

From the Information Soapbox: Information Handling Dialectically Considered

By ELIZABETH B. WOOD

Winterthur Museum

ONLY the most acute of "first"-watchers would have recognized the occasion at Atlanta as an event of historic significance. The last session of the last day of the 1966 annual meeting of the Society of American Archivists was a panel addressed to the topic "Application of Automation to the Control of Archives and Manuscripts." Under the chairmanship of Frank Evans of the National Archives, five panelists presented case histories of experience with machine-assisted indexing of archives, manuscript collections, or publications. At least one previous session had dealt generally with information retrieval;¹ but the Atlanta panel represented the first time a full session of the annual meeting reported accomplishments as distinguished from speculation with regard to automated information handling.

The particular significance of the session at Atlanta is that it was held and was well attended. Archivists² have been characteristically conservative with regard to the new technology of information handling. Automated information handling is a *fait accompli* for librarians and records managers in many areas of business, industry, and science; and it is increasingly used in government. The so-called "information explosion" has forced such organizations to turn to automation as the only means of dealing with a very real problem. Thus far, most archivists have been able to ignore the possibility of a problem. In scheduling a session on automation, archivists—as represented by the Society—have publicly recognized that automation may have some relevance for them.

As one of the panelists, I have been asked to contribute to this

Mrs. Wood, now privately employed, was formerly Librarian of the Joseph Downs Manuscript and Microfilm Collection of The Henry Francis du Pont Winterthur Museum, Winterthur, Del. Her paper, prepared especially for the *American Archivist*, was developed after her participation as a panelist in the Society session to which she refers.

¹ A workshop session at the annual meeting at Raleigh, N.C., on Oct. 2, 1963. Calvin Mooers of the Zator Corp. was chairman.

² In the interest of simplicity I shall use the term archivist throughout this paper to include manuscript curator, document librarian, and any other title used to designate one who has charge of the arrangement, description, and care of manuscripts and public records.

special issue of *American Archivist* devoted to automation and information retrieval techniques in our field. While I assume that my fellow panelists—Frank Burke, Rita Campbell, Sister Claudia, and Russell Smith—will deal more closely with their experiences with computers, my own work with the Termatrix, an optic-coincidence system of information retrieval, has been recounted elsewhere.³ Rather than repeat what is already available, I have chosen to deal in a philosophical mode with some questions concerning information handling that I feel ought to be considered by archivists.

I do not pretend that any of the thoughts that follow are original. In fact I have given special consideration to those areas where one or more of the other panelists mentioned experiences similar to my own or where my experiences have been dissimilar. On the other hand I take full responsibility for my statements. They represent my own views. I have not fully discussed any of the ideas with the other panelists or with any of the numerous information scientists.

I believe that both the technology and the science of information retrieval are still in that stage of development where many points of view may properly be entertained. Although information specialists and science librarians have carried the field to much higher levels of sophistication than I am capable of understanding, much less expounding, it would serve no purpose to attempt to put my ideas into the technical language that has evolved. In coming into the subject much later than professionals in other disciplines, archivists are necessarily confronted with a serious and difficult language barrier. Perhaps the greatest danger may be that archivists will too readily accept the concepts advanced by others without recognizing that adequate solutions to their own information problems may require the development of different concepts.

The following questions are raised and candidly discussed with the desire of evoking vociferous rebuttal at future convocations of archivists. Well considered debate, bolstered with experience, can contribute much toward developing information handling techniques suitable for archives and manuscript collections.

³ "A New Method of Indexing Manuscripts," in *American Archivist*, 25:331-340 (July 1962); also available in Xerox from U.S. Department of Commerce, Office of Technical Services Sales Office, Washington, D.C., 20230, order no. PB 164 357, \$1.10. See also my *Report on Project History Retrieval; Tests and Demonstrations of an Optic-Coincidence System of Information Retrieval for Historical Materials (Drexel Library School Series, no. 14, 1966, 122 p.)*, available from Drexel Bookstore, Drexel Institute, Philadelphia, or from Winterthur Bookstore, Winterthur, Del., \$3 plus 25¢ postage. Termatrix is a product of Jonker Business Machines, Inc.

Are we discussing the same subject under a multiplicity of terms—information handling, information retrieval, automation, computer programing, mechanization, etc., etc.?

Yes, Virginia, there is a Santa Claus. Write him a letter addressed to Uncle Sam and ask for a good five-cent glossary. It would appear that we are all discussing various aspects of the same subject. *Information handling* would seem to be the most general term. *Information science* is logically the most desirable term to cover the entire theory and practice of information handling, but in reality the term has been too much restricted to computer-oriented theory.⁴ *Information retrieval* is more often used to include noncomputer specialists, but when expanded to *information storage and retrieval* the term frequently refers to microfilm systems. Information handling systems may be manual, machine-assisted, or automated.

Merriam-Webster in the 1963 *Collegiate Dictionary* defines automation as "3: automatically controlled operation of an apparatus, process or system of mechanical or electronic devices that take the place of human organs of observation, effort, and decision." I am never certain just how the term is meant with regard to information handling. Of course, Webster does not always have the last word, but if the phrase "take the place of" is to be interpreted literally, it seems unlikely that automation will ever be popular with archivists. Furthermore, those who have had experience will admit that it is not really true. For the most part, I think that the term *automation* applies to the use of machines, particularly electronic ones, to reduce *repetitive human activities* of "observation, effort, and decision."

Hereafter I shall mostly confine myself to the term information handling, which I prefer as the most inclusive term. Under it I would classify, as applicable to it, all the terms: information science, information storage and retrieval, information systems, data processing, document storage and retrieval, automation, automated indexing, automated abstracting, computers, and all other devices and methods to facilitate the handling of information. Each of these terms has its own specific and exclusive meanings; each may properly be used in its own place. I simply wish to give my discussion the broadest possible interpretation.

⁴ Isaac D. Welt, "Editorial: Information Science—Science Information," in *American Documentation*, 15:249 (Oct. 1964).

How will archival practice be affected by improved information handling technology?

Asked to justify his profession, the archivist will always maintain that man needs access to the records of the past as an aid to making decisions for the present and the future. This statement is as acceptable a justification for improved information handling techniques as it has been for any other archival procedures. Ultimately, through logical extension of this axiom, the recent technological developments in information handling will necessitate a complete reevaluation and readjustment of many of our present archival practices.

Archivists working with current records know that the computer is almost as great a threat to future historical research as the telephone. On the one hand computers produce tremendous masses of repetitive print-out (much of it on paper of dubious permanency). The physical mass, irregular size, and poor quality of the paper product present critical problems in storage and preservation. Furthermore, very little of the print-out is comprehensible except through use of the machine and the program that produced it; and eventually most of these machines and programs will be discarded as obsolete.

Archivists must begin to play a more active role as guardians of historical information. They will be forced to evaluate and to discard most of the masses of machine print-out; but they must also learn to understand the machines in order to know how to demand permanent records of value to the future. Heretofore archivists have gratefully received records as handed to them. I believe that eventually they must be empowered to determine the form of permanent records. This will be from considerations of history, not of economy.

Archival procedures having to do with all our currently inactive records will change more slowly because the problems seem less pressing, but they will change. More than 25 years passed from the time of the development of a workable typewriter until it was generally accepted as a tool for libraries; but today few librarians would produce a catalog in manuscript. Once archivists understand the advantages of machine processing, they will revise and abandon more tedious practices.

What factors should be considered in planning an information handling system?

A point that was emphasized and reemphasized by the panelists

THE AMERICAN ARCHIVIST

at Atlanta was the importance of presystem analysis and planning. The success of a system will depend upon how completely and how truthfully two questions are answered: (1) *What is the staff really doing?* and (2) *What should the staff be doing, given the means?*

The answers to these questions will be determined by:

1. *The size and nature of the collections.* If a collection is very large, very active, and poorly organized, the staff will be spending much time looking for and bringing out material. Logically the staff should be organizing and describing the collections in order to cut down on wasted time and effort. The mass of a collection does not, in itself, determine the difficulty or ease with which order is imposed. This accepted archival principle is equally essential to effective machine-assisted indexing on any level. Various levels of access, such as item access, record group access, and inventory access were discussed by the other panelists in Atlanta. Regardless of the system utilized, we must not lose sight of the fact that organization is the key to the effectiveness of any information handling system—manual or automated.

2. *The funds available.* Archivists need not be told how important funds are to effective production. Funds determine the size and quality of accommodations, the size and quality of staff, the availability of suitable equipment, and the quantity and quality of creative output. They will be the deciding factor in the selection of information handling systems.

3. *The users.* Archivists usually feel that they know the users of their material rather well, and they usually do. They are, however, inclined to evaluate users by the questions they ask rather than by the questions they would ask if they could expect to receive answers. We all ridicule the seemingly silly topics selected by Ph. D. candidates. We do not realize that many and more significant topics are abandoned because the information retrieval in archives is so slow that such topics require more time and funds than the doctoral candidate or even the "mature" scholar can afford.

4. *The interpretation placed upon the duties and obligations of the organization to its public.* This is always the most important policy-determining factor in any reevaluation or reorganization.

In the planning stages, information handling should not be confused with information retrieval. An effective information handling system will take all four of the stated factors into account from the planning stage onward. It will reconsider them through each step of internal procedure: from the time a collection passes in through the door until it has been reshelfed after use; from the time the archivist writes a begging letter in quest of a desirable collection through the closing hour when the janitor sweeps the reading room, shuts off the lights, and locks the door.

What are the qualifications of an acceptable information retrieval system?

At the outset of *Project History Retrieval* I outlined six desirable attributes of an effective information retrieval system: (1) controllable uniformity of cataloging procedure, (2) completeness of retrieval, (3) speed and simplicity of cataloging procedure, (4) speed and dependability of retrieval, (5) compactness of records and cataloging equipment, and (6) reasonable cost of equipping, installing, and operating the system.

To these attributes may be added some further qualifications. Dr. Campbell asserted, and other panelists agreed, that a machine retrieval system is a point of departure rather than a faster or more economical way of doing what an archives is already doing. It probably will not require fewer people or cost less. With the same staff and for approximately the same cost, one hopes that a machine retrieval system will do more, do it faster, and do it better than a conventional procedure.

Two negative qualities are to be avoided.

A system should not impose an arrangement on a collection. Ideally it will be flexible enough so that, when new technological developments come, the whole body of coded information can be transferred into a new system automatically without moving collections or disrupting operations.

A retrieval system should not be machine bound. That is, if the equipment selected will not accommodate the concepts to be processed at the outset, it will never be satisfactory regardless of how cheaply or quickly it does what it was designed to do. One must, however, be cautious in evaluating equipment. Neither speed nor complexity in a piece of equipment is a reliable index to its usefulness.

How costly are information retrieval systems? Who will bear the cost?

The cost of any information handling system can be broken down into cost for staff, equipment, supplies, and accommodations. In *Project History Retrieval* I tried to demonstrate that the most primitive procedure is not necessarily the least expensive. Staff is the most costly part of any operation and is a cost that will continue to increase as archival agencies continue to strive for the salary standards of other professions. Any procedure that wastes staff time is costly.

Russell Smith has voiced the opinion that a collection should be

THE AMERICAN ARCHIVIST

large and heavily used to justify the cost of the more sophisticated computer systems. This backs up my own assertion that archives should consider the less expensive, less highly automated systems as a first step toward better information handling.

The cost of an information handling system is something that must be measured in terms of every problem. The eventual cost of a procedure, good or poor, is obviously borne by the repository. Archivists, however, might well follow the lead of business organizations in expecting the system to pay for itself, either in reduced operational costs or in increased service. They might also borrow some practices from science librarians, who have pointed the way toward sharing part of the cost burden through centralized information centers and through transferring some of the cost to the user.

Archives and manuscript repositories have always been accepted as service organizations. Although they have never charged users for services connected with reference, they have charged for photocopying services. Now, when the more worthwhile projects of scholars are being increasingly supported by foundation grants, it seems probable that many of these grants would cover at-cost reference fees if archives were to offer increased information-searching services with reliable results at reasonable costs. Isn't it possible that by doing more "spoon-feeding" of researchers, archives might spoon more of the gravy into their own platters? The money is available; someone is getting it; why not archives?

How important is time in evaluating information handling procedures?

Information science literature is full of references to "real time" and "real cost." Without pretending to be conversant with this literature, I can say that one need not have very much experience with retrieval systems to realize that there is more to the value of time than is obvious at first consideration. I have spoken of wasted staff time because I believe that most archivists would be appalled to realize how little they get for their salary dollar, through no fault but poor organization. Although wasted time is important to archival budgets, speed is not necessarily so essential as some equipment manufacturers would have us believe. Historians do not work under the same time pressures as scientists; therefore archives can select systems that may work more slowly but more effectively for archival purposes.

In another sense, speed may not mean time saved. For example, many operations that use computers set up a search schedule. It is

not unusual to schedule a certain type of search once a week. When the search is finally made it may require only minutes. A manual search on a Termatrix system can usually be made immediately. While the total searching time may require several minutes for each relevant item, adding up to as much as hours for heavy documentation, still the researcher can have his answers within the day. I make this point not to detract from the value of computers but to emphasize the importance of deciding what it is that is really wanted from a system. For purely reference work the Termatrix may be preferable to any computer made, but it does not manipulate data or prepare print-out. If either of these operations is desirable the Termatrix is not suitable.

What are the problems that arise in setting up machine-assisted information systems?

Probably very few problems will come up during conversion to a machine system that were not previously present (though possibly unnoticed). The most serious problem to be expected is that once a project is launched everyone concerned will have ideas. Eventually the project will outgrow itself; the planners will try to do too much at once.

The value of a pilot project cannot be too often or too vehemently recommended. Select a reasonable, typical part of the collections; if possible, have it processed without commitments to the purchase of equipment. Allow plenty of time for this pilot project and be prepared to start over several times. The staff will necessarily have a learning period. If sufficient time is allowed for the pilot project, learning errors will fall within the project time and will be much less costly than if the same errors occurred in the final product.

To what degree will terminology be a difficulty? Is there a need for a uniform terminology among archives?

The more complex a system grows, the more important control of terminology becomes. In the sciences, information specialists have spent considerable effort in producing thesauri. It has been suggested that this is a need for archives as well. While I do not doubt that we will begin to borrow lists of terms from one another, I doubt whether it will ever be possible or desirable to establish a uniform thesaurus for archives. The needs, use, collections, and funds of archives vary too much for that. Each equipment manufacturer has recommendations to offer on how to handle terminology. Planners will be well advised to devote considerable at-

tention to understanding all the implications of these recommendations.

We need, however, to come to some agreement on the information-handling terms we shall use. If archivists can evolve a basic set of information control terms, researchers will be able to go from one archive to another, quickly learning to use systems that will vary widely in equipment and output but will operate on common basic principles.

For example, I have been using the term "categories of information." Robert H. Bahmer suggested a relevant term—"common access points." At Atlanta I learned that my colleagues who are devising systems for collection control have used other terms to mean the same thing, but *we are all collecting essentially the same pieces of information.*

I have described common access points that I thought would be useful to all users of historic materials. They are: type of document, date or time period, place, subject, and proper names. Dr. Campbell quite rightly pointed out the need to distinguish between subject and substantive content. She also dealt with the difference between common descriptors and precise descriptors. Allowing for these distinctions and various common collection control terms that we are using or shall use, it seems probable that the list of common access points, categories of information, or whatever we wish to call them will never exceed 10 or 15.

Working individually as we are, we are also using different terms to mean the same basic concepts in the system instruction area. For the old library term "cross-reference" we have variously used such instructions as "see also," "post on," "add on," "tracer," etc. There may be subtle differences among these terms, but it would be well to work out common meanings for them. Similarly, we all have occasion to use the instructions "and," "and/or," and "and/not." The Termatrix system is simple enough to require only infrequent use of the terms, but they would have occasional value. Before adopting them I wish to know whether someone else is using other terms that might be preferable. I have been using, for instance, the term "negative evidence." This applies to those instances where a false-coincidence of terms reveals evidence to indicate why positive evidence has not turned up. Have others experienced this phenomenon, and what do they call it?

What byproducts may be expected from machine-assisted information handling?

The first obvious result of machine-assisted processing is that it

forces us to be more systematic. To those who complain about the boredom of being systematic, we who have had experience can counter that it isn't boring at all. Through the very process of accumulating information about our collections we learn unexpected things about them, and so do our users. This serves as the starting point for new ideas. Through being systematic we are also enabled to achieve hitherto impossible continuity. Though good archivists will always be at a premium, it will no longer be so tragic when one passes on, because much more of his effort will be recorded in usable form. This continuity can also extend to interarchival cooperative efforts such as union indexes or projects where several archives would carry out individual parts and share them with one another. The final result of this continuity will be the better, more comprehensive histories we need to have. If our initial justification for archives is valid, sounder decisions will be the result of better histories.

How will the dignity of the individual archivist be affected by automation?

Fear of the machine is not confined to laborers; it is felt by educated men as well. The archivists who have advocated machine processing have been surprised by the reactions (or lack of reactions) of some of their colleagues. Those who have actually used the machines can only wonder at this distrust of the machine.

No machine has ever replaced, and it seems unlikely that any machine ever will replace, that unique phenomenon that is man. Though more is known about its nature than ever before, the unfathomable mystery of life and intellect is as deep a mystery as ever. All the machines that man has devised are merely extensions built onto the miraculous person of life-containing man. Man has given himself wings, more hands, steadier hands, a longer and more complex memory, a longer and healthier life. He can through the use of intellect extend his intellect, but he cannot create intellect any more than he can create life.

None of the great ages of technological advancement have added to or taken from the dignity of man. The bronze age did not, nor did the classical age, the industrial revolution, and the air age; the age of automated information handling will not either. A man who thinks great thoughts and acts great acts has always been a great man and always will be; fools will always do fools' work.