	231
1	86	86
1	58	58
1	41	41
2	29	58
1	28	28
1	27	27
1	26	26
3	23	69
1	21	21
1	20	20
2	19	38
4	17	68
3	16	48
4	15	60
3	14	42
5	13	65
11	12	132
15	11	165
24	10	240
25	9	225
39	8	312
42	7	294
69	6	414
134	5	670
221	4	884
436	3	1308
1191	2	2382
4921	1	4921
7162	TOTAL	12933

separate titles (heavily anthologies) on reserve in six of the nine schools in the study. Other popular authors are William Shakespeare, William Faulkner, and Edward Said.

A second dimension of analysis of authors in the database views the overlap of authors across schools. **Appendix 3** is a list of authors (and their represented titles) clustered by the extent of overlap. The list shows that two authors (Edward Said and William Shakespeare) have a total of 17 titles on reserve at seven of the nine schools in the study. Six schools have 59 works by 4 authors on reserve. Five schools have 45 works by 13 authors on reserve. Four schools have 164 titles by 47 authors on reserve in the year 2000. Some 278 works by 123 authors are on reserve in three schools during the combined Spring and Fall terms.

Clustering the titles in the combined Spring/Fall 2000 database in terms of works by single authors does not necessarily yield useful findings. Without significant assessment

of lists of works in author clusters, it is not possible to make qualitative assessments of the works on reserve or to distinguish between single works, anthologies, and edited or abridged versions. The database does not yield information about the presence or absence on reserve lists of authoritative or scholarly editions of a work.

Publishers

Perhaps the single most striking findings of the study derive from assessment of information on publishers. The study found that a total of 375 separate publishers produced the nearly 13,000 books on reserve in the nine schools during 2000. University presses play a particularly significant role in supporting undergraduate teaching in history and English literature. **Table 5** illustrates this role by displaying the most frequent publishers of titles on reserve in Spring and Fall 2000 in nine schools and indicating university presses with an asterisk. This list of 51 publishers accounts for 65.5 percent (8,468) of the total group of titles in the database. Of the ten most frequent publishers on this list, eight (80%) are university presses. Of the twenty most frequent publishers, eleven (55%) are university presses. The ten most frequently appearing university presses account for 3,789 titles or 29.3 percent of the total number of titles in the database. University presses account for 4,613 titles (54.5 %) of the publishers with fifty or more titles.

Table 5: Publishers of 50 or More Titles, 2000

Rank	Count	Publishers of 50 or More Titles	University
1	921	Oxford University Press	*
2	528	Cambridge University Press	*
3	458	Harvard University Press	*
4	456	Penguin	
5	379	University of California Press	*
6	341	Princeton University Press	*
7	309	University of Chicago Press	*
8	304	Norton	
9	295	Cornell University Press	*
10	263	Yale University Press	*
11	238	Routledge	
12	212	Harper	
13	203	St Martin's Press	
14	200	Knopf	
15	185	Vintage	
16	167	Johns Hopkins University Press	*
17	164	Random House	

18	128	Longman	
19	128	University of North Carolina Press	*
20	121	Stanford University Press	*
21	118	Indiana University Press	*
22	118	Macmillan	
23	114	Doubleday	
24	114	Houghton Mifflin	
25	112	Blackwell	
26	105	Viking	
27	104	Prentice-Hall	
28	99	Columbia University Press	*
29	92	Basic Books	
30	92	University of Wisconsin Press	*
31	91	Harcourt, Brace, Jovanovich	
32	86	Pantheon	
33	83	Hill and Wang	
34	80	University of Illinois Press	*
35	77	Duke University Press	*
36	74	Little, Brown	
37	74	McGraw-Hill	
38	73	Rutgers University Press	*
39	73	Scribner's	
40	69	Methuen	
41	64	Farrar Straus Giroux	
42	60	Verso	
43	59	Simon and Schuster	
44	58	Chelsea House	
45	57	New York University Press	*
46	56	Hackett	
47	55	University of Pennsylvania Press	*
48	54	New American Library	
49	53	Free Press	
50	52	Greenwood Press	
51	52	University of Minnesota Press	*

8468 Total Count of 51 Publishers

The implications of this finding form the basis for possible next steps of the BYTES project. The BYTES steering group that reviewed the study's findings agreed that some of the most fertile ground for follow-on research and demonstration projects associated involved the value of book content now owned and managed by university presses. The group discussed the various ways in which a group of research libraries could begin to develop relationships with content providers, particularly university presses. The three models considered briefly included:

1. digitization "just-in-case" of large bodies of content (from back lists of out of print titles?) based on expected need identified from ongoing BYTES-like analyses;
2. consortial copyright clearance activities conducted in tandem with university presses; and,
3. on-demand provision of all or parts of a work through consortium-focused digitization projects.

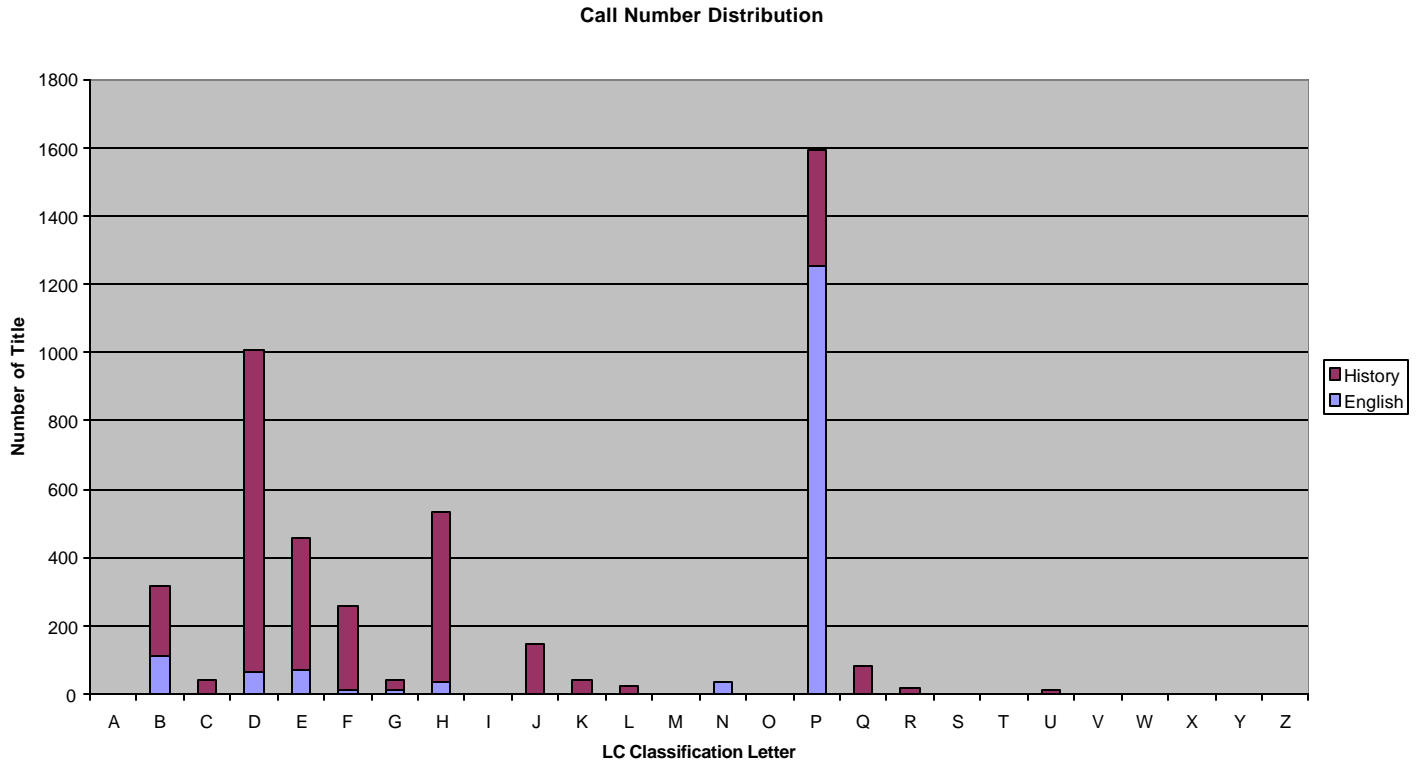
Concept/Subject

If money were no object in a research project and a study could extend indefinitely, exploring issues as they arise, then significant work could be done to identify and assess clusters of titles on reserve by concept or subject. In a one-year exploratory BYTES project, the analysis team decided to take a shortcut. Instructions to the nine schools that agreed to submit reserves data for the Fall 2000 term asked for the inclusion of call number information for each title submitted. For some schools this request presented insurmountable burdens because call number data would have had to have been compiled and entered by hand for each title on reserve. Other schools simply shifted their data extraction criteria and derived call number data from the library's online catalog. Ultimately, call numbers were submitted for 4,673 (74.7%) of the 6,251 Fall 2000 titles represented in the study.

Appendix 4 is a table that displays the distribution of titles across the Library of Congress classification scheme. The table shows familiar clustering of titles in some predictable ways. For example, titles classed in LC categories D, E, and F, the principal history classes, account for some 1573 titles, or just over half of the total number of titles placed on reserve for history courses. Perhaps more interesting is the fact that the other titles are distributed across the remaining LC classification categories. Additionally, some 340 titles classed in traditional literature categories (P) are on reserve for history classes. A similar cross-over or wide distribution pattern holds for materials on reserve for English classes, although the amount of cross-over is not as great.

Chart 1 is a representation of the call number distribution across a list of categories reduced to the letters of the alphabet. The Chart displays the number of titles classed into each call number category, broken out by the disciplines (History or English) for which the titles were on reserve.

The chart demonstrates the particularly strong presence on both history and English reserve lists for the nine research libraries of books classed within the literature category (“P”). Over one-third of the titles on reserve in Fall 2000 are literary works or works of literary criticism.

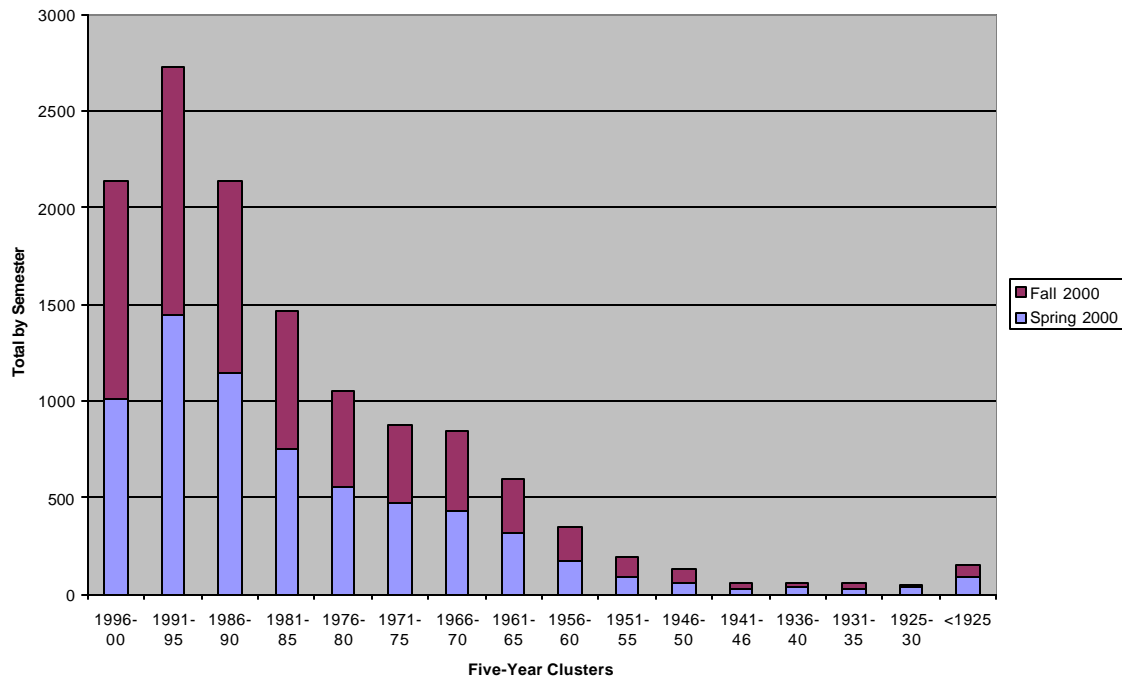


Imprint Date

The imprint date for each title on reserve during 2000 was recorded in the combined database. The imprint date reflects the actual date of publication of the item on reserve. In some small number of cases this may slightly obscure the date of original publication. For example, a 1995 reprint of a book originally published in 1895 is recorded in the database as 1995. The analysis team conducted a visual inspection of the combined database to detect the extent to which this practice distorts the findings; it is convinced that the incidence of modern reprints in the database is too small to undermine the overall patterns of publication date.

Chart 2 is a visual depiction of the publication patterns for books on reserve during the Spring and Fall terms of 2000. The table illustrates the extent to which recent publications dominate reserve lists of research libraries in the study. The patterns of Spring and Fall 2000 do not vary appreciably. The small spike at of the far right column of the chart is the total number of titles in the study published before 1925 and are, therefore, without copyright restrictions.

Publication Date Distribution



A closer look at publication dates emphasizes the patterns evident in the chart. **Table 6** clusters the titles in the combined Spring/Fall 2000 database in five-year increments and lists the total number of titles by semester and combined. The table shows that 2,143 titles (54.3%) on reserve in the nine research libraries were published from 1990 to 2000. At the other end of the spectrum, only 1.2 percent of the titles in the study are old enough to give confidence that they are out of copyright. In between these end points, fully 90 percent of the titles in the study were published since 1960.

Table 6: Publication Date Distribution

Half-Decade	Spring 2000	Percent	Fall 2000	Percent	Combined	Cumulative
1996-00	1013	15.20%	1130	18.10%	2143	16.60%
1991-95	1446	21.70%	1281	20.50%	2727	37.73%
1986-90	1144	17.20%	999	16.00%	2143	54.34%
1981-85	752	11.30%	712	11.40%	1464	65.68%
1976-80	556	8.40%	497	8.00%	1053	73.84%
1971-75	475	7.10%	405	6.50%	880	80.66%
1966-70	427	6.40%	416	6.70%	843	87.19%
1961-65	317	4.80%	277	4.40%	594	91.79%
1956-60	172	2.60%	173	2.80%	345	94.47%
1951-55	88	1.30%	110	1.80%	198	96.00%
1946-50	54	0.80%	75	1.20%	129	97.00%
1941-46	26	0.40%	36	0.60%	62	97.48%
1936-40	35	0.50%	29	0.50%	64	97.98%
1931-35	24	0.40%	30	0.50%	54	98.40%
1925-30	33	0.50%	19	0.30%	52	98.80%
<1925	93	1.40%	62	1.00%	155	98.80%
TOTAL	6655	100.00%	6251	100.00%	12906	100.00%

Journal Use

Apparently, journal articles do not play a large role in courses in history and literature and possible other courses in the humanities. During the Spring 2000 term, a total of 402 articles were on reserve in nine participating schools. These articles were drawn from 205 titles. **Table 7** illustrates the distribution of articles across journal titles during the Spring term. Given the relatively low incidence of journal titles in the combined database, the analysis team did not conduct an assessment of the Fall 2000 term. In the table, for example, one title accounts for 22 articles; two separate titles each have seven articles recorded in the database; fifteen titles each have three articles represented. Some 139 of the 205 journal titles (67%) had only one article on reserve. As with so many of the data distributions in the study, most journal articles exhibit a strong tendency to cluster in a few journal titles. One third of the journal titles account for approximately two thirds of the articles on reserve.

Table 7: Journal Titles and Articles, Spring 2000 Semester

Discrete Titles	Articles per Title	Total Articles	Cumulative Total
1	22	22	22
1	17	17	39
1	13	13	52
1	11	11	63
1	9	9	72
1	8	8	80
2	7	14	94
2	6	12	106
4	5	20	126
9	4	36	162
15	3	45	207
28	2	56	263
139	1	139	402
205	TOTAL	402	

An examination of the reasons for the relatively low count of journal articles on reserve was out of scope for this study. Speculation might center on three areas: (1) the lack of complete bibliographic control of journal articles in the source files used to generate the data from participating schools; (2) the tendency of faculty to cluster journal articles and other non-book reading matter in course packets; and, (3) the increasing availability of journal content in electronic form on university campuses. The analysis team did encounter up to forty course packets listed in the individual databases submitted by participating schools. The team decided to exclude course packets from the analysis due to the complexity of tracking adequate bibliographic information about the content of these packets.

Appendix 5 is a set of lists that characterize the journal titles on reserve during the Spring 2000 term. The first list is the titles sorted alphabetically by title. Additional lists describe the electronic availability of the content of journal titles. Included is information

on the aggregators that provide subscriptions to a given journal title and the date span of electronic availability.

The publishers of journal titles with articles on reserve in Spring 2000 are a diverse lot. Some 82 titles (40%) are produced by university presses. Another 42 journal titles (20%) are produced by non-profit organizations not directly affiliated with a university. Publishers in this group are mostly scholarly or professional associations. The remaining 81 publishers (40%) are commercial entities.

Electronic Availability

Looking ahead to the increasing availability of books in electronic formats, the BYTES project sought to anticipate trends in availability by examining the availability of journal content in electronic form. Steve McCracken and his staff at Serials Solutions generously offered to compare the list of 205 journal titles with reserve holdings in the Spring 2000 term against the extensive list of e-content aggregators tracked by Serials Solutions. The particular search and matching work done for the BYTES project is not a standard product offered by Serials Solutions, but instead was done as a special service to test the viability of broadscale matching.

Serials Solutions tracks the electronic content of some 69 aggregators, which provide licensed access to bundles of electronic journals. Examples of journal aggregators include Ebsco, Gale, ProQuest, and SwetsNet. Twenty-eight of the 69 aggregators on the Serials Solutions list have one or more journals listed in the BYTES database. As many as 164 of the 205 journal titles (78.5%) are available electronically in at least one aggregator's products. Typically a given title is available through an average of four aggregators. One title (*Journal of Women's History*) is available electronically through eleven separate aggregators.

The extent to which the BYTES journals are represented electronically varies widely by aggregator. The following seven aggregators have the greatest number of BYTES titles available in their bundled products. This list shows the number (and percentage) of BYTES journal titles provided electronically by the aggregator, the number of titles with relatively deep coverage, and the earliest date represented by the electronic version of BYTES titles in the aggregator's database.

Aggregator	Titles	Percent	Most Coverage	Earliest Date
Ebsco	117 titles	57.0%	33 titles	1984
OCLC	111 titles	54.1%	13 titles	1988
Gale	79 titles	38.5%	38 titles	1983
ProQuest	70 titles	34.1%	18 titles	1984
HW Wilson	46 titles	22.4%		1994
SwetsNet	44 titles	21.5%	16 titles	1995
Information Quest	35 titles	17.0%		1996

Of all the aggregators on the Serials Solutions list, JSTOR clearly went back the furthest in coverage time for any title listed with multiple providers for which JSTOR was one. The broadest time-span coverage for BYTES journals in JSOTR is 1880 to 1995 for *Science*, 1983 to 1996 for *Representations*, and 1895 through 2000 for *American Journal of Sociology*.

The bottom line of a simple assessment of journal content availability is that the breadth of coverage across journal titles needed for reserves is wide and increasing while the depth of coverage is, at the present time, far from adequate to support the learning needs of university students in the humanities. Additionally, barely half of the journal titles needed for undergraduate reserves are available from even the most comprehensive aggregators.

INTERPRETATIONS and RECOMMENDATIONS

There is a high interest on the part of those who have heard of the BYTES project in seeing the lists of titles and learning about the findings. Every reader may come to somewhat different conclusions. For the project team, several obvious points emerge:

1. *Reserve readings focus on secondary literature.* Given that the Project has been tracking reserve readings, it comes as no surprise that BYTES has identified far more "secondary" literature -- that is, modern scholarship and criticism -- than primary texts, which are the original works of literature that students more often purchase in paperback.
2. *The dates of the works cluster strongly towards the present.* Even when literary texts long out of copyright are taught, the unmistakable teacher preference for readings appears to be for recent editions. This has strong implications for any library or publisher initiated program to develop innovative forms of access: copyright holders, in many cases of them commercial publishers, will have to be players in any such initiative, and so far they have been very cautious players indeed.
3. *The books of American and British university presses account for 2/3 of the supplemental book readings assigned during the study period.* As has been indicated, this phase of the BYTES project was able to study supplementary (reserves), rather than required course readings. Such supplementary readings offer analysis, commentary, and interpretation of the historical and literary works that the primary focus of study. And since such analysis is undertaken more often in academia than in other settings; since academics gain recognition on the basis of their scholarly work; and because university presses are the dominant publishers of such cultural analysis, our finding about the key role of university presses in the undergraduate learning environment should have come as no surprise. Nonetheless, the extent to which teachers rely on university press publications was revelatory to all on the BYTES team. (The preponderance of a small number of the leading university presses suggests as well that those

enterprises have been successful both in finding their readers and in finding the authors to address those readers in ways that are widely welcome.)

4. *Pedigreed anthologies are important in undergraduate readings.* The abundance of titles listed under Yale/NYU literary critic Harold Bloom is an interesting indicator of an existing textual practice in the print publishing market. Many of Professor Bloom's titles are not his own original works of criticism, but come from a Chelsea House series that Bloom edited a decade and more ago. The series comprises dozens of volumes of collected scholarly and critical essays about major works of literature, from A.R. Ammons to William Butler Yeats. The existence and influence of this series suggests that the task of selecting and packaging critical literature for academic use was already recognized and plays an important part in the publishing system. The digital equivalent of such a process remains to be invented. The difficulties recently experienced by the editors of the *Norton Anthology of Theory and Criticism*⁵ offer a similar lesson -- selecting and compressing material for teaching within the traditional bounds of print is increasingly difficult and scholars and editors are increasingly restive.
5. *Works of highly regarded academics are important in undergraduate teaching.* For example, the list of most popular authors holds few other surprises. M.H. Abrams of Cornell is one of the most eminent of living literary critics and, at age 89, has held that eminence since at least the publication of *The Mirror and the Lamp* in 1953. Peter Brown of Princeton single-handedly ignited the burgeoning study of late antiquity in the English-speaking world with books published in 1967 and 1971 and has taught for most of that time at Berkeley and Princeton.
6. *The reserve reading lists indicate a high degree of customization and individuality in teaching* of what are in many cases standard undergraduate topics in history and literature. While Columbia University, one of the BYTES participants, has a highly structured program in the humanities, one that is required of all first-year students, there are no such comprehensive programs in the other eight participating universities. This variety is beneficial, for it introduces a certain amount of diversity into undergraduate education. At the same time, our study found that there is enough overlap, even within one calendar year in a limited number of neighboring institutions, that a case can be made for making certain works ubiquitously available in colleges and universities.
7. *Reserve services, both print and electronic, are labor-intensive.* Most librarians will agree about the need to find more efficient and effective ways to deliver such services. The labor to create sophisticated electronic reserves services, can pay off in that students are increasingly willing and able to access information online (even if they will print longer items for ease and convenience of use) and the readings become available anytime, anywhere. At the same time, electronic reserves services require the library to create and re-create readings each and

⁵ Scott McLemee, "Making the Cut," *Chronicle of Higher Education*, 4 May 2001.

- every semester, as well as to go through the same intensive permissions processes required in print for readings that are needed in multiple copies or for multiple semesters. A truly efficient e-reserves system will be able to draw upon extensive repositories of useful scholarly and teaching materials in electronic form. No such critical mass of digitized books exists to serve students and teachers.
8. *At this time, there exists no effective e-books business model that can support electronic reserves services in academic libraries, certainly not in any way that can scale to multiple courses over multiple semesters over the years.* On evidence from the BYTES study, we believe there is a need, not only for an e-book repository, but also for an appropriate business model to accompany it. An effective e-reserve system would be able to dip into such a repository for content and to pay a reasonable fee for any books or parts of books that are used in a class during a given semester. The library would require no ownership of the book (ownership is a not uncommon business model among e-book publishers, in which the print or print-plus price of a book is charged for each e-copy); nor would it pay an access fee in perpetuity for readings that are not desired in perpetuity. The models offered by netLibrary.com, Questia, and e-Brary, for example, do not meet classroom readings needs.
 9. *Faculty as teachers are at this time more interested in e-books than faculty as scholars.* Limited conversations with faculty during this project suggest that faculty can be very enthusiastic about easy access to books for their classes, whether these books are required or supplementary, and that some believe that electronic availability could meet their teaching needs very well. Not all the faculty who are enthusiastic about e-readings for students are so enthusiastic about e-readings for their own work. That is not to say they would not use such books themselves, but that for various reasons the medium may offer them far less at this time (fewer choices of books, low quality of mark-up, less experience in reading on screen). It seems that for e-books to take hold in academia, they should be pitched toward students.
 10. *Electronic access to frequently used texts is an important unmet need.* Readers should be able to click on the URLs in an online catalog and check out the titles for a given amount of time; access (from the publisher's point of view) should be at least as secure as a paper copy, thus meeting the publisher's need to take in even more revenue than with paper alone. The community is served by increased and remote access to the title as well as enhancements to access – dictionary lookup, text to speech, etc. But what we need, for purpose of teaching, is subscription access to needed titles so that multiple students in a single course could access the titles or portions of them simultaneously.

QUESTIONS FOR POSSIBLE FUTURE STUDY

It is irresistible, given the content and interest of this project, to think about some of the wider questions that can begin to be addressed by this kind of inquiry.

A. How do e-books relate to other digital forms and formats of information (e-journals, images, course materials developed for the class)? How do they complement these materials? Does the internal structure of books play a role in their online use? This question is best explored through engagement with selected and interested faculty.

B. Is there a possible consortial component to the collective management and shared access to a large core of high us books in electronic form?

C. What are the steps to bring librarians and faculty together in defining the future of electronic books and other formats for teaching support? What is the role of e-books in classroom pedagogy and student learning, insofar as it can be determined at this time?

D. What is the relationship between academic institutions, particularly the libraries that deliver much of the information, and publishers or third party vendors like netLibrary and all the rest that are springing up? How would we think about this question?

E. If we begin devising new strategies for supplying required and recommended readings for course use, can we affect the way faculty think about what they teach and how they teach it? By making materials more available, can we give faculty additional teaching opportunities?

BYTES (Books You Teach Every Semester)

List of Attachments

1. Instructions for Completing Database Template. The assessment team prepared this document to assist the nine participating schools collect comparable and complete data for the Fall 2000 term.
2. Duplicate Standard Titles. This is a list of the discrete titles that are on reserve in two or schools in 2000. The list is sorted alphabetically by title. For each title, included are the author, the year(s) of publication, the publisher(s), the standard number chosen to identify the title (ISBN or OCLC identifier), and the schools that hold the title on reserve.
3. Authors by Extent of Overlap. This report lists authors and their works on reserve at three or more schools in 2000. The authors are clustered according to the number of schools that share works by a given author. For example, Edward Said and William Shakespeare have works on reserve in seven schools in 2000.
4. Distribution of Titles Across Library of Congress Classification. This table provides details that are graphically summarized in Chart 2.
5. Journal Titles and Electronic Availability. The information in this appendix lists the 205 journal titles from which one or more titles are on reserve in Spring 2000. The appendix also shows how many of these titles are available electronically and the source or sources of the electronic version, including the extant years of electronic publication.