MONASTIC SCRIBES
AND MICROCOMPUTERS:
BRIDGING THE GAP
BETWEEN MEDIEVAL SCIENTISTS
AND MODERN SCHOLARS*

BY NAN HAHN

Dr. Hahn is Assistant Research Professor in the History Department at Rutgers

THERE has long been a basic need in the area of medieval studies for solid, scholarly, critical editions of many unpublished works that were written, copied and read during the Middle Ages. The importance of these materials in the understanding of medieval thought and activities hardly needs elaboration. Medieval scholars themselves, such as Albertus Magnus, Roger Bacon and Robert Grosseteste, recognized the importance of preserving their works and those of their predecessors and frequently exchanged copies of earlier and current works through an extensive network of communication by travel, correspondence, and widespread circulation of their writings. The editions of letters and treatises of the medieval scholars are as basic to our understanding of their period as proper editions of Aristotle were to their understanding of the Greek world. (Scholars such as Thomas Aquinas went to great lengths to get reliable editions of Aristotle.) Yet the preparation of such editions today is a long, tedious and frequently expensive task. The scholar must first locate the often scarce catalogues of various European manuscript collections, comb these for citations of the desired work, attempt to verify their accuracy, and finally either order the treatise on microfilm or go to the location and use the original document there. Considerable time, travel and expense must be invested in the bibliographical search alone—even before one sees the source document. Furthermore, the references from catalogues are often incorrect, and researchers frequently find that they have invested heavily in a "dead end" search.

These problems make it evident that a systematic and readily available compendium of information on manuscripts would greatly facilitate progress in medieval studies. In one area, that of scientific manuscripts written in Latin, such a goal is being realized. The Benjamin Data Bank of Medieval Scientific Manuscripts in Latin, a part of Rutgers University, was planned in 1974 and has been rapidly developing since 1978.[1] Its purpose is to record data regarding the contents of manuscripts, using information taken from the manuscripts or microfilms of them, from catalogues and other printed sources, and most importantly from collections of manuscript descriptions offered to the Data Bank by scholars in the field. Large and small collections, describing more than 25,000 manuscripts, have been received from scholars in recent years, and more are expected.

The result will be a retrieval service for scholars through which the project will provide bibliographic citations of all recorded copies of a given source document, with information on the copies location and availability, and references in the literature on the subject. The citations will also include the names of scholars currently working in the specific area and the locations of microfilm copies that may be borrowed. (Some of this information is available from professional newsletters and other publications, but most is gathered by the director from correspondence and personal contacts with other scholars at meetings, libraries and other traditional sites of historical information exchange.) Users will have at their disposal the information, accurate and current, that they need to begin basic research, as well and they will learn whether or not they will be duplicating work already in progress.

The initial body of materials being incorporated in the Data Bank is a collection of some 14,000 manuscript descriptions gathered by the late Francis S. Benjamin of Emory University in research spanning more than three decades. Nan L. Hahn, a former student of Benjamin, began working with him on this project in the last years of her graduate work and assumed responsibility for the collection and its development when Benjamin died in 1973. She soon enlisted another former Benjamin student, Wesley M. Stevens of the University of Winnipeg, as a collaborator and consultant. Stevens' specialization in early-medieval science

is an appropriate complement to Hahn's work in the High Middle Ages.

Originally it was planned to enter the data onto magnetic tape by keypunching cards and reading them onto tape for editing and storage, and much work was done by this method. Although at the outset of the project such a system appeared economically feasible, it became increasingly apparent that it would eventually become very cumbersome and expensive. By 1978, when the project acquired its present base at Rutgers, The State University of New Jersey, it had become clear that an extensive reassessment of methodology was needed in order to maximize efficiency, simplicity and ease of modification. Three key areas of need were identified: a reliable and economical means of data entry; a group of programs to manipulate the data for search and retrieval on the university computer once they are entered; and a set of encoding rules that would permit the entry of widely diverse forms of data into a system from which they could be retrieved by the search programs.

The first problem addressed, therefore, was the method of data entry. Technological advances in computer science had by 1978 made it feasible to eliminate the very expensive keypunching and verifying steps. It was determined that a small staff with a microcomputer could handle ninety per cent of the entry tasks in the project office under the immediate supervision of the director. The remainder could be done by using the editing systems on the university computer at a higher but feasible cost. The project director and several consultants investigated a number of types of equipment and decided that the most efficient kind of entry system would be two work stations, each with a dual-drive microcomputer and a cathode ray tube terminal, and a shared hard-copy printer-terminal.

The value, size, diversity and varied potential uses of the holdings, already expanded by other collections by 1978, demanded that software for the project be very carefully designed. It was felt that four major criteria were felt basic to the success of the Data Bank:

(1) Universality in the entry format, to allow for entry, preservation, and retrieval of data collections coming to us from a multitude of sources, each with its own style or styles of record-keeping.

(2) Access to the data from a number of levels and at multiple points within these levels, including access by logical selection within one level and among various levels.

(3) Expandability for continued acquisitions of new material.
(4) Flexibility in the system for ease in correcting errors, updating information and adding new data within each file as items become more precisely described.

The second area of need, that of search and retrieval programs, was met most fortunately. Professor John B. Smith of The Pennsylvania State University offered the use of his Bibliographic and Grouping System (BAG/2), a group of programs which could be adapted to the data readily and at minimal cost. BAG/2 was installed, and pilot runs were completed. It was made compatible with the university computer system and will fulfill all needs for search and retrieval very adequately in the coming years.

The third area of need was for encoding rules to organize the data into a format that could be read by computer and stored in machine readable form. We needed this to be in a form that would be acceptable by the microcomputer for entry and usable by BAG/2 for search and retrieval. This was by far the most tedious task; it required some fifteen months and numerous test runs and revisions. The result was two-fold: a twenty-six-page booklet of rules covering all forseeable situations, being in effect the staff's reference work for the more difficult problems of categorizing data and a training manual which covers some eighty per cent of the encoders' questions and contains sample entry sheets, a glossary, and other items of use to the beginning encoder. These materials, coupled with Professor Smith's BAG/2, form a solid base of methodology that can be shared with researchers doing similar work in related fields. Already we have shared the entire package with several projects, and we expect to share the system with other such projects in the near future.[2] We have recently combined these materials with a case history of the project to form one volume for other potential users.[3] The advantages provided by such sharing are described above in our initial criteria; in addition, the recipients have testified to the saving—

[2] Professor Richard D. Monaghan of the Université de Montréal is using the system to build the index to the *Codices Manuscripti Operum Thomae de Aquino* (Roma, 1967-), an adjunct to the Leonine Commission's publication of *Opera Omnia Thomae de Aquino* (Roma, 1883-). Professor Jeffrey Huntsman of the Medieval English Lexicography Project at the University of Indiana is seeking to make his project's system compatible with ours for the easy and efficient exchange of data.

[3] The updated, one-volume version is available from Hahn in an inexpensive, computer printout form for cost plus handling charges.

in salaries, time and frustration—of more than a year's work in each instance.

It is becoming increasingly apparent that the methodology can be shared by various projects in medieval studies, even when they are not clearly related; such sharing must be sought and utilized extensively if scholars in medieval studies are to reach the expanded horizons visible in the age of computer science. We are corresponding with scholars involved in other manuscript projects in the United States, West Germany and Belgium, with the hope that we can arrive at an agreement on a format for machine-readable manuscript descriptions; this will be necessary ultimately if information exchange is to be efficient. Individual scholars in both Europe and North America have expressed interest in adopting our system for their own, manually-recorded manuscript notes. This use of a common format seems especially wise for younger scholars who will doubtless find themselves able (and probably expected) to use computers for research and record-keeping during the bulk of their careers. If they begin their note-taking systematically in a machine-readable form, they should be able to transfer their early work into computer storage efficiently when the opportunity arises.

In the year following installation of the equipment in October 1979, the staff entered enough data to run several pilot tests verifying the reliability of the encoding rules and their compatibility with BAG/2. The entire system is now working well. Since September 1980 the staff has been entering the Benjamin papers according to the final version of the encoding rules. The director prepares the more difficult material for encoding, and the encoders (now two on part time schedules) enter it. They can maintain an entry rate of 2,500 lines (eighty characters per line) per week when all equipment is working satisfactorily, but that is not the usual situation. A halftime proofreader handles all proofing, and the editing and corrections require additional time from the encoders. At the rate that has been established, we expect to complete the Benjamin papers by late 1983. At that time we will pause and build the first portion of our standardized spelling lists, described below, and based on these lists, several subject indices.

In the following three years, 1983-1986, we plan to enter three other important collections already in hand and perhaps two others that have been promised. Other remaining tasks are described below.

It should be noted that we are entering the data from the manu-

scripts just as they appear, with hundreds of variant spellings preserved. These diverse spellings, one of our thorniest problems, must be preserved for the text to be authentic; but somehow the words must be standardized to be retrievable. This problem will be handled in two ways. The words will continue to be recorded exactly as they appear in manuscript, but in addition a list of standard spellings will be developed using the following guidelines. Both proper and common words of classical Latin or Greek origin will retain the classical Latin spellings as their standard form. The standard spelling adopted for non-classical words will be a common spelling found in the sources that reflects the presumed vernacular source of the words. In these matters the director acknowledges her debt to R. E. Latham, editor of the *Dictionary of Medieval Latin from British Sources*, whose guidelines for variant spellings she has adopted.[4] His system is being used for two very important reasons: First, in an area as controversial as medieval spelling this system appears sensible, scholarly and totally feasible; secondly, adoption of this system makes the Data Bank compatible with the *Dictionary of Medieval Latin from British Sources* and thus with the international medieval dictionary of which the British work is projected to become a part.

   After the data are entered, the computer will produce an alphabetized list of all key words in the titles, incipits, explicits and other parts of the text that have been recorded. From this list the project director will produce a list of standard spellings, based on the Latham guidelines; a computer program will cause the standard spelling of each word to be inserted in brackets next to the variant spelling. From similar key word lists the director will also produce subject lists, such as astronomy, medicine and surveying, as time permits. Such lists will be very useful to the scholar working on a treatise which appears with various incipits, each of which, however, retains one or two of the key words. They will also be useful in locating fragments of a treatise when the beginning portion is missing.

   Variant spellings in names of authors are being preserved as the data are entered, by inserting an accepted standard spelling in the author category of the description and retaining the alternate spelling in a separate category within the citation. Thus a search for an author by the

---

[4] *Fascicule I A-B* (London: Oxford University Press, 1975), pp. xii-xiii.

standard spelling will produce all citations in the fields, but each cita-
tion will also present its own peculiar spelling as a part of its records.[5]

Data are entered into a table-model microcomputer with a dual drive,
through a cathode ray tube terminal, which allows communication with
both the microcomputer and the university computer. The encoder
makes corrections on the terminal screen as she spots errors and then
feeds the data in 200-line files onto a diskette (floppy disk) in the
microcomputer. Each of these files is labelled and can be brought back
up on the screen for editing after proofreading. Periodically, the encoder
inserts a disk into the second drive, makes a copy of her work and
"locks" it into the second disk to prevent accidental loss of entries. A
hard copy of a group of files is produced on a printer-terminal for proof-
reading by the other encoder. After proofreading, the file is brought back
up on the terminal screen, and corrections are made; the corrected file
is then duplicated. The material is then transported via telephone to the
computer center where it is read onto magnetic tape for storage. At the
computer center, a second copy of the data is placed on a back-up tape
for security, and the first tape is stored until needed for search and
retrieval. When a retrieval is required for a user, it is brought up from
storage to disks (current, immediate-access storage units) for search by
BAG/2, which can be supervised by the project director, the encoders,
or designated personnel at the computer center. BAG/2, using the uni-
versity computer, produces a printout of the needed data, written in
readable English, in standard bibliographic form. Each session on the
computer is monitored for cost, and the charge appears on the printout.
(The Data Bank will add a small surcharge to this to cover updating
and maintenance plus mailing costs, but the total cost will be modest.)

The primary method of access by the user will be a request, usually
by mail. This request can be for an author, a title, an incipit of a work, or
for any other of the categories under which information is filed. One
can also specify a combination of criteria, as positives and/or negatives,
as in the following example:

---

[5] The choices of our standard spellings are being made by searches through standard
reference works such as Lynn Thorndike and Pearl Kibre, *A Catalogue of Incipits of
Medieval Scientific Writings in Latin* (Cambridge: The Mediaeval Academy of Amer-
ica, rev. ed., 1963), George Sarton, *Introduction to the History of Science* (Balti-
more: The Williams and Wilkins Company, 1927-1948), and such scholarly works as
Marshall Clagett, *Archimedes in the Middle Ages* (Madison: The University of
Wisconsin Press, and Philadelphia: The American Philosophical Society, 1964-).

List all manuscripts that contain the following specifics:

(1) Sacrobosco, *De Sphaera*,
(2) Messahala (any of his works),
(3) in fourteenth- and fifteenth-century manuscripts only,
    OR
(3) in all *but* fifteenth-century manuscripts,
(4) now housed in London.

Such options allow the user to eliminate all excess material and pay only for that which is useful. The user may also elect to receive either the listing of the entire table of contents of the relevant manuscripts or a shorter, and therefore cheaper, version of only the bibliographic information plus the single item containing the desired treatise.

As soon as the Benjamin papers are entered, the director will publish notices on use of the Data Bank in appropriate journals and newsletters. The notices will list the categories under which information is stored, explain the retrieval system with its multiple-option possibilities, and give directions for use of the service. Examples of searches and their costs to the user will be included. This information will be available in pamphlet form to inquirers.

Publication of the entire body of data is not planned. The purpose of the system is to make this vast amount of data available in small, manageable lots. A second purpose is to keep the information in the system current and make additions over the years. Both of these goals would be defeated by publication of the complete file.

The need for this international repository and retrieval system is clear. Cooperation and support, from scholars in medieval science at a score of institutions in several countries, have been demonstrated repeatedly by donations of time, reviews of proposals and pilot studies, and contributions of manuscript descriptions. The conceptualizing and development of software needs, hardware needs and methodology were cooperative ventures of scholars from several universities in the United States and Canada. Cooperation with other similar projects in Europe is underway, and the development of a broad-based, international consortium appears to be practical. The time seems to have come for the twentieth-century implementation of a practice broadly accepted among thirteenth-century intellectuals. Roger Bacon, Robert Grosseteste and

other medieval scientists, who themselves travelled widely and maintained an international network of scientific information, would clearly be pleased if they could witness the reestablishment, seven centuries later, of an efficient international communication system to preserve and promote their ideas and accomplishments.