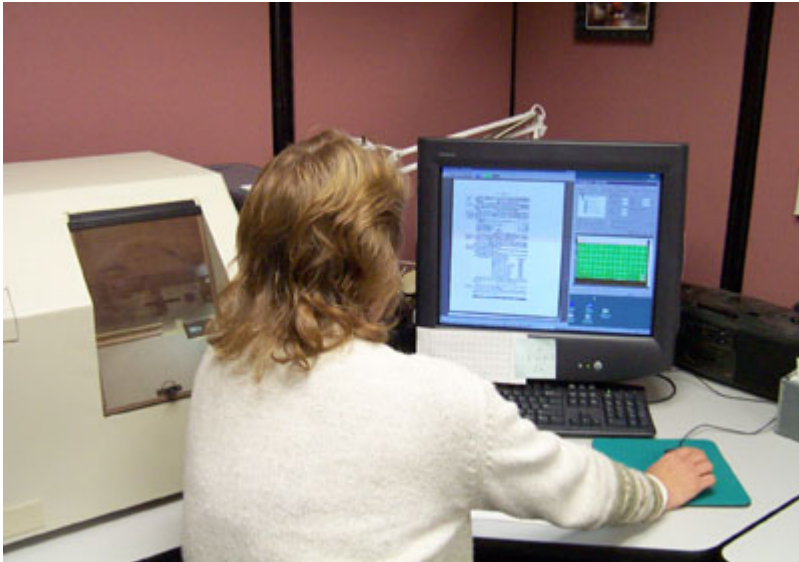


# American Originals



## Part I

Drive up Main Street in Chester, Vermont, and you will see all the staple ingredients of a small New England town, and then some. Its four churches tell a tale of waves of growth and change: a white clapboard Congregational, a red brick Baptist, a pale gray and blue Victorian Episcopal, and the Scots-influenced old stone. From the Romanesque library to the Pollyanna-esque Country Girl Diner, the town is dense with the various sizes, shapes, and periods of New England's past.



Fig. 1. The front of NewsBank's Chester, Vermont, production facility, where Evans Early American Imprints, Series I, once took form and Evans Digital Edition is now coming to life. (The entire NewsBank facility is actually much larger.) All photographs and screen shots of Evans for this article were made by Katherine Stebbins McCaffrey.

The Main Street production facility of NewsBank, Inc., seems right at home in

these surroundings: the company resides in a white clapboard building that, long ago, was the Adams Funeral Home. Out back, the printing shop runs off color copies and other small orders for local customers. Yet, inside, among a generic assemblage of office cubes, a couple dozen people and a handful of computers labor at a decidedly nonlocal project: ushering the Readex Microprint Corporation's microform collection of *Early American Imprints, Series I, 1639-1800* into the Digital Age. Here, photographic copies of thirty-six thousand plus of the nation's oldest printed documents are dissolving steadily into ones and zeroes that will dramatically reshape the study of early American history and life.



Fig. 2. Kelly Lauren monitors input from the Mekel scanner after feeding fiche through the clear plastic portal on the left. The green square on the right represents the entire microfiche, and the larger image shown on the left is one page scanned from it.

NewsBank began the fascinating process of creating the *Evans Digital Edition* (as this new incarnation is called) in July 2002, and the project is scheduled to be in production through July 2004. Yet, the creation of *Evans Digital* encompasses another history that lies between the immediate one of digitization and the distant one of the founding of the American nation. Three men, in particular, toiled in that middle ground, and their grand plans still dot the landscape of scholarly inquiry today, as well as profoundly shape the structure of the *Evans Digital Edition*.

Librarian and longtime cataloger Charles Evans laid one foundation in 1902, when he aimed to compile a "chronological dictionary of all books, pamphlets, and periodical publications printed in the United States of America from the Genesis of printing in 1639 down to and including the year 1820, with bibliographical and biographical notes" (as his advertisement trumpeted). Supporting himself and his family almost entirely on subscriptions to the *American Bibliography*, the aging Evans obsessively went after his quarry armed with pencils and corset boxes bursting with note cards, working alternately out of his home in Chicago and numerous East Coast archives. He died in 1935 after tracking down 35,854 imprints and publishing twelve volumes,

but before he could complete the listing of publications for the year 1799.

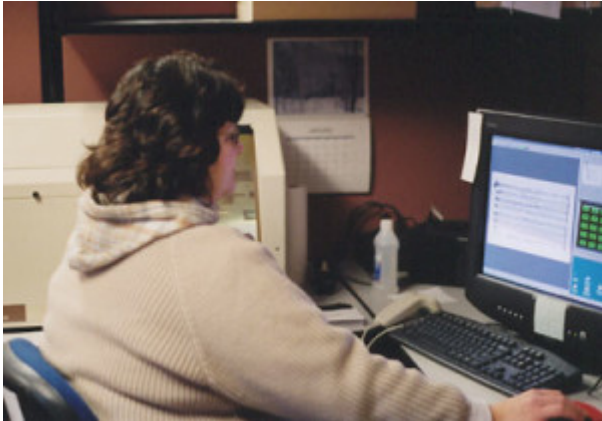


Fig. 3. Cindy Tufts scrutinizes a page of sheet music scanned by the Mekel. Each page lights up as a smaller green square as it scans (far right), until the entire fiche becomes green, signaling the last page

As Evans's astonishing and extraordinary work gained prominence, he found himself torn between the archivists who sent lists of errata for and oversights in early volumes and the librarians who excitedly pressed him for the next volumes in the series. During this period, one fellow librarian consoled Evans, "No one has ever prepared the perfect bibliography and no one ever will." Happily, though, someone kept trying.

That someone was Clifford Kenyon Shipton, the early American scholar and head librarian at the American Antiquarian Society in Worcester, Massachusetts, who, in 1954, built a new level on Evans's base. Shipton completed the catalog of imprints for 1799 and 1800 and initiated the AAS's publication of the Evans volumes in a series of indexed books. He also contracted with Readex Microprint Corporation to shrink all the books in the Evans series down onto Microprint in an ambitious effort to begin carefully correcting the bibliography and widely disseminating the film for every extant imprint listed. As if all that were not time consuming enough, he created "target cards" (each with the basic bibliographic information about a document) for every title.



Fig. 4. Standing solo in another room, the computer running the OCR software “reads” and translates text at a rate of ten seconds per page. The rectangles with large black characters show close-ups of pages while the rectangles with smaller gray characters next to them display the ASCII text as it is generated.

In choosing Readex, Shipton tapped another man who knew something of fantastically large, impossibly idealistic projects: Albert Boni, founder of the Modern Library. Boni had let Vermont’s granite ledges and millstreams lure him from the New York publishing world to Chester in 1945. There, while Luther Evans and Verner Clapp experimented with space-saving methods for converting hundreds of thousands of titles to microfilm at the Library of Congress in Washington D.C., Boni tinkered away trying to solve what he understood to be the problem at the heart of microreproduction: microfilm itself.

From the start, Boni’s objectives for his technology ran counter to the vein unearthed by Nicholson Baker in *Double Fold* (Random House, 2001). Boni neither rejected paper as a storage device nor crusaded for clearing library shelves. He actually hoped to convince librarians to buy *more* titles printed on *more* paper, albeit in the modest form of six-by nine-inch cards, each of which could hold one hundred pages of a standard book. The problem with microfilm, Boni thought, was that it was inordinately expensive. Miniscule reproductions on special quality paper would be much more affordable both to produce and to acquire, and could last up to two hundred years, according to one U.S. Bureau of Standards’ estimate. Moreover, Boni did not seek to replace titles, but to induce librarians to supplement their collections with cabinets full of some of the world’s greatest and rarest works—and, of course, the patented machines necessary to read them.



Fig. 5. Debbie Swisher compares printed copies of eighteenth-century North Carolina law with the microfiche version, adjusting the computer image for maximum quality and tagging the file with “metadata,” including cross-references.

As a collector, Boni had plenty of curiosity about photographic processes but little experience with them. Attempts to read early versions of Microprint with existing magnifying lenses and lamps only produced micropiles of ashes because of the enormous amount of light (and thus heat) required. But the fifty-three-year-old innovator could not be deterred by illegibility and incineration, and by his fifty-eighth year he had perfected his process and machine: pages reduced on microfilm were printed and reprinted like so many copies of a photograph onto emulsion-coated cards at the Chester printing shop and then read with a cheap lens that filtered out heat and a lamp with the intensity and power of an automobile headlight.

In June 1950, Boni opened the doors to Readex and began work on the entire series of British House of Commons Sessional Papers under the sponsorship of the American Historical Association. Soon the company was filming and printing the *New York Times* for the New York Public Library, as well as the Annual Subject Catalog for the Library of Congress and the declassified papers of the Atomic Energy Commission. Five years later, when Clifford Shipton contacted him regarding Evans’s *Imprints*, Boni must have relished the opportunity to erect not only a monument to America’s past and Evans’s effort, but to the usefulness of his own invention. Readex promptly rented space in the basement of the AAS, and began issuing the Microprint versions of *Early American Imprints*. It finished the first series of *Early American Imprints* in 1968, the year after Shipton retired from the AAS.

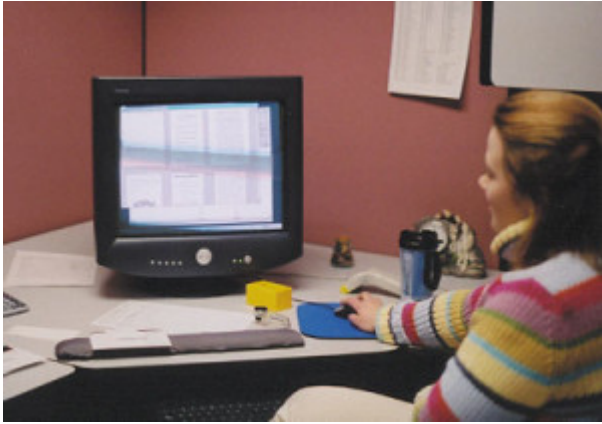


Fig. 6. Data Conversion Manager Korrie Heiden checks several pages in a document and links them together electronically. Boxes of forty-year-old microfiche—the source material for this project—sit on the desk to the lower left of the picture.

## Part II

It took thirteen years to create Evans's *Early American Imprints*, but it took only a decade for Boni's Microprint technology to fall out of favor with librarians and their patrons. Although Microprint remains indispensable to historical research today (despite the fact that machines that can read the cards are now few and far between and nearly impossible to repair), by the 1980s institutions favored microfiche and microfilm, which were read on relatively easy-to-use machines that also offered printing capabilities. In answer to the demands of the market, Readex switched to making archival quality duplicates of the original film used for producing Microprint cards, and in 1983, when Connecticut-based NewsBank, which specialized in archiving and filming newspapers, purchased Readex, the company simply expanded Boni's old printing shop in Chester.

Newsbank still occupies space in the basement of the AAS, in a dimly lit room crammed with file cabinets, a precariously overstacked desk, and—the very heart of the operation—a gigantic, Recordak (Kodak) microfilm camera. This is where, for the past forty years, Stanley Shapiro, who bears a remarkable resemblance to a trim St. Nick, has run the microform operation, carrying on the work started by Evans, Shipton, and Boni.



Fig. 7. Evans Digital Edition's stylish main page. Patrons may search in a variety of ways—this page requests all materials authored by Benjamin Franklin.

Today, Shapiro and his staff still follow the same basic procedure in filming the twelve hundred AAS imprints to be added to Evans Digital. Whatever is to be filmed—from pamphlets to schoolbooks to broadsides—is positioned carefully in a large, jointed cradle under a wide pane of glass. Gently the item is pushed upward with the help of a bicycle chain and a crank until it is pressed flat against the bottom of the glass, at which point the operator photographs it with a camera situated about three feet above. But the film created now is only a master copy to be sent to Chester to be digitized.

Naturally, the architecture of *Evans Digital Edition* resembles in some respects its predecessors'. To begin with, NewsBank's technicians in Chester have chosen to use the non-archival microfiche "originals" from the 1950s and 1960s as source material, since they yield better quality scans, though at a slower rate, than the archival copies made in the 1980s. To effect the transformation to sleek, modern binary code, staff members feed each carefully cleaned fiche into a high-quality Mekel scanner. The Mekel looks like an unassuming beige plastic rectangle attached to a fairly ordinary Dell desktop. Upon loading, a small green replica of the fiche appears on the monitor allowing operators to track each page as the computer scans the image at 400 dots per inch (dpi). The resolution of 400 dpi seems modest, below what consumer scanners offer now, until one considers that, on fiche, each page measures about seven-eighths (or less) of an inch wide. To achieve a final, enlarged resolution of 400 dpi, the scanner is acquiring an astonishingly dense 6,400 dpi, all while separating documents page by page, converting them to grayscale, and tagging them with a unique file name that will help trace their every step toward inclusion in the Evans Digital database.



Fig. 8. Franklin—as a bookseller—can be located in the "History of Printing" section of the site.

Because each page is independent in the system, in the next phase of

production, a computer running optical character recognition (OCR) software can pluck multiple scans from a holding computer's memory to generate ASCII text. To "read" a scanned page, the OCR machine runs a ten-second automated sequence that makes the image high contrast, adjusts the angle of the page, transcribes the text, and reconverts the image to grayscale. If left running day and night, the machine would probably finish perusing the thirty-six thousand books in Evans's *Imprints* in about nine months.

Even at the highly technical OCR stage, though, a tension between present and past technologies exists. Colonial American documents present a particular problem for OCR because of the quality of originals (many have broken type or show the effects of uneven hits on the press), variations in typefaces (such as the old style *s* that looks like an *f* to twenty-first century eyes), abrupt and inconsistent abbreviations (like *Quest.*

for *question* or *play'd* and *playd* for *played*), and orthography (consider *republick* or *rejoyce* or *booke*). Therefore, before a page travels back to storage, filters "clean" the text with a series of algorithms designed to find and fix errors. While NewsBank places fierce pride in these filters, real limits to the abilities of OCR to master the impressions made by seventeenth- and eighteenth-century metal type remain, evidence of which lies in the Help Section of the Evans Digital portal. There, patrons are instructed to abandon formatting their searches with the character *s* and instead insert a wildcard character like *?*—as in *Ma??achu?ett?* for *Massachusetts*.



Fig. 9. A screen showing multiple matches for the author search for "Franklin, Benjamin."

After being machine read and cleaned, the page image again rests in storage, where it awaits a quality control review and the addition of "metadata," which includes bibliographic data and cross-references. In the final stages of production, staff members call up a fiche's worth of pages looking for Shipton's target cards, which were also filmed and positioned on the fiche at the beginning of each document. Early on, using Evans's *American Bibliography* as a strict guide, the editorial staff for Evans Digital compiled



categories for the information; in addition to creating a structure for the archive's digital portal, these classifications have become part of the metadata that organizes the database. Here, a page from the *Bay Psalm Book*, for example, would be marked with not only its page number but also the fact that it belongs with the other pages of *The Whole Book of Psalmes*, a Psalter printed in 1640 by Matthew and Stephen Day of Cambridge, Massachusetts, and electronically filed under cross-references like "Music in Churches" and "Psalmody." (Since every title in *Evans Digital Edition* is indexed to the AAS's catalog, they are all also keyed to subject headings.)

At this point, too, other staff members review pages, comparing them to the microfiche versions and manually adjusting the image for maximum readability. Staffers may reject pages marred by cutoffs, bad skews, or light or broken text. In each case, a rejection means staff will pull the microfiche again and try to improve the scan.



Fig. 10. An example of the citation page for Franklin's "The Way to Wealth." Patrons can also opt to see individual pages or look through the table of contents.

So far, each month, NewsBank's staff has captured about a terabyte—that's over one trillion bytes—of colonial thought—more information than can be stored on ten ordinary desktops. When the project is finished, seven drawers of fiche will need to be digitally stored on the equivalent of over three hundred desktops. But don't go to Chester hoping to see armies of computers assembled to deploy America's founding documents; NewsBank built from scratch several high-end computers for storing the ever-growing *Evans*, which is also backed up on at least \$16,000 worth of digital tape. While on the material level, the digitization of *Evans* means the shifting of information from one type of plastic to other types of plastic (and some metal), on the experiential level, the change effected is nothing short of enchanting.

The interface for *Evans Digital Edition*, developed with the assistance of a panel of librarians, looks like commercial Websites, with stylish lettering, a tasteful palette, and an uncluttered layout. Screens offer browsing and

searching capacities coupled handily together on one page, guided by a comfortably familiar row of browsing tabs. If an institution decides to invest in the expanded cataloging software developed by the AAS—offered, like access to Evans Digital, at different pricing levels to match the sizes and missions of various purchasers—any hit on an Evans reference in a library catalog would produce a direct link to the electronic image and text of the document. Once a patron locates an item, she can look at full citations, move through the book page-by-page, choose from two different formats for printing, use the “Table of Contents” feature to skip around or locate pages that matched her search criterion, and scale the images up to 300 percent or down to 25 percent in Alice-in-Wonderland fashion. Future improvements in searching and printing designed to assist teachers and researchers may also be integrated. In addition, NewsBank recently announced a text-creation partnership with the University of Michigan whereby the text of six thousand imprints, selected by the AAS, will be hand-keyed into another database (available separately, directly from the University of Michigan)—an initiative that will produce as close to 100 percent accurate searchable text as is humanly possible for some of the period’s most widely used or historically significant documents.



Fig. 11. The title page for “The Way to Wealth.” Users can scale this image (shown at 50%) down to 25% and up to 300%, as well as choose from two different formats (TIFF or PDF) for printing.

As Newsbank staff members readily admit, the construction of *Evans Digital Edition* means the market for Evans’s *Early American Imprints* on microfiche has been demolished. Already fifty-one institutions, large (Columbia, Ohio State, and the University of California system) and small (Williams, Hanover, and Calvin), are on board, and NewsBank has halted fiche duplication. Libraries that own *Evans Imprints* and that subsequently opt for *Evans Digital* must face a difficult decision: what to do with all these little plastic artifacts of twentieth-century Americana and the file cabinets that house them? Inevitably, some libraries will simply discard the fiche and print, as machines available to read them fall into disrepair and users voice their preference for the digital edition’s seductive searchability. Still, perhaps some librarians will

take a cue from NewsBank's Chester surroundings and let the past accumulate organically in all its forms—books, Microprint, microfiche, and terabytes—cheek by jowl, like so many churches in a small New England town.

### Further Reading:

Those interested in the history of microreproduction may consult Nicholson Baker's *Double Fold: Libraries and the Assault on Paper* (New York, 2001), but should keep in mind that Evans's *Imprints* and similar initiatives had a profoundly different set of objectives than did the projects Baker chronicles. In addition, the Association of Research Libraries has collected [many responses](#) to *Double Fold's* assertions. Although a few links are now broken or expired, this remains an excellent source for a broader picture of the debate surrounding microreproduction, digitization, and the multiple missions of libraries in the twenty-first century. Edward G. Holley wrote a truly fascinating biography of Charles Evans entitled *Charles Evans: American Bibliographer* (Champaign, Ill., 1963)—despite its age, it is still a terrific read and has a lot of interesting information about Evans's project and the creation of Readex's *Imprints, Series I*. Those interested in Albert Boni's early career should refer to Jay Satterfield, *The World's Best Books: Taste, Culture, and the Modern Library* (Amherst, 2002). The Whiting Library, in Chester, Vermont, maintains a Vermont Room, which is also an outstanding source for articles about the history of Chester and the Readex Microprint Corporation (many thanks to the librarians for their assistance). Readers interested in the history of the American Antiquarian Society and its librarians should click [here](#) or consult *Under Its Generous Dome: The Collections and Programs of the American Antiquarian Society*, 2nd edition (Worcester, 1992). Many thanks to NewsBank staff members Ken Dufort, Georgia Frederick, Kelly Lauren, Steve Osterland, Caroline Reyes, Stanley Shapiro, Debbie Swisher, Cindy Tufts, and Mike Walker for taking time out of their busy schedules to show me *Evans Digital Edition*; special thanks to Vicky Gardner, Korrie Heiden, and Jim Hornstra for making time on many occasions to answer my repeated queries. Thanks, too, to the librarians who helped give me a sense of how this product will be used, and to Crowley Micrographics for information about the Mekel scanner.

This article originally appeared in issue 3.3 (April, 2003).

---

Katherine Stebbins McCaffrey was born in Vermont, teaches at Boston University, and is at work on a history of spectacle use in America.