

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Documentary Editing: Journal of the Association  
for Documentary Editing (1979-2011)

Documentary Editing, Association for

---


1980

# Comprehensive Text Processing and the Papers of Henry Laurens

David R. Chesnutt

*University of South Carolina*

Follow this and additional works at: <http://digitalcommons.unl.edu/docedit>

 Part of the [Digital Humanities Commons](#), [Other Arts and Humanities Commons](#), [Reading and Language Commons](#), and the [Technical and Professional Writing Commons](#)

---

Chesnutt, David R., "Comprehensive Text Processing and the Papers of Henry Laurens" (1980). *Documentary Editing: Journal of the Association for Documentary Editing (1979-2011)*. Paper 126.

<http://digitalcommons.unl.edu/docedit/126>

This Article is brought to you for free and open access by the Documentary Editing, Association for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in *Documentary Editing: Journal of the Association for Documentary Editing (1979-2011)* by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# *Comprehensive Text Processing and the Papers of Henry Laurens*

DAVID R. CHESNUTT\*

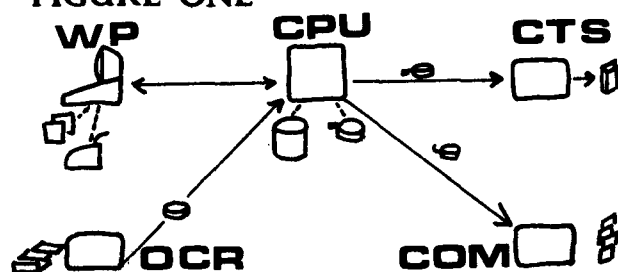
The idea of using the computer to perform routine editing procedures has attracted the attention of a number of scholars in the last few years. Among those who names immediately come to mind are David Trask and Miriam and Peter Shillingsburg. Trask revolutionized the editing of the ongoing series *Foreign Relations* when he introduced the use of microcomputers. Under Trask's leadership, the State Department historian's office now

supplies the Government Printing Office with machine-readable files instead of the traditional typescripts. This allows the type for the printed volumes to be composed without rekeying and saves time throughout the production cycle of each volume. The Shillingsburgs' Thackeray project at Mississippi State University uses a large, central computer and sophisticated computer programs to carry out many editorial tasks. And, like the *Foreign Relations* series, the Thackeray volumes will be typeset from machine-readable files rather than typescripts.

\*David R. Chesnutt is co-editor of the Laurens papers at the University of South Carolina. This paper was presented at the Association's 1979 meeting in Princeton, New Jersey.

The *Laurens Papers* will publish its last traditional, hot type volume about fifteen months from now. With the

FIGURE ONE



publication of volume 9 in February 1981, the project will have completed an almost comprehensive edition of Laurens's papers prior to the Revolution. Yet these nine volumes contain only one-third of the Laurens papers. Another eight thousand documents remain to be published. The initial publication plan developed in the 1960s called for a series of highly selected volumes to cover the Revolution and later periods. However, the importance of those materials led to a revised publication plan calling for six printed volumes and a yet to be determined number of microfiche. We plan to complete the publication of those eight thousand documents in about the same length of time it took to publish the first nine volumes — about ten years. What makes the new publication format and schedule feasible is the computer, or to be more accurate, the series of computers that we will harness for our needs.

Our basic objective is to go from an eighteenth-century holograph to a published document in a single typing. Reaching this goal requires considerable planning, because the new editing system affects every step of the editorial process. The most pronounced changes will be in the areas of 1) transcribing; 2) preparation of the editorial apparatus; and 3) production of the volumes and fiche. The savings in each of these areas are the savings that result from the elimination of retyping and proofreading. In terms of staff time, we estimate a savings of about ten years; in terms of money, we estimate a savings of about a million dollars.

#### *The Text Processing System*

The components of the Laurens text processing system include a word processor; an optical character reader; a large, central computer; a computer typesetter; and a computer-output microfiche unit. Three of these components will be off-site; that is, we will contract for the services of an optical character reader, a computer typesetter, and a microfiche unit. Although we will only work directly with the word processor and the central computer, we have had to consider very carefully how our

two units could be linked with the outside components (see figure 1).

Our system's most important component is the word processor, a microcomputer designed for manipulating textual materials. With so many word processors on the market, we had to spend several months evaluating various units. We finally chose a WANG Model 5-II. Our decision was based on ease of operation; compatibility with the university's central computer; and cost. We will use the WANG as a kind of super-typewriter to transcribe the last six thousand documents in the series.

The optical character reader (OCR) in our system is a SCANDATA 2250/1, a computerized character reader that can be programmed to read any typewritten material with an accuracy rate of between 95 and 99 percent. We have contracted to have about two thousand documents put on computer tape. These documents have already been transcribed but they have not been verified. Rather than retype the documents on the word processor, we chose to use the OCR as a means of saving time and money. To read the six thousand pages of these two thousand documents will cost about \$1200. Once SCANDATA produces a computer tape of the documents, the documents will go to the central computer and then to the word processor for verification and editing.

The central computer at the University of South Carolina is an Amdahl 470-V6. For all practical purposes the Amdahl is about the same as an IBM 370, which is one of the most common academic computers. We will use the Amdahl primarily as a means of storing our documents and our editorial apparatus. And we will use the Amdahl as a means of producing the computer tapes that will drive the other two components of the system — the computer typesetter and the computer microfiche unit.

Among the wide variety of computer typesetters now available, most can produce the Baskerville typeface used in our series. With more and more printers upgrading their equipment we are confident our university press will have a number of printers to choose from. This should enable us to keep our production costs down. More important, in supplying the printer with a computer tape, we will be able to eliminate the keying, or retyping, that now takes place when a printer is furnished with a typescript. This, in turn, will reduce substantially the amount of proofreading required in the production phase.

The microfiche unit is the final component of our text processing system. Again, the actual production of the fiche supplements will be handled by an outside contractor; this will assure compliance with archival and industry standards for the fiche supplements. As with the printed volumes, our "manuscript" will again be a computer tape. I should note that the use of computer tape for both printed volumes and fiche will enable us to move a document easily as the final process of selection is carried out.

### *How the System Works*

With this general overview of the hardware in our text processing system, we can discuss the processing of an actual document. Our example is a document from the file of items scheduled to be converted by the optical character reader, documents from the period December 1774 to August 1778 in which Laurens was heavily involved in the politics of the Revolution. When SCANDATA returns the 9-track computer tape to us, the tape will be read into the university's central computer and stored in the computer's mass storage system. The documents will then be ready for editing. Because the documents have already been transcribed, our first step will be to verify those documents in the 1774-1778 file. Here we focus our attention on the 1774-1775 file.

To verify the documents of the 1774-1775 file, we must first move a copy of that file from the central computer to the word processor. This is accomplished by a command given to the central computer from the terminal of the word processor. Note that the central computer retains the original file. This provides us with a security copy. If we accidentally erase the file on the word processor or if the word processor is struck by lightning, we are still in business. (I should also note that the OCR documents are backed up with a security copy in the form of the original computer tape; we always maintain at least three copies of every computer file we have.) But let me return to the problem at hand. To store the 1774-1775 file of documents requires about fifteen of the word processor's floppy disks. (The 1774-1775 file contains 443 documents totalling about twelve hundred pages; each diskette holds about eighty pages.) We begin the verification process by calling up the first document in the first diskette: a twelve-page letter written to John Laurens after Henry's return to South Carolina from England.

Once we have the document on call, we look at the first page of the transcription on our 24-line television-like screen. One person reads from the photocopy of the letter; another person follows the copy on the screen. Corrections in capitalization and punctuation, deletions, and insertions are made as necessary on the screen. The process is repeated page-by-page until the letter has been completely verified. If the verification team is interrupted, they simply mark their place, save the changes they have already made, and refile the document on the word processor diskette. The diskette's copy is updated with the changes made that day. If the proofreaders are uncertain about a particular reading, they can temporarily mark the passage so that other members of the staff can be consulted. Once all of the transcription questions have been resolved, the security copy of the letter in the central computer storage is updated with the corrected file.

Pre-selection is the second step in handling Henry's letter to John. Will the letter be included in the printed

volume, or will it be placed on fiche? If the letter is to be printed, annotation will be required. Our letter from Henry to John is one of those letters which is to be published in volume 10. A paper copy of the letter is made on the word processor's printer and turned over to an editor for annotation decisions.

Once the annotation decisions are made, the research and drafting of notes takes place in a very traditional fashion . . . almost. One of the problems in this letter is a reference to Samuel Brailsford's son whose first name is not given. The researcher wonders if the Brailsford boy is mentioned in a later letter. Perhaps his name is given there. Instead of reading through the letters, the researcher uses the search function on the word processor, keying the search on the word "Brailsford" and whatever variant spellings are likely to be used. As it happens, young Brailsford is mentioned in several later letters, but his first name is not given. Nevertheless, the researcher has a much clearer idea of how young Brailsford fits into the story. (As new letters are added to the annotation stack, it becomes difficult to remember whether or not we have already assigned annotation on a particular person. Here again, the search function of the word processor can be used to resolve our question quickly.) For the most part, however, annotation is less affected than other editorial tasks by the text processing system until it is time to write the finished notes. Once the draft notes have been completed, they are entered on the word processor. The draft is printed out and reviewed and the notes are modified accordingly. The footnotes do not have to be retyped, merely edited on the screen of the word processor. (In many cases in the past, we have gone through two or three typings before we got the notes in final form.) With our annotation on the letter from Henry to John now in final form, we are ready to move on to other tasks.

After we have edited the documents in the 1774-1775 file, and perhaps those of the 1776 file, we will reach the point of finally deciding what is to go in volume 10. Our letter from Henry to John will probably survive the final cut, but others may not. We will undoubtedly change our minds about certain letters that no longer have the importance they seemed to have at the outset. To remove a document from the file-to-be-published and put it in the file-to-be-published-in-fiche will be a simple matter of moving it from one master file in the central computer to another master file. After the print and fiche files have been determined, we can use our word processor to generate automatically a table of contents. Other parts of our editorial apparatus — like the volume introduction, the list of abbreviations, the principal dates of HL's life, appendices, and the index — will be added to the master print file. Upon completion of the master print file, we will move that file to computer tape. The computer tape, in turn, will be given to a printing contractor for typesetting. (To be continued)