

Electronic Collection Development in the Harvard College Library

Gary King*

6 December 1996

This document describes the organization, decision processes, and operating procedures by which the Harvard College Library (HCL) acquires electronic data.¹ I first explain HCL's long-established book collection procedures and then turn to electronic data acquisitions. The document concludes with some possible issues for discussion. The information herein was gathered from interviews with library staff, all of whom were exceedingly helpful.

Book Selection and Purchasing

0.1 Organization

HCL spends approximately \$9.3 million on acquisitions annually, of which Widener's portion (devoted to the social sciences and humanities, as well as general reference materials) is \$5.7 million. The remainder is divided up among a variety of much more specialized libraries, some of which have collections that overlap Widener's.

To maintain efficient decision-making at this high level of purchasing, the organization of the Widener book selection division parallels that of the publishing trade. Thus, book selection groups are divided into regions within

*Department of Government, Littauer Center North Yard, GKing@Harvard.Edu, <http://GKing.Harvard.Edu>, 617-495-2027. My thanks to Nancy Cline, Dale Flecker, Barbara Halporn, Ned Keenan, and Sidney Verba for comments on an earlier version.

¹HCL includes these libraries: Widener, Cabot Science, Fine Arts, Harvard-Yenching, Hilles, Houghton, Kummel Library of the Geological Sciences, Lamont, Littauer, Eda Kuhn Loeb Music, and Tozzer. The Faculty of Arts and Sciences (FAS) includes over 40 independent libraries outside of the HCL system, such as the libraries in the natural and physical sciences controlled by individual departments. A list of all 54 FAS libraries can be found at <http://hul.harvard.edu/libinfo/faculty/fas.html>.

which publishers are located.² This enables the bibliographers to develop detailed knowledge of specific publishers, to track their catalogs and other advertisements, to review national bibliographies when available, and to keep apprised of new suppliers or other changes among existing presses in their areas of responsibility.

This organization is sometimes described as “language based,” but this is not accurate, despite the correlation between the two descriptions. For example, the decision to purchase an English-language book published by Scandinavian publishers is made by the Scandinavian section. (For purchases of very expensive and complex materials, more than one section may be involved.)³

The non-Widener libraries in FAS make purchasing decisions without consulting Widener’s staff. Since these are substantively focused libraries with smaller acquisitions budgets, most of their purchasing is organized by subject area. All Harvard University Libraries, including those in schools outside FAS, list their book acquisitions in HOLLIS.⁴ This “bibliographic control” via HOLLIS prevents some unnecessary duplication in purchasing, but more importantly provides a centralized facility for users to locate the books they desire.

²The sections of the Widener book selection division include American and English; East, South, and Southeast Asia; Latin America, Spanish, and Portuguese; French and Italian; German and Dutch; Scandinavian; Sub-Sahara African. In addition, sections on Judaica, the Middle East, and Slavic are grouped separately under a separate department, the main difference being that they also have responsibility for cataloging. Modern Greek is an additional separate department; the Map room and Government Documents are Widener collections.

³Organizing the library’s massive book buying effort according to the distribution of book suppliers would appear to reduce transaction costs. Similarly, when most people do personal chores, they also reduce transaction costs by going to the supermarket and buying everything needed, and then going to the clothing store to buy anything required there. Other plausible organizations of the library’s buying, such as by disciplinary category, are might be less efficient — as it would be for individuals to buy everything they need to eat and wear on Thursday, and then to make a separate trip to the supermarket and clothing store to meet all their needs for Friday. However, as reasonable as this justification may seem, other major libraries follow a diverse range of this and other organizing principles.

⁴*HOLLIS*, the Harvard On-Line library Information System, is available at <telnet://Hollis.Harvard.Edu>. It is very fast, but has a primitive and relatively unfriendly user interface. HOLLIS also includes modules, not directly available to the public, for managing the library’s book purchasing, management, and circulation functions. *HOLLIS Plus* is a web site with a growing collection of networked electronic resources provided by a variety of suppliers; it is available at <http://hplus.harvard.edu>. The central administration’s Harvard University Library (HUL) office, which manages HOLLIS and HOLLIS Plus, plans to replace both with *HOLLIS II* by approximately January of 1999. Information on HOLLIS II plans can be found at <http://hul.harvard.edu/hollis2/>.

0.2 Decision Making

Each year, the head of the Collections Development Department (Barbara Halporn) decides how to allocate the acquisitions budget to the six selection units in Widener. Parallel decisions are made by the other purchasing areas within Widener. In practice, these budget allocations change only incrementally from year to year, with small marginal adjustments between sections due to changes in the publishing industries in each area. Once this budgetary decision is made, the specific decisions about which materials to acquire are almost entirely decentralized. That is, with rare exceptions, all non-electronic purchasing decisions are made by the library's individual professional bibliographers, covering their publishing regions. The decisions to purchase very expensive microfilm collections are made by the individual bibliographers, who have a knowledge of the research needs of the faculty and students. They occasionally seek advice of the faculty for assistance.

The goal of collections development is to anticipate library research needs. Bibliographers value the advice of Harvard faculty knowledgeable in their areas of responsibility, and (although there are exceptions) book purchase requests from faculty are generally approved. Faculty requests for specific items are quite rare, especially relative to the number of books they regularly purchase (e.g., a total of 26,616 volumes were purchased last year by the American and English section, which comes to about 50 books a day, every working day, for each of the two bibliographers), although influence from the faculty in the general directions taken by collection development is substantial. Scholars in the humanities seem to have substantially closer relationships with the bibliographers than do scholars in the social sciences. As one librarian described it, the book selectors and humanities scholars understand each others needs so well that they are able to finish each other's sentences. In contrast, social scientists typically have not met any book selectors.

Although individual book purchasing decisions are easily influenced by the faculty, and the library tries to cover all current book needs, most buying by the library is independent of who is presently on the faculty, what research they are now doing, or what classes are being taught. As a research library (or, as it is often called, the "library of record"), HCL has developed extensive collections in areas that are of no obvious interest to current faculty or students. This archival function can benefit the faculty in the long run when some of these areas suddenly become the subject of scholarly inquiry.

Electronic Data Acquisition

Electronic resources include online full text and image collections, bibliographic databases, fully numerical databases, and encyclopedias and other reference tools. Some of these materials imitate or are intended to replace printed publications while others represent new information products. Some of these materials require proprietary operating software, specific hardware, or a selection of network platforms in order to function. Some are stand-alone products; others can be made available through the web more generally.

Setting up a fixed organizational structure within the library to purchase electronic resources is difficult because of the rapid growth and massive changes in electronic publishing. Not only are the suppliers changing, but their products, the delivery mechanisms, and the equipment to decipher what we buy changes frequently. Buying a book requires that we have space on a shelf in a building with appropriate climate control. The book must be written in a language, format, alphabet, and font that can be understood, and preferably published on paper that can be preserved in the long run. Book publishers have had hundreds of years to stabilize these and other aspects of their industry. There is little doubt that something like this will happen in electronic publishing, but we are not there yet.

At the moment, the electronic materials the library collects fall into two main categories: *CD-ROMs* and *networked data*. Since CD-ROMs are a physical product that, along with documentation, are about the size of a book, CD-ROM purchases are already part of our current book acquisition efforts. Individual bibliographers make decisions about CDs roughly as they do about books. Although the bibliographers do not always have the expertise to evaluate electronic media, they are learning by taking classes, self-instruction, and helping each other. At present, they do not routinely request and evaluate demonstration copies of electronic data bases, although they receive input and advice upon request from the HCL automation group. Before purchasing, they ensure that each product can be run on HCL equipment, and they spend some time, along with the automation and reference groups, in trying to understand the product after it arrives.

Through the Widener approach to book purchasing, the library has purchased very few CDs. Indeed, only 500 CDs have been purchased by all of Harvard's libraries. (In addition, roughly 100 have been acquired and another 100 printed by the Harvard-MIT Data Center; a list can be found at <http://hdc-www.harvard.edu/hdc/cdlist.html>.) This relatively small number may be partly accounted for by CDs being what appears to be a

transitional media, the successor to which has already been negotiated and approved by the industry. Moreover, many electronic resources are now networked materials. The industry appears to be moving very quickly towards these networked resources. Most CD-ROMs are produced in the United States and United Kingdom, and so it is the American and English Section of the book collection division that purchases most of these materials, although other sections sometimes make contributions and some other HCL libraries, such as Fine Arts and Music, also make some purchases.

The library has responded flexibly to the pressures from this new industry. A half-dozen decision-making mechanisms have already been tried, modified, discarded, or reinvented. One of the problems is that the plans that are designed to maximize cooperation among Harvard units in order to save the most money and guarantee the widest access are often those with the highest transaction costs. Thus, a reduced payment to a supplier might be counterbalanced by a higher cost in library staff time and effort. Many of the ways to make this task easier will need to be implemented by the suppliers who appear to be responding to the strong incentives to achieve this.

Although they are sometimes described as being identical, the organizational structures and decision making authority for buying books (and CD-ROMs) as compared to networked data are fundamentally different: Whereas book purchasing decisions are almost entirely *decentralized*, networked data purchasing decisions within HCL are almost completely *centralized*. The new head librarian (Nancy Cline) is currently reviewing these organizational structures, and may make changes, but, at least until now individual decisions have all been made by the head librarian and the associate librarians (Lawrence Dowler, Associate Librarian for Public Services and Susan Lee, Associate Librarian for Administration and Finance). The most expensive purchases (such as on-going negotiations with Lexis-Nexis, Current Contents, and ABI Inform) and those requiring immediate decisions (such as when special deals arise because of HCL's relationship with other Harvard libraries or cross-university consortia) are made by these senior library staff alone. For the more usual case of moderately priced, less visible data bases, the library has established an *Electronic Resources Council* (ERC).⁵ This group of mid-level professionals from different FAS libraries

⁵Current members of the ERC include Michael Blake, Cabot Reference; Ellen Cohen, HCL Director of Financial Services; Rod Goins, Automation Coordinator for HCL; Ann-Marie Breaux, Technical Services, Lamont Library and ERC Coordinator; Michael Fitzgerald, Electronic Texts Librarian, Widener and ERC Chair; Steve Love, Reference, Hilles; Judy Warnement, Librarian, Botany Libraries; Lynne Schmelz, FAS Science Li-

meets every month or two to collect and review all the proposals for electronic purchases that have been forwarded by the book selectors, faculty members, and other sources. They then rank the proposals based on academic program need, cost, reviews, and feedback from library colleagues at other universities. The only titles considered are those for which the library presently owns the computer hardware and operating systems necessary to run them.

At present, the ERC does not make generally available, such as via the web, a list of items currently under review. Suggestions from faculty in person to a librarian, via the web, or through email are responded to, but there are no standardized procedures for follow-up. Those making suggestions do not always receive notification that their request is being granted, and no explanation or notification is given for requests not granted. Although there are no standardized operating procedures for these matters, all library staff involved in the process are quite open to discussing electronic resource purchases at any stage in the decision process if they are contacted directly.

The ERC has no acquisitions budget and its members have almost no knowledge of how much has been or will be spent on electronic resources. Their job is to collect information, rank the proposals, and forward them on up. This ranking is typically not accepted without substantial modification. In both theory and practice, decisions in this area have been made by the head librarian and the two associates.

Funds for networked resources are now being set aside in an “electronic resources fund,” which is a separate budget category rather than a tax on other library functions. This year, this fund is set at \$275,000 of the \$9.2 million acquisition budget.⁶ The library recognizes that expenditures for electronic resources will need to grow, and it has worked out a tentative plan to increase this budget while not hurting its book purchasing effort. The plan at least for the next several years is for the above-inflation “unrestricted” budget increases from FAS (now slated at 1.5% per annum) to go into the

braries Coordinator, Librarian Cabot and Tozzer.

⁶The electronic resources fund does not include some expenditures by the departmental science libraries (often as organized by Cabot), despite the fact that some science databases are partly paid for by this fund. Science library decisions about networked resource purchases are also made centrally in informal meetings of the science library staffs, but because of the close relationship between these libraries and their constituents there appears to be little ambiguity or potential for disagreement about priorities and preferences. Unfortunately, coordination in purchasing electronic (and nonelectronic) resources for the physical and natural sciences appears to be much more difficult than in the humanities and social sciences because of the lack of a centralized science library analogous to Widener.

electronic resources fund. Of course, all planning for electronic resources may change as needed.

HCL often negotiates with Harvard libraries outside of FAS (usually as arranged by the University Library and Director Sidney Verba) to pick up parts of site license fees. Various decision-making mechanisms have been developed for this purpose as well. However, if HCL wishes to purchase an electronic resource, it is almost always willing to, and often does, pay the entire cost of the item, even if none of the other schools agree to participate.

It would appear from the choices the library has made so far that they have the most expertise in the areas of bibliographic databases and reference works. They appear to have collected fewer full text databases and far less of available numerical collections. Government Documents (headed by Diane Garner) has purchased a variety of numerical collections. However, Government Documents are often forced, by lack of funds or by the difficulty in negotiating joint agreements with other libraries, to purchase stand-alone versions of software instead of the full networked version that everyone can take advantage of. Some of this material appears on the HCL local area network, which is accessible only from the reference areas of the Widener, Lamont, and Hillis libraries.⁷ Most of it is only available on isolated workstations in the library.

Unlike books, there exists no centralized bibliographic control facility for electronic materials. Networked resources purchased in part by HCL or for which HUL is involved are usually distributed via HOLLIS Plus, and they are listed in the HOLLIS index.⁸ But numerous other electronic resources have been purchased by other libraries and non-library units all over the university for which no information is included in HOLLIS, HOLLIS Plus, or any other index. The lack of bibliographic control, or information at the time of purchasing, has at least three consequences.

First, it is hard to know who purchases networked or networkable resources, where they are available, and who has access. When one faculty member at Harvard buys a book, it has no ramifications for the rest of the university. But when one person buys a piece of software or an electronic database, it can have major consequences. New purchasers might not be tak-

⁷Materials on the HCL local area network presently include the Social Sciences Citation Index; US Statistical Abstracts; America: History and Life; Arts and Humanities Citation Index; Gale's Dictionary of Quotations.

⁸The current policy of HUL is that to be included in HOLLIS Plus, a networked resource must be of scholarly interest to a broad segment of the Harvard community, perform reasonably reliably, and have an acceptable interface. HOLLIS Plus also includes materials produced by a Harvard unit or funded by a Harvard library.

ing advantage of University site licenses. They also might not know that, in some cases, a small additional contribution could give the entire university access. After discussions with some members of our joint subcommittee, I asked Ann Margulies, who heads project ADAPT (the university-level administrative data project), whether ADAPT could be adapted to track software and other electronic purchases. She found that it would not be difficult and agreed to do it. We do not know what the ultimate ADAPT product will be like, or how well it will work, but in theory, whenever someone buys something sitting at their PC, the system would be able to indicate whether anyone else has bought a copy, whether a site license exists, and even whether enough people are requesting copies that negotiating a site license might be a good idea. ADAPT will need to deal with issues of privacy and not delaying purchases of individuals to accomplish these broader goals, but this may help to some degree.

Second, a related issue is that TPC (the Technology Product Center), a university-owned store that sells directly to Harvard faculty, students and staff, provides no information about site licenses. So not infrequently, Harvard personnel wind up paying full price for electronic resources (and other software products) for which they could have discounted prices or even free access. I talked with TPC director Frank Urso about this. Although this problem is not common among high volume purchasers, it does happen with individuals. TPC will now post signs in the store about the availability of site licenses next to the full-price items. They also plan to modify their web site so it is easy to find information about all site licenses. The subcommittee might look into this after the changes are made.

Finally, negotiating site licenses for networked resources can take a great deal of time. Part of the problem is that the industry is still quite immature, but even when a supplier is easy to deal with arranging what is essentially an international treaty among the various units at Harvard can still be very difficult. This puts a major burden on many units within Harvard, especially Dale Flecker at HUL. Solving this problem, which stems from the extremely decentralized way Harvard is organized, will ultimately require solving problems bigger than the library system.

It turns out that Anne Margulies had a proposal in to the central administration to hire a full time site license negotiator, but only for software, not networked resources. She has recently agreed to our request to rewrite the job description for this position to include responsibility for databases and networked software. This coordinator will have many diverse responsibilities and so is unlikely to come close to solving the more general problem, but at least this will be an additional resource that members of the university

community can call upon.

Issues

I list here some questions for discussion. Some are issues at Harvard; many are probably also issues at other major university libraries. Few have obvious answers, and some have been considered or were originally raised in one way or another by library staff.

1. How might we increase communication between faculty and librarians about how the library makes purchasing decisions and how to influence purchasing requests? Is it worth developing a web site that lists new acquisitions and perhaps even the status of new requests? Should there be a standard weekly list in the *Gazette* of HOLLIS Plus items? How can HOLLIS II solve these communication problems more formally?
2. Should HUL exercise bibliographic control over electronic materials through HOLLIS even for purchases outside of HCL, perhaps even for those outside of the library system?
3. Is there some way to ease or institutionalize the negotiation of site licenses for networked resources?
4. How can the library gain the expertise necessary to build a more complete collection of numerical databases? Should Government Documents have representation on the ERC?
5. Books are purchased for current scholarly use as well as for apparent long-term archival value. In contrast, networked resources are only purchased based on current need at present. Should we begin to buy networked resources for their archival value too?
6. Networked resources do not cover all current needs of faculty and students, but book purchases are intended to cover all current book-needs and many other areas of no current interest. Given this, when is it appropriate to turn down a request from a faculty member's for an electronic resource in order to buy books (or expensive microfilm collections) that no faculty member currently wants?
7. Much effort at Harvard and elsewhere has gone into preserving paper materials. At present, little is being done to preserve electronic data,

although many groups, consortia of universities, and nonprofit organizations are talking about the problem. Should we begin to study whether anything can be done in this area, or when it will become feasible to do something?

8. Should we pass up on electronic materials that are difficult or impossible to preserve? (Preservation of electronic materials includes concerns not only about the physical longevity of the media on which the data is stored, but also whether the format, operating systems, and the computers necessary to run the software or access the data will remain available.)
9. Some electronic materials can only be run on one PC. Should we pass up on purchasing these in favor of networked resources? Should we also not buy items that have especially unfriendly user interfaces?
10. Should we focus on acquiring only those electronic materials that can run on the library's computers, or should materials that can be run on other computers around campus be considered as well? To put it another way, should the library be prepared to provide access to computer hardware and software for all electronic media it acquires? The library does not provide translation services for foreign language books; should it provide user support for unfamiliar electronic services?
11. Some electronic databases come with their own front-ends that can only run on one PC, but their data are accessible directly. That means that it is possible, although in some cases very difficult, to reprogram them to be accessible on the network. Should the library allocate funds to do this reprogramming or wait until the industry responds to this need? (At present, the library does no reprogramming of this kind, although HASCS does a small amount when requested for specific classes.) What if the industry does not respond for some data bases?
12. Should the library spend additional money or effort to extend site licenses so that students and faculty can access these resources off-campus?