'COMPUTERS AND ARCHIVES – SOME RANDOM THOUGHTS'

by

P.D. Wilson, B.A., Dip. Lib.

During 1971, while a student at Canberra College of Advanced Education, I was exposed to a course of instruction which included both archival work and a basic course in computing. The inevitable mixture of ideas resulting from these courses has led me to give particular attention to the relationship between computer technology and archives. This article is, as the title states, no more than a random collection of thoughts, meant as a tentative basis for some additional work which I hope to do later. It is anything but a well-researched paper on a complex subject, but it is humbly offered as an Aunt Sally, should anyone wish to add any observations upon a relatively new aspect of archival work in Australia.

There are two important areas of interest in the subject – the first is the use to be made of computers in improving information retrieval in archival work; the second is the consideration of problems raised by the need to conserve, as archives, certain products of computing processes, in the same way in which other archival material is preserved. In other words, it is reasonable to ask 'What is automation doing *to* us and what can automation do *for* us?'.[1] I hope to deal with the information retrieval aspect by looking at developments overseas and then offering some suggestions on what we might consider doing here. The matter of archival storage of computer product does not seem yet to have attracted as much attention as one might expect in any part of the world, at least to my limited knowledge. I should be glad, however, to be proved wrong on this point.

Before considering retrieval aspects in detail, there are two main generalisations to be made. Computer hardware and software developments are in a constant state of flux and what is true today may not be true next Tuesday. As well, all the real development in computer-based information retrieval has been carried out on what I would call historical manuscript material and not on archival material defined by Jenkinsonian tenets. In general, I think it is fair to say that initial expectations, or perhaps misconceptions, which were engendered by overselling in the computer industry in the late 'fifties and through the 'sixties have not been realised as far as computing on a world-wide basis is concerned. This is not the fault of the machines themselves, or, indeed of the people who design and operate them. It is mainly a result of uninformed media-inspired generalisations about computers, leading the non-expert to believe that a computer, virtually on its own, can do almost anything we might want it to do. I think if we ask those of our librarian colleagues who have been closely associated with automation work about this point, they would readily agree that rosy

1.   Rhoads, James B. 'Programs of the National Archives'. Illinois libraries, v.52 (2), Feb 1970, p. 140.

expectation has given way to cold respect for the things computers can do extremely well, and a clear understanding of their inevitable limitations. Computers are truly marvellous manifestations of twentieth-century technology, but, in the main, they are still extraordinarily dumb clerical workers, by comparative human terms, who refuse to answer when spoken to and who need considerable babying by their keepers to get them to do their allotted tasks. On the other hand, their capability for storage of information is far in excess of the capacity and the accuracy of the human brain and it is this aspect which makes them extremely useful in the field of information retrieval. Finally on this point, it is only fair to the industry to remark upon the rapid developments taking place in hardware production, especially in the field of smaller, specialised units. Indeed, it may not be so very long before inexpensive tabletop units of considerable storage and computing power are available, replacing the monsters occupying whole floors of buildings, which we are accustomed to seeing in popular treatments of the subject. This miniaturizing, brought about by radical progress in component design, may have a bearing in the future upon the cost and convenience aspect of the use of computers in all phases of information retrieval. If hardware development has been rapid, software development, for highly specialised tasks such as archival and library work, has not. Again I would call as witnesses some of our librarian colleagues to testify that system analysis and design, program writing and implementation have been the main things which keep them awake in the small hours and not the ability of the hardware to carry out its job. It is necessary to point out that the hard work in computing is transferring your particular problem and its associated material into a form which a machine can readily digest and process. I feel that in saying that, my instructors in the Computing Studies School at CCAE will see that their efforts last year were not entirely wasted!

Turning now from the cold hard world of scientific reality to the hazardous and ill-defined one of archival definitions, I must labour the point I mentioned earlier about the application of computer techniques so far to manuscript collections rather that to archives. I write with some trepidation here, as I feel that a neat, easily defined, line of demarcation between manuscript material and archives is something of a Golden Fleece which we Jasons are not going to find easily. I hope I can get some agreement with the statement that most of us have no trouble deciding which is which in 90 to 95 per cent of cases — but in the remaining 5 per cent, say, it becomes anything but easy. I, for one, would be glad to hear from any Einstein amongst us who can provide a simple $e=mc^2$ for solution of this problem. Having said all that, I must return to the central point, that manuscript material is more manageable in extent and subject coverage, for information retrieval purposes, than is archival material, once again emphasizing that I am a devotee of Sir Hilary's pronouncements on this subject. Of course, the above remarks are made under the realisation that computers will never remove the human aspect in any kind of indexing effort, except possibly in some future system of designated word recognition which could replace human selection of indexing terms. As one American expert says, '... only after arrangement and description have been professionally tended to can the archivist hope

to begin to apply computer techniques to aid him in the field of control and reference'.[2] This applies equally to manuscript librarians as to archivists and will, I hope, finally scotch any Utopian ideas about the ability of present generation computers to do all our work for us whilst we lean back and indulge in light banter with the humbly grateful researchers who are waiting the few minutes it takes for the machine to cough out all the answers they need for this week's PhD thesis. So although the fact remains that all or most of the work done so far has been on manuscript collections, it remains to look at the obvious possibilities of the application of computer techniques to archival work. One problem associated with this aspect, and indeed with the production of any aid or guide, is the consideration of the needs and expectations of the researcher. I propose to be thoroughly cowardly and slide around this by saying that any large automated scheme must take this into consideration in its earliest planning. We should beware of producing a system which 'describes records in a structure which is probably totally irrelevant to the approach the researcher wants to take'.[3]

Looking overseas at developments in the manuscript and archival fields, which are usually lumped together under the latter term in the literature, one's attention is immediately drawn to the United States. This is not surprising, in that the US has a very large and sophisticated computer industry and in that our librarian/archivist colleagues there have never been backward in trying new techniques. Taken on a national basis, the most imposing program so far has been Spindex and Spindex II. I regret not being able to give very recent information on the project, but briefly it is a system designed by the staff of the National Archives to control and make available the material in the Presidential Libraries under that authority's control. It had its beginnings as early as 1958,[4] and has been supported by special financial grants. Its aim was to index down to folder level in the collections and then retrieve, from a data base consisting of the aggregate folder entries, individual references keyed by a fixed set of entry terms selected from the contents of the whole collections. One could imagine, for example, that a request to the Spindex II system to give a printout list of all references to Hoover, Edgar J. would produce quite an avalanche of research leads! The end product of the system is designed to give the user a reference consisting of Institution name, Collection name, Series name, Box indication and then Folder numeration.[5] Remembering that the Presidential Libraries are not all in Washington, but are dispersed as far away as Texas, the advantages of such a system for researchers can easily be seen. Although I have seen nothing in print to justify my saying so, I think that the National Archives has firm intentions to expand such a system as Spindex II into a much larger undertaking

2.   Burke, F.G. 'Spindex II: an aspect of archival information retrieval'. Records management journal, v.8 (2), Summer 1970, p.22.

3.   Burke, F.G. 'Computer techniques for the National Archives'. Computers and the humanities, v.4 (1), Sept 1969, p.13.

4.   Burke, F.G. 'Automation and historical research'. Libri, v.19 (2), 1969, p.82.

5.   Burke, F.G. 'Spindex II: an aspect of archival information retrieval', p.20.

to include the contents of Federal archives and possibly those of the States as well. There are some fleeting references to the possibility of setting up a national data base at a geographically central location such as St. Louis[6]. Presumably this data base would operate firstly on record group indexing, providing researchers with an initial coverage of both manuscript references under the Spindex II type approach, and also would provide references to relevant record groups in Federal and State archives. This latter description is mostly speculation and I can only recommend a close watch on the professional literature for further concrete details of such a scheme. Burke summed up what he considered to be the eventual aims of the project when he wrote:

> ... we are looking forward to an on-line, real-time computer system which will permit direct and explicit access to a body of information about the holdings of the major archival and manuscript institutions in the country.[7]

Across the border in Canada, the Public Archives of Canada has followed a similar line of progression from initial work on manuscript material to exploratory work on archival holdings.[8] Canadian aspirations are set out clearly in a printed report on the activities of the Public Archives covering the years 1959—1969:

> The use of electronic data processes (EDP) to produce indexes to the papers of Prime Ministers was studied in 1965 and a program was developed in co-operation with the Department of National Revenue's Data Center. In 1969 the first complete index was produced. A study is now being undertaken to examine the feasibility of the use of automation on a broader scale to improve information retrieval, particularly for public records.[9]

It is hard to believe that some high quality developmental work is not going on along similar lines in the Soviet Union. With the highly organized structure of central, regional and Republic archival institutions, there would seem to be ample opportunity for automated systems of retrieval. I have been unable to find any information produced in English on the subject and I would be most grateful for any references which may have come to notice of any reader. Finally in my brief survey of overseas developments, there is need to mention

6.   Burke, F.G. 'Computer techniques for the National Archives', p.17.

7.   Burke, F.G. 'Automation and historical research', p.86.

8.   See Atherton, J. 'The application of mechanization to manuscript catalogue production in the Public Archives of Canada'. Canadian Archivist, v.1 (10), 1966, pps.3—7.

9.   Public Archives of Canada 1959—1969. Dominion Printer, 1970, p.6.

the work done in the United Kingdom on automated indexing systems in Northern Ireland and Sussex. [10] As far as I can ascertain, the Public Record Office has not so far shown any great interest in the problem, but I hope that my ignorance of the subject is not the cause of unfounded criticism of that great institution.

This sketchy survey of overseas developments brings us to some theoretical consideration of what might happen here in Australia in the future. I think it is unreal at this stage, when so many of us have more pressing problems of control of material, shortage of staff and finance, and unsatisfactory accommodation for our archives, to worry about what automation we are going to implement tomorrow. However, there is nothing to be lost, and perhaps something to be gained, in a crystal-ball type examination of the future. None of us is in any doubt, I think, that the present standard of our aids and guides is not sufficient to give the best possible service to researchers. New South Wales and Tasmania have given us an exemplary lead, and I feel sure that the Commonwealth Archives Office will extend into the field as soon as possible. However, we all need to do a great deal more for the 'customer' if our services are to be placed at their true position in the whole field of information provision. In this connection, then, it would seem a good idea to give some joint consideration to a national data base scheme, incorporating the resources of the State historical libraries, the National Library of Australia, the Commonwealth Archives Office, the State Archives, the University Archives, the special archival collections and the tertiary libraries holding manuscript collections. The creation of such a data base, located probably in Canberra, would pose some thorny problems. However, we should be able to learn something from the experience of overseas institutions and such a wide-ranging program should not be impossible. If the large scheme were to be considered too expensive or premature, the Commonwealth and State Archives could still provide a useful service by the creation of a data base incorporating some kind of sub-series indexing, certainly not to document level, but possibly to file level. What is important is not so much the type, extent or aim of any system as the vital need for as much consultation and resultant compatibility as possible. Once systems are established in isolation, cost factors in compatibility become an inhibiting factor in any future combination and the only way to make sure things are going to 'fit' is to agree beforehand. It is quite possible that we shall not see such a data base established before the year 2020. I hope we can do better than that, but I do think there is some point in at least raising the issues now even if there is little liklihood of action, if only on the premise that the longer things are talked about, the less likely it is that there will be precipitate and ill-considered decisions taken adversely affecting such a projected national system.

10. See Darwin, Kenneth. 'The use of the computer in indexing records'. Journal of the Society of Archivists, v.4 (3), Apr 1971, pps.218–229, and 'Computers, modern records, and public relations'. Notes and news, Journal of the Society of Archivists, v.4 (3), Apr 1971, p.234–235.

I am not abandoning my beliefs when I say that there are a great many other things to be done in both the archival and manuscript fields in Australia. There would no doubt be some who would argue that printed guides such as the *Guide to collections of manuscripts* are quite sufficient. I can only counter argue that this level of coverage will not always be adequate and with the ever-increasing complexity of information which must be considered as legitimate raw material for a national historical data base, nothing but an automated information retrieval system will eventually be able to keep up with the researcher demand. I fully realise that a great deal of higher level work must be done in the nation's archives before we can readily think of indexing at the sub-series level, but I think it will have to be done eventually and I believe that we must be ready to use the technological advances available at the time we feel we are ready for such a step. As Archivist within the State Library of Queensland, I am only too painfully aware of the magnitude of the sorting and indexing tasks which awaits us and can only agree that '... the computer will not provide information that does not exist in some form of a finding aid.'[11]

This preoccupation with immediate problems should not prevent us from looking further to research on automation aspects which may make our jobs more complicated but which may make those of our professional successors much easier. We should not be frightened of automation, but we should see it as yet another aid in our basic task of preserving information and making it readily available.

Having dealt at some length with the retrieval problem, it now remains to say but a few words on the practical problems which are going to confront us as archivists in the preservation and storage of computer produced information. At present, there are two main types of computer product, magnetic tapes and print-out, which we need to be concerned about. In such a rapidly changing technology, this situation could change quickly, but these two will do for the current discussion. Taking tapes first, they are at present made of Mylar, which is a plastic material, with magnetic configurations imposed upon a coating of metallic substance on the active side of the tape. In other works, they closely resemble ordinary tape we have come to know through the increasing use of tape recorders in everyday life. What poses a problem for us is that eventually some tapes will need to be kept for archival purposes, providing as they do the cheapest method of conserving machine readable information. As the usefulness of the tape depends directly on the permanence of the magnetic configurations it carries, it is obvious that some care will have to be taken with its storage so that the tape retains its original form. Special non-magnetic conditions may have to be provided, in even temperature/humidity conditions (similar to those recommended for the storage of microfilm) for archival preservation. Most readers will be familiar with computer printout sheets. These come in endless concertina form, usually about 15 x 11 inches in size. The computer may be a dumb clerk, but it is also an extremely inefficient one where paper use is concerned. Any clerk

11.   Burke, F.G. 'Spindex II: an aspect of archival information retrieval', p.21.

as prodigal of paper as the average computer would be given the sack very early in his career. The main problem of computer printout for the archivist is sheer bulk. In bound form, some aggregations of printout make even some nineteenth century registers look like pocket books. Microfilming of printout for preservation is an accepted practice in the industry, but the problem does not end there as there are a number of unknowns in the archival preservation of microfilm as well. The problems do not excuse us from tackling them, however, and we must bear in mind Burke's reminder:

> The archivist is as much responsible for these computerized records as he is for the more traditional forms, but there are problems which have not yet been fully glimpsed and the solutions have not yet been forthcoming.[12]

Let me leave you with one sobering thought. Is it going to be necessary, at our present rate of technical progress, for an archival institution to preserve not only the tapes and microfilm product of a given computer installation, but also sufficient obsolete equipment so that the contents of the stored material may be retrieved? The 21st century archivist may need a B.Sc. or a B.E. rather than a B.A!! and his archives building may well resemble a technological museum!

At this stage, in summary, we can say that we Australian archivists have not yet taken much interest in computer techniques as they may apply to our holdings. Overseas experience, so far, has not shown any development beyond the level of identifying and locating collections, but there is no doubt that indexing to document level is theoretically possible using computer assistance. We have to face up to the fact that the products of computer processes in the field of public records will present us with considerable problems in storage and preservation. Taken together, these observations point to a need for us to be fully aware of the overseas developments in computer use in archives and to keep ourselves informed on what can be done with this tremendously useful piece of modern technology.

In addition, to my footnote references, the following bibliography may be of interest:—

'Automation in archives and manuscripts collections'. American Archivist, v.30 (2), Apr 1967 (Complete issue devoted to this subject).

Bell, Lionel. 'Controlled vocabulary subject indexing of archives'. Journal of the Society of Archivists, v.4 (4), Oct 1971, pps. 285–299.

Burke, F.G. 'Report on a survey of automation activities in archives and manuscript repositories in the United States and Canada'. American Archivist, v.31 (2), Apr 1968, pps. 208–10.

12.   Burke, F.G. 'Automation and historical research', p.89.

Campbell, Rita R. 'Machine retrieval in the Herbert Hoover Archives'. American Archivist, v.29 (2), Apr 1966, pps. 298–302.

Fishbein, Meyer H. 'Appraising information in machine language form'. American Archivist, v.35 (1), Jan 1972, pps. 35–44.

Fisher, Barbara. 'Archivist and the computer'. Pacific and northwest Library Association Quarterly, v.32, Jan 1968, pps. 8–12.

Fisher, Barbara. 'Byproducts of computer processing'. American Archivist, v.32 (3), Jul 1969, pps. 215–224.

Report. Discussion 'Automation and archives'. Journal of the Society of Archivists, v.3 (1), Apr 1965, p. 41.

Rhoads, James B. 'The historian and the new technology'. American Archivist, v.32 (3), Jul 1969, pps. 209–214.

Rieger, M. 'Archives and automation'. American Archivist, v.29 (1), Jan 1966, pps. 109–110.

Shelley, Fred. 'The Presidential Papers program in the Library of Congress'. American Archivist, v.25 (4), Oct 1962, pps. 429–432.