

Archives and the Computer

Michael Cook

University Archivist, University of Liverpool

Archivists have been investigating the archival application of the computer for nearly twenty years, but so far there have been few manuals which analyse these applications in a straightforward, relatively non-technical way. Mr Cook discusses the nature of automated systems, including records management systems, and describes a number of the important ones currently in use by archivists; he finally examines the problems and scope of machine-readable archives. Emphasis is placed on the general description and listing of archival materials by computer systems rather than the exploitation of archives in research.

This book is addressed both to the archivists with little experience of computers and to students of archives administration.

Contents

- Preface
- 1 Computer systems
- 2 Systems for records management
- 3 Systems for archival description, retrieval and management
- 4 Machine-readable archives
- Glossary of technical terms
- Appendixes
- Notes and references
- Abbreviations
- Select bibliography
- Index

ISBN 0 408 10734 0

Archives and the Computer

Michael Cook

Butterworths

Archives and the Computer

Michael Cook

Archivist, University of Liverpool

Butterworths

London Boston

Sydney Wellington Durban Toronto

United Kingdom Butterworth & Co (Publishers) Ltd
London 88 Kingsway WC2B 6AB

Australia Butterworths Pty Ltd
Sydney 586 Pacific Highway, Chatswood, NSW 2067
Also at Melbourne, Brisbane, Adelaide and Perth

Canada Butterworth & Co (Canada) Ltd
Toronto 2265 Midland Avenue, Scarborough, Ontario
M1P 4S1

New Zealand Butterworths of New Zealand Ltd
Wellington T & W Young Building, 77-85 Customhouse Quay, 1,
CPO Box 472

South Africa Butterworth & Co (South Africa) (Pty) Ltd
Durban 152-154 Gale Street

USA Butterworths (Publishers) Inc
Boston 10 Tower Office Park, Woburn, Mass. 01801

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, without the written permission of the copyright holder, application for which should be addressed to the Publishers. Such written permission must also be obtained before any part of this publication is stored in a retrieval system of any nature.

This book is sold subject to the Standard Conditions of Sale of Net Books and may not be re-sold in the UK below the net price given by the Publishers in their current price list.

First published 1980

© Michael Cook 1980

ISBN 0 408 10734 0

British Library Cataloguing in Publication Data
<p><i>Cook, Michael</i> b. 1931 <i>Archives and the computer.</i> 1. Archives – Data processing I. Title 025. 17'1 CD971 80-41286</p>
ISBN 0 408 10734 0

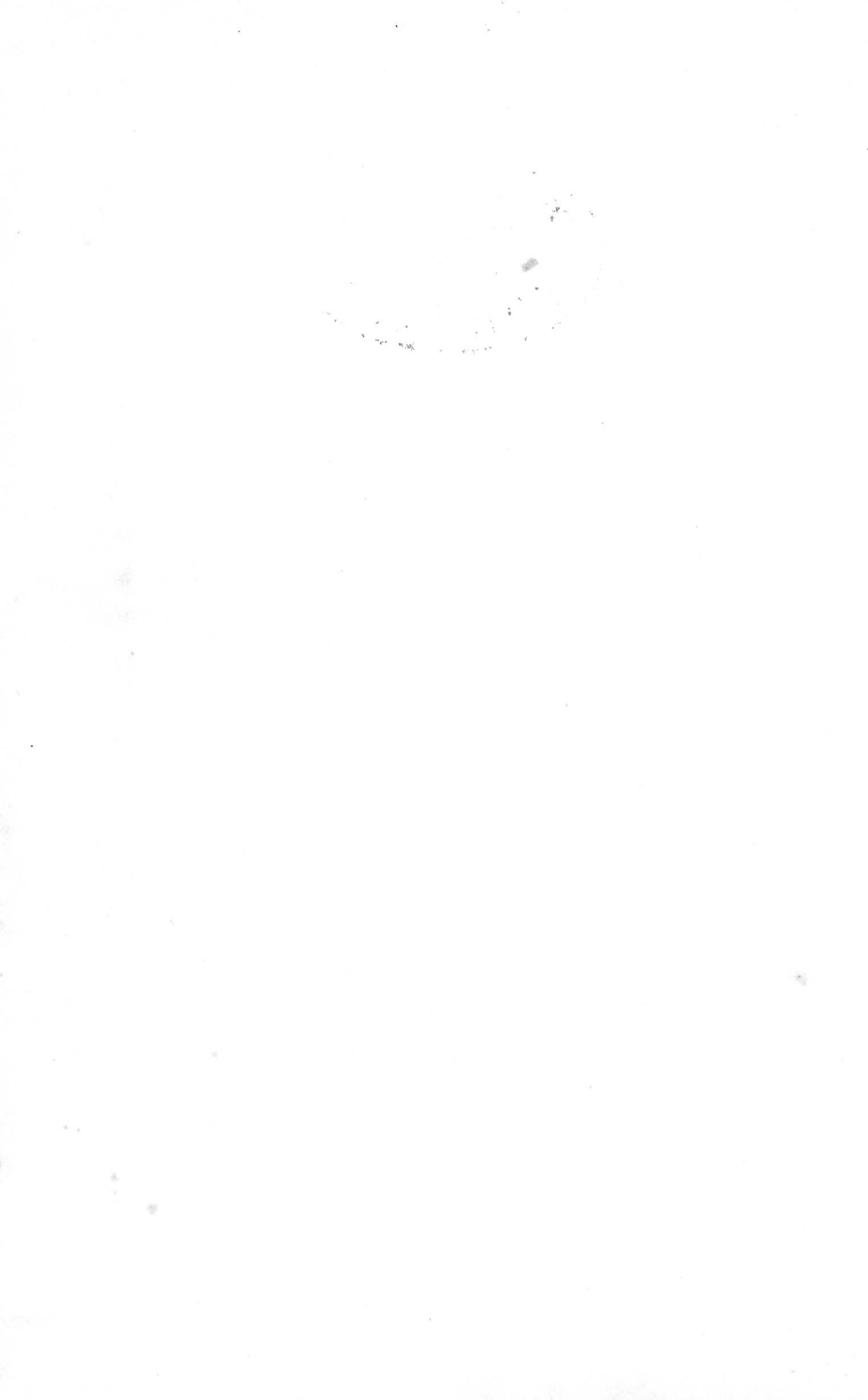
Printed and bound in Great Britain
by W & J Mackay Limited, Chatham

Contents

List of figures	7
Preface	9
1 Computer systems	13
2 Systems for records management	55
3 Systems for archival description, retrieval and management	76
4 Machine-readable archives	107
Glossary of technical terms	121
Appendixes	
A Select directory of archival systems	130
B Guidelines for selection of machine-readable and related records for permanent preservation	133
C Access to files in the machine-readable archives	134
Notes and references	135
Select bibliography	137
Index	146

Figures

1	Steps in the production of an archival inventory	page 18
2	An example of a proof report, showing data base arrangement during input	28
3	An example of output. The data base has been sorted and arranged on the page	28
4	kwic and kwoc machine-generated indexes of file titles	37
5	A punched card, showing the punch combinations for letters and numerals	42
6	Input data form (CAR/1)	59
7	Records management processes (ARMS)	62
8	Amendment to retention schedules input forms (ARMS)	63
9	Records retention schedules (ARMS RM02)	66
10	Monthly review schedule (ARMS RM07)	67
11	Disposal notification (ARMS RM08)	68
12	Monthly master list (ARMS RM09)	69
13	Yearly report on record use (ARMS RM12)	70
14	Simplified flow-chart: data capture processes in PROSPEC	79
15	NARS A-1 report	84
16	SAGE entry as arranged in the master list	92
17	Final version of a SAGE input, in printed form	92
18	Name/subject index printed out as master list (Smithsonian Institution Archives)	94
19	Inventory of machine-readable records (Public Record Office)	110
20	Instructions for completing inventory (Public Record Office)	111
21	The process of appraisal (NARS)	113
22	Checklist of requirements for storage of magnetic tape records (NARS)	115



Preface

It is time that there was a manual of archival applications for the computer. Archivists have been investigating the subject and seeking practical computer development since early in the 1960s. Dr Califano's paper to the International Congress on Archives at Brussels in 1964 marks the first co-ordinated effort by the profession to agree on the principles of automatic data processing as applied to archival materials. Since then the subject has been developed rather patchily by archivists. Beginning at about the same time, librarians and documentalists have made better progress, and, using the advantages of their greater numbers, their greater need for current awareness and real time working, and their generally better access to organizations which run large computer systems, have devised several practical and elegant systems for cataloguing, indexing, information retrieval and general management of their offices. None of these amounts at present to a widely recognized national or international network, and all display individual characteristics which mean that each system has specific advantages and drawbacks. (Perhaps all human systems will ultimately prove to have this disadvantage.) Nevertheless these systems are now available as models and packages, and archivists may hope to learn from the experience of their near colleagues.

This may suggest that there has been no advance in the development of purely archival automatic systems. This is not so: the leading national archives services of the developed world have established or promoted sensible practical systems which, in some cases, are already being offered as the basis of co-operative projects or as embryonic networks within their own countries. There have also been a few independent innovations as the result of the work of smaller archives services and one or two from the Third World.

It is well known that computers and everything associated with them are changing and developing so rapidly that writers cannot

keep up with them. There are also fashions in thinking about them. At present we seem to have moved from a period of excessive optimism to a period in which the limitations and drawbacks of automated systems are much to the front of users' minds; no doubt we should aim at a point somewhere between these extremes. In spite of the drawbacks and of the risk of systems failure, computers do offer the most important new development in archives work since the opening up of archive-based research in public education in the second half of the nineteenth century. This is a good time therefore to attempt an assessment of what has been done so far, and to offer a simply worded explanation to those who have hitherto not taken much notice of these developments. The time has now passed when the archives profession as a whole can ignore the subject.

The present study is addressed primarily to two groups of people. It is hoped that archivists who have had no experience with computers, and with little mathematical or scientific background in their education, will find it a useful guide to this rather complex new world. At the very least, it may help them to master its new vocabulary; and with the new vocabulary the new battery of concepts which has had such a vitalizing influence on information work generally. The second group is that of students who are about to enter upon careers in archives administration which will stretch well into the next century. The need to reach the minds and interests of students at the universities of both Liverpool and Ghana is the immediate inspiration of this book, and I would like to offer public thanks to all the students with whom I have discussed computer applications over the last five years or so. They have, as they may recognize, provided many of the ideas.

It is important to state clearly what is included in the present study and what is left out of it. It is clear that the theory and practice of archives administration as it exists nowadays must be the co-operative work of many people. No one individual or published work should try to cover too wide a spread of subjects. This book therefore concentrates upon giving a general introduction to the computer and its possibilities for operating as a tool in archival management. After a general discussion on the nature of automated systems, and their relation to manual ones, a select group of important systems is described in as much detail as may be needed for a reasonably secure understanding of their operation, and with enough information to guide further study. Records management systems are included, for this is an area in which there is good opening for versatile and practical systems, which if properly planned go far to solve problems never really tackled heretofore. Finally, a brief discussion of the problems involved in the treatment of machine-readable archives offers a

field of action which is open to many archivists working in large organizations, and highlights a problem which has good pioneering studies and a body of practice, but which is as yet little known to the profession at large.

What is not attempted here is any treatment of computer systems which aim to use the contents of archives. This book is about records management, archives administration, the conduct of archives and services, the building up of finding aids to archives (including indexes to these). It is not about the exploitation of archives in research; that interesting and worthwhile subject has a literature of its own, but still lacks a practical manual. This book also avoids any very detailed and thoughtful study of indexing and its associated world of thesauri, syntactical and semantic structures and contextual data retrieval; there is a very considerable literature already upon this, and it would indeed be desirable if more archivists interested themselves in the technicalities of indexing. But for the moment the possibilities inherent in systems for the general description and listing of materials is taken as the central point of interest of this book.

I would like to offer my most sincere thanks to many who have helped to produce this study. I have taken their advice, used their facilities or claimed their support. The book is much the better for it, but for its errors and omissions I am solely responsible. I would particularly like to thank the following: Lionel Bell at the British Library; S. C. Newton, at that time at the British Steel Corporation; Annette Kennett and her staff at Chester City Record Office; Guy Cangah and his staff at the Archives Nationales of the Côte d'Ivoire (with kind recollections of my visit in 1977); Maureen Patch and Glyn Evans at the Dyfed Record Office; Margaret Whittick and (at the relevant time) Ron Chesterman at the East Sussex Record Office; H. S. Cobb at the House of Lords Record Office; F. Burke, C. M. Dollar, and S. E. Hannestad at the National Archives and Records Service of the U.S.A.; M. E. Carroll, H. Naugler, and (at the time) Hugh A. Taylor at the Public Archives of Canada; Michael Roper at the Public Record Office; A. L. Bain at the Smithsonian Institution, Washington; David Butler, Bruce Jackson and Alan Seaman at the Tyne and Wear Archives Department; Sam Kotei at the University of Ghana; Maynard Brichford at the University of Illinois; and at the University of Liverpool a whole army of colleagues, including Adrian Allan, Peter Baker, Stephen Foga, Andrea Rudd, Zafar Siddiqui in the University Archives, and John Martin of the University Computer Laboratory. I am also grateful to acknowledge the kindness of the following institutions who have allowed publication of material from their own systems: Chester City Record Office, Dyfed Record Office, East Sussex Record Office, the House of Lords Record Office, the National Archives of the Côte

d'Ivoire, the National Archives and Records Service of the U.S.A., the Public Archives of Canada, the Public Record Office, the Smithsonian Institution, Tyne and Wear County Record Office, and the University of Liverpool Computer Laboratories.

The University of Liverpool

Michael Cook
October 1979

1 Computer Systems

A computer is 'a mechanical or electrical device for processing information'. This recent definition makes it quite clear that a computer is something relevant to the work of those who deal with information. It is not an instrument only, or primarily, of use to mathematicians or natural scientists, or to accountants, whose main concern is computing figures. The term 'computer' (which, we are told, arose accidentally) does not describe the most characteristic action of the computer in its outward relationships, though it may describe the internal methods which it uses. The French term *ordinateur* – implying a machine which arranges collections of disparate items into a significant order – is more suitable to the computer in its everyday use as a processor of information.

Archives offices, or archives units in libraries and museums, are, for the purposes of the present argument, examples of specialized data banks. On the analogy of a monetary bank, they are places where large amounts of a certain kind of information are stored in safe keeping and can be utilized, because it is felt that the information in them is of significance to society. Generally, the difficulty experienced by the archivist is that it is a slow and technically demanding job to extract usable information from the documents, or to construct finding aids that are complete enough to allow others to do so. Work on archives is usually funded from public money, or from money set aside for a cultural rather than a utilitarian purpose. This is rightly so, for, as the Grigg Report laid down in 1954, it is the duty of a civilized state to make provision for its archives.¹ But it means that there are normally not enough people or resources to do the job under ideal conditions, and that it is important to provide results within quite a short time. This

predicament is even more acute in countries of the Third World, where there are too few resources to devote to the problems of archives, but where the information contained in them is a particularly valuable part of the general stock of information, and is needed to support development work. It is also compounded by the great bulk and rapid accretion of most archive accumulations, which means that, if they are not controlled rapidly and effectively when received, the problem of doing so later will be insoluble.

Archivists therefore are under pressure to provide results rapidly, and not to get swamped by unlisted mountains of paper. The computer has provided an instrument which for the first time offers the possibility of improving radically both the rate of output of finding aids and the depth of access they give to the raw material. With their aid, the same type of manually constructed material already produced by archivists could be used to give a vastly more productive output. The all-important factor is the design of the system.

In this situation the computer does nothing to impair or replace the personal and professional skills of the archivist. It is there simply as a tool, rather as microform systems are. Like them, the computer is part of a complex system, in which a number of operations is combined. The archivist is needed to plan and define this system, and it will not work without him, or without the use of his skills and judgement.

The Scope of Automated Systems in Archives

A rapid survey of existing archival applications shows that there are several areas in which automated processes could be or have been introduced.

- 1 In records management, computer systems can be used to control the inflow of records into custody, the identification of documents required by users, the control of documents issued to users and their return (circulation control), and the identification of records due for disposal at a particular time.
- 2 In archival management, the control of archives passing through processes (sorting, fumigation, repair, description, storage, etc.); description and listing of archival material; production of systems of finding aids; retrieval of documents and/or information from the stock; and indexing.
- 3 Special archival projects, such as the construction of large-scale indexes, the control of microforms, publishing, or the exploitation of archival sources in research.
- 4 The management of archives offices and records centres. Applications here include compilation of user statistics, ordering-up procedures for moving documents to and from

users, controlling the allocation of storage space, and control of stocks of consumables. Probably this group of applications will be most apt for larger institutions.

- 5 A separate group of activities concerns the management of machine-readable archives and records. Computers are clearly necessary in order to make use of, or even to store on a long-term basis, valuable archives which have been generated by computer-based projects. This is a problem which is beginning to impinge upon the work of many archivists, but which is still regarded as a rather specialized area.

These are processes which have been developed or suggested as suitable for computerization. If an archives office is interested in doing this for any of them, it will have to carry out an enquiry into its own methods, objectives and needs, so that it can determine to what extent an automated system could be found to suit it. It will be necessary to carry out an analysis of its current systems. On the one hand this may clarify the logical steps in existing processes and the objectives they are aimed at; on the other hand it will take account of possibilities offered by a computerized version.

These possibilities are quite various. A choice may be made, for instance, between specialized piecemeal operations and general ones. In the former case, the computer carries out a particular series of tasks (such as indexing personal names) in a general context of manual systems. The latter situation attempts to automate whole processes and to combine a number of different operations. The more complex the computerized system, the more difficult it is to fit all the bits together, the greater the disruption of the established method, and the greater the possibility of failure. But also the possibility of radical improvement in access to archival information is the greater, and this must be a main aim in introducing computers. It is one which, if successful, justifies the cost and trouble involved.

Archivists are not accustomed to regard their services as similar in kind to those offered by, for example, manufacturing industries. Yet, from a systems point of view, an archives office has many of the same characteristics as the business control of such an industry. Consequently archives offices may hope to benefit from technological innovations which were conceived as aids to efficient business administration. The development of data-base management systems is a good example of this. Like any administration, an archives service works by carrying out a series of different operations upon a data base which holds all the information it needs. In a business administration, this data base includes lists of creditors and debtors, together with details of their financial operations, and stock lists; in a university administration, the data base consists basically of lists of students and of staff, with

relevant information. Once the necessary system has been installed, these data bases can be manipulated so that they will automatically provide the most useful pieces of data when and where they are needed. In the case of an archives service, the data base consists of lists of archival holdings, together with all the detailed information which goes with each item. The system is required to arrange these data in a variety of particular orders, and to produce sections at every necessary point. For example, the repository staff will need shelf or location lists; the conservation staff will need lists of items in need of repair or which should be inspected; the searchroom staff need to produce guide entries, inventories, handbooks and indexes; the publications section wishes to publish selected descriptive entries; and all sections, as well as the users, wish to have access to particular archival documents held in the store. It will be seen, then, that an archives office can organize itself around a central data base just as a business organization can. Computer control of that data base will involve much careful planning, but is inherently likely to provide a much more effective control than any form of manual operation. It is also likely, incidentally, to alter the working conditions of the archives office staff for the better. Many of the tasks which have to be done in manual systems of archival management are dull and repetitive. In a computerized system much of the repetitiveness is eliminated, and instead there are problems which are intellectually taxing and interesting. A well-designed computer system is a tonic to the staff.

If computerization of a general system is possible at all, it carries possibilities for the radical alteration of the product. So far, it has been assumed that computers are an alternative way of executing what has been effected in the past by manual methods: the establishment of administrative control over archival processes, and the production of traditionally defined finding aids like guide entries or inventories. But in fact the computer has the potential to produce quite new products. Out of a single data base (basic descriptive lists of archival holdings) it can, for example, produce an infinite number of specially selected finding aids in specific orders; this possibility has been referred to by K. Darwin.² Finding aids therefore could be tailored to any specific user's needs. Beyond this, any computerized data base can be used for two kinds of purpose. It can print out various descriptions in various orders; these descriptions may be sorted and selected by the system in order to achieve a particular objective and may include updated information. It can also (if the system is so designed) allow a searching procedure, in which the user interrogates the system until he identifies exactly those documents held within it which answer his enquiry. Both these possibilities carry important