

## The Chicago Digital Distribution Center and BiblioVault Repository

*Ventures to Improve the Dissemination of Scholarly Books  
and to Enhance the Operations of University Presses  
A Case Study of Collaboration among University Presses*

Delivered at the Symposium  
*The Economics of Digitization:  
Toward Sustainability and Institutional Collaboration*  
May 17, 2004

*Introduction* ..... 1  
*University Presses: Their Roles and Challenges*..... 1  
*Developing the Chicago Digital Distribution Center and BiblioVault*..... 3  
    The Concept ..... 3  
    University Press Adoption of the CDDC and BiblioVault ..... 6  
*Accomplishments*..... 7  
    BiblioVault File Standards ..... 9  
    Backlist Files—Conversion of Titles Lacking Digital Files ..... 10  
    Frontlist Files—Preparation and Deposit of Books with Digital Files ..... 10  
    BiblioVault Repository ..... 11  
*Conclusion*..... 13

## Introduction

I am pleased to be able to talk with you today about the Chicago Digital Distribution Center and its BiblioVault repository. This venture is structured as a component of the University of Chicago Press but it involves substantial and growing collaboration within the scholarly community, including other presses, libraries, and scholars.

Begun in 2001, this initiative involves the use of digitization to aid the economics of scholarly communication for a growing group of university presses and, hence, for the scholarly community as a whole. Nearly thirty university presses now participate in the venture and we expect several more presses to join by next spring.

In the first phase of this initiative, which has just ended, the University of Michigan Digital Library Production Service provided digitization services.

The BiblioVault Advisory Board that includes university press directors, library directors, and information technology specialists will meet for the first time in early June.

In short while this venture is being developed under the auspices of the University of Chicago, it is truly a creature of the scholarly community. In order to enable it to evolve to be as useful as it might be to this community, we will work to increase collaboration among presses, libraries, and scholars.

## University Presses: Their Roles and Challenges

University presses are a critical element in the ecosphere of teaching, research, and learning in the western world, particularly for monographic books. I could consume far more time than is allotted for this talk in describing the role that university presses play in this ecosphere. Suffice it to say that through their decisions on which monographs to publish, university presses help determine which scholars are hired and promoted in the academy, particularly in the humanities and social sciences. These scholars rely on university presses, and a few commercial presses, to publish their career-building books.

The economics of university presses are increasingly dire. Traditionally university presses have balanced their books with the assistance of subsidies from their parent institutions and from other organizations that are interested in their publications. Not long ago subsidies from parent institutions averaged \$500,000 per year per press.

In general, a small fraction of the books published by a university press are the works of students or faculty at that university. For example, at the University of Chicago Press in recent years this share has been about 10 percent. Thus, a university press

supports the mission of disseminating research produced at its own university—and enhancing the stature of its students and faculty and, thus, the university itself—to only a modest degree. In difficult times, university administrators begin to wonder why they should be subsidizing this enterprise. There are many other ways that those funds could be spent to further a university's mission, including for programs that would more directly benefit the university.

With state budgets under great pressure in the last few years, their grants to their public universities have eroded. Earlier this month the press at one Midwestern state university reported that its administration was seeking to cut spending university-wide and planning to cut the press's subsidy by nearly 40 percent. The University of Idaho Press is slated to close this year.

Similarly private universities' endowments and their yield suffered with the decline in the stock markets and these universities have been less willing to underwrite their presses. Earlier this spring Northeastern University announced that its press would either close its doors or become part of a consortium.

The demand for university press books was never large and it has eroded substantially over the past several years. Among the factors contributing to this decline in demand are the following:

- 1) The strains on university and college library budgets caused by the huge increases in the cost of STM serials and the need to invest in information technology,
- 2) Tight budgets in academic institutions overall leading to increases in libraries' acquisitions budgets that are smaller than the increase in the cost of materials,
- 3) The narrow specialization of the typical university press monograph, and
- 4) The increased efficiency of the used book market.

Currently most new scholarly monographs sell fewer than 1,000 copies in their lifetimes, and many sell fewer than 400 copies to libraries. Authors and presses hope that their books will be adopted by courses, particularly at the graduate level. In fact, relatively few of them are adopted by more than a handful of courses to such a degree that students regularly purchase them. Easily half of the copies of books shipped to college and university bookstores for sale to students are regularly returned. Many students buy used copies, share copies with friends, or use library copies. In addition frequently instructors will assign a chapter or two from several titles and the students will not need to purchase these books.

## Developing the Chicago Digital Distribution Center and BiblioVault

Initially the CDDC was envisioned as a provider of services to presses that use the services of the Chicago Distribution Center (CDC). At present the CDC provides warehousing and related distribution services to nearly thirty presses with over 26,000 active ISBNs. The warehouse contains about 11 million units. The average title has over 400 copies in stock. In the most recent 12 months the CDC shipped about 4.25 million units (for an average of about 160 units per active ISBN) and managed the return of over 1 million units. Over 70 percent of the ISBNs at the CDC sold fewer than 200 units in the last year, and the majority of the titles sold no more than 50 units.

Page 3 of 15

### *The Concept*

In 2000, the University of Chicago Press began to develop the concept of the Chicago Digital Distribution Center (CDDC) as a short-run digital printing (SRDP) center located in space adjacent to the CDC warehouse. The CDC's sophisticated inventory software would be modified to trigger a print order when a book stored at the CDC ran out of stock. A preset number of copies would be printed within two days. The needed units would be shipped with the rest of the book order that triggered the printing. The remaining units from the printing would go into inventory.

This system would enable CDC-client presses to do the following:

- 1) Keep in print or bring back into print paperback books for which demand existed but was too low to warrant continued offset printing;
- 2) Cut their risk of over-printing titles for which demand was uncertain;
- 3) Improve their cash flow situation; and
- 4) Provide copies immediately when an offset printing was on order but the title was out of stock, thus saving course-adoption of titles, the bread and butter of university press sales.

As planning advanced, the CDDC came to include the BiblioVault repository for digital book files which would feed the CDDC digital book center and other digital and offset printers and marketing agents around the world. The BiblioVault repository would include scanned image files for titles for which no digital files existed, printers' files for recent and new books, and related metadata.

With the development of the concept of the BiblioVault came the related concept that this repository could be the foundation of a program for providing scholars with online access, directly or indirectly, to the books in the BiblioVault. This concept led to the expansion of our program of digitizing the older titles. The CDDC development effort would include not only scanning the books to provide image files

for use in printing but also OCRing and page level metadata to move the initiative one step closer to being able to deliver these books in electronic form to the scholarly community.

*Page 4 of 15*

However, the expanded vision of the BiblioVault made it clear that this repository could provide valuable services to all university presses regardless of the location of their distribution center. The economies of scale of such a repository are such that it made sense for university presses to work together to develop this resource for mutual benefit.

Electronic access to the books in the BiblioVault will be possible when the BiblioVault, participating presses, and other related entities have resolved issues of rights clearing, business models, e-commerce, subscription fulfillment, marketing to libraries and to online service providers, and delivery platform. We will work intensively with presses, libraries, scholars, and information technologists to resolve these issues over the months ahead.

In September 2001 The Andrew W. Mellon Foundation granted the University of Chicago \$1.5 million to fund initial development of the Chicago Digital Distribution Center and its BiblioVault repository for digital book files. Last December the Foundation granted another \$1.25 million to further the development of the CDDC's infrastructure, to expand the number of presses involved, and to increase the number of titles in the BiblioVault repository. By summer 2005 the BiblioVault will contain digital files for over 12,000 titles, including files for about 5,000 titles that were digitized with grant funds.

Great strides have been made in the design and development of the CDDC, its digital publishing services, the BiblioVault repository, involvement of university presses, and digitization and deposit of books. Fundamental components are in place to meet the needs of university presses—and those who use their books—in the years ahead.

The CDDC initiative has two key goals in the short term.

**First, it seeks to improve the dissemination of books published by university presses, including making available titles for which demand is low or sporadic.**

Libraries and scholars lament that books they would like to acquire or assign in classes are no longer in print. Presses, on the other hand, confront a marketplace in which most of their titles sell fewer than fifty copies a year and virtually all of them sell fewer than 200 copies a year. When any of those titles go out of stock, printing them in the traditional manner (offset) is uneconomical. Due to the substantial make ready activity in offset printing, printers seek print runs of at least 600 copies.

In the past, university presses have allowed such books to go out of print. In fact, university presses have typically allowed about as many titles to go out of print as

they publish each year. For example, Chicago had about 5,000 active ISBNs for several years even though it publishes approximately 250 titles annually. Only with the availability of short-run digital printing and, in particular the CDDC, has Chicago's list of active ISBNs begun to grow again.

The CDDC will allow university presses not only to maintain their current titles in print for as long as there is demand but also to bring back into print titles that libraries and scholars seek to purchase.

**Second, the CDDC initiative seeks to enhance the operations and the finances of university presses.**

Clients of the Chicago Distribution Center derive the greatest benefit from participating in the Chicago Digital Distribution Center, but the BiblioVault repository provides valuable services to other university presses as well. These presses can use the BiblioVault to manage the digital files for the cover or jacket and interior of their books, printing decisions, and marketing.

As envisioned, the CDDC is well integrated with the CDC's information systems. A CDC press designates a book that has demand that is insufficient to justify offset printing and is out of stock or running low on stock as a CDDC SRDP title. In doing so, the press sets the standard print run, reviews its price and discount class, and then lets the system dictate the printing of that title. If there are no digital files for that title, the CDDC manages the digitization of the text and the cover of the book, and places the resultant files in the BiblioVault repository. When an order requires more units of the title than are in stock, the CDDC system triggers a printing, usually of 24 units. The digital printing center fetches the necessary files from the BiblioVault. Within two days the units are printed and the order is ready to go out the door. Any units beyond the number needed for that order go on the shelf to be used for subsequent orders.

No longer does the publisher need to devote scarce staff time to reviewing the stock situation of these titles that sell in modest quantities, deciding whether to print them again, and placing printing orders. No longer does the publisher tie up funds in inventory that will take several years to sell if they sell at all. After all the few professors who have been assigning that title may retire or change their reading lists.

The CDDC enables participating presses to continue to offer their books and to generate modest income streams from them, while investing only modest upfront management time and little cash over time.

Currently presses are using the digital book center primarily for older in-print titles for which demand is too low to justify reprinting the titles via offset printing. However, presses can also use it to advantage when a newer title is running out of

stock and it is not clear whether demand will warrant a traditional offset printing. Presses can also use SRDP for the initial paperback printing of a title for which the demand is uncertain. If the title enjoys unanticipated demand, an offset printing can be ordered. The digital book center is also valuable in filling orders until an offset reprinting for a title reaches the warehouse.

Short-run digital printing will have a greater role in scholarly book publishing when publishers begin to make more strategic print run decisions and when distribution centers charge storage fees that reflect the cost of holding these overstocks. As current stocks of scholarly books are depleted through sales or pulping of excess inventory, short-run digital printing will become the norm for the vast majority of university press titles.

### *University Press Adoption of the CDDC and BiblioVault*

At present the CDDC provides digital printing services—now limited to the digital book center at the Chicago warehouse facility and Edwards Brothers' facility in Ann Arbor—and the services of the BiblioVault repository, including conversion of books to digital formats for deposit in the BiblioVault.

In late 2001, the CDDC invited seventeen CDC-client presses to participate in this new initiative. The service agreement offered each press free digitization and deposit in the BiblioVault of a number of older (backlist) books determined by the size of its book publishing program, financial assistance with obtaining electronic files for recently and newly published books and free deposit of those files in the BiblioVault, and free storage for all of these files in the BiblioVault through the end of the grant period. In return, the participating press agreed to retain its files in the BiblioVault and to make them available for full-text searching via the Scholars' Portal. If a press chose to withdraw the files for a book from the BiblioVault, it would repay the expense of digitizing and depositing that book. The presses were given the opportunity to use the digital book center for printing but they were not obliged to do so. All seventeen of these presses became CDDC participants.

From its inception, the vision behind the CDDC attracted the interest of university presses that were not clients of the Chicago Distribution Center. By spring 2003, the BiblioVault database and the Publishers' Portal (a secure Web site through which publishers manage their electronic files) were sufficiently developed that additional presses could be accommodated. In addition, the initial participants had been able to deposit fewer files for new books than anticipated. As a result we were able to use grant funds to provide BiblioVault services and to fund obtaining and depositing recent titles from several additional university presses.

In the second half of 2003, publishers at the CIC (Committee on Institutional Cooperation) universities—Illinois, Indiana, Michigan State, Pennsylvania State, and Purdue—that were not CDC clients were offered participation in the BiblioVault. All but Illinois are participants. Illinois is too busy with other activities to take on this new endeavor at this time.

The University of Virginia Press began participating in fall 2003. In December 2003 Harvard, New Mexico, and Yale decided to participate in the BiblioVault. All of them have begun programs for depositing files for recent books and for pinpointing backlist titles to be digitized.

In February 2004 the University of Texas Press decided to participate; the University of North Carolina Press has just begun to deposit titles. We anticipate that several other university presses will become involved in the BiblioVault later this year.

Chicago is working with scholarly publishers and particularly their production and business managers to delineate the long-term benefits of smaller print runs and more efficient cash management. Such education occurs in the form of discussions with individual presses and also through presentations at industry meetings. In early summer 2004 as part of the current grant activity, we will issue a white paper on this topic.

In July 2004, the CDC will begin to charge two cent per book unit per month for inventory that exceeds a five-year demand at the current rate of sales. This charge reflects the cost of that warehouse space. CDC client presses are beginning to pulp excess inventory in order to avoid this charge.

Plans for the CDDC anticipated a modest adoption rate. We knew that university presses commonly allow as many books to go out of print each year as they publish, and our allocations of conversion subsidies were based on this potential demand for SRDP. As mentioned earlier, presses generally lack manpower to tackle new operating initiatives with alacrity.

### **Accomplishments**

The CDDC initiative has accomplished the following:

- Establishing a short-run digital printing center adjacent to the Chicago Distribution Center (CDC) warehouse, including developing a smoothly functioning operation in concert with Edwards Brothers, the printer that is operating this digital book center.
- Modifying the software controlling the Chicago Distribution Center fulfillment operations in order to manage the digital printing center's activity.

- Designing and developing the BiblioVault—a multifaceted state-of-the-art digital file storage system—to support short-run digital printing, conventional offset printing, marketing initiatives, and future delivery of partial and full books online to libraries and other intermediaries as well as scholars.
- Working with scholarly publishers, both those who use the services of the Chicago Distribution Center and those that use other distribution centers, to develop the features of the CDDC and to begin using either its full set of services or just the BiblioVault repository.
- Working with participating presses to identify the older and new titles that they would deposit in the BiblioVault and designate for short-run printing, to adopt standards for the systems they would use in participating in the CDDC and the BiblioVault, and to obtain books and files for conversion and deposit.
- Designing and launching the Publishers' Portal to the BiblioVault through which participating publishers can securely deposit and access their book files and information about those files.
- Designing and developing various tools that enable operations staff to work with files in the BiblioVault efficiently.
- Designing and launching the Scholars' Portal to the BiblioVault that allows the public to learn about the contents of the books in the repository.
- Working with Edwards Brothers and design and production experts to set standards for files to be deposited in the BiblioVault.
- Developing an operation to manage the BiblioVault, to efficiently bring files into the repository, to manage those files after they are deposited, and to work with presses to make the BiblioVault valuable to them.
- Developing and distributing guides to file standards and to use of the Publishers' Portal for the BiblioVault.
- Digitizing and depositing in the BiblioVault 2,500 backlist titles and working with presses to bring digital files for recently and newly published books into the BiblioVault.
- Developing relationships with digitization vendors.
- Working with university presses to adopt financial and management practices that would enable them to use new technologies to keep scholarly books in print, to use the new technology to reduce their risk and improve their cash flow, to bring out-of-print books for which interest

continues back into print, and to disseminate information about their books to a global research community.

- Developing an initial understanding of the challenges that author contracts and permissions place on scholarly presses that would like to make their books available online and a framework for thinking about how to confront these challenges successfully.

*Page 9 of 15*

A few statistics will highlight the distance this initiative has traveled from its start-up in fall 2001.

- Eighteen CDC client presses are using the CDDC. Among the larger of these presses are Chicago, Michigan, Minnesota, Pittsburgh, Stanford, and Wisconsin.
- The 10 non-CDC presses that are participating at this time are Harvard, Indiana, Michigan State, New Mexico, North Carolina, Penn State, Purdue, Texas, Virginia, and Yale.
- Electronic files for over 4,000 books have been digitized and/or deposited in the BiblioVault repository. Sixty percent of the titles are backlist titles that were scanned to produce raster (image) PDF files and OCR'd to produce ASCII files that could be used to search the contents of the books. The rest of the books are recent and new titles for which presses had or could produce or obtain vector PDF (printer's) files.
- December 2003 was the peak month to date for the CDDC's digital book center due to orders by college and university bookstores for the second semester of the academic year. In that month the CDDC digital printing center produced over 16,000 units, over five times the number printed a year earlier. In 2003 the CDDC's digital printing center produced nearly three times the target units. In 2004 to date the digital printing center has been operating at or above its contractual target.

### ***BiblioVault File Standards***

In order for the BiblioVault repository to offer the university press community a good balance of cost and benefit, it must have standards for files that presses and their vendors will adopt for all books published, now and in the future. Such standards are also needed for preparation of files for older books that are digitized so that, regardless of the vendor that undertakes digitization, the files will be consistent and can be used successfully in digital and offset printing and for other intended uses (described below).

Early in the development of the BiblioVault, we issued standards for the printers' PDF files for the text and for book jackets so that files for recent books could be deposited in the BiblioVault with minimal effort. We found, however, that Quark and other desktop publishing systems—not PDF—files were often all that university presses had for their books. These publishers needed help in converting these files to PDF for deposit in the BiblioVault. Staff from the design and production departments at Chicago worked with the BiblioVault operations coordinator to produce guides for conversion of such files into vector PDF files; we have distributed these guides to participating presses. This team is working closely with BiblioVault presses so that they understand and follow these procedures and deposit their files routinely.

### ***Backlist Files—Conversion of Titles Lacking Digital Files***

The CDDC process for the conversion of books for which no usable digital files exist (which we call backlist titles) involves several steps: (1) 600 dpi bitonal scanning of book text pages to produce TIFF images; (2) 600 dpi grayscale scanning of halftones (an optional step initially that will be standard from now on); (3) 600 dpi four-color scanning of the elements of the book jacket or cover, manipulation of those elements, color correction, and finally creation of a PDF for use in digital or offset printing of the cover or jacket; (4) optical character recognition (OCR) scanning of the TIFF images to create an ASCII file that can be used in a full-text search; (5) addition of several elements of page-level metadata to the ASCII file; (6) creation of a PDF of the TIFF images for the book. This process produces files that can be used for various printing and marketing purposes and which can be delivered as PDFs online after simple down sampling of the PDF files and the addition of bookmarks.

Approximately 2,500 backlist books were digitized fully and input into the BiblioVault.

### ***Frontlist Files—Preparation and Deposit of Books with Digital Files***

The CDDC staff, supplemented by experts from the production and design staff of the University of Chicago Press Books Division, is working with CDDC presses to deposit vector PDF (printer's) files for both covers, or book jackets, and interiors of recently and newly published titles in the BiblioVault in standard formats. Those specifications have been provided to participating presses. The presses are beginning to supply those specifications to their vendors and to request that their printers provide them with final files for a book that conform to those standards. The presses then send these files to the BiblioVault for deposit.

Vector PDF files have several advantages over the raster PDF files that the BiblioVault is producing for backlist titles for which such files do not exist. These

include (1) the text of the books can be modified with new editions and a new PDF file created; (2) the file is natively searchable; (3) the files are smaller; (4) the quality of the printed products from these files is better; and (5) these are simply copies of digital files resulting from the production of book so there is no incremental cost of producing these files. However, to produce such files for backlist books—for which no digital files were initially created—would be several times more costly than producing the combination of raster PDF files and ASCII text files that the BiblioVault is using to provide printing and searching functionality.

Even in recent years, many university presses have not acquired or retained copies of these final files for either the interiors or the covers and jackets as their books were published. In addition, most university presses have limited production staff and use freelance designers and typesetters for their books and the typesetting files for the books are not housed at the press itself. Assembling the files for a book, often in desktop publishing software, and converting them to vector PDFs has been a challenge for the presses and for the BiblioVault staff that has been working with their production managers to develop the appropriate processes.

### ***BiblioVault Repository***

The first Mellon grant enabled the initial design and programming of the BiblioVault as an electronic repository for book files. The BiblioVault system includes (1) a server and database for maintaining the files themselves; (2) a secure Web interface, the Publishers' Portal, through which participating publishers can securely deposit and access their book files and information about those files; (3) various tools that enable operations staff to work with these files efficiently; and (4) a Scholars' Portal that allows the public to learn about the contents of the books in the repository.

The BiblioVault repository provides a system for persistent storage of electronic book files and related data; it includes a relational database, file storage on a UNIX file system, and a Publishers' Portal that gives BiblioVault staff and staff of the participating presses access to the information and files in the repository.

In the current BiblioVault design, the primary content object is a "title." Each title can have up to three "book" forms: clothbound, paperback, and e-book. Each form has a distinct identifier, or ISBN. Descriptive metadata, such as author, title, copyright year, and subject headings, is stored in the database and is keyed to each ISBN. The database is also used to record events, or steps, in the process of ingesting book files and to manage batch-processing of BiblioVault files and metadata.

The BiblioVault system was constructed primarily from open source software components. It uses standard protocols for communication and a database accessible to users via the Web.

Both the Publishers' Portal and the Scholars' Portal are accessible at <http://www.bibliovault.org/>; the latter is now the front page for the BiblioVault. The features of the Publishers' Portal from the viewpoint of a participating press can be reviewed by using "demo" as both a login and password; that will provide access to the system for Midway Plaisance Press, a dummy press created for this purpose.

Page 12 of 15

The BiblioVault Publishers' Portal is designed to allow access to descriptive title information (metadata) and to electronic files on the part of individuals who have password identifiers: publisher/owners of the book files, BiblioVault and CDDC staff, and authorized vendors (e.g. printers or marketing service providers). From the portal, authorized users are able to do some or all of the following tasks depending on their status: (1) add new titles to the BiblioVault collection; (2) upload files for archiving and conversion or transmission to vendors; (3) view and modify metadata records harvested from library catalogs; (4) manage user accounts; and (5) track the history of book files in the repository.

While the Publishers' Portal is user friendly, staff members of participating presses are supported in their use of the BiblioVault by a printed and online handbook *Using the BiblioVault*. In addition, they have been supplied standards for the files that are to be submitted to the BiblioVault and instructions on how to create such files.

In January 2004, the Scholars' Portal became accessible to the public via the Internet in beta test. It offers (1) full-text searches on all BiblioVault content; (2) field searches on author, title, ISBN, copyright year, and Library of Congress subject headings; and (3) browse features that allow users to find books by title, author, or publisher. The Scholars' Portal features direct links to the publishers' Web sites and to their shopping services. At this time the results of Scholars' Portal searches are the titles and basic information about the books in the BiblioVault repository that include the terms or other variables searched. Searches do not currently yield text from the books. In the current grant period, we will investigate means of making selected, relevant portions of book front matter or text available as part of the results of a search of the Scholars' Portal.

The marketing department of the University of Chicago Press Books Division will assist BiblioVault staff and participants in promoting the Scholars' Portal using low cost, Web-based marketing techniques, including notices to relevant library and scholarly list serves, a logo that participating presses can put on their Web sites and printed brochures, and the like. We will roll out this campaign slowly as the contents of the Scholars' Portal become richer.

During 2002, the BiblioVault team reviewed metadata classification efforts. After evaluating the element sets and definitions, we assembled a nominal metadata set that enables us to relate information about titles in the BiblioVault to other metadata systems. We chose to store BiblioVault metadata information in tables in a relational database, to facilitate import from and export to a variety of online systems. We will strengthen this capability later this year.

*Page 13 of 15*

## **Conclusion**

The Chicago Digital Distribution Center, and particularly its BiblioVault repository initiative, is an example of a collaborative venture that involves many members of the scholarly community as participants and advisors. The venture is managed by the University of Chicago Press but the longer term realization of the greatest potential of this venture will require the agreement of university presses, libraries, and scholars that the vision of the CDDC for delivery of book content is valuable and fair to all elements of the community.