

# THE POTENTIAL OF ONLINE BOOKS IN THE SCHOLARLY WORLD

*From the Columbia University Online Books Evaluation Project*

By Mary Summerfield, Carol Mandel, and Paul Kantor, Consultant

December 1999

## INTRODUCTION

>From winter 1995 to autumn 1999, the Online Books Evaluation Project at Columbia University studied online books as resources in the academic world. The project analyzed (1) the Columbia community's adoption of and reaction to online books; (2) the relative life cycle costs of producing and owning online books and their print counterparts; and (3) the implications of traditions of scholarly communications and publisher and marketplace reactions for online books. The project was supported by Andrew W. Mellon Foundation and by four participating publishers.

This paper summarizes our key findings and draws upon other work in this field to support judgments on the potential role of online books in the scholarly world and on various marketplace issues. *The Online Books Project, Columbia University: Final Report*, available at <http://www.columbia.edu/cu/libraries/digital/texts/about.html>, gives the full results of the project. Many other reports and papers, including the full research protocol,<sup>1</sup> are also linked to that Web page.

Like many other projects experimenting with emerging electronic media, the Online Books Project had to build the collection of scholarly online books with which the concept was to be evaluated. In most cases, publishers did not have electronic versions of their books from which we could easily create an HTML version.<sup>2</sup> Thus development of our collection took longer and, at the end of the study, the set of books was smaller than anticipated.

## THE FUTURE OF SCHOLARLY BOOKS

Overall we found that electronic books, in particular online books, can have a growing place in the scholarly world over the years ahead. Data tracking the use of online books showed that distinct need for certain books, e.g., for reference or for a course reading,

---

<sup>1</sup> Any of these reports and papers may be downloaded or printed out from this site for personal study. Please contact Elaine Sloan, University Librarian at Columbia, to request permission to use any of them in full or in part in any other way.

<sup>2</sup> This remains a problem in late 1999 as publishers and vendors attempt to develop a business of online books. An article (Lisa Bransten, *Start-Up Firm NetLibrary Targets Readers Seeking Data*) in the November 4, 1999 *Interactive Wall Street Journal* referred to netLibrary keyboarding and scanning books to make them part of their collection.

drew scholars to the online form of the book even at this early stage in its development. In individual and group interviews, scholars expressed considerable optimism that online books would help them to be more efficient, and possibly even more effective, in their teaching, course work, and research. They recognized that our small collection did not yet reflect the full potential of this medium for delivery of scholarly works. College and university librarians recognized that they might be able to provide their scholars with much greater access to books via the online format, with easier access to materials that would otherwise be available only in print form and better access to books now available only through interlibrary loan or visits to another library. Librarians expected to continue to provide some books in print form, but that they might decide to provide others, such as reference works or texts that are quickly outdated, only in online format.

Given our research findings and the current growth of the hardware and software components of the industry, we are confident that, within a few years, the book software community (authors, scholarly and textbook publishers, vendors, libraries and readers) will be providing stable electronic books and the hardware community will be creating multi-functional, portable devices that make the reading and manipulation of texts on-screen satisfactory for most users of such texts.

In that new world, new scholarly books will remain available in print form, especially in major libraries and on a print-on-demand basis at bookstores, but the electronic versions of current reference works, textbooks, scholarly monographs, and collections of essays will serve most scholars for most uses of these works. Publishers may decide that the economics of printing are such that they will publish in traditional print format only books expected to sell thousands of copies.

Students will use online, on-disk (e.g., CD-ROM or DVD), or e-paper versions<sup>3</sup> of books for much of their course work, manipulating the data, and using analytic software or multi-media information within a textbook. New devices will be much more portable, so that reading can be done on the subway, with better screen ergonomics, and better tools enabling the scholar to manipulate the contents of the book.

Scholars doing research may then do most of their preliminary and intermediate level browsing and reading in books and journals online. Only to read at length would most scholars seek out a portable or print copy of a book available in online form.<sup>4</sup> Over the years as reading devices become better, fewer scholars will feel the need for the print form of a scholarly book even for extensive reading.

In this vision, the design of individual online books and of whole collections must facilitate scholars' use of the books as well as protect copyright. Scholars and librarians advocated all of the following design features in interviews. A scholar will be able to search across a whole subject category or a specified subset of online books, seeking

---

<sup>3</sup> E-paper is a concept that researchers at MIT and at other research institutes are exploring for displaying digital images – text or graphics – temporarily. It involves electrostatic charges through the material that makes up the pages and which rotates micro-balls that are white on one side and black on the other, hence producing the images.

<sup>4</sup> Our research found that most of the use of scholarly books does not involve such extended reading.

terms relevant to his research. Online books will be included in a Web catalog with hyperlinks to the books themselves. Each book will have a stable Web address so that a library can easily design a presentation scheme without layers of publisher pages. Online book vendors will provide libraries with catalog records and standard statistics on use of the online books. Vendors will also migrate and refresh the online books over time so that the books continue to be accessible to scholars with changing local hardware.

The table of contents will include a comprehensive set of chapter subheading links. Thumbnails of various graphical images will be offered as a group, linked to their presentation in normal size and location. A click will take the scholar to an interesting image and the related text. Movement back and forth among pages of the book will be smooth. Footnote text will be presented in a block adjacent to the linked text in the book. The user will be able to choose to look at two pages. The book will be paginated as a print book to facilitate both citing material and finding cited material. References will be hyper-linked if the referenced book or article is available online. The book will be linked to a dictionary so scholars can look up terms. Scholars want pages to have adjustable fonts, margins, and the like so that each reader can choose the look he likes best. Each user will be able to maintain a file of annotations and highlightings of the book. Those annotations could be shared among a group of readers, say an instructor and a class.

This vision relies on the availability in the online format of a large portion of the materials that a set of scholars wants to use. If they must go to the library and search the stacks to find many of the books and journals that they need, these scholars will find much more limited value in the online format. Students who rely on current textbooks, relatively recent monographs and books of essays, and other readings placed on reserve are likely to be earlier adopters of online books. These materials are likely to be available electronically sooner than a massive general collection of scholarly monographs. In general, advanced graduate students and faculty are likely to adopt this format more slowly, as fewer of their books will be available in it in the early years. These more senior scholars will stay with print books more extensively for a longer time and will move to the online format only when it provides substantial advantages in size and currency of collection and convenience of use.

Market conditions will determine the availability of online books to individual scholars. Will a scholar be denied all access to an online book if his library has not purchased it in some sense? Or will he be allowed to search all titles and to briefly browse interesting titles free (equivalent to looking at a book in the bookstore) with a payment required from someone (library, department, scholar) only when he wants to use a book at length?

Within a few years, publishers should realize the potential of online books to enhance their profitability. They will redesign their production processes from directions-to-authors-on-formats-to-submit through editing through typesetting to produce a digital version that can be put online easily. In this world, new works will be put online as soon as the print versions are produced, if not sooner. However, older monographic scholarly works will be republished only if publishers perceive sufficient demand to cover the costs of producing and distributing that version. We estimate a period of a decade or so for the collection of scholarly online monographs to grow to where more senior scholars will

markedly change their patterns of research behavior. However, such scholars might soon adopt the online format for review of newly published books, if a significant number of the major works in their field become available in that format.

## **PROJECT CONCEPTUAL FRAMEWORK**

As noted in the introduction, the Online Books Evaluation Project looked at (1) scholars' behavior and reactions to online books, (2) lifecycle costs of traditional print books and online books for publishers and libraries, and (3) marketplace reactions to the concept of online books. All of these were assessed in the context of environmental developments in the relevant areas of academia, scholarly communication, and the computing and Internet worlds. A wide variety of tools, including server data, a variety of online, mailed, and hand-distributed surveys, and individual and group interviews, were used in this evaluation. They were described in detail in the project research plan and in the final report; both are at <http://www.columbia.edu/cu/libraries/digital/texts/about.html>

## **DEVELOPMENT AND DESIGN OF THE ONLINE BOOKS COLLECTION**

The Online Books Project began formal activity in January 1995. However, discussions with publishers began two years earlier. When this project was proposed, the World Wide Web was just emerging, and we expected to develop custom SGML browsers, as other online publishing projects were doing. However, once the project was underway, we judged the Web to be the best delivery system. First, scholars would have access to the Web from university locations and many, if not most, would be able to access it soon from home via modem dial-in to campus servers, or ISPs. Second, the Web would maximize the value to scholars, as the greatest potential added value from online books would come with truly digital books.<sup>5</sup> Only this online format would allow development of truly interactive books taking advantage of current and anticipated capabilities of Web technology, such as links to other online resources, and inclusion of sound and video, data files and software for manipulating data. Perhaps only such enhanced online books would offer sufficient advantages over traditional print format that scholars would substitute them for the print format in some or all of their modes of use and for some or all classes of books.<sup>6</sup>

As of June 1999, the end of our data gathering, the collection totaled 168 online texts, including six reference works and 54 classical texts in social thought. Four publishers provided modern books to the collection – Columbia University Press, Oxford University Press, Garland Press, Simon and Schuster Higher Education. The contemporary works were in six subject areas: biography, literary criticism, earth and environmental science, philosophy, political science/international affairs, and social work. A few of these books

---

<sup>5</sup> Many projects working with books that were already published have used a system of scanning a book's pages and using OCR to create an index of its contents.

<sup>6</sup> The project was not able to test these enhancements to the standard print book format as we were converting existing books not creating new products. Other systems such as PDF now have greater capabilities than they used to have. Thus, publishers, vendors, and libraries would be wise to evaluate the trade-offs of the various digital book production and display options.

were designed as college textbooks; others were scholarly monographs or collections of essays. Each of these books was in the Libraries' collection in print form, circulating from the regular collection or Reserves, or non-circulating in Reference, as well as in one or more online formats.

Presenting a set of online books and the individual books was a challenge. Information about the books was presented to the Columbia community in several ways:

1. In CLIO, the Columbia online catalog, initially in the standard version and then in late 1998 in a Web version that allowed the scholar to click on the URL and move directly to the book;
2. In a Web site devoted to the books with an A-Z listing by author;
3. In a set of major subject category groupings on another Web page.

Scholars could search for terms relevant to their work across the whole collection or within the subject groupings.

At the outset, the project editorial team decided that a book would be mounted with each major element as a separate file, e.g., a table of contents-title page, the introduction, each chapter, the bibliography, the index. The user could click on a chapter title in the table of contents to go to that chapter. Footnote numbers were linked to the footnote contents; a *back* link returned the reader to the text. The browser's *find* feature located words within a chapter. A pagination tool took the scholar to a known page number (e.g., from the index or from a citation). Page numbers were not used within the text, so scholars could not cite a page number in a reference to the online book. (Even by the end of the project, scholars had not accepted this practice. They wanted the online books to be paginated as their print counterparts.) To move among chapters the scholar returned to the table of contents and clicked on the new chapter heading.<sup>7</sup>

Other digital journals and books projects also faced the challenge of deciding what features to include in a system design and under what conditions to modify that design. Substantial changes are likely to confuse existing users of a collection. Such confusion often leads to misuse, resultant dissatisfaction, and a smaller chance that a scholar will return to the collection. On the other hand, if such modifications are true improvements, they should result in greater user satisfaction overall and more use in the long run.<sup>8</sup>

## CONTEXT OF THE ONLINE BOOKS EVALUATION PROJECT

As anticipated, during the six years in which we planned and executed the project, relevant environmental elements changed substantially. Developments in the Internet

---

<sup>7</sup> Web enhancements over the five years of the project would have allowed us to add features. However, we could not economically modify the books already in the collection. Our designers decided that consistency within the collection was more important than adding features for only the more recent books.

<sup>8</sup> See, for example, Bishop, A.P., Neumann, L.J., Star, S.L., Merkel, C., Ignacio, E., and Sandusky, R.J., *Digital Libraries: Situating Use in Changing Information Infrastructure*, to appear in upcoming **JASIS** special issue on digital libraries. This reports on the NSF/DARPA/NASA Digital Libraries Initiative project at the University of Illinois. This project used a test bed of recent engineering journals.

and the World Wide Web, in computer literacy and access within all levels of academia, in the development of digital libraries, and in scholarly publishing are outlined below.

### **National Environment – Expanded Potential Access to Online Resources**

The national computing environment increasingly favored scholars' adopting computers, the Internet, and online resources. Prominent mainstream newspapers like *The New York Times* featured computers, the Internet, and related topics daily. The price/power relationship for personal computers improved enormously, with adequate computers available for under \$1,000 since 1998. Penetration of personal computers and use of the Internet grew throughout American society, with over half of households owning one or more computers by mid-1999. At that time, over 100 million adults in the U.S., or about half of all adults, were using the Internet, up from 65 million a year earlier. Researchers also found that users in 1999 were heavier and more sophisticated users of email and the Web.<sup>9</sup> But after 1997, the typical domestic user experienced little improvement in available speed of access to the Internet (56K at best) or in prices for ISP accounts.

### **Columbia Environment – Increased Access to & Use of Computing & the Web**

At Columbia University, even as it continued to grow, the modem pool ran at near peak capacity utilization, leaving scholars often frustrated in their attempts to dial-in to the campus network for email or electronic resources.<sup>10</sup> Student ownership of computers grew until that most students reported possession of a computer in their campus residences and easy access to the Web. Students living in Columbia residence halls had the luxury of Ethernet connections that made their Web connection vastly superior to that of the typical graduate student or faculty member working at home and using an ISP or other telephone link.

### **Scholars & Libraries – A Potentially Frustrating Relationship**

As knowledge and publication of journals and books expand, scholars face increasing difficulty in locating articles and books that will be useful to their course preparation and research. At the same time, libraries struggle with the challenges of acquiring the materials that their scholars need and keeping those materials readily available for students and faculty. Scholars are often frustrated by limitations of the print library. They often find that:<sup>11</sup>

- their libraries have not acquired the books they want to use;
- their libraries acquired those books but they are not yet on the shelf;

---

<sup>9</sup> Estimates by the Strategis Group reported in *The New York Times on the Web*, November 10, 1999. <http://www.nytimes.com/library/tech/99/11/biztech/articles/10net.html>.

<sup>10</sup> Many Columbia scholars residing off-campus had ISP accounts as well as the ability to dial-in to Columbia directly. The online books were accessed via a sign-in with one's Columbia ID, so that a scholar could reach them using any Internet route.

<sup>11</sup> Analysis of these problems is discussed in *Objective Performance Measures for Academic and Research Libraries*, Paul Kantor, ARL (1984).

- those books entered circulation and are not now on the shelf and that the recall process will take many days;
- the online catalog lists those books as on the shelf but no one can find them there or anywhere else in the library;
- those books are on reserve and, hence, available at best for a period of time too brief to allow extended reading and review;
- that browsing through the various areas of the stacks that might have books relevant to their work takes substantial time and can only be done during the hours that the library is open.

In addition, scholars often cannot determine which books may be useful from library catalogs alone. The typical catalog record includes too little data to tell the scholar whether the book includes relevant information on a narrow topic.

Scholars are enthusiastic about online books, which many believe will solve all of these problems, except possibly the first.

### **Continuing Roll-Out & Expanded Use of Electronic Scholarly Resources**

Coincident with the growing penetration of the Internet and development of the Web came the early digital scholarly resources such as abstracts and indices (A & I), electronic journals, and reference works. The digital A & I resources are so popular that publishers have nearly ceased production of paper versions. Electronic journals, both backfiles and current issues, are increasing in availability and popularity over time. As time passes, more scholars are aware of what is available in electronic form, find the increasingly large collections of value, and chose to use them.<sup>12</sup> This familiarity with electronic versions of these two key categories of scholarly library resources leads to scholars recommending them to their colleagues and to expanded interest in all types of electronic resources. In the first seven months of 1999, Columbia scholars had 37 percent more accesses of the growing JSTOR collection of journal backfiles than they had in all of 1998. In the peak month, April, JSTOR use was equivalent to about one access per Columbia community member.

Reference works are so clearly suited for the online format and were so quickly adopted online that we focused this project on other forms of scholarly books, such as collections of essays, monographs, and textbooks. For example, in 1994 the text-based *Concise Columbia Electronic Encyclopedia* had over 15,000 sessions at Columbia.<sup>13</sup> But as more sophisticated, Web-based substitutes became available after 1994, use of this

---

<sup>12</sup> In general, electronic journals projects have found that the breadth and depth of a collection are critical to user adoption. In addition, it simply takes time for scholars to become aware of the availability of these resources and to find an opportunity to use them. The Bishop et al paper cited above discusses these issues. So does *Summary of SuperJournal Findings: Readers*, Draft April 26, 1999, <http://www.superjournal.ac.uk/sj/findread.htm>.

<sup>13</sup> This short entry, text resource was available only on the CWIS-gopher platform CNET. Thus, we cannot count hits.

encyclopedia dropped at an increasing rate, with sessions down to about 2,700 in 1997, the last year for which we had complete data. Typically, use of the more sophisticated resources increased over time. For example, Columbia scholars made nearly 50,000 hits on documents in the online *Encyclopedia Britannica* in 1998.<sup>14</sup> Usage of the Columbia Web version of *The Oxford English Dictionary* nearly doubled from the first half of 1998 to the first half of 1999 to nearly 29,000 hits. The number of users of *The OED* increased by 28 percent from the last half of 1998 to the first half of 1999 to 1,370 distinct users.

### **Initial Roll-Out of Electronic Reading Devices & Commercial E-Books**

Since 1998 several firms have been fine-tuning and marketing electronic book reading devices, e.g., Rocket eBook and SoftBook, and books to read with these devices. At this point, each such device has a proprietary format so a consumer cannot buy an electronic book designed for the Rocket eBook and read it with the SoftBook. However, this is an untenable situation and the industry has taken steps to introduce a standard format for the texts to be read with these devices. Over the next few years it is likely that the industry will try many hardware options with various ergonomic and pricing characteristics. We expect that the primary hardware format to evolve will be a multi-purpose one that has general computing and communicating capabilities and is well designed for reading and manipulating the contents of a book.

## **FINDINGS OF THE ONLINE BOOKS EVALUATION PROJECT**

### **Economics of Scholarly Book Publishing & Online Books**

A key facet of the Online Books Evaluation Project was analysis of the lifecycle costs of scholarly books in print and online format. What does it cost to publish and provide these books to scholars via libraries? Would online books be more or less costly than the traditional print format?

The online book format has developed slowly; as a result the industry of online publishing and the terms of providing online books are still in their infancy. With technology evolving as well, the various systems for and costs of publishing, distributing, maintaining, and owning online books are at the beginning of their developmental cycles. Nor is it common to measure lifecycle costs for print books. But the available evidence permits order-of-magnitude comparisons.

### ***Online Books Integrated into Publishing Process***

The model likely to evolve is one in which production of online book files will become part of the publishing process, somewhere between the author's creation-of-a-manuscript-in-a-word-processing-program and the typesetter's creation-of-the-film-for-offset-printing.<sup>15</sup> Then publishers or vendors will develop collections of online books that they

---

<sup>14</sup> Britannica's statistics were faulty for four months of 1999.

<sup>15</sup> This last step might be eliminated for a class of low demand books for which desktop publishing and on-demand printing might become standard.

will maintain on central servers and offer as individual titles or groups of titles to libraries and individual scholars. Some publishers, e.g., Chadwyck-Healey, have begun to offer online texts in this way. NetLibrary, with a single-copy circulation model, is the first major vendor in the general library online book market.

As online books are not yet a standard publisher's output, our model assumed conversion of a print book, ASCII or other file format to HTML or SGML<sup>16</sup> as an add-on cost to the traditional publishing process. The University of Pennsylvania is employing another system to mount Oxford University Press books that involves distilling PostScript files into PDF files with Adobe Acrobat and reassembling the chapters into a single file. It uses Compose, a plugin, to build bookmarks and links from the book index to the pages.<sup>17</sup>

Publishers or vendors would incur additional costs in cataloging, creating a permanent URL, maintaining a central server, and transmitting the online books to the reader via the Web or an alternative secured Internet mode. An alternative system for providing some electronic books, e.g., textbooks, might be CD-ROM or DVD.

### ***Print Books – Analysis of Costs to Scholarly Publishers***

Exhibit 1 gives pro forma revenues and costs for five print books from a major university press and estimates of the costs of producing and maintaining online versions of those books. We should note that these five books were expected to have large sales compared to the scholarly monographs about which there has been so much concern in recent years. These five were expected to have sales of several thousand units each, while relatively few scholarly monographs sell more than 500 copies. In addition, these books were published by one of the larger university presses and, hence, should have enjoyed relative economies of scale. We analyzed the revenue and costs for these books in print form at length in a 1998 white paper on the economics of scholarly communication.<sup>18</sup>

These pro forma financial statements estimated that these five scholarly monographs would sell a total of 13,650 copies (3,450 cloth copies and 10,200 paper copies) at an average net price to the publisher of \$17. The projected surplus (total net income minus total cost) for these books was \$12,668, or \$0.93 per copy sold.<sup>19</sup>

---

<sup>16</sup> If a book were marked up in SGML, the standard is to then convert it to HTML for Web viewing *on the fly*. John Price-Wilkin at the University of Michigan championed this system.

<sup>17</sup> Roy Heinz, Director, Information Systems, University of Pennsylvania Library, provided information on their process for conversion and the costs that they are experiencing.

<sup>18</sup> Mary Summerfield, *Issues in the Economics of Scholarly Communication. A White Paper Supporting The Andrew W. Mellon Foundation-Funded Projects – The Online Books Evaluation Project & Columbia International Affairs Online*, Revised March 1998. This paper is available online at: <http://www.columbia.edu/cu/libraries/digital/texts/econpap.html>.

<sup>19</sup> This press, like many university presses, receives free space and utilities from its parent university so this analysis does not reflect all economic costs of this enterprise.

## Exhibit 1. Scholarly Book Publishing: Projected Revenues & Costs

### Sample of Books in Print & Electronic Format

	<b>Book 1</b>	<b>Book 2</b>	<b>Book 3</b>	<b>Book 4</b>	<b>Book 5</b>	<b>All 5 Books</b>
Book Pages	296	320	300	232	280	1,428
Cloth Copies Printed (#)	1,000	500	500	1,500	400	3,900
Cloth Copies Sold (#)	900	400	450	1,350	350	3,450
% Sold of Printed Copies	90%	80%	90%	90%	88%	88%
List Price - Cloth	\$45.00	\$42.00	\$50.00	\$32.50	\$49.50	\$43.80
Net Margin - Cloth	80%	80%	75%	65%	80%	76%
Net Price - Cloth	\$36.00	\$33.60	\$37.50	\$21.13	\$39.60	\$33.29
<b>Net Income - Cloth</b>	<b>\$32,400</b>	<b>\$13,440</b>	<b>\$16,875</b>	<b>\$28,519</b>	<b>\$13,860</b>	<b>\$105,094</b>
Paper Copies Printed (#)	0	2,500	3,500	na	3,000	na
Paper Copies Sold (#)	0	2,300	3,300	1,800	2,800	10,200
% Sold of Printed Copies	na	92%	94%	nc	93%	na
List Price - Paper	na	\$17.50	\$15.50	\$17.50	\$17.50	\$17.00
Net Margin - Paper	na	65%	75%	77%	80%	74%
Net Price - Paper	na	\$11.38	\$11.63	\$13.48	\$14.00	\$12.62
<b>Net Income - Paper</b>	<b>\$0</b>	<b>\$26,163</b>	<b>\$38,363</b>	<b>\$24,255</b>	<b>\$39,200</b>	<b>\$127,980</b>
Total Copies Printed	1,000	3,000	4,000	na	3,400	na
Total Copies Sold	900	2,700	3,750	3,150	3,150	13,650
<b>Total Net Income</b>	<b>\$32,400</b>	<b>\$39,603</b>	<b>\$55,238</b>	<b>\$52,774</b>	<b>\$53,060</b>	<b>\$233,074</b>
<b>Costs</b>						
Plant (Typesetting)	\$4,903	\$4,962	\$5,936	\$4,085	\$4,089	\$23,975
Paper, Printing, Binding	\$3,633	\$7,451	\$7,887	\$8,312	\$8,015	\$35,298
Royalty - % cloth	0%	7%	10%	7%	7%	6%
Royalty - % paper	na	7%	10%	7%	7%	8%
Royalty Amount	\$0	\$2,772	\$5,524	\$3,694	\$3,714	\$15,704
Others (Contributors Payments, etc.)	\$0	\$0	\$3,600	\$0	\$0	\$3,600
<b>Total Cost of Sales</b>	<b>\$8,536</b>	<b>\$15,185</b>	<b>\$22,947</b>	<b>\$16,091</b>	<b>\$15,818</b>	<b>\$78,577</b>
<b>Gross Margin (Income-Cost of Sales)</b>	<b>\$23,864</b>	<b>\$24,417</b>	<b>\$32,291</b>	<b>\$36,683</b>	<b>\$37,242</b>	<b>\$154,496</b>
Fixed Overhead	\$11,268	\$11,268	\$16,179	\$16,179	\$11,268	\$66,162
Variable Overhead	\$10,692	\$13,069	\$18,228	\$16,168	\$17,509	\$75,666
<b>Total Overhead</b>	<b>\$21,960</b>	<b>\$24,337</b>	<b>\$34,407</b>	<b>\$32,347</b>	<b>\$28,777</b>	<b>\$141,828</b>
<b>Total Cost (of Sales + Overheads)</b>	<b>\$30,496</b>	<b>\$39,522</b>	<b>\$57,354</b>	<b>\$48,438</b>	<b>\$44,595</b>	<b>\$220,405</b>
<b>Surplus (Total Net Income - Total Cost)</b>	<b>\$1,904</b>	<b>\$80</b>	<b>-\$2,116</b>	<b>\$4,336</b>	<b>\$8,465</b>	<b>\$12,668</b>
Source: Major University Press. This press reports stability in these costs.						
<b>Electronic Book Production Costs</b>						
<b>Using Print Text:</b> Scan, OCR, Encoding, All Proofing (UMI-HTI): ~\$1.51/pg.	\$447	\$483	\$453	\$350	\$423	\$2,156

<b>Using E-Files:</b>						
HTML coding for ASCII pg: \$0.36/1K char.: ~\$1/pg.	\$296	\$320	\$300	\$232	\$280	\$1,428
HTML coding for Quark pg:~\$2.15/pg.	\$636	\$688	\$645	\$499	\$602	\$3,070
E-File Proofing & Fine- tuning, Graphics: ~\$0.42/pg.	\$124	\$134	\$126	\$97	\$118	\$600
PostScript to PDF for Web: \$0.04/pg. (best case)	\$12	\$13	\$12	\$9	\$11	\$57
E-File Management: ~\$1,000 for contract for 50 books	\$20	\$20	\$20	\$20	\$20	\$100
Total Cost of Maintaining Books on Server: ~\$1/MB/yr.: ~44pg/MB w/graphics:~\$0.023/pg/yr. PV:~\$0.35/pg for 30 yrs.	\$104	\$112	\$105	\$81	\$98	\$500
Cataloging (CU)	\$25	\$25	\$25	\$25	\$25	\$125
<b>Costs of Online Version for 30 year period (except refreshing &amp; migrating):</b>						
Print to SGML	\$596	\$620	\$583	\$457	\$546	\$2,881
ASCII to HTML	\$569	\$591	\$556	\$436	\$521	\$2,753
Quark to HTML	\$889	\$959	\$901	\$702	\$843	\$4,395
PostScript to PDF	\$161	\$170	\$162	\$135	\$154	\$782

### ***Online Books – Modest Incremental Costs to Publishers & Vendors***

Hitherto, with few exceptions, online books have been a secondary product after the production of print versions was completed. Thus, the costs available for analysis are those of converting a print book to online format – generally either by scanning and further processing a print book or by taking an electronic file and converting it to SGML, HTML, or PDF.

As a result of its Humanities Text Initiative, the University of Michigan has extensive experience in the first of these methods. They estimate that the process of scanning, running the scanned text through optical character recognition software, encoding in SGML, and proof reading at all stages costs an average of about \$1.51 per page. This cost does not include management of the project or the cost of information systems.

In mid-1998 Columbia contracted out HTML coding of the electronic files of books for this project. This coding was done based on a set of standards established by the Columbia digital library editing staff. Some of these files were in ASCII format; others were in Quark. The cost of coding for ASCII was \$0.36 per 1,000 characters or an average of about \$1 per page for this set of books. Quark conversion was more complicated and more than twice as expensive – an average of \$2.15 per page for this set of books. Back at Columbia the process of proof reading and fine-tuning the returned files and adding graphics cost an average of about \$0.42 per page. Managing the contract for this conversion cost about \$1,000 in staff time, or about \$20 per book.

The University of Pennsylvania has found that in the best of cases, i.e., when it receives clean PostScript files with fonts that its software can interpret easily, its system of converting these files to PDF costs only four cents a page. A student worker can create a final, web-ready PDF file of a 300-page book in an hour. A troublesome book can take five times as long if not more, for a per page cost of twenty cents or more. None of these costs include management supervision or other overhead costs, but it is important to note how much lower they are than those for the two methods just described. This supports our theory that publishing technology can change to incorporate electronic books at a low incremental cost.

Columbia's Academic Information Systems staff calculated the cost of maintaining books on a server, including hardware, software, and related labor, as about \$1 per MB per year. Books vary in size, but with some graphics a book might have about 44 pages per MB, for an annual cost of \$0.023 per page per year. If a publisher or vendor were to maintain a book on a server for 30 years, the present value of the cost for those 30 years would be about \$0.35 per page, or roughly \$81 to \$112 for each of the five books in our sample.<sup>20</sup> We do not have estimates of the costs of migrating these books over time, but it might be reasonable to assume that they would be in the same range.

Cataloging an online book when a catalog record for a print version already exists cost Columbia about \$25 in professional staff time including benefits. This cost would fall substantially over time, as cataloging became a case of copy cataloging undertaken by non-professional staff instead of a case of original cataloging undertaken by professional librarians. Librarians advocated that publishers or vendors produce and provide to them the catalog records for online books.

As Exhibit 1 shows, the costs of the online version vary with the number of pages in the book as well as the method of conversion. They would also vary with the complexity of the books, i.e., the amount of graphics, multimedia, and links to other online resources, but we have assumed uniformity in this analysis. For our five books, the present value of the lifecycle costs of production and maintenance might be approximately as follows:

	<u>Base</u>	<u>W/ Migration</u>	<u>Break-Even Quantity</u>
From Print to SGML	\$2,881	\$3,381	199
>From ASCII to HTML	\$2,753	\$3,253	191
From Quark to HTML	\$4,395	\$4,885	287
From PostScript to PDF	\$782	\$1,282	75

It is much less costly to convert books from print or ASCII than from Quark. The PostScript to PDF method seems to win by a tremendous margin. However, we did not analyze the differences in the utility of the electronic books resulting from these methods. The weighted average net income per copy sold for these five books was projected at about \$17. To cover the incremental costs of producing, maintaining, refreshing, and

<sup>20</sup> This present value estimate assumes a real cost of money of 5 percent per year as well as a 30 year time span. These parameters are used in the later estimates of the lifecycle costs of a book for a library as well.

migrating the online versions of these books, the publisher would need to sell the above quantities of these books in online format at this average price.

This \$17 net income is based on the mix of cloth cover and paperback copies of these titles that this press expected to sell. If the online books were sold at the cloth cover price (a weighted average of \$30.46), the break-even quantity would be much smaller. On the other hand, if they were sold at the paperback price (a weighted average of \$12.55), the break-even quantity would be much larger. These values are two-thirds or more of the total sales for many specialized monographs. However, they are much lower than the projected sales for the five books included in our model.

The publisher or intermediary might charge the purchasing library a service fee for maintaining the book online, migrating it regularly, and the like, as netLibrary is doing.) If so, these costs could be recouped over time rather than through initial sales.<sup>21</sup>

### ***Online Books' Lifecycle Costs for Libraries Lower than Print Books***

One of the project's implicit hypotheses was that online books would have lower lifecycle costs. This hypothesis was based on two expectations: (1) Print books require ever-more-expensive manpower for acquisition, processing, and circulation as well as the cost of storage space; (2) Online books would use ever-less-costly computing hardware and would require little staff time.

We estimated the present value of a library's lifecycle costs for both formats, assuming a 30-year lifetime and a five percent annual real interest rate for each of the cost elements involved in purchasing and owning a book. Exhibit 2 lays out basic cost elements and estimates the present value of the costs at Columbia for both types of books. Assuming \$50 purchase prices for the print book and the online book, the present value of the total lifecycle stream of costs is about \$156 for the print copy and \$127 for the online book, for a 19 percent savings with the online version.<sup>22</sup> These data support our original hypothesis.

A scholarly library incurs costs for the Internet and computing infrastructure that allows it to provide electronic resources, such as online journals and books, to its scholars.<sup>23</sup> However, institutions have incurred these costs already, even though their libraries have not provided online books. As a result, we have not included these sunk infrastructure costs in our calculations of the costs of providing individual online books. However, if an institution would assume other unique infrastructure costs in providing its community with online books, an analysis of this sort should include those costs. One such cost

---

<sup>21</sup> The vendor could charge higher rates in the early years than support actual costs (knowing that the book had greater value to the libraries while it was relatively new) and use the excess funds to fund those activities in the later years. The vendor could promise to maintain books that become stale quickly, e.g., programming titles, for only a relatively short period and then take them down when an insufficient number of libraries were willing to support them.

<sup>22</sup> This model assumes original cataloging by libraries for online books – as Columbia experienced -- rather than the likely case of copy cataloging like that typical for print books. With cataloging costs for online books at the level of copy cataloging, the savings would be substantially greater.

<sup>23</sup> A college or university uses these resources for many purposes beyond library services.

might be that of expanding printing capacity so that scholars could print out portions of online books. Universities and their libraries must decide the extent to which they will pass on these costs to their scholars.

<b>Exhibit 2. Libraries' Lifecycle Costs of Book Ownership</b>		
	<b>Print Book</b>	<b>Online Book</b>
Purchase Price, Average	\$50.00	\$50.00
Selection	\$3.59	\$3.59
Processing:		
Ordering		\$2.00
Locate & Handle Bibliographic Record		\$5.92
Receive Physical Item	\$43.67	\$0.00
Payment		\$2.00
Initial Physical Processing		\$0.00
Cataloguing		\$25.00
Storage	\$4.61	\$0.00
Average Cost of Circulation	\$43.97	
Stack Maintenance	\$5.47	\$38.43
Collection Maintenance	\$1.90	
Repair/Rebind	\$0.28	\$0.00
Replace -- New Book & Processing	\$2.08	\$0.00
<b>Total</b>	<b>\$155.57</b>	<b>\$126.94</b>
<p>Selection: Estimated from share of hours spent by librarians &amp; assistants at relevant salary &amp; fringe rates divided by number of new items. (Assumes average librarian salary of \$45,000; staff assistant salary of \$22,000; and student assistant wage of \$8 per hour.) Most books are purchased via approval plans, not individually selected and ordered.</p> <p>Storage: Present Value of 30 years at \$0.30 per year with 5% interest rate. Malcolm Getz estimated annual cost of storing one volume in off-site storage (the marginal method) as \$0.30.</p> <p>Cost of Circulation for Print: Present Value of 30 years with 5% interest rate at \$2.86 per circulation and an average of one circulation per year.</p> <p>Stack Maintenance: Includes shelfreading, shifting.</p> <p>Collection Maintenance: Includes searching for and tracking missing books.</p> <p>Replace Print Book: 2,500 volumes lost and 116,000 purchased annually: 2.16% loss rate, assumed value replaced over 30 year period in purchase price and processing.</p> <p>The online model assumes that books will be bought via a system with terms negotiated with one or several publishers or intermediaries, e.g., approval plan, user selection &amp; mass availability, user selection &amp; individual availability, etc. Costs of ordering the books will be similar to those via approval plans for print books. However, cataloging costs are estimated at the original cataloging level experienced in the Columbia experiment, rather than at the copy cataloging level that would prevail in the long run.</p>		

## Use & User Reactions

### ***Online Reference Works Were More Used Than Their Print Counterparts; Format & Timeliness Were Critical***

Our collection of six reference works saw varying patterns of use, but all were apparently used more often in online form than in print form.<sup>24</sup> Use of *Columbia Concise Encyclopedia* and *Columbia Grangers World of Poetry*, two older works not aimed at a university audience, declined substantially to a few hundred sessions per semester, but reference librarians reported that those resources are used only a few times each month in paper format. Use of *The Oxford English Dictionary* grew significantly; with 1,370 unique scholars executing almost 29,000 hits on its Web version in spring 1999. *The OED* was the most used resource in the online books collection. Use of *African-American Women*, *Native American Women*, *Chaucer Name Dictionary* fluctuated from semester to semester, always in the several hundred hits or sessions range.<sup>25</sup>

### ***Monographs & Humanities Texts Had Modest but Growing Use***

Use of the three Chadwyck-Healey humanities full-text databases (*English Poetry Database*, *English Verse Drama*, *Patrologia Latina*) grew over time, but remained relatively modest. During spring 1999, 52 to 122 scholars used each of these resources.

Use of the 54 *Past Masters* classic texts in social thought declined by half from 6,632 hits in 1996 to 3,384 hits in 1998. However, in spring 1999, hits rose 58 percent compared to spring 1998. Use of these texts was largely concentrated in a small share of the titles that were used in classes in political philosophy and theory. In the period July 1996 to June 1999, one-seventh of the texts received two-thirds of the total hits.

The online books collection included 36 monographic (non-reference) titles as of July 1997, 55 as of July 1998, 68 as of year-end 1998, and 108 as of June 1999. Twenty-five of these books were assigned reading in one or more courses for one or more semesters during this period.<sup>26</sup> These monographic books received 3,542 hits in 1997, 4,885 hits in 1998, and 2,919 hits in the first half of 1999. In spring 1999, in 806 cases an individual scholar used one of these titles one or more times. Some individuals used more than one online book so the total number of individual users was smaller.

### ***Books Were Used Significantly in the Online Format***

Books that were available online may have been used by more scholars in online format

---

<sup>24</sup> We lacked a firm count of use of reference resources that were kept on open shelves as scholars could access them and reshelve them at will. Reference librarians could track of use of books kept behind their service center.

<sup>25</sup> These resources were available in two formats for which the measuring systems varied. It is impossible to tell for certain if declines in use of the text-based CNet versions, measured in sessions, were offset by gains in use of the browser-accessible CWeb versions, measured in hits.

<sup>26</sup> This count is based on faculty having put a book on reserve for a course. We had no way of determining if an instructor used a book in a course but did not put it on reserve.

than in paper format.<sup>27</sup> In spring 1999, nearly three times as many scholars clicked on the average online monographic book as circulated its print version. Using the so-called *Principle of Use Until Satisfaction*, we assumed that any encounter between a scholar and a book was equally likely to represent a complete use event. That is, while some encounters were longer than others, in each case the user could continue until satisfied. In the use of a traditional paper book, that use could have been as little as a browse of its Table of Contents while standing at the stacks or as much as checking it out and reading the entire book. The average number of hits per monographic book user, per half year, hovered at four to five throughout the study period. Thus, typically use of the online books had depth beyond clicking on and looking at the Title Page-Table of Contents file.

Only one-fifth of the users of monographic online books used more than one or two titles throughout the study period. This is not surprising, given that we lacked the critical mass of books to attract a scholar repeatedly.

### ***Online Book Use Occurred in A Mixed Format Environment***

Surveys and interviews indicated that scholars were not simply reading books online. They tended to browse the books online and then to print out relevant portions or to look for print copies for extended reading. Sometimes they referred to the online version to track down a quotation or a citation. Other studies have found that this pattern is common in using journals as well.<sup>28</sup>

### ***Online Books Use Varied by University Cohort***

The distribution of use among Columbia cohorts varied by book category. For some books, undergraduates were the primary users, for others graduate students in the Arts and Sciences or in professional programs were dominant. Faculty members were never more than a few percent of users.

### ***Online Books Did Not Displace Personal or Library Copies***

Our in-class survey, administered when a reading in the online collection was assigned in a course, revealed that a growing share of students used the online versions in some way, but few read online. We asked how the students had read the assignment. By spring 1999, 39 percent indicated some form of online book was used, the same share as used their own copies of the book. The student's own copy of a book was both the most common single method of reading an assignment and the single most preferred method.<sup>29</sup> However, by spring 1999, 43 percent preferred some form of online book use while only

---

<sup>27</sup> We could not track use of print-on-paper books that did not include checking them out so we lack a total count of use of books in that format.

<sup>28</sup> See the articles on the University of Illinois DLI project and U.K. SuperJournal project that were cited earlier.

<sup>29</sup> The responses to these questions would depend on the nature of the assignment. If only one chapter of a book was required reading for a course, our understanding is that few students would be likely to purchase that book. On the other hand, if the whole book was assigned and the book was available at a reasonable price, a much greater share of the students would be likely to purchase it.

a third preferred to use their own copy.

### ***Cost and the Use of Books***

Supposing for the moment that a scholar is familiar with online books, has easy access to them, and incurs no cost from use of an online book, we can illustrate the way that scholars seemed to feel about these books in terms of a simple preference table. There are many modes of using a book, but for simplicity we divide them into two: read much of the book and read little of the book. Similarly, there are many ways that we can describe the cost of a paper version, and for simplicity we simply divide them into two cases: low and high. If a scholar intended to read little of a book, the ability to locate things within it would be more important, while comfort of reading, annotation, etc. would be less important. In this case an online book would be preferred, without regard to the price of the paper version. However, if the scholar intended to read much of the book, the inconvenience of online reading would be the dominant factor, forcing the scholar into a *buy versus borrow* decision. The following simple table shows these four cases.

<b>Scholar's Preferences for Book Access</b>		
	<b>Read Much</b>	<b>Read Little</b>
Low Cost Book	Buy	Online
High Cost Book	Borrow	Online

This table makes it clear that, if a library wishes to respond to its scholars' preferences, its collection of paper books should continue to exist to serve the substantial reading, high cost situation. Two of the quadrants demand that the library provide online access to electronic books. As the readability of online books improves, and scholars' habits evolve, the meaning of *read much* will change, so that the two quadrants on the right of the table will steadily expand as a share of total preference for use of scholarly books. If the library is to remain a key resource for scholarly access to monographs, its online holdings must expand to keep pace with these evolving preferences.

### **Librarians Find Appeal In Online Books**

In discussions from summer 1998 to fall 1999, college and university librarians expressed great interest in and considerable optimism about the potential for online books in their collections. They viewed reference works as having particularly great utility in online form. But they also saw value in having books that are in high demand, of transient topicality, or not part of their print collections available in this format. Fewer were willing to contemplate acquiring online versions of books that are in their standard research collections, mainly because they believed that they could not afford to pay for such a book twice.

Librarians were concerned about how the marketplace for scholarly books would evolve and particularly how online books would be provided. How would online books be packaged and priced? What guarantees of availability in the short and long run,

preservation, format updating and the like would publishers or vendors provide? What conditions of use would publishers and vendors seek? What would happen to the important copyright concept of *fair use*? Librarians also hoped to see cataloging and usage statistics provided by vendors, as well as user-friendly design of both the general interface to online books and the individual books.

### **Potential Market Arrangements**

Our findings on scholars' interests in online books and on the costs of providing online books suggest that scholarly publishers should test new product models that combine print and online availability of books. With the right mix of offerings, the overall market should expand. The goal of these models would be to increase availability of books to scholars and profitability to scholarly publishers. Without market experimentation, we cannot know which combination would most enhance the social good.

Mixed product models could expand research library markets, develop new academic library markets, and attract more individual scholars as buyers. Possible options include:<sup>30</sup>

- Retain and expand the research library market with:
  - A free online version, i.e., an enhanced product package, or an online version available at a small incremental price, when the library purchases the print version.
  - Modestly priced online collections offered to libraries that buy a substantial share of the titles in print version. This would likely expand sales of the print copies somewhat.
- Expand sale of scholarly monographs into the libraries in the U.S. and abroad that purchase few such books. Price online collections for sale to consortia. Make an on-demand electronic version available as an attractive alternative to interlibrary loan.
- Encourage individual scholar's purchases of monographs with:
  - Pay per electronic view. This could be attractive to scholars not having library access to print or electronic copies.
  - Online ordering of a print copy at a discount from the site of a library's online version.
  - Lowered prices for the print copies to reflect their modest marginal cost.

The processes of designing, producing, and marketing online books are in their infancy. Some scholars who have begun to author books with an online format in mind find it much more challenging than anticipated. How should primary documents be included? How should they be incorporated into the analysis? Will the text automatically update itself? Such issues do not arise in authoring traditional scholarly monographs.

---

<sup>30</sup> These market models are discussed at greater length in Mandel, C. and Summerfield, M., *Scholarly Monographs Online: Potentialities and Realities Suggested by the Columbia University Online Books Evaluation Project*, January 1998, at <http://www.arl.org/scomm/epub/papers/mandel.html>.

Over the next several years the concept of online books will resolve itself so that scholars, authors, publishers, vendors, and libraries will be able to decide what combinations of modes and models will be most beneficial for the scholarly community as a whole, including authors, publishers, vendors, libraries and readers. In the final analysis the entire system is one by which authors communicate to readers across spans of both space and time. The present system has evolved in a way that supports value-adding transformations at several points: the publisher adds quality and authentication; the vendor adds distributional services; the library adds organization and preservation, as well as further quality certification. With electronic books there is no reason to sacrifice any of these values. However, the final economic driver will be the value delivered to readers, which is the ultimate quid pro quo that leads to the injection of a revenue stream into the process. As technology permits us to blend several of the traditional stages of the communication chain, ultimately economic models will more accurately reflect that value delivered to readers, and will correspond more closely to their allocation of attention to the works, whether in reading or in citation. Writers, editors, and publishers in the world of online magazines like *Salon* are already finding that they can assess specific articles, and even headlines, by logging the amount of attention that readers pay to them. The scholarly world is also very much concerned with the capture of attention, as measured by citation analysis. Thus it is likely that such analyses will be introduced at several points in the chain of communication, and will, eventually, become the basis for assignment of costs and prices.