

# The use of sampling techniques in the retention of records: A RAMP study with guidelines

PGI-81/WS/26

General Information Programme and UNISIST

United Nations Educational,  
Scientific and Cultural Organization

Paris, 1981

PGI-81/WS/26

Paris, 1981

THE USE OF SAMPLING TECHNIQUES  
IN THE RETENTION OF RECORDS:  
A RAMP STUDY WITH GUIDELINES

prepared by

Felix HULL

General Information Programme and UNISIST

United Nations Educational,  
Scientific and Cultural Organization

Recommended catalogue entry:

Hull, Felix

The Use of sampling techniques in the retention of records: a RAMP study with guidelines / prepared by Felix Hull [for the] General Information Programme and UNISIST. - Paris: Unesco, 1981. - iii, 106 p., 30 cm. - (PCI-81/IS/26)

- I - Title
- II - Unesco. General Information Programme and UNISIST
- III - RAMP (Records and Archives Management Programme)

CONTENTS

	Page
Foreword	i i
1. The Appraisal of Records: A Frame of Reference	1
2. Sampling: Some Theoretical Considerations	9
3. Sampling of Conventional (textual) Records: Practical Experiences	18
4. Sampling of Non-conventional Records: Cartographic, Audio-Visual and Machine-readable Records	31
5. Problems and Alternatives to Sampling	38
6. Conclusions and Guidelines	47
Bibliography	57
Appendix 1. Tabulation of Guidelines	60
Appendix 2. Sampling by Canadian Social Insurance Numbers	63
Appendix 3. Part of a table of Random Numbers	64

\* \* \* \* \*

## FOREWORD

Among the various techniques adopted for reducing the bulk of certain classes of records is 'sampling', a method, which can vary from a purely subjective choice of examples through a variety of procedures to an exact statistical process, providing an ideal objective answer for the student involved in the quantitative analysis of data. In so far as this is a special procedure appropriate only in special circumstances and requiring careful assessment of method to be employed, size of sample required and precise evaluation for the purposes of research, it has been considered desirable to treat it as the subject for a special study. This study, therefore, is directed to one particular aspect only of records disposal and to one technique of appraisal, which should only be applied when circumstances indicate that a particular need is present.

Nevertheless, there is a certain misunderstanding prevalent regarding sampling. Terminology has tended, in the past, to be less than precise and the whole question of the use of sampling has given rise to much uncertainty and some misgivings among archivists. Here is a process which by definition leads to the destruction of a high proportion of the total documentation involved: can we be sure that the right material is preserved? Is the proper statistical sample required for accurate analysis by computer necessarily what the archivist looks for as part of his records, or indeed, what the historian or sociologist really needs? To what extent, if selection is inevitable, must purely objective criteria dominate; do we require the ordinary or the extraordinary in our archival sample?

It is because of these many uncertainties; of questions of principle which disturb those concerned with sampling as it affects archives; and the desire for some assessment of procedures and for suggestions for suitable basic guidelines, that this study has been prepared for UNESCO in co-operation with the International Council on Archives. It approaches the subject in two ways: (a) by consideration of theoretical principles and methods of sampling; and (b) by an examination of the experience of a number of national and other repositories where sampling has been practised. On the basis of these assessments an attempt has been made to draw up some essential principles for the application of different methods, even if it has not proved wholly possible to state categorically what should or should not be done. In this study, too, it has not been overlooked that many repositories are increasingly concerned with non-conventional archives and that pictures (still and motion) sound archives and machine-readable records all play a part in the contemporary scene. Although still more difficult to assess within the terms of this study, these newer forms of records have not been ignored.

I am, naturally, indebted to a great many people in respect of the data which had to be collected for this study, and, in particular would wish to thank the Australian Archives, the Public Archives of Canada, the Archivni Sprava, Prague, Czechoslovakia, the Direction des Archives de France - Cité des Archives Contemporaines, the Social Insurance Institution Archive, Finland; the Bundesarchiv, Coblenz, German Federal Republic, the Public Record Office and the Scottish Record Office, the National Archives and Records Service, Washington D.C.; U.S.A.; the **Nordrhein-Westfälisches** Hauptstaatsarchiv, Dusseldorf and the Stadt Köln

iii.

Historisches Archiv, the British Architectural Library, the British Broadcasting Corporation, the Imperial War Museum, London, the Meteorological office, Bracknell, the National Film Archives, London, the National Maritime Museum, Greenwich, and the Victoria and Albert Museum, London; Dr. C.M. Dollar, Dr. Frank Evans, Dr. Eckhart G. Franz, Dr. W. Kennedy, Mr. Carl Newton, Miss A. Oakley, Dr. S.C. Pearce, and Mr. P. Walne, and also members of the staff of the Kent County Archives Office, Maidstone, in particular Mr. M. Carter and Miss A. Revell who helped with the translation of technical letters and papers and Mr. P. Bloomfield who provided examples of current practice both in that office and through the Working Party of the Society of Archivists. I am especially indebted to Miss Ann Wale, my former secretary, who typed the final version of the text.

Felix Hull  
Bearsted, Kent

April 1981

1. The Appraisal of Records: A Frame of Reference

- 1.1 'The appraisal of modern records is one of the most important and sensitive problems of archival science.' (1). Thus began a study of the Appraisal of Contemporary Records, prepared for the VIIIth International Congress on Archives, held at Washington D.C. in 1976, yet it is only within the past half-century that archivists have become fully aware of the necessity for their involvement in this activity. As late as 1922 (to be reiterated in the second edition in 1937), Hilary Jenkinson could write in his famous Manual, with complete authority, that 'if we look at the objections . . . to the possible methods of destroying portions of our older Archive collections, we shall see that all are based on the difficulties that arise when the Archivist and the Historian are given . . . a share in the creation of those archives which it is their true business only to keep and to use respectively . . . We can only add here that if they (sc. the objections) were insuperable in regard to the Archives of the past, they must be equally so in regard to those of the present and future.' (2)
- 1.2 This argument was essentially based on the fear that involvement in selection for preservation would inevitably result in subjective decisions by the archivist concerned and that, in so doing, that impartiality of the record, which Jenkinson was so anxious to preserve, would be impaired. Some indication of the change in outlook which has occurred is indicated by a sentence from a paper on 'Statistical Sampling' written in 1975 by an American archivist. He wrote: 'As archivists we know that we, too, only preserve and protect a sample of the total record that describes the work of the creating agency. "The record" is never complete even with the best records keeping techniques and personnel . . . Archivists should be able to sample without undue anguish about the integrity of the records.' (3) Still more, one might say, is this true of the wider issue of appraisal, which among its various techniques embraces sampling and it is fully apparent that a radical change has come over archival thinking since the early years of this century.
- 1.3 Indeed it may be said that for nearly 6,000 years there is evidence of the place of the archivist as the conserver of the record in its various physical forms, clay tablet, papyrus, parchment or paper, and that throughout this period of time, although aware of those factors damaging to archives, the custodian never thought of himself as the destroyer or selector of those precious items passing into his care. Agarde, a seventeenth century British archivist, could write of the hazards of fire and water, rats and mice, and could even acknowledge the possibility of human error in misplacing, but the concept that he, as an archivist, played any part in the deliberate selection of the records he served would have been wholly foreign to him.
- 1.4 It was not until the nineteenth century that considerations of destruction as anything but the random risk of age became a conscious concern of those involved in record keeping, but even so, it was in a passive rather than an active role that the archivist saw himself and it was often for others - the creators of the archives - to determine what, if anything, should be discarded. Criteria for selection which were issued as early as 1835 by the French Ministry of the Interior(4) and various forms of general instruction, specific destruction schedules or, perhaps, lists of records to be regarded as 'permanent', remained a feature of European administration until recent years. This outlook

was encapsulated by the well-known dictum of Jenkinson that, 'for an Administrative body to destroy what it no longer needs is a matter entirely within its competence and an action which future ages . . . cannot possibly construe as illegitimate or as affecting the status of the remaining Archives. '(5)

- 1.5 The 'great divide' resulted from three significant changes which began to affect the written record over a century ago, but which accelerated rapidly after 1914 so that today there is no independent state which is not conscious of the outcome. These changes were: (a) the rapid expansion of government in terms of function and in the complexity of administrative organization resulting therefrom; (b) the increasing ability over the period to proliferate records by means of carbon copies, photographic and xerographic copying and, latterly, all the paraphernalia of machine-readable documentation; (c) the growing tendency for public access for research purposes to be speeded-up, and the increase in open government in many, though not in all, areas of activity. As a corollary of this it should also be noted that just as governmental growth (national, local or in the sense of the control of big-business), results in a growth of bureaucracy, so, too, records, in whatever form, are the life-blood of that bureaucracy, so that, irrespective of the means of easy proliferation, the mere fact of more government would itself tend to create more documentation.
- 1.6 The study prepared for the 1976 I.C.A. Washington Congress cited some examples for archival and record growth which are sufficient evidence in themselves and stress why action has become essential. It refers to the U.S.A. as holding 930,000 cubic metres of archives, growing annually by 165,000 cubic metres; the Soviet Union in 1975 reported holding 35 million records with an annual intake of an estimated 4,000,000 'retained records'; and France it recorded held in the Archives Nationales and the departments some 1,370,000 shelf metres with an annual growth rate of 51,000 metres.(6)
- 1.7 It is essentially this astronomic growth in quantities which has precipitated the archivist into the world of records control as well as that of records preservation and it is perhaps understandable that those countries most immediately concerned with this new activity were those with a vast current documentation problem, rather than those immersed in a tradition reaching back to the Middle Ages and beyond. The science of records management thus was born in the United States and the first serious theoretical work on that subject arose from the same source.
- 1.8 There is, however, another factor to be considered before details of appraisal as a method can be examined. This is the pressure on archive services created by shorter closed periods associated with the vast growth of documentation. This fact has had the significant effect that for the administration to be responsible for selection is no longer wholly effective or acceptable because there must be some room for the assessment of research potential when records are appraised. This change of outlook has been referred to as the fundamental change of concept from archives as 'the arsenal of law' to archives as 'the arsenal of history' (7) and it is noteworthy that in a study on the training of archivists, the explanation was offered that records management was the basis from the archivist's point of view of 'the appraisal of the potential research value of the documents' .(8)
- 1.9 It was around such concepts that battle was fully joined during the 1950's with Jenkinson as proponent of the conservative view and Theodore Schellenberg emphasizing the imperative need for involvement; and while the one, as we have



seen, eschewed all archival selection, as opposed to that carried out by the administration, the other insisted on the need for the archivist to determine those secondary research values which accrue to archives as their administrative life diminishes.

- 1.10 Nevertheless, appraisal is but the core, the very heart, of the larger concept of records management and over the years an entirely new emphasis has arisen which relates to the creation, maintenance and use of records rather than only to their destruction, so that, in current terms, the control to be exercised over records covers their whole life-cycle. From the point of view of this study, however, attention is focussed on that central issue at which stage the archivist and maybe the records manager and researcher, are concerned with determining what shall be retained on a long-term or indefinite basis, not to use the common form 'permanent'. The key factors behind whatever decisions are taken will be both economic and cultural: i.e. the sheer cost of the retention of an enormous and growing bulk - which is the essential record problem; and the impossibility of research use in terms of those astronomic quantities of records. To these must be added the requirement to assess in as objective a manner as possible those classes and series of records which contain the maximum evidential and informational content.
- 1.11 A most valuable recent paper on appraisal(9) emphasises the point that the 'retirement of records as they become non-current is inevitable. The only alternative is between unplanned chaotic retirement on the one hand and considered systematic retirement on the other in which retirement decisions are made on the basis of record values.' It goes on to analyse the various procedures which lead to the retirement of records and in this respect identifies (a) the preparation of inventories or surveys of the records of each agency concerned; (b) the analysis of each survey to determine retirement policy; (c) the preparation of control documents for action to be taken; (d) 'the careful appraisal of the specific retirement proposals made in the retirement instruments - i.e. their review and their approval or disapproval, on the basis of objective criteria of value' ; (e) the implementation of decisions agreed. It should be noted, too, that it sees item (d) as being essentially carried out by the archival agency and not by any other body, thus firmly endorsing the view of archival responsibilities set out by Schellenberg.
- 1.12 There has, however, been some measure of variation in terms of practical policy between the various national and other archive services of the world. The Washington Congress study indicated in 1976 that methods varied and that the appraisal of records was controlled by differing kinds of schedules or 'retirement instruments'. Emphasis upon the general schedule covering all classes of records created by an agency and indicating appraisal decisions reached between the agency and the archival body concerned is central to thought in the U.S.A. and in Australia; on the other hand, the outcome of the Report of the Grigg Committee in the United Kingdom in 1954(11) established the concept of periodic reviews of classes and series, partly administrative (a reflection of the Jenkinsonian position) and partly by the Public Record Office through the activities of inspecting officers. Other countries within the British Commonwealth have tended to follow this pattern, though with their own variations. In fact the difference between these two patterns of control is more apparent than real, for the two approaches are not irreconcilable and the real issue is more a question of the degree of archival control at all stages of the record life-cycle. Indeed this basic issue may be stated as one of responsibility - to what extent is it the archivist who determines policy and action; is he an advisory consultant or the person with the determining voice in the final analysis?

- 1.13 Whatever the methodology to be adopted, or by whatever means the instruments of retirement or schedules of destruction or retention are determined, the person or persons responsible for decision-taking require some guidance in the form of specific criteria for making their assessment. The older view that all appraisal was the function of the administration, or even the view that all appraisal was upon administrative criteria alone was relatively easy for it could be established without question that if there was continuing legal, fiscal or other administrative need this alone would determine the selection of the records as the archives of the future; and it is still true that if these factors are the grounds for retention then they will normally be accepted as the determinants for the action to be taken. It is also true, of course, that much purely ephemeral documentation can be discarded at a very early stage by the officer responsible for it, so that, it is true as previously noted, that the record is never wholly complete, even if everything remaining on the administrator's desk or in his filing system is transferred direct to the archives. In appraisal one must never forget the large element of what the French refer to as 'waste basket papers' . (12)
- 1.14 The author of a recent Society of American Archivists' Handbook on appraisal and accessioning has emphasized the importance and the difficulty of appraisal in the words: 'Records appraisal is best considered as a process that requires extensive staff preparation, a thorough analysis of the origin and characteristics of record series, a knowledge of techniques for the segregation and selection of records, an awareness of the development of research methodologies and needs, and a sequential consideration of administrative, research and archival values.' It is from this position that a more detailed examination of criteria must be attempted for the key to appraisal lies in that 'sequential consideration' with which he concludes his statement. Accepting the administrative element, many have attempted to determine 'historical criteria'. In the United Kingdom the Public Record Office and some local government offices have approached the problem in this manner and one such set of criteria recognizes the following elements: administrative needs, administrative history (i.e. documents having an informational content regarding the development of the organisation), genealogical and demographic value, topographical value, and an amorphous area of documentation which relates to the implementation of national policy at whatever level is relevant. Inevitably, however, such criteria reflect the characteristics of the agency concerned and it is desirable, therefore, that a more generalised statement should be established and, once again, the archivists of the U.S.A. help to provide an answer. As long ago as 1944, one State archivist attempted to lay down a set of rules(15) and this, of course, was done even before Schellenberg prepared his major study of the whole modern records issue.(16) More recently one of the most comprehensive statements of criteria has been published, an analysis so valuable that it requires extensive quotation:

'Determination of archival value is an act of judgement and therefore necessarily more or less subjective. But it is possible to minimize such objectivity by defining the objectives and criteria of appraisal, i.e. by setting forth standards of value in terms of which the appraiser makes his judgments. It should be stressed that the appraiser is not concerned with 'primary values', those of the records for the originating agency in connection with the implementation of its basic functions, but rather with 'secondary values' , the residual values the records may possess once they have become non-current. These secondary values fall into the following principal categories:

- 5 -

- (a) Continuing administrative, legal, and/or financial values for the agency of origin and/or other agencies even after the records have lost their primary value for current operations.
- (b) Values that the records may possess which serve to protect the civic, legal, property and other rights of individual citizens or the public at large - the 'individual rights' value.
- (c) 'Evidentiary' or 'functional documentation' values - i.e. those reflecting the historical evolution of the creating agency in terms of its powers and functions as established by law or other regulation, its organizational structure, its programmed, policies, procedures, decisions and significant operations. Key records having such values are needed over the long term as evidence of the agency's stewardship of the powers and responsibilities officially assigned to it. Moreover, by exploiting the experience embodied in these records, future officials whose work lies in the same area of responsibility may benefit from both the successes and failures of their predecessors. Basically records of this kind help to give continuity and consistency to the actions of the creating agency over time.
- (d) Informational value i.e. those contributing substantially to research and scholarship in any field of knowledge. The appraisal process here weighs the potential research values of records for the historian, political scientist, economist, sociologist, geographer, statistician - indeed for social and physical scientists and students of the humanities in every discipline.'

The author concludes: 'clearly appraisal is not an easy task; indeed it is probably the most demanding of all the professional responsibilities of the archivist . . . It is the key phase by virtue of the fact that appraisal decisions irrevocably determine the content of the archives to be preserved permanently and therefore the nature of the sources available for use in the future by government, the research community and the public generally. '

1.15 There is yet another factor to be considered in this preliminary examination of appraisal. The archivist, traditionally, has been concerned with individual pieces or items irrespective of format. He has listed volumes, rolls, parchments or paper sheets as they fell within the class or series under examination. Since the middle of the last century the standard pattern for administrative documentation has been the paper file: a collection of items, usually in chronological order, ideally, but not always, relating to a single topic, person or place, and threaded, clipped or otherwise attached within a cover. So long as this was still the occasional usage it was possible for the archivist to consider each item within the cover, but in the circumstances of contemporary government (in the widest possible sense), this is no longer practical and whole series of files must be assessed together, the individual file normally replacing the individual item in any consideration. Appraisal of conventional records, therefore, has become a matter of consideration of files, individually at the lowest level of selection, but more usually of series as a whole amounting, perhaps, to thousands of actual files. Later in this study there will be cause to refer to series such as the promotion files for soldiers in the German

Federal Republic, amounting to some 130,000 individual files; or to the sickness insurance files in Finland where a single series includes 2,802,000 files.(18) It is in relation to quantities such as these that the above criteria must be considered and it is in such cases, too, that the archivist faces decisions which may include the destruction of the entire class or series. The appraisal challenge, therefore, rests essentially upon the nature of the actual files under consideration: of what do they consist? have they been properly maintained - indeed has the filing system, itself, been adequately thought-out in the first place? are they individual and variable in nature or do they form a homogeneous mass of papers concerned with one function only and essentially similar in character?

- 1.16 It is of special interest to note how clearly has been stated in recent years the distinction between administrative and research criteria and it may be argued that so far from casting aside earlier precept, what has happened is that Schellenberg's thesis has been added to that of Jenkinson, so that administrative control of records so far from being denigrated by the growth of management techniques, remains the primary factor in all appraisal supported by the secondary study of research values. It is necessary, therefore, to consider briefly the timing of appraisal programmed, a matter which has already been touched on in discussing the variations of methodology. It has been seen that there is a tendency to approach appraisal from one of two standpoints: either it is carried out on the basis of, and as a direct result of, scheduling in which case each category of material will be evaluated on the basis of the time scale agreed during the drawing up of that schedule; or, alternatively, all records, irrespective of inter-class variations, will be reviewed at set times during the life-cycle of the records. In this regard it is interesting to note that the American studies cited above, and indeed the appraisal regulations of the U.S. National Archives and Records Service, have little or nothing to say about times and seasons; whereas the Public Record Office instructions are based on the concept of reviews five and twenty-five years after closure of the file.
- 1.17 In either event the primary evaluation will be on the basis of administrative need and certain archival classes (i.e. classes, or series, of known administrative, legal or fiscal, evidential or informational worth to the agency) will be identified and will join those earlier archives already accepted as of indefinite life (i.e. permanent so far as anything can be). This basic analysis which may be carried out by the agency itself, or by the agency in association with the records service, will normally lead to only one of two decisions: retention or destruction. There is usually little room for any other decision for it is rare for the organisation concerned to have need for only a partial retention of a particular class.
- 1.18 When, however, secondary values are under consideration the position is somewhat different. Here the archivist is not dealing with series of known archives so much as re-examining what one might term 'the discard pile'. Those records which have been selected by primary criteria are unlikely to be re-assessed, but those which the agency has rejected may nevertheless be of concern to the archivist and to his clients. In this regard, therefore, there are three possible answers to the appraisal question: retention, destruction and partial retention. It is this last with which this study is especially concerned, but it is a decision which may take one of two forms. In the first place it may involve the detailed examination of individual files and a process variously termed weeding, stripping, purging or screening. This is essentially

the removal from the file of ephemeral papers, or the selection from a file of those particular forms of records which are regarded as of long term value, though the bulk is dross. It has been rightly observed that 'normally, the processor cannot justify taking the time required for weeding in a correspondence file, but this does not prevent a decision to cull certain types of material ...' (20) The decision to retain less than the whole may however involve a very different process, the selection of part of the material on the basis of a scientific statistical method or on the basis of criteria of quality and it is this process which is termed 'archival sampling' and with which this study is primarily concerned.

- 1.19 In fine, therefore, the appraisal of records is essentially a phenomenon of the present century, forced upon archivists by the sheer bulk of documentation created by present-day governments. Methods of approach may, and will, vary from country or from archive service to country or service, but there are few archival services worthy of the name which do not now employ some measure of records management and therefore some form of appraisal. National legislation may determine the time-scale for evaluation and the period before access to the public is permitted, but the problem, both economic and cultural, will be the same in that bulk presents impossible storage costs and impracticable research media; so that, although the scholar and the archivist may each deplore the need for destruction they must per force come together in a determination that only that is kept which will really be worthy of future consideration whether by government or for the purposes of research. In this already complex area the added problems of machine-readable records and the other results of modern technology play an increasing role. As a result it is fully apparent not only that not all records can be retained but that it may be possible to keep a part only of series which would otherwise be lost. It is also postulated that such portions of the whole can nonetheless provide as satisfactory a unit, both as the record and as a research tool, as would be found were the whole class or series to have been retained. It is the purpose of the remainder of this study to probe into that suggestion.

#### Notes

1. A. Kromnow, 'The Appraisal of Contemporary Records', report for International Council on Archives, 1976
2. Hilary Jenkinson, Manual of Archive Administration, p. 149
3. R.M. McReynolds, 'Statistical Sampling of Archives', National Archives and Records Service, 1975
4. M.J. Brichford, Archives & Manuscripts: Appraisal and Accessioning, Society of American Archivists, 1977, p.1
5. Jenkinson, op.cit., p.149
6. Kromnow, Op.cit., p.2
7. F.B. Evans, 'Modern Concepts of Archive Administration and Records Management', UNESCO Bull. Libr., vol. XXIV, No.5, September-October 1970.

8. L. Bell, 'The Professional Training of Archivists' UNESCO Bull. Libr., vol. XXV, No.4, July-August 1971.
9. M. Reiger, 'Modern Records retirement and appraisal practice', UJISLAA, vol.1, No.3, 1979
10. Kromnow, op.cit.
11. Grigg Committee Report on the Public Records (Cmd. 9163), 1954
12. Brichford, op.cit., p.17
13. Ibid., p.2
14. 'Guide for Officers in Charge of Modern County Council Records', Kent County Council, England, 1979 revision
15. Norton on Archives, ed. Thornton W. Mitchell (1975), P.240
16. T. Schellenberg, Modern Archives, principles and techniques, 1956
17. Reiger, op.cit.
18. See pp.34-5
19. An example of an agency retaining only part of a series will be found with the report on Canadian practice on p.32; I am also indebted to Dr. F.B. Evans for information of a somewhat similar case in Pennsylvania.
20. Brichford, op.cit., p.16

2. Sampling: Some Theoretical Considerations

- 2.1 A useful, general and non-technical definition describes sampling as 'a small separated part of something illustrating the qualities of the mass' and the taking of a sample is said to enable one to 'get a representative experience' or to 'test the qualities' of the subject of the exercise.(1) It follows therefore that sampling, in the archives sense, is the deliberate selection of certain documents (files, volumes or whatever is the normal format within the class or series), as an example of that class or series for permanent preservation.
- 2.2 There is of course a sense in which all selection at whatever level is a form of sampling, and that the moment a decision is taken to retain less than the whole, this process comes into operation. But this actually begs the question, for sampling is a very specific method of selection employed by the statistician to answer problems in quantitative analysis, and also a determinable process whereby the archivist attempts to reduce bulk, not at an early stage of the management procedure, but as part of the final stage of control. In other words, sampling is not a check imposed by the administration which created the record, but a technique related to reference and research needs and one aspect of the final determination of the future of the record based on historical and similar criteria. It is true that a recommendation to sample may find its way onto a disposal or retention schedule as part of the ultimate disposal plan, but it should only be adopted after clearly defined principles have been laid down by the archivist, not by the records manager (unless they be one and the same person), and possibly with the advice of an expert in statistical analysis or a scholar in the field most likely to find use for the result of the sampling process.
- 2.3 It follows that this procedure is particularly suitable, at least in theory, for those records generically referred to as special category or particular instance papers. Schellenberg defined these as: 'recurrent records - the kind that can be appraised on the basis of their informational content, that pertain to specific actions which are usually definitely terminable, and that can ordinarily be retired by schedule for disposition'.(2) That may indeed be true, but as the archivist knows only too well, often enough they also consist of bulky classes of sensitive or confidential papers which must be kept for a considerable period of time, even as long as 50 or 100 years, before they can be destroyed, let alone made available for research and which, while they may contain significant information, especially of a personal character, are so bulky as to preclude retention as a whole. The question arises, however, whether such a selection can ever be wholly satisfactory both in terms of archival principle and research need, and, if so, how this can be carried out. Sampling it is postulated can provide an answer to these questions though whether to the full satisfaction of all concerned is part of the problem lying behind the concept of this study. Ideally, all should be kept: if this proves impossible or impracticable, how best can a proper selection be attached?
- 2.4 Sampling, as will become clear, can involve various methods, each of which is suspect to some degree and in every type of sample and every kind of methodology there are imponderable which rightly give rise to much uncertainty. In very general terms a sample can aim at one of two goals: either it can attempt to preserve what appears on subjective grounds to be the most significant or important element of the series; or it can aim at achieving a totally objective

selection in which actual content is subservient to the principle that archives must be Impartial. Each of these approaches has its particular virtues, each gives rise to serious misgivings, but the real Issue in archival terms lies in the element of destruction. In every case, to take a sample of a record class or series implies, and normally involves, destruction of the residue.(3) This is the divide which separates the research scholar from the archivist and which makes sampling so difficult and dangerous an archival skill.

- 2.5 When a scholar deems that a sample is required for his statistical purposes, he selects, on an appropriate basis, a portion of the whole, but leaving that whole to be reconstituted and re-used time after time. In contrast, when the archivist samples, he leaves the portion so selected as the future whole; the residue has gone and can never be used again. In 1969 the British Records Association chose 'Sampling' as the topic for their Annual Conference seeking advice from Cambridge scholars in this regard. The example chosen by the opener of the debate was those probate inventories found among British probate records especially in the period between the sixteenth and eighteenth centuries. No one would suggest or agree to the proposition that these records should be sampled archivally and the residue destroyed, so that whatever value the discourse had in explaining research method, it failed utterly to assist the archivist faced with the vast series of today, the content of which is only peripherally of value in comparison with the example chosen. The proper application of any sampling technique to records must result in the destruction of some records or it is pointless as an exercise, and ipso facto to the loss of some knowledge, however limited.
- 2.6 It is apparent therefore, that while appraisal always implies some destruction based in the first instance on administrative grounds - the only basis upon which Jenkinson in his day would countenance selection - sampling is normally a method of applying non-administrative criteria and always involves a measure of destruction.(k) Inevitably both purpose and application are open to question: the archivist is, and must be, concerned with the quality of the record, its integrity and its impartiality; the researcher, while acknowledging this aspect of the matter, has also his specific need which may, or may not, have been covered to his satisfaction by the action which has been taken. A sample which is subjective may be too partisan, the purer objective sample may in effect emasculate the class to which it is applied. It is therefore necessary to consider the alternative methods of sampling and to assess in theory their strengths and weaknesses.
- 2.7 Although terminology varies somewhat between nations there are four recognized methods of sampling which can be readily identified:
- (a) the taking of examples or specimens;
  - (b) purposive sampling or qualitative sampling on a pre-determined pattern or basis;
  - (c) systematic sampling also on a pre-determined basis; and
  - (d) random sampling employing specific and scientific objective process;
- both (c) and (d) may at times be referred to as quantitative sampling and since each form requires its own pattern of constraints, each must be examined in some depths.



- 2.8 The example. This is a usage which has little if any true archival merit and may also be said to be of uncertain value for research purposes, yet, occasionally, it may be adopted with some Justice. It is simply the selection from an ephemeral series of papers of one or more specimens to illustrate administrative practice at a particular date. It may be applied to any series of records which will otherwise be destroyed in toto and even occasionally unused forms may be retained as examples of what was once done. It is a usage which may well be applied in the case of those housekeeping records, e.g. postage books and other minor financial records, which will not be kept for administrative purposes and have no known research potential, but where the retention of an example or specimen of a class will at least help to complete the picture of a former administration and may help to provide precedent for future action.(5) The resultant example remains an archive in the technical sense, but has very limited historical or other use since it can only be cited as an indicator and in no way used for comparative or statistical study. It is possible to retain a number of examples - a 'representative sample' - in which case there is a similarity to some methods of systematic sampling, or it may be that examples are chosen to illustrate moments of change in the methods of document keeping. Whatever is done it is clear that all finding aids should state plainly what has been retained and the basis upon which the examples were chosen. In British Local Government experience this type of sampling has been used on occasions with the records of the former Boards of Guardians, 1836-1930, simply to show that certain classes were once maintained.
- 2.9 Purposive sampling. Only one remove from the taking of examples is the more systematized method of qualitative or purposive sampling and nothing gives rise to more uncertainty than this technique, for this process is, and must be, essentially subjective. It takes place when a selection is made on a pre-conceived set of criteria, the intention being to retain the most significant or important records of a class or series. The similarity to the example is clearly apparent, but the process is carried out, not because it seems desirable to retain a specimen, but because certain material, in an otherwise destructible class, is considered to be of special value for research. The difficulty is, of course, that although the sample thus taken may have a certain subjective validity, it provides a biased or skewed picture of the class in question and has no worth in terms of statistical analysis except within very circumscribed limits. Nevertheless this kind of sample has been taken by archivists from time to time simply because it offers a means of retaining information of a specific kind on a variety of topics, where the class as a whole is of little general significance.
- 2.10 Before wholesale condemnation is accorded, it is proper to reflect that when an administrator selects on his own criteria, something very like purposive sampling takes place, the question being whether administrative criteria are to be regarded as archivally valid while historical, sociological and other bases of selection are not so regarded.(6) Any form of selection other than some types of systematic sampling and all random sampling will, in fact, be subjective even though the criteria laid down may be legal, fiscal or administrative in character, but whereas in theory it is perfectly proper for the officer of an agency to determine what will be useful to him and his successors for the purposes of the agency served, if a third party determines selection with research interests in mind then questions of validity and subjectivity inevitably arise.
- 2.11 It may be useful, in this general discussion, to consider one or two examples from British experience In the past simply to illustrate occasions when this

method has been used and to judge the problems surrounding it. Among the Destruction Schedules issued over the years have been ones concerned with the records of the former courts of Quarter Session and with the records of the Coroner. In each case the actual records, although designated Public Records, are held locally and administered at that level. So far as the records of the Courts of Quarter Sessions were concerned, schedules of 15 December 1941 and of 13 February 1968 each dealt extensively with the criminal aspect of the courts' activity, the only essential difference being that the former permitted records of later date than 1837 to be destroyed, the latter only those more recent than 1889. In each case indictments and depositions of witnesses were scheduled for destruction after a period of years except, in the first case, for those relating to riots, a category extended in the second instance to include 'cases relating to riots and to matters of public interest such as commons and waste lands'? The fact that this decision' taken at the centre could lead to the destruction of practically all records of a criminal nature concerning the Courts between 1837 and 1970 is perhaps of less moment to the present argument than that one schedule limits retention to a very small element ignoring all other matters of special interest while the other leaves a very wide area of discretion to the custodian of the records. It is clear enough that this kind of instruction may lead to a very uneven coverage of an important series of local documentation.

- 2.12 The case of the records of the Coroner's jurisdiction is even more difficult to justify.(8) In this case the Home Office Circular No.250/1967 indicated that only registers and papers relating to treasure trove were to be retained, but it then went on to indicate that because cases of sudden death such as were brought before the coroner could contain matters of special historical interest the custodian might choose to keep other unspecified case files. In this instance it is apparent that the archivist, or whoever is responsible for the records, has the widest possible discretion to select whatever seems to him to be of 'interest', so that, if the officer in question happens to be interested in industrial matters cases of accidents in factories may be kept, but if that is not his particular concern they may equally well be destroyed.
- 2.13 One other case from experience may be cited, for it both indicates the value and the weakness of this kind of archival sampling and shows well enough why the researcher may be thankful that a purposive sample has been kept, while deploring the method and its affect upon the series as a whole. In this special case a series of records was found concerning the working of the appellate tribunals established under the British Military Service Acts, 1.916.(9) These records consisted of registers giving brief particulars of all cases heard and files of the more detailed case papers. Consideration was given to the destruction of the latter, but it was found that while the vast majority dealt with brief extensions given to tradesmen and others before call-up to the Forces, a small number dealt with conscientious objectors, a matter of far greater social and historical interest. The result was that a small systematic sample was taken of the case files as a series, but that all files concerning conscientious objection were retained. Ignoring for the moment the systematic element, the purposive element was a clear case of the retention by this method of a small group of records having some specific research potential. It was valid on two counts: first that it could be used within the constraints of records in general concerning pacifism during the First World War; second, that thanks to the existence of the registers, it was still possible to make an overall statistical analysis of the class and to assess thereby the nature and quality of the work of the Tribunal. On the other hand it could be argued

that survival depended upon the personal opinion or subjective judgement of the officer determining the selection and that that fact, *ipso facto*, invalidated the whole exercise. No one is ever likely to be wholly satisfied with purposive sampling, yet it is one method of ensuring that records containing information of some special or peculiar interest or value are retained from a large and predominantly ephemeral class or series.

2.14 Systematic sampling. As in all matters concerning records management, it is important that terminology should be exact and certainly in the U.K. there has been considerable confusion until recently over the difference between systematic sampling and random sampling. It is only relatively recently that the statistician has persuaded the archivist to adopt his methods and terminology with the result that, for many, any sample based on numerical series or chronological series was regarded as 'random' As will be seen later 'random' sampling not only involves a haphazard selection but also requires that every piece in a series has an equal chance of being selected, the very anithesis of 'systematic' sampling. As the name implies, systematic sampling depends upon the establishment of a particular pattern of selection in that either every nth. file is preserved, or else all the papers for a particular month, year or other chronological unit. It used to be considered that, especially in the former case, the result was sufficiently haphazard for statistical analysis, but this is not necessarily the case and researchers have complained in the past of the wrong papers being retained by this method. In the chronological case, there has been a tendency to retain census year records, or for records to be kept for every fifth year. In the case of British rating records a fifth year retention has been supplemented by the records of years in which particular legislation affecting the assessing of rateable values became effective, but this too has found as many critics as supporters.(n) This system has In fact been used for other classes of financial records where their content is basically similar and of very little moment, so that either every nth. volume could be kept or that the volumes for particular years are retained. It is apparent that while a reasonable cross section may result with a large series of files, when this system is applied to volumes the keeping of a systematic sample may result in something very close to the keeping of specimens or examples, and that in those circumstances some of the criticism which has arisen is readily understood.

2.15 The following quotation from a letter received in 1970 will illustrate just how confused some archivists have been regarding this method and also show how unsatisfactory the outcome can be:

'I see the P.R.O. Schedule suggests random sampling of post-1889 indictments. Is it important how this is done or can it be entirely random providing nothing of real interest and importance is discarded? The reason I ask is this: I tried several methods with one huge bundle of about 200 summary convictions and depositions: by saving every sixth summary convictions and every fourth deposition I found I had selected 100 per cent female offenders: The whole bundle didn't suggest this at all.'

This illustrates not merely a complete confusion regarding the meaning of 'random', but also the very curious way in which a deliberate attempt at randomness can result in an atypical selection. It again, perhaps, underlines why scholars have been less than satisfied with this as a method of selection despite the current opinion that systematic methods approximate to randomness. It also, of course,

stresses the archivist's acute dilemma between the purely objective sample and the destruction of that which is, or which appears to be of 'interest'. Even when papers are retained at, for example, five or ten year intervals it is still possible for the information desired by the searcher to be lost by reason of the elimination of the records of intervening years. That being so, the fact that scholars also find this method unsatisfactory for the provision of an adequate objective data base, determines that considerable doubt must be expressed regarding it. Yet it remains a firm favourite among archivists largely, one suspects, because it is so easy to apply.

2.16 Before leaving systematic sampling, however, one must briefly refer to two related methods referred to as 'the Turnus system' and 'regional sampling', each of which is a form of representative sampling. Each of these methods refers to the case where there are a number of regional or district offices of the same central agency. In the first case, records will be kept in a form of rotation, i.e. district A will retain records of year 1, district B of year 2, and so forth. In the second case the records of one or more of the district or regional units are retained as representative of the whole. Neither system is entirely satisfactory for the reasons previously indicated, though the 'Turnus' pattern is the more acceptable for statistical purposes and the latter method has the advantage of keeping all the records for the particular units selected.

2.17 Random sampling. In the world of statistics, randomness implies that every unit in a series has an equal chance of representing that series - the sample is essentially haphazard and in no way depends upon patterns or periods as does systematic sampling. This, therefore, is the method adopted for any poll taken purely on a statistical basis. Any special features are eliminated from the sample, which must be as ideal in an objective sense as it is possible to make it. One study on statistical method refers to the fact that a poll based on a telephone directory automatically omits those persons without a telephone or ex-directory and therefore is not truly random except within that constraint and many similar examples could be quoted.

In the case of archives, therefore, all those considerations which have been examined for purposive and systematic sampling are strictly eschewed. The true random sample must avoid at all costs any suspicion of special cases or special pleading, neither may it relate to nth. items or years or in any other way take particular interests into consideration. If all these matters are rigorously excluded it is then argued that the result will be acceptable for quantitative analysis or for any other statistical purpose. The unspoken corollary is, of course, that all the residue of documents is destroyed! When a ten per cent poll is taken for political or other purposes the remaining ninety per cent of the population is not thereby condemned! That, be it added, although a fact of archival sampling is by way of a caveat, it does not of necessity invalidate the method used.

2.18 Everything to do with this process must be of a random nature and it cannot therefore be carried out without some sophisticated elements. The basic need is for a random number table. This is a table of numbers 'found by a kind of electronic roulette wheel and . . . checked in many different ways to ensure that they are as random as possible'. As a result they may be read to left or right, up or down and as single digits, in pairs or triplets according to the total number of items in the parent group to be sampled. The second - but time consuming and thus costly - requisite is that the whole class or series of papers to be processed must be numbered throughout so that each piece is identifiable by an individual number unique in itself.

- 2.19 In using the table it has been suggested that the starting point be determined by the drop of a pin and the reading direction by reference to the minute hand of a watch or clock. In this regard the following explanation has been given: 'Read the numbers upwards if the time is between 7 1/2 minutes to the hour and 7 1/2 minutes past the hour; read the numbers to the right if the minute hand is pointing to the right anywhere in the quadrant between 7 1/2 minutes past the hour and 22 1/2 minutes past the hour ...' and so forth.(15) If therefore a series of 900 files is to be sampled and a ten percent sample is required once the starting point and direction have been determined, the numbers will be read off in triplets until the full sample of 90 pieces has been made, any number above 900 being ignored en route.
- 2.20 Despite the merits which are apparent in random sampling, the inherent complications indicate that it is not a task to be lightly undertaken and it may well be that expert advice should be sought from those likely to use the result before embarking on the project. In any event the size of the sample is a matter of concern both to the archivist and to the user.
- 2.21 Size of sample. In all sampling the size of the resultant group of pieces is a matter of some concern, and although the archivist in determining size will also be concerned with the continuing nature of the class involved and the future needs of that class in however mutilated a form, in the case of random sampling certain other factors will also apply. It is one thing to decide on the retention of certain papers or volumes because of their content or because they cover particular years, but neither of these criteria is valid for quantitative analysis. In this instance the archivist must bear in mind that the larger the sample the more accurate will be the result in terms of showing the nature of the parent group, and that the degree of accuracy required will therefore be a determining factor in deciding the size of the sample. It is important to appreciate that absolute size is more significant statistically than the proportion, i.e. if a class of 100,000 pieces is sampled, a sample of one hundredth may contain as much essential information as one of a tenth, provided the class is homogeneous. In other words, the more homogeneous the base the smaller in absolute terms the sample can be to incorporate essential information. Plainly detailed knowledge of the content is a prior condition before sampling methods can be determined and this again may be a reason for consulting the potential users of the records before a determination is made.
- 2.22 Obviously such considerations do not apply to any of the other methods, for the fundamental criteria are wholly different. In the case of the Military Service Act, 1916, the whole of a particular element was sampled partly because of its intrinsic interest and partly because it was a small enough element to be preserved as a whole. A ten per cent systematic sample is likely to be determined largely on the size of the whole or because census years or some other term of years is likely to give the most generally useful result. The archivist is constantly made aware of the basic problem that sampling involves the destruction of information, and that whatever system or possibly combination of systems he chooses to use must provide the best possible answer in terms of the record to be sampled and also in terms of his clients who will have to use the result. Despite the need for objectivity some personal judgement must enter into all kinds of sampling, not least in the determination of the size of sample. Even in the case of the random sample a decision must be made which is totally different in kind from the sample of

individuals or of their opinions, for someone is deciding to keep a set quantity on the basis of the likely research value of what is kept and that decision whether taken by the archivist or the scholar will be a subjective judgement.

- 2.23 The size of sample will depend upon the nature of the record, its content and apparent significance, and upon the purposes to which the sample once obtained is likely to be put. A ten per cent sample may appear to be a simple standard but there may be cases where, on account of the size of the series and its content, only five per cent or even one per cent will be a valid exercise. Equally, it may be that purely quantitative assessments are immaterial, and that entirely different criteria commend themselves to the person responsible for decision taking. Whether this should be so will be considered later; it has been so in the past and it is well to consider international practice in this contentious area.

### Notes

1. Concise Oxford English Dictionary.
2. Modern Archives, P.108
- 3\* The point at issue is that various methods of sampling may be applied to the same archive unit, as for example was done in the case of the records of the Registrar General of Shipping in the United Kingdom. In such a case, although there will ultimately be a measure of destruction, the selection for preservation element will be larger than the original sampling implied.
4. See also the discussion on P.17-8. F.B. Evans cites a case in Pennsylvania where sampling took place in order to illustrate procedure. This 'representative sample' is a form of example valuable for establishing precedent and as **having** evidential value for the agency concerned.
5. The example given in **fn.** 4 above is of this kind. One might regard the compilation of abstracts a not dissimilar process, though it is not, of course, sampling as we understand it.
6. It has been pointed out that it is for this reason that the British pattern of a 'first (administrative) review' has been rejected in the U.S.A. and it is noteworthy that the present investigation into the British methods are partly the result of the risks of 'administrative bias' . (The report of the Committee headed by Sir Duncan Wilson on current British practice, Modern Public Records (Commd. 8204), H.M.S.O. , was published Just too late to be considered in detail for this study).
7. Home Office Circulars. (a) Public Records: Clerk of the Peace, Second Schedule, 30 Oct. 1941, ref. 832,538/11. (b) ref. 38/1968 Public Record Acts, 1958 and 1967. It is acknowledged that such schedules tended to be capricious and are no longer in favour, but problems have arisen in application.

8. Home Office Circular, ref. 250/1967 regarding Coroners.
9. 5 & 6 Geo. V, c.104 continued by 6& 7 Gee. V, c.15
10. It can of course be argued that neither the taking of examples nor purposive sampling (i.e. selection of most important material) are true sampling since in neither case is the totality of the class or series represented by the selection made. Nevertheless this common terminology has been and is still used and the practise may play an important part in an overall 'sampling' procedure. See also P.78, fn. 1.
11. A leaflet on this subject was issued by the British Record Association.
12. A. Kromnow, op. cit., p.20.
13. Russell Langley, Practical Statistics, (Pan Books, 1970)
14. Ibid., pp. 42 and 43.
15. Ditto, pp.42, 44.

3\* Sampling of Conventional (textual) Records - Practical Experience

- 3.1 The initial approach to this study involved the seeking of information from a number of established archive services by the means of a simple questionnaire. This asked whether the repository had used sampling techniques; in what circumstances they had been used and what methods had been employed; what size of sample had been taken and upon what criteria the choice had been made; and, finally, whether difficulties had arisen either because the methodology had proved unsatisfactory or because of searcher reaction. Unfortunately not all those archive services approached replied and of those that did, although uniformly very helpful, few actually answered the questionnaire in such terms that any comparative analysis proved possible.
- 3.2 Nevertheless the information forthcoming was extremely relevant to the matter under consideration and is therefore well worth setting out in a condensed form. It is also of interest to note that some confusion over methodology and terminology is still apparent and there is a tendency for those countries or agencies which have a long documented historical tradition to be less sure of the complete validity of the statistical methods than those lands with a shorter documented history, though with massive contemporary archives. Although it is clear that similar concern over methods has been experienced in the New World as in the Old, there appears to be a tendency for modern technology to be more readily assimilated there than in Europe.
- 3.3 The evidence submitted falls into two categories: that including practical examples and that essentially of general principle and theory. To this has been added some additional information culled from recent articles in professional literature. This chapter, therefore, takes the form of a resumé of each body of evidence arranged by country and where appropriate, by agency followed, if it seems necessary, by a very brief commentary on what has gone before. It also needs to be stated that where detailed and practical evidence has been received, in nearly every case it relates to series of files concerning welfare, health or unemployment insurance thus indicating a common factor in the determination of material to be sampled; and also that while this is so, there is always an imponderable factor of national individuality, so that in terms of evidential or informational content each series of records will vary according to the country of origin.
- 3.4 AUSTRALIA: Australian Archives. Canberra. The Australian Archives service has accepted sampling as a valid element of its records disposal programme with appropriate instructions for those carrying out such actions.(1) The decision to sample is normally embodied in the disposal instructions, forming an element of the General Disposal Schedule governing each class. It would appear that these methods are limited to the example, purposive sampling and systematic sampling and that although the word 'random' is used this in no case implies purely statistical sampling methods. Two distinct cases are identified:
- (a) The case of 'temporary value records' where the aim is to retain such specimens or representative samples as will indicate what such series once contained. 'The sample attempts to reproduce essential characteristics, functions or features of the series; . . . and the number of items selected . . . varies according to (1) the nature of the information or activity contained In or reflected by the series or class; (2) the rate of accumulation of record items; (3) the format of the series or class; (4) the significance of the series in relation to the records maintained by the agency! In this way a simple series, e.g. Treasury forms, may be represented by a single item to show the form of the record and the reamer of use' .



(b) In the case of selective sampling of case files series 'criteria should be established for the selection of those components which have research potential'. This scheme will be developed jointly between the Archives and the agency and selection will be 'random' (i.e. 'taken at random and including the first and last item in a series' ). The actual selection process on the basis of the Disposal Schedule will be carried out by agency staff, though if temporary records are held in archive custody, as opposed to agency, then archives staff will be responsible.

3.5 The purpose of all sampling is to reduce quantity; the choice will arise from the constraints inherent in the specific circumstances of each record situation; and the method is 'to take a structured sample on a continuing basis' . It is accepted that a ten per cent sample is adequate. Examples are provided of Sponsored Students' Files and War Veterans Case Files. The former is an example of a topographical systematic system:-

There are three elements to be distinguished. (a) The particular educational scheme (one of' nine); (b) the country of origin of the student; (c) the course to be taken. Since the schemes were designed to cater for particular groups of countries a ten per cent sample from each country ensures that the different schemes will be represented, that the countries will also be proportionately represented, and that the variety of courses will be covered. This selection is agency controlled before transfer to the archives on the basis of 'files due for sentencing annually rather than the number created each year' .

3.6 This interesting report emphasises the traditional view of sampling and the desire to evolve criteria of significance. It should be compared with the report from Canada where a much more radical approach has been established. There appears to be little evidence of an appreciation of randomness in the statistical sense and there is much emphasis on the value of specimens or examples of series and classes.

3.7 CANADA. Public Archives. The Canadian archives include extensive series of case files which, in the past, were 'usually selected only with a view to reflecting, if possible, all aspects of a particular program'. The realisation that this approach, while reasonably acceptable in terms of traditional research, was totally inadequate for the social scientist and for those using quantitative methods, has led to a closer examination and greater use of random statistical methods. The National Archives are now 'committed to developing methodologically sound sampling techniques' and have used the **records** of the Unemployment Insurance Commission as a means of establishing satisfactory criteria.

3.8 Unemployment Insurance Commission. The records of this body are held both centrally and at a local level and some paper records are retained at both levels. The Central Pay Office, however, also holds a computer file, microfilm copies of correspondence of forms detailing specific problems and of computer tapes. The class of benefit claim files, 1971-9, alone amounts to twenty million files, held on microfilm cassettes. The Unemployment Insurance Commission, itself, 'maintains a continuous longitudinal A.D.P. sample' of these claims.(2) Because all this material is linked with other taxation

and employment records by the use of an identifiable Social Insurance Number, if has been possible to take a 10 per cent statistical sample based on those files which bear a S.I.N. number with a terminal digit 5. This application can not only be used in terms of creating a data base within the computerized system, but can also apply to paper files bearing the final digit 5. A second stage or refinement in this sampling exercise is based on the distribution of names so that ultimately the sample will be of files, whether computerized or paper with the terminal digit 5 and a name beginning with a particular selected letter chosen on the basis of alphabetical distribution.

- 3.9 It was also stated that experience suggested that in the case of established and fairly static programmes there may be an advantage in the taking of 'periodic' samples rather than continuous samples. 'Preserving hundreds of metres of randomly sampled case files containing essentially the same information over a given period of time overburdens the researcher and needlessly squanders archival resources. It may be more worthwhile in certain situations to sample periodically the records to 'tie in' with 'research activities or document a changing program or aspect of the societal environment'.
- 3.10 This is an important example in that it is regarded by those using it as acceptable in all respects. It will therefore require further examination because it is the most radical single system reported and because of these claims that it satisfies both traditional research and quantitative analysis.
- 3.11 CZECHO-SLOVAKIA. Archivin Sprava, Prague. The Czech Socialist Republic does not use sampling methods to any great extent and the uses of such methods are defined in broad terms in the Regulations concerning records disposition. Nevertheless in its view there are three circumstances in which some form of sampling might apply:
- (a) 'when the character of documents suggests that these might by their contents, supplement or illustrate information contained in other documents judged as archival documents;
  - (b) when the character of documents does not exclude a possibility that these might be judged as archival documents some time in the future, although they do not have such a value under criteria currently applied;
  - (c) when the character of documents suggests that they will be of interest to researchers' .
- 3.12 In respect of the methodology this appears to fall into the category of qualitative sampling and is cited as (1) cases where a few single documents (examples) are taken - this appears to be the most frequent usage; (2) where part of a set or series of records concerning 'one specific sphere of action' can be identified; and (3) cases where it seems appropriate for a complete set of records to be retained (i.e. all documents originating from the activities of the organization concerned'). This suggests that the preference lies between some form of example or a purposive sample on one hand, or something approaching regional sampling on the other. There appears to be no use of any method which could properly be regarded as statistically random. The important comment is added that 'the application' of sampling 'in practice, however) in all respects depends on the Archivist in charge of the records selection procedure. Such a subjective approach may result in a lack of perfection, insufficient precision or poor results ...'

- 3.13 It is somewhat difficult to determine exactly how this methodology is applied, because of the uncertainty over the exact interpretation of technical terms. Very kindly the reply to the questionnaire was in English, but even so, some difficulty remains. It appears clear, however, that nothing approaching statistical sampling is, so far, in use or is contemplated.
- 3.14 FINLAND. Social Insurance Institution, Helsinki. The operations of this institution create three considerable series of personal files - pension, sickness insurance, and rehabilitation files - which are retained, partly in local offices and partly at the centre. In all, some 4,373,363 files were retained in local offices in 1979. These records are kept for ten years after the death of the insured person and are then automatically destroyed, but, in view of the potential value of the information for research purposes in the fields of social security, social politics, history and medicine, a sample is retained. It has been considered that 'a representative picture of the insured population could be obtained through continuous selection of files of all insured persons born on certain days each month'. As a result the files for all those persons born on the 8th, 18th and 28th of each month are kept.
- 3.15 This selection has the advantage that (a) the day of birth also forms the first two digits of the population register number; (b) the outcome is approximately a ten per cent sample; (c) the day of birth has no direct connection with the insured person's morbidity; (d) the like technique can also be applied to burial ground claims; and (e) the same system is also applied to hospital and social welfare records. In addition, however, the earliest 1,000 claims for national invalidity and old age pensions are retained on historical grounds as are files which involve appeals to the Insurance Court.
- 3.16 Sickness Benefit Vouchers. Because these vouchers contain details of treatment, medicines and travel expenses, a representative sample of one per cent per annum is kept. This is obtained by the retention of records for certain districts on certain workdays in February, May, August and November. For this purpose, too, the cities of Helsinki, Turku, Tampere and Oulu were chosen plus offices from representative insurance and hospital regions and the choice depended upon location and number of refunds, coupled with consideration of the number of insured persons and the proportional size of rural areas in relation to the towns and cities involved.
- 3.17 Clearly this is a very sophisticated procedure and results in a systematic sample of insurance files which is statistically valid. It is less easy to Judge the validity of the selection of vouchers because the criteria of selection appears very complex and one wonders whether the result is as valuable for research as is hoped.
- 3.18 FRANCE. Direction des Archives de France - Cité des Archives Contemporaines. It is apparent that considerable thought has been given to the problem both at the centre and in the Departments. Articles by Michel Duchien(3) and F. Burckard(4) indicate an attempt to reach a consensus of opinion on methodology by means of a questionnaire to the various archive repositories. The former reveals that of sixty-three replies, forty (63.4 per cent) favoured statistical sampling over other methods because, although random, it was not arbitrary and it was easy to apply. On the other hand it was criticised as being only useful for numerous homogeneous categories of record and that as a method it was 'blind' (aveugle), and also that a high sample was required in order to provide adequate evidence of the series.

- 3.19 The other method discussed was that of 'specimen groups or series' (fonds témoins) and in this case thirty-one replies supported the method and twenty-seven were opposed to it. The objections to this system were many. In particular it was considered to sacrifice local history interests for an illusory uniformity which could lead to arbitrary specialisation by repositories. The point was also made that where fonds témoins were used it was necessary for special information to be available regarding the holdings of repositories, perhaps by the dissemination of microfilm lists of such holdings.
- 3.20 Details were provided of three sampling exercises carried out at the Cité des Archives Contemporaines, each case being a form of systematic sampling rather than true random, statistical sampling.
- (1) Papiers of Santé-Travail (Health and Employment Insurance). In theory it was agreed that the records of one entire year in every thirteen would be kept as a whole and that the first year of deposit (1966) would be appropriate for this purpose. During the subsequent twelve years the papers of one month only would be retained, but that the month would change year by year, e.g. 1967 January, 1968, February, 1969 March, and so forth, but with a proviso that the year to be retained as a whole would be dependent upon any particular political or legislative change which had taken place. In the event La Direction des Archives de France ordered that because of the change of President in 1969, that year should be retained, so that January, 1966, February 1967 and March 1968 were kept along with the whole of 1969.
  - (2) Dossiers for Miners Social Security (Securité Sociale dans les mines). These records of the careers of individual miners have been sampled by keeping a ten per cent sample and also in retaining the key documents relating to each individual (la feuille recapitulative de renseignements).
  - (3) The records of three separate professions which have ceased to exist, viz. harness makers and saddlers, dyers - launderers, and cycle and car makers. In each case a tenth sample of the files is taken.
- 3.21 The methods reported are essentially systematic and are typical of the outlook in Europe. Some comment will follow regarding the mixing of a statistically acceptable chronological system with a measure of qualitative sampling such as occurs with the papers of Santé-Travail. It is important to note that, mainly within the constraints of more traditional methods, the French have probed deeply into the theory of sampling and the dangers inherent in some systems.
- 3.22 GERMAN FEDERAL REPUBLIC. Replies were received from the Bundesarchiv and also from Nordrhein - Westfalisches Hauptstaatarhiv and Stadt Köln Historisches Archiv. The general view expressed was that random or systematic methods were acceptable for material after 1945 in date, but that in the special circumstances of the Republic what could be done depended to a degree upon the relationship of the Federal Government to its constituent elements and it is apparent that local conditions are of significance in interpreting sampling experience.
- 3.23 Bundesarchiv, Coblenz. Three separate examples were provided:
- (1) Professional promotion files of soldiers (Beruforderungsakten von Soldaten). This is a class of 130,000 numbered files which is still growing. A small sample of 0.1 per cent (i.e. one file in each

thousand) is to be retained the size being determined by the fact that there is much duplication of information in other records so that a larger sample is unnecessary.

- (2) Civilian employees files needed to establish service benefits (Personalakten von Zivilbediensteten). In this instance the main series is not retained but those cases in which the Minister of the Interior has been concerned are kept. Selection is determined by reference to the peculiarities of each office and the files are therefore retained on a group basis for particular offices or for special salary categories. There is also a sampling of individual files selected on a time basis, i.e. if the time limits of the case so far as it applies to preservation in the administration runs through years which have the final digit 0 (1970, 1980, etc.). In these cases the files are then listed and are offered to the Bundesarchiv. A final sample will be based on number series or initial letter and will also take into consideration the special management features of the particular local office concerned.
- (3) Immediate Aid Action files (Soforthilfe-Leistungsakten der Lastenausgleichsverwaltung). These are papers concerning war-damage to property and a selection is made by the Administration concerned, the Federal Equalisation Administration, on the basis of one file for each type of action within established and constant criteria. As a result approximately 0.1 per cent of the class is retained.

- 3.24 Nordrhein-Westfälisches Hauptstaatsarchiv. Selection is made essentially upon qualitative grounds. Although it is claimed that the sampling is 'often at random' it appears that the methodology is either purposive or systematic rather than random in the statistical sense. 'We select, often at random, sometimes choosing case files which are especially characteristic of the administrative pattern, or on a duration basis, or those files which strike us because of the names of prominent people which occur'. Another usage is the selection of the records of a particular office or court which can be regarded as typical or the whole, e.g. 'there, too, we mostly use random methods, but are careful to choose an office or court whose authority does not have too small an extent'.
- 3.25 When individual files are sampled the methods used are by choice of years, or of months, or on the basis of the initial letter of a name, e.g. the letter H is a favourite choice. Other methods which have been employed include random spot checks on a fixed percentage basis and also selection on a regional basis, especially in the case of Dusseldorf which may be regarded as reasonably representative of the whole District.
- 3.26 Stadt Köln Historisches Archiv. Large series of uniform case files are sampled by retaining those for persons whose name begins with the letter H. This method provides about 8.5 per cent sample. It is a method recommended about twenty years ago as statistically valid and despite some objections has been adhered to ever since. It has the inherent weakness in Germany that names of Romance language origin do not commence with H, but this objection is partially overcome because the records of the Office of Aliens are retained in full. The method has the considerable advantage that it is manageable for the archives concerned, and practical in administrative terms. It is considered that a numerical system would be more complex to apply.

- 3.27 The somewhat limited use of more sophisticated methods is to be noted and also that (subject to difficulties of translation) the German usage of 'random' appears to be far from specific. Further comment will be made regarding alphabetical sampling, but it is noteworthy that there is a recognition that the use of one specific initial letter may have special local or linguistic significance.
- 3.28 UNITED KINGDOM. Information was received from the Public Record Office, the Scottish Record Office, the British Broadcasting Corporation, a working party of the Society of Archivists and the Kent County Archives Office. Methods vary considerably but the Public Record Office, in particular, is tending towards random techniques though still less sophisticated than Canada or the U.S.A.
- 3.29 Public Record Office, Kew. Sampling methods have been used from time to time simply to reduce the bulk of long series of papers 'with some research potential'. The method adopted has been 'random' but avoiding the use of random number tables, a percentage serial basis or a periodic time-series has been preferred. There is some concern regarding the outcome of sampling methods and while the comment 'any sample is better than none' has been received, a consensus view appears to be developing that where the retention of a time series is not an option, it would be preferable to keep the whole of fewer series rather than imperfect examples of many. Future policy in this field will depend upon the outcome of the Wilson Committee currently examining British policy regarding Public Records.(5)
- 3.30 Two valuable instruction leaflets were prepared in 1975 regarding 'Selection of documents for preservation' and the 'Sampling of records for statistical use'. The latter paper includes the comment, 'the taking of statistical samples does not preclude the selection for preservation of cases with an individual historical value provided the latter is done after the statistical sample has been drawn and the two selections are clearly differentiated' . No actual examples of sampling were submitted.(6)
- 3.31 Scottish Record Office, Edinburgh. Because of the nature of the holdings and the extent to which long series of administrative records are held by the Public Record Office rather than in Edinburgh, the problem has been less acute, but on occasions an ad hoc decision has been taken usually as the result of academic advice. In all appropriate cases, policy or precedent files have been retained and any sample, if taken, has been primarily to illustrate the working of policy decisions. The methods used are either the retention of census years or by means of selected territorial coverage, i.e. a regional selection. One class so treated are the probation service records where files relating to cases in representative areas are chosen so as to provide examples of all types of local background.
- 3.32 British Broadcasting Corporation. This organisation creates both conventional and non-conventional records but a report was only received regarding written records and archives. Pre-1955 records are held at the Written Archives Centre, Caversham Park, Reading, and more recent non-current papers at a records management centre in London which is still evolving procedures. Nevertheless, the pre-1955 papers have been examined both from the point of view of general appraisal and from that of sampling with the following results so far as sampling is concerned:
- (a) Policy papers are regarded as unsuitable for sampling;

- (b) Routine administrative papers are not sampled, they are normally retained;
  - (c) Programme and constituent files 'appeared to be the nearest equivalent to case files, but on closer investigation it became obvious that some programmed and some people were more important than others ...' As a result programme files are retained on the basis of merit, i.e. 'if the content is controversial . . or of potential historical or social interest, if it illustrates the use of a new technology or equipment, then the file should be kept. In the case of contributors it is felt that sampling is impossible though some measure of weeding or stripping may be necessary.
- 3.33 The comment is made that for the B.B.C. 'any kind of statistical or random sampling has no advantage and indeed may have positive disadvantages' . Either the whole series must be kept or the whole may be destroyed or other internal factors make sampling inappropriate. The Corporation is aware that the methods employed 'contain a large element of subjective choice' , but given the unique nature of the material involved selection on an individual file basis appears the best solution.
- 3.34 This example of procedure is of interest for three reasons: (a) it is an example of a very specialist area; (b) it is a good example of a case in which the only basis for sampling, where that applies at all, is qualitative and therefore the only example quoted is of purposive sampling; and (c) it is an excellent sample of the problem of non-homogeneous series which in many respects are more comparable to non-conventional records than to the traditional series of particular instance papers of government departments.
- 3.35 Society of Archivists Working Party on Local Government Retention Schedules. This body of archivists from representative local authority offices has been meeting regularly over the past three years. It has considered in detail the application of all sampling methods as well as of microphotography to the records of Social Services Departments, comprising many series continuously growing, but where there may be considerable variation in terms of record keeping between different authorities. Their considered comment is as follows:
- (a) The taking of examples - with series which have no financial implications and are of slight continuing interest the records should be reviewed annually and one file retained as a record of the less significant activities of the department.
  - (b) Purposive sampling - a useful expedient for files which reflect an interesting policy, but the files to be retained need to be identified by the administration so that they may be kept in addition to any main sample.
  - (c) Systematic sampling - a useful method and easy to apply: e.g. the choice of every fifth or tenth file or the selection of the records of one or two years in every decade. There is an adverse comment that long-term trends may be ignored by the use of this technique.

- (d) Random, statistical, sampling - in general regarded as too complex and difficult to apply.
- (e) Microphotography as an alternative to sampling - although a few authorities already microfilm social service records, in general the costs of the hardware involved and the uncertain duration of the life of the film must be seriously considered. If, however, the administration accepts microphotography as a valid and cost effective method of handling information, then it can prove a useful alternative to sampling.
- 3.36 Kent County Council Archives Office, Maidstone. Various methods have been used over the past twenty years especially forms of purposive and systematic sampling. Recently, however, a working party has been set up once again to consider the records of the Social Services Department. A careful analysis of the many series of files has been carried out and where appropriate it was agreed, in January 1980, that the Department itself would take a five or ten per cent sample at the end of the administrative life of the files, legal considerations having been also taken into consideration. In some cases a purposive sample would also be taken by the identification of cases of particular significance before transfer to the Archives Office.
- 3.37 In addition, however, other experimental work has been carried out and the Archives Office with the agreement of the Department has undertaken a threefold examination of 3,747 case files concerning children born between 1948 and 1965. This process entails (a) the taking of a one per cent systematic sample by the retention of files numbered 47, 147, 247, etc.; (b) a purposive sample of 'interesting or controversial cases', including those of sudden death; (c) the retention of key documents containing basic information regarding individuals (usually this action entails the removal of one or two documents from each file so treated). This exercise is still in progress but although no decisions on its efficacy can as yet be taken, it is already considered that the weeding or stripping element is too time consuming with massive series to be processed.
- 3.38 There is yet one other element affecting the records of the same Department. The files referred to above are those held originally by the central office of the Department. Because of the structure of English local government there are also a number of Divisional Offices within the county and the retention of the records of one or more of these 'local offices' is also under consideration, though the problem of duplication between central and local offices has to be taken into account.
- 3.39 The complexity of the British experience reflects not merely the relative ease of obtaining information, but also the nature of archive services within the country. It is of special interest to consider the attitude of the British Broadcasting Corporation with what one might term a specialist repository, where a qualitative approach appears to be more appropriate than any other system and also that the question of micro-photography as an alternative to sampling has been given some consideration. The very interesting comment from the Public Record Office that the retention of fewer complete series rather than unsatisfactory samples of many should particularly be noted.
- 3.40 UNITED STATES OF AMERICA. National Archives and Records Service. Over the years N.A.R.S. has used both qualitative and quantitative methods of sampling and has tended to favour systematic techniques in which every tenth or hundredth, etc., case is selected according to a fixed numerical system. Qualitative



criteria have included cases referred to or discussed in the Annual Reports of specific agencies; cases involving extensive litigation; cases setting major legal or policy precedents; cases receiving marked attention by the mass media; cases involving significant Congressional interest or investigation; and cases regarded by authorities in that particular field of study as being of some special importance or in some way unique.

- 3.41 The most frequent cause of difficulty has been the fact that sampling instructions are usually incorporated in those duties to be carried out before transfer to the archive service. Agencies, although given instructions have failed to implement these or, on occasions, have failed 'to do an adequate job'. There has also been some staff resistance within agencies in this respect.
- 3.42 An important survey of methods was carried out by R.M. McReynolds in 1975, (7) in which systematic systems based on a numerical pattern and true random sampling are preferred. The author makes the point, however, that not all records can be successfully sampled, only those which are essentially homogeneous in character are really well adapted to sampling methods.
- 3.43 Other experience within the U.S.A. Sampling has been the subject of considerable debate and study within the United States and, in particular, three recent articles published in the American Archivist are worthy of note as illustrating different methodologies.

- (1) Michigan Historical Collections.(8) The case considered was that of a large collection of archives of Hannah Lay & Co. a lumbering concern. It was agreed that minutes, correspondence and special reports should be retained, but these only accounted for a very small part of the whole. Financial records, which were the major problem were sampled as follows: (a) by the retention of selected years, e.g. the census years between 1860 and 1930 plus the mid-decade years; and (b) because of the reorganisation of the industry, the years 1882-4 were retained as a whole. Although the method used is referred to as being quantitative it would appear from the article in question that the records were selected from those areas which the persons determining the process could regard as providing a good sample. Problems arose because of the time overlap found to occur in the case of volumes and because ledgers based on personal accounts were also difficult to relate to time-series. It was therefore decided that such documents should be microfilmed in preference to sampling.
- (2) Random Sampling of Congressional Papers in the Mid-West.(9) This exercise was the outcome of local discussions and the technique was designed with a view to sampling the papers of individual Congressmen found in the State repositories in the Mid-West, in such a manner that sufficient material would be retained for the use of scholars; to present a true picture of the function of a typical Congressman's office; and to serve various unspecified avenues of research, including possible future interests not yet established. After a wide discussion both inter-professional and with interested academics an experiment was carried out in Wisconsin and a twenty per cent sample of certain large homogeneous series was attempted. The size of the sample was determined on the basis that scholars normally require a sample of two to three per cent for their purposes and that this relatively large sample would provide wide scope for a number of independent projects. The basic

problem met was to establish criteria for the initial appraisal which would decide that a series was worthy of sampling. As a precaution the residual eighty per cent of the whole was retained initially so that a qualitative assessment could be made and any material considered significant kept. Ideally it was felt that this should not be necessary, and that if the twenty per cent sample proved satisfactory this second stage could be stopped.

(3) National Archives and Records Service. Public Welfare Case Records.

In handling this vast class of records it was determined that only sampling and microphotography were possible ways of controlling the bulk. It was also considered that while sampling 'reduces the possibility of linking case records', microfilm was only practicable where there were sufficient resources available. At the same time microphotography, combined with computer technology, would provide the best solution of all and would also make it easy to maintain a documentary anonymity. The greatest foreseeable problem in this kind of exercise would be the availability of suitable expertise.

- 3.44 Recently there has been further significant work done in the United States both locally and nationally. An important publication was issued in 1980 regarding the appraisal and sampling of the files of the Massachusetts Superior Court, 1859-1959; and the National Archives and Records Service has received a court order to examine the records of the Federal Bureau of Investigation.(n) Arising from this latter case, however, an important principle has been enunciated that it is desirable to select two samples: (a) a statistically valid sample; and (b) 'an exceptional case or purposive sample to catch the historically significant case files' .
- 3.45 While the information provided from these various sources is far from determining methods or giving an even coverage of archival practice, it does underline the very mixed methods in use at the present time and the uncertainty which still exists among archivists. All methods in use appear to arouse a certain anxiety among the users and while qualitative and quantitative methods have been adopted in various places and at various times, no one is wholly happy about the validity of the former nor entirely satisfied with the complete authenticity of the latter. While it would appear that the large national repositories are becoming more willing to utilise statistical methods, there is still a large body of opinion which in practice favours systematic sampling, largely because of the relative simplicity of use. It is also noteworthy that wherever it has been possible to utilise resources incorporating microfilm and E.D.P., these are found to provide a useful alternative to normal sampling techniques combined with a highly flexible information retrieval system which itself provides the availability of samples of information without consequent destruction of the residue. On the other hand, as has been observed, microphotography is probably 'too costly a procedure for many state archives', though it 'may still represent a feasible solution to the problem of bulk' .(12) This is the case especially when it is undertaken by the creating agency as an administrative record keeping procedure.
- 3.46 It cannot be said at this present time that the information received offers more than an indication of the varied opinion and practice to be found in Europe, America and Australia, though it does show some resistance towards new methods and a positive need to find solutions to the all pervading problem of bulk, especially where extensive case records are involved. Appraisal by individual

pieces is prohibitive in this area because of bulk in terms of numbers of files and bulk in terms of storage. The alternative of indiscriminate destruction, even though the archivist can only see a very minimal information content, is generally unacceptable both to research and to the custodian alike. The archivist is therefore forced to determine whether to retain part of the whole, and, if so, how? Before considering further, however, the theory and philosophy of this matter, some consideration of appraisal and sampling as they affect non-traditional archives is needful.

- 3.47 In the original letter of enquiry it was specifically asked whether there had been any noticeable searcher reaction to methods employed, but essentially this question elicited no replies. What was forthcoming, however, was a comment from Australia and from Westphalia that the systems were too recent in application for a reaction to be recorded. It was also apparent that there was a realisation that the attitude of academics to sampling programmed was important and that consultation was a valuable element in the programme. This point was specially made by Canada and by the Scottish Record Office. Finally there is the viewpoint expressed to the Public Record Office, Kew, that 'any sample is better than none'. This reflects a very real concern among academics that destruction, even of ephemera, may result in the loss of significant information and the growing fear that the conservative view of administrative criteria applied by the administrator will be detrimental to the validity and comprehensiveness of the surviving archives. The pressure would appear to be in two directions: (a) for the retention of more classes or series as a whole; and (b) if sampling must take place then quantitative needs should take preference over qualitative interests and this is, of course, essentially the position taken by the Canadian Public Archives.

#### Notes

1. Australian Admin. Instruction 1977/3. Procedure for the Destruction/Review of Time-expired Accessions.
2. A longitudinal sample is one relating to a particular function selected numerically or chronologically on a continuing basis.
3. La Gazette des archives, No.77 - 2e trimestre, 1972, PP.82-3
4. La Gazette des archives, NO.108 - 1er trimestre, 1980 - p.12
- 5\* Modern Public Records (Cmnd. 8204), H.M.S.O. 1981
6. RAD Occasional Papers, Nos. 4 & 5, P.R.O., 1975. It may also be noted that the Wellcome Foundation is currently funding the design and execution of a random sample of the records of the Department of Health and Social Security.
7. R. Michael McReynolds, 'Statistical Sampling of Archives', NARS, Nov. 1975
8. Larry Steck and Francis Blouin, 'Hannah Lay & Co.: Sampling Records of a Century of Lumbering in Michigan', American Archivist, 39 No.1, Jan.1976, pp.15-2P.

9. Eleanor McKay, ' Random Sampling Techniques: A method of reducing large, homogeneous series of Congressional Papers', American Archivist, 41 No.3, July 1978, pp. 181-9
10. R. Joseph Anderson, Public Welfare Case Records: A Study of Archival Practices', American Archivist, 43 No.2, Spring 1980, pp.169-179
11. M.S. Hindus et. al., The Files of the Massachusetts Superior Court, 1859-1959: An analysis and a plan of action. I am indebted to Mr. R.M. McReynolds for this information which was received during the final revision of the text.
12. Anderson, op. cit., p.176

4. The Sampling of lion-conventional Records: Cartographic. Audio-Visual and Machine-readable Records
- 4.1 The terms of reference for this study included not only those traditional categories of archives and other records which have already been considered but whether sampling procedures were, or could be, applied to (a) cartographic records; (b) audio-visual records; and (c) machine-readable records. While it has proved much more difficult to obtain practical evidence of sampling usages in these fields than with textual or conventional records, it is important that each of these special categories be examined. From the limited evidence submitted it would appear that there is a fundamental difference of approach to appraisal in the first instance and that in all but a few very special cases this may automatically rule out the use of those sampling methods beloved of statisticians, though occasionally other methods may be appropriate. Two additional factors which have to be taken into account are that in some very specialised fields the custodians have not really come to terms with archival theory as it is usually accepted; they regard themselves more in the profession of curator than archivist and this colours the attitude towards methodology. The other fact is that the newest technology is only to be found in the largest of repositories and that only recently have the people concerned really been faced with a problem which may lead to the adoption of sampling techniques.
- 4.2 A reply received from the Imperial War Museum, London, identifies the problem. It reads: 'In this Department (of Documents), as in the Department of Film, Photographic and Sound Records, any sampling is done at the pre-acquisition stage, that is by selecting or rejecting material on offer according to its relevance to our field of interest ...' Very much the same attitude is shown by the British Architectural Library in respect of the acquisition of architects' drawings and papers and the Public Record office stated: 'It is difficult to see how one can select still and motion pictures, sound recordings or cartographic records other than on an item by item basis'.
- 4.3 In effect the key issue is the nature of the informational element in the items falling within these categories and the lack of homogeneity comparable with that found in series of case files and other material sampled in the traditional archive area. In these circumstances it is difficult to understand how quantitative methods, such as random statistical sampling can be applied, though clearly it is possible to carry out a process comparable to that involved in purposive sampling. Once again there is a distinction between initial appraisal of material to determine whether it is potentially significant for future reference only and the determination that only part of what has been so appraised shall be kept. In other words the informational content is more significant than the evidential content. A body such as the Imperial War Museum will assess the significance of a series of photographs, slides or films on a qualitative basis and will discard unwanted or duplicate material in advance of formal receipt - it will not, so far, assess those series subsequently with a view to retaining only part of what was initially accepted.
- 4.4 In this respect selection and sampling are almost coincidental, though certainly the sampling element bears little or no relationship to the concept of statistical evaluation. The U.S. records service 'sampled' both still and motion picture files in this fashion(1) and the Scottish Record Office, which has received massive deposits of engineering records, has described the problem of the appraisal of engineering drawings in the following terms:

'The main problems are the sheer bulk of these collections, with drawings counted in hundreds of thousands, and the physical form of the records . . . They do not conveniently fit on to standard archive shelving nor, for that matter, do they conform to the normal criteria for selection. In the case of shipbuilding . . . where construction is more or less on a "one off" basis, there will generally be a complete set of drawings in respect of every contract if not in fact every vessel built . . . The aim must be to preserve permanently the principal drawings - the "as fitted" key drawings wherever possible . . . Where a product is mass produced . . . then it is not unreasonable to be required to keep a complete set of drawings for every model made . . . But where a range of similar products is made . . . these may be assembled from information selected from a "bank" of drawings . . . There is then no alternative to preserving the "bank" of drawings completely and, for each order also a list of drawings used.' (2)

It will be noticed in this detailed description that although there is a qualitative element at work, nevertheless this cannot be said to be true sampling. It does, however, illustrate the 'grey area' which exists between primary selection and any subsequent sampling activity.

- 4.5 Cartographic records (maps, plans, architectural drawings, etc.) In general terms, maps by their very nature are unsuitable for sampling. Except that duplication can occur or photographic enlargement or reduction, each map (just as any other record) will be individual in that it contains some specific and unique information. So, too, maps will either be found in horizontal series, i.e. all relating to the same point in time, but to a series of territorial areas; or they will be in a vertical time series, usually formed artificially, but concerning one area and illustrating changes at different times. Normally either series type will provide a single and unique informational study which could be retained as a whole or discarded as irrelevant, but for which sampling would create a still greater irrelevance because of the destruction of any evidential or informational linkage which might originally be present. It is possible with the horizontal series to limit retention to those maps relating to a particular geographical area and to discard the residue or to offer it to another repository for which it has greater relevance, but such action would in no way produce a sample acceptable under any of the headings used.
- 4.6 There are, however, some special cases worthy of consideration where large series of maps are being constantly produced rather as case files are continuously created. Such an example is provided by the technical records produced by and for the British Meteorological Office. The Central Forecasting Office creates or receives no fewer than one hundred and six series of hand-drawn or computer produced charts. Of these series, twenty-one are assessed as permanent archives and three other series are subject to a special review after ten years. All the others are accepted as expendable and are scheduled for destruction on a time scale varying from two months to twenty-five years. In four of these cases, however, the records produced for one year in ten are to be retained as a sample. In fact, twice in the regulations governing retention, a form of systematic sampling is advised: once in that a copy of every chart used in the Central Forecasting Office on 15 January each year is to be retained and in the second case (the four series already noted) the regulations state that copies of 'surface and upper air prognostic charts' for 1960 and for each tenth year subsequently are to be retained permanently.(3)

- 4.7 Although these series of charts might superficially be regarded as comparable to case files, no particular advantage would accrue from the application of statistical methods in the form of true random samples and a form of time series sampling is the only method which will produce an acceptable result. It is also to be noted that so few of the many series created are even considered for this kind of treatment, even with the four categories which are to be sampled the residue is to be destroyed after the relatively short period of five years, an indication that possible research use rather than likely administrative reference was the criterion behind the decision.
- 4.8 Two other special cases refer to the problem arising when a series of maps and plans has been microfilmed and consideration is given to the destruction of the originals. The case may be cited of the architectural record drawings retained by the County Architect of Kent, England. Here was a case in which there was administrative and cost advantage to the authority for the 28,000 plans and drawings to be filmed, but these documents covered all the buildings for which the County Council had been responsible. The Archives Office, therefore, on being consulted, sought two provisos: (a) that the whole series be transferred to the Archives Office and retained as a whole unit the micro-filming had been successfully carried out; (b) that the two Departments co-operate in determining what selection, if any, might be made of significant drawings. In the event, the value of the first of these provisos was proved when serious difficulties arose with some of the filming process, while in the second case it was decided (1) that all site plans of major projects (schools, police houses, etc.) should be retained; and (2) that sets of plans should be kept for buildings of special architectural merit or those which illustrated the first of a series of buildings indicating a major change in architectural policy. It should, of course, be remembered that the official record remained intact, but in a different medium, though at the same time a qualitative sample was regarded as having a certain intrinsic validity and also as providing the searcher with some evidence in a form more easy to handle.
- 4.9 A second case of a somewhat similar kind relates to the records of a small former Rural District Council in south-east England where a policy of micro-filming all planning applications was determined. After the changes resulting from the Local Government Act of 1972, (4) the original papers for the period 1947-74 which it was intended should be destroyed, but which in fact survived, were unwanted by the successor authority because it had the microfilms. On consideration both a systematic and a purposive sample were taken, the first on the basis of application numbers (50,100,150, etc.) and the second on a pattern of specific criteria including: important buildings or those listed as historically significant by the Department of the Environment; historic or conservation areas; major or controversial developments; churches, schools and inns; a small cross-section of housing types; agricultural developments and industrial or commercial developments. Whatever the validity of this selection it may, perhaps, be regarded more as an example of a regional or representative sample than as a means of handling cartographic records as such.
- 4.10 Audio-visual records  
Many of the problems affecting cartographic records also apply in the case of audio-visual records. Decisions will be taken by recipient agencies on the basis of the relevance of such records to their holdings. Occasionally legal or evidential value may reside in the material, more often, perhaps it will have a hi@ informational content.(b) In these circumstances although it may be possible to appraise photograph by photograph or film by film, discarding those which add little or nothing to the sum of knowledge, there appears to

be no way in which statistical methodology could or should apply and, unless a series of illustrations of the same place were recorded at set intervals of time, even a systematic sample would appear to be of little substance. Purposive sampling, however, is a possibility with audio-visual records especially if the decision is taken essentially on the relevance of the information contained in each picture for the subject to which it relates. It must not be overlooked that such material is not of necessity an isolated class. It can and often does, form part of the corpus of a traditional archival group and, as such, will be treated as an integral part of that group, though there may, occasionally, be an overriding case for its retention although the other related papers are destroyed. The comments here, therefore, are more particularly relevant for specialist collections of illustrative records not directly associated with other forms of documentation.

- 4.11 Motion pictures provide similar problems and appraisal can only properly be taken on the basis of the entire film or on the entire product of a copying agency, or of a year's deposit under copyright and not on the value of individual frames. It would surely be improper for the archivist or custodian to tamper with the content of the film, though a decision to retain or destroy could properly be taken. Once again it is most difficult to see how any kind of statistical or systematic sampling could be applied, except in so far as a series of similar films on the same topic might hypothetically be taken over a period of years and a later decision also be agreed that only those for every xth year would be retained.
- 4.12 The British Film Institute: National Film Archive determines selection by means of specialist committees. They do not sample, but the comment is made as regards an early assessment that not only is it very difficult for any type of material, but that 'in the case of film and television, and particularly the latter, if you don't make such an assessment quickly the material disappears.'
- 4.13 Sound archives present a still more complex problem, not least in view of the ease with which tapes may be scrubbed of content and then re-used, a factor of economics which is not to be ignored. It is also true that, at least in the United Kingdom, such records are not of necessity handed over to any archival agency and therefore their survival, and indeed appraisal, is less controlled at the present time. In archival terms, their appraisal resides essentially in the informational content of each tape, and although questions of bulk storage will arise, in comparison with traditional records sound archives are relatively less bulky and therefore, on that count alone, easier to provide for. As with a series of large paper files, decisions can be taken on the basis of file or tape titles and in that sense, in theory, a sampling technique could be applied. It is doubtful, however, whether any advantage would accrue - each tape must be assessed, aurally, if its real worth is to be discovered and each will, therefore, stand on its own merits.
- 4.14 The Public Record Office reports that it considered carefully the retention of tapes of court proceedings, but determined otherwise on the grounds that copies would be available elsewhere. It was also thought that perhaps the tape for the final day of each law-term might be worth preservation, but it was realised that 'we should be preserving nothing more than the flavour of a court hearing - and only part of a trial case anyway . . .' The report adds that to have selected on the basis of the intrinsic interest of individual tapes was out of the question, for the tapes contained 30 tracks covering up to 50 courts a day. A selection of tapes on the basis of informational value was made from those prepared by the British Central Office of Information and its predecessors for



radio broadcasting for governmental purposes. In this assessment both lists of titles and subjects were examined and tapes were examined by hearing.

- 4.15 Once again the individual item is the key element in the appraisal pattern and it is doubtful whether a series of tapes could be effectively assessed in any other manner, or whether the retention of every nth tape or of a random sample of tapes could have any virtue whatsoever. It could rightly be determined that the content of a tape was trivial and therefore unworthy of retention, but that would have little or no bearing on the content of any other tape even in a comparable series and no evidence of sampling procedures for such records has been forthcoming.
- 4.16 Machine-readable records  
The dilemma posed throughout this chapter has been the selection of once-off items and the basic question of whether destruction in such circumstances can be offset to a degree by the application of sampling methods is shown still more clearly by a consideration of machine-readable records. That computer files represent a very special case was indicated in 1976.(6) The report on 'Appraisal' for the 8th I.C.A. Congress reads that 'it seems to be a unanimous opinion that, in principle, these types of records should be subject to the same regulations which govern conventional material' and then adds, 'in practice, however, a certain amount of special treatment for machine-readable records seems to be unavoidable. The information retained by a magnetic tape is not completely comparable with that held by conventional records' .
- 4.17 In the United Kingdom, the Public Record Office normally reviews tapes 'file by file' on the basis of their individual value for statistical, scientific, historical or other research purposes and, in line with the general argument already established here, having so determined on retention, they 'would not consider it worthwhile reducing . . . size by keeping only a sample of the data . . . although of course, many files, especially those relating to surveys, contain only sample data' . This is very similar to the position already suggested for sound archives, namely that the tape must be appraised as a unit and that a later, secondary, selection process involving sampling is purposeless.
- 4.18 Nevertheless, the Public Record Office can identify three circumstances in which the files selected would, themselves, constitute a sample. These are:
- (1) sample files from a longtime-series of large files (e.g. from a series of monthly files extending over several years in which monthly variations were not significant);
  - (2) sample files from a sequential off-line data base, subject to periodic updating over a long period;
  - (3) a sequential file as a sample taken at one period of time as a special exercise from a non-sequential on-line data base, subject to continuous updating over a long period' .

In fact, it appears that so far the P.R.O. has only been faced with the circumstances of example (2), so that to a degree the categories are hypothetical.

- 4.19 Another British example of a special kind is, again, the Meteorological Office(7) which, apart from those charts referred to already, creates or receives seventy-four categories of machine-readable records. The great

majority of these are classed as of permanent value and in no case is anything like sampling advised, though there are one or two cases where the data so received will be transferred to an archive data set. For example, one of the instructions reads: 'Assistant Directors should consider carefully whether any experimental data may have general value, over and above that for the project for which it was obtained. If so, it should be carefully documented and kept for an appropriate period'. This, of course, is not samplings such, though it relates closely to the kind of qualitative assessment which may be adopted with purposive sampling.

- 4.20 The views expressed by the British Public Record Office are not very different from those of the National Archives and Record Service of the U.S.A. It points out that, in technical terms, it would be very easy to take a sample of machine-readable records but doubts the value of the exercise, since random sampling in this context 'seldom reflect the real records/information world, since in most instances the distribution of the records, say, in a case file (sc. machine-readable) is arbitrary' . It suggests, therefore, that a stratified sample, i.e. one covering such matters as sex and marital status, would be required, but considers that the cost of the operation would be prohibitive and unjustified. Nevertheless it identifies three distinct types of computer file and comments on each in relation to sampling. In the case of the one-time study it suggests that 'it would be more cost-effective to forget about sampling and save the entire file' - a comment which is in line with the general argument regarding unit appraisal and selection, the burden of much of this chapter. In the case of non-cumulative, up-dated, files, which, because of their frequent up-dating, provide a snapshot effect, it is suggested that an annual or other time-interval copying might be carried out so that, in time, an historical file would be created. In the third case, it refers to recurring (non-mathematical) files, such as master tax-payer lists, which again might be retained on a the-series basis.
- 4.21 An important article, written in 1978, on 'Appraising Machine-Readable Records'(8), stresses in particular the significant changes which are likely to affect computer technology in the immediate future. Still more relevant for this study, it outlines the various appraisal decisions which must be taken before a computer file can be accepted for retention by the U.S. National Archive and Records Service, identifying both the technical and the archival considerations which must be borne in mind. Although these headings present a valuable technique which could with Justice be adapted for conventional archives and records, they leave little room for anything approaching a sampling method and, in view of the complex nature of computer files, it is very difficult to see how any system remotely approaching a true random sample could be satisfactorily applied.
- 4.22 Nevertheless, that such is a possible and apparently feasible and cost effective practice is suggested by the report received from Canada(9), which refers to the Unemployment Insurance Commission as not only holding a computer file but also conducting 'a continuous longitudinal E.D.P. sample of benefit claims' . The method adopted based on the Social Insurance Number with a terminal digit 5 has already been discussed. The same report also refers to periodical analysis of this data by Revenue Canada and Statistics Canada.
- 4.23 It is of interest to note that it is the administrative agency and not the archives which has instituted the measure of sampling, hoping thereby to establish a satisfactory data base which will be valid for administrative, statis-

tical and research purposes. Its validity in this last respect is obvious once the computerised sample becomes open to examination, and as the report of the Public Archives of Canada states, 'in the case of the U.I.C. benefit claim files we will pull a very minute continuous sample of paper files as there is a much larger and more easily researched data base in E.D.P. records' .

- 4.24 In a sense, however, this merely confuses the issue. A searchable data base provides a vast area within which research samples can be taken, but to state that fact does nothing to justify, or otherwise, the actual sampling of E.D.P. records. This must remain an uncertain project, especially as the only direct example appears to be one taken by the agency itself and not the archivist. If the administration can Justify the cost of sampling its computer files, then the archivist may benefit from that process, but he is unlikely to apply sampling techniques to machine-readable records which have been closely appraised before being transferred to archival storage.
- 4.25 In sum the applicability of sophisticated sampling techniques to non-traditional archives must remain 'not proven' . They form record material which is suitable for unit appraisal in most cases, and it is only in very exceptional circumstances that a systematic or qualitative sample can be taken and that that decision can be regarded as sound in methodological terms.

#### Notes

1. Information supplied by F.B. Evans
2. Information supplied by J.H. Sime of the Scottish Record Office
3. Report of Meteorological Office Data Review Board, 1980 - Annex 7C: Retention of Technical Records.
4. Eliz. II, 1972, c.70
5. Walter Rundell, junr., 'Photographs as Historical Evidence - Early Texan Oil' , American Archivist, No.41, No.4, Oct. 1978, pp. 373-98.
6. A. Kromnow, 'The Appraisal of Contemporary Records' , I.C.A., 1976
7. Meteorological Office, op.cit.
8. Charles M. Dollar, 'Appraising Machine-Readable Records', American Archivist, vol. 41, No.4, Oct. 1978, pp.423-30
9. See PP.32-3 for details

## 5. Problems and Alternatives to Sampling

- 5.1 It has already been suggested that all selection is a form of sampling, that whenever we determine what shall be preserved and what destroyed we are inevitably retaining less than the whole and that therefore a form of sampling is taking place. There is, however, a clear distinction between appraisal as an initial activity and appraisal as a final stage in the management process, as there also is between retention of records on the basis of their intrinsic qualities and the keeping of a representative part of the whole. Any agency will maintain a variety of classes of records - those regarding policy making or, perhaps, legislation; those of a fiscal character be they budgetary or house-keeping in nature; those concerning personnel; and those of an executive nature indicating the implementation of policy decisions. At an early stage in the life of these records decisions will be taken which can be codified in a disposal or retention schedule. Destruction recommendations at this level are likely to, indeed should, relate to entire series of records within the overall class structure and decisions will be taken according to the opinion of those concerned, be they administrators, management experts, archivists or academics as to the significance of the series in (a) the continuing needs of the agency itself; (b) their relationship to the work and development of the agency, i.e. its history as a vehicle of government; and (c) any wider informational value which the records may be deemed to contain.
- 5.2 This is of course a perfectly proper element of decision taking and appraisal and it will reduce the number of series to be retained. It therefore also assists in the reduction of bulk for storage and it is primarily for this reason that the process was initiated. As a result, however, the archives of the agency will be that much less than complete for future purposes, one can only say that considered opinion indicates that despite the destruction proposed, no serious damage has been done to those archives as a whole nor to the needs of scholarship. This process, although incorporating a key element of appraisal is nevertheless in no sense true sampling; it is dealing with whole series and is decision taking at the highest level so far as the records themselves are concerned.
- 5.3 In contrast to this, if the surviving series are then reconsidered in greater detail and a determination is made to retain less than the whole of any particular series, then, that process does involve sampling. It has to be asked therefore why such a course should be taken in the first place, to whose benefit it is taken and whether the outcome will prove to provide what was anticipated? It further needs to be asked whether the exercise was in any event worthwhile and whether there was not an acceptable alternative.
- 5.4 The archivist is properly alarmed by destruction, his training has been geared to the preservation of the record, not to its elimination and there must therefore be some overriding consideration at work if material of potential value is not to be retained. Forms of appraisal will take place at all stages of the life of records and in this way the significance of the individual piece, document or item, quite as much as the composite file will have been determined in respect of its value to the administration it serves and the actual purpose for which it was created. Thus much documentation, including some drafts, casual notes and memoranda will be cast into the wastepaper basket by the officer who wrote or received them almost immediately after use. Problems of the morality of destruction do not arise unless the material is being kept in toto and in steadily increasing quantities or, where a third party becomes interested in the question of preservation, probably for reasons remote from those for which the papers in question were created. It is therefore apparent that problems of bulk and cost of storage and also of the potential use of the records

for research purposes may play a determining role in decisions including those recommending sampling. It follows, too, that where a series is being retained for legal or fiscal reasons for a period of years, after which destruction will be automatic, no problem will arise unless someone other than the administrator or the archivist questions the validity of that decision. Sampling only arises as a practical exercise as we face the problem of what we can afford to keep in perpetuity on an economic basis, and whether the residue to be kept will provide valid acceptable evidence or information for purposes of research. This matter is made more acute by the realisation that classes of documentation, especially case files of all descriptions, form much of the essential research base for social scientists and for all scholars using quantitative methods of analysis.

- 5.6 It is at this stage that the archivist is faced with the question whether the series to be appraised is to be retained as a whole for evidential or informational purposes; whether the anticipated use of the series will justify the cost of storage - especially if the series is a continuing one and an annual growth must be taken into account; and whether, especially if the answer to the latter question is negative, it will be none the less possible to retain the essence of the series by means of a sampling technique.
- 5.7 It is clear from the evidence submitted that there is a marked difference of approach between the Old World and the New in this respect. It is not merely that the level of sophistication varies, but that there may well be a basic distinction in archival approach. The archivist's difficulty lies in his knowledge and understanding of the written history of his own particular society and culture. For example, no United Kingdom local archivist can afford to ignore the significance of those incomplete survivals of individual case records in local courts from, say, the seventeenth century, and this experience inevitably colours his approach to comparable records of today. Although aware of the overwhelming problems of bulk and storage costs, he is nonetheless deeply concerned with the content of the individual file. Where such relatively early papers do not form such an important element in the cultural background it may possibly be that much easier to accept the pure statistical sample as the proper answer to all difficulties. On the other hand, it is also apparent that the present approach in the U.S.A. and Canada is the outcome of a much more rapid acceptance of modern research methodology than has been the case in Europe. Thus the interest in the Old World examples rests in the varied attempts to provide systematic samples which can be regarded, and, locally, are regarded, as statistically valid and correct.(1)
- 5.8 This, however, highlights the problem that the varying solutions are themselves exceedingly local, not only in operation but also in applicability. For example the method adopted in Köln for the sampling of social administration records is the retention of all files for persons whose names begin with the letter H. This method it is reported provides an 8.5 per cent sample and was recommended as statistically valid. Nevertheless comment is made of the objections raised as to this method, not least that it succeeds in eliminating nearly all names of Romance origin from the sample. After receipt of this report an experiment along the same lines was tried for the files of the Social Services Department of the Kent County Council, England. If the letter H was chosen it was found that a six per cent sample would result, but that this would entirely omit individuals of West Indian origin, a particularly important element within the series in question. To have selected the letter A or the letter O which in either case would include a significant West

Indian element would result in either a fourteen or a twelve per cent sample. Something of the same kind was apparent in Canada where it was found that if immigrants from Vietnam were to be included in a sample based on alphabetical distribution, then the letter N would have to be used. Despite the apparent validity of the initial choice for the locality in question it would appear that the resultant sample was less than wholly valid, that as a method it could not be easily transplanted and that in another context it might easily omit essential informational factors. This in fact underlines the archivist's dilemma in respect of all sampling methods - namely that he is trying to provide an objective corpus of information for the researcher who needs to use statistical analysis, while at the same time retaining elements of special information which he knows to be present within the record series.

- 5.9 Put another way, a sample in order to be acceptable must embrace the essence of the whole - it must not be biased, yet it must contain all key elements of evidence or information known to be incorporated in the corpus of the series of class. It is therefore apparent that there are at least two conflicting concepts at work in the attempt to achieve a satisfactory sample. The scholar, anxious to use material ( usually of modest informational value), may well consider that statistical analysis will provide an answer or a lead in respect of questions he is posing. If the series is considered unworthy of preservation as a whole, then by the use of random methods that scholar can obtain answers to his questions comparable to those which he would receive if the series were retained and he, in his research, took samples of information from the whole series. On the other hand the archivist may well be aware that even among the dross of the discard of a repetitive or relatively insignificant series, there may lie items of supreme importance for some specific projects, but he is also conscious that if those items are then removed and added to the sample the validity of that sample and its random quality are affected and its use for quantitative methods essentially destroyed.
- 5.10 This example is not altogether unrealistic; it is an actual dilemma which has faced archivists from time to time. The case of the Registrar General of Shipping in the United Kingdom illustrates the problem. The crew lists and cognate records for the period 1865 to 1913 were initially deemed to be unworthy of retention as a whole on grounds of bulk and of content. After consultation it was agreed to retain a 10 per cent random sample, statistically acceptable.(2) Subsequently, however, and following various representations, a 10 per cent systematic sample based on census years was also taken. Then a quite different factor entered the picture and the records of certain important vessels were offered to the shipping companies concerned and the local archivists of maritime counties were permitted to salvage papers relating to their own localities. This example is not given in any sense of criticism but purely because it stresses the dilemma of retaining the general picture as opposed to retaining features relating to special cases. Archivaly speaking those papers retrieved by local archivists in this case can have little true validity; they are not even comprehensive in a purely local sense, they simply represent the haphazard survival from the discard pile; yet for the student of shipping history during the period covered and especially for the coastal shipping of the time they can provide essential, even if limited, information. It may be argued that the two official samples taken would provide comparable information and a statistical base for discussion yet here is again the key question: the random sample is pure, statistically objective and acceptable; all other samples are suspect to a degree, yet all may contain and retain vital information which would otherwise be lost. What should be done?

- 5.11 There is no doubt that if the scholar is dependent upon quantitative analysis then the methodology of the random sample is essential; if, however, he is seeking flesh with which to clothe the bare bones of his statistics, what then? Even the researcher using quantitative methods may find it desirable to check inferences derived from random sampling against out-of-sample data.
- 5.12 It is doubtful whether a wholly acceptable answer can be devised. Scholars have tended to look askance at sampling methods and sometimes to claim that what they required had been destroyed and this once again raises the question of whether the archivist should ever diminish a series or class deliberately. After all, if nothing is kept then at least the loss must be accepted, though perhaps with some regret; whenever we sample a series the question will always arise: what was destroyed, how was it selected, and ought the whole to have been saved? If only statistics are required and the method is irreproachable, fair enough; but once the scholar looks beyond that point, even though he may be thankful for what has been retained, questions will be raised in his mind. This is not so where time and chance have done their worst: it always arises where man is known to have interfered with the survival pattern.
- 5.13 It was essentially on these grounds that Jenkinson wrote so forcibly over fifty years ago of the dangers of archival selection as opposed to administrative choice and even if his thesis can no longer be held to offer the final answer, his criticism must be faced and answered.(3) If it cannot be answered to our satisfaction then we must examine whether some alternative approach may not provide the answer we seek.
- 5.14 There is, of course, one other question which must be considered: namely whose should be the determining voice in this matter, is the user to be the prime mover in this respect or is there some other factor which will decide what is done? The more sophisticated the method to be adopted, the more likely it will be that some specialist knowledge will be required in determining both method and its application, and in the case of random sampling it is not unusual for the potential user to assist in deciding the exact method and the size of sample. This involvement of the specialist is not of itself to be deplored and indeed a measure of consultation may well be a most valuable factor in deciding appraisal policy and the techniques most suitable for the cases under examination. It is of interest to note that the systematic methods in use in France, Finland and the German Federal Republic have all been examined and approved, if that is the right word, by academics(k), and In the case of the Registrar General of Shipping, too, as also the General Accounts Office of the U.S.A., the advice of scholars was sought before the initial random sample was taken. It has been seen, however, that all these methods may fail to produce the ideal result and in the circumstances it cannot be stressed firmly enough that the final decision must lie with the archivist himself. No one can prognosticate accurately the ways in which research will develop, and experts, however well intentioned, can and often will present a special case applicable to their own field of study, though not necessarily to that of anyone else. Although this discussion leads once again to the basic dilemma of theoretical objectivity against actual informational needs and therefore in no way resolves the problem of sampling, it does stress the need for the final determination to be in the hands of someone not open to special pleading. Take advice from all quarters, but decide on archival grounds as to the action to be taken. We and no one else, are aware of the implications of storage and related costs as well as of the wider potential of the records to be appraised and in this regard our decision must be final.

- 5\*15 It seems, therefore, that we must either accept sampling as a valid, but imperfect, course of action, which alone can permit the retention of at least part of the documentary information in our care, but which will never be a wholly satisfactory practice because in archival terms it is not and cannot be solely a statistical exercise; or we must find some generally acceptable alternatives: (a) the retention of the whole series; (b) the use of more than one sampling technique in relation to a single case; (c) the use of microphotography, i.e. miniaturisation, to provide essentially the same answer as in (a); and (d) the computerisation, i.e. data composition, of significant data.
- 5.16 (a) Retention of whole series: from the point of view of the user this is, of course, the ideal solution and there are some scholars who would urge it despite the inevitable problems which it would create. In his most important article written in 1967, Pierre Boisard discussed the methods employed in sampling the Income Tax records housed in the Archives de la Seine and stated categorically that 'the only satisfactory conclusion . . . would have been to keep everything; but given the quantities, this was impossible'.(5) Something of the same conclusion was expressed in a comment from the British Public Record Office, that it was considered that in many cases it would be preferable to keep the whole of fewer series rather than imperfect examples of many. Nevertheless this is the solution which has already been discarded on grounds of cost and informational content at the initial appraisal. If cost and storage capacity linked with the continuing growth of the series form the overriding factor in determining future action, then this solution, ideal though it may be, must be discounted. If it had been considered either economically possible or academically desirable, then the sampling proposal should not have been put forward in the first place. Sampling is not an appraisal technique which is applicable to any class or series the content of which is deemed to justify permanent retention as a whole.
- 5.17 (b) The use of combined methods: it was indicated in chapter 2, that it is possible to sample on a qualitative basis, where a pre-determined set of criteria were to be applied to the records in question. It was, however, accepted that this could give rise to undesirable and subjective judgments which would make the result invalid for statistical purposes, although it might prove useful in special cases for particular areas of research and in that sense provide that element of specific information required to qualify or amplify the statistical analysis. There can be a case, therefore, for adopting some purposive or qualitative form of sampling in parallel with the purely statistical sample. so, too, forms of systematic sampling may provide a satisfactory second element to be used side by side with the random sample. If, however, combined systems are agreed then three corollaries apply: (1) that in all cases the purely random sample must be taken first in order that it shall be statistically inviolable - it also follows that the two or more samples must be treated as separate entities and that this fact, together with full note of method used must appear in any list prepared of the sampled material; (2) that if it has been decided that only a sample of x per cent is necessary, then, in fact a considerably larger fraction of the whole is being retained and that the economic factors which played a part in the initial decision to sample must be considered and (3) that the second sample will inevitably be somewhat skewed by reason of the taking of the first.



In the case of combined systems there is also a place for what statisticians refer to as 'transverse' and 'longitudinal' samples. Once the principal statistical sample has been taken, it is permissible to retain the records of a single unit, i.e. what has been termed a regional sample, as an example of the complete functioning of one unit of administration; or, alternatively, to retain the documentation of a particular function across the board. Examples of this kind occur with the meteorological records and with the Canadian system.

- 5.19 (c) The use of microphotography: because of the considerable reduction in size resulting from a microphotographic programme it should be possible to retain the entire series in a minute space compared with the original written record. This can most certainly apply with some series of maps and plans where the bulk involved is a major factor even in the retention of a sample. In general terms, however, the overriding consideration must be one of cost combined with the strictly limited informational value of the papers to be filmed or sampled. It is not necessary here to state all those factors which must be considered before a microphotographic programme can be accepted as a record management tool, but certain considerations are of immediate significance. If a series has already been filmed and the original papers destroyed, then the general primary appraisal factors apply, and if the outcome is marginally in favour of retention, the sort of situation in which sampling would usually be considered, then the very fact that filming has already taken place should make it more possible to retain the whole series, in preference to keeping only a part. If, however, the records are still in their traditional form the question is quite different and questions to be asked include - is the series of such a format that even in sampling the use of micro methods is imperative, e.g. certain cartographic records? Is the series of such a format that microphotography can be carried out with the minimum of problems of preparation and the minimum of operational skill? Is the material to be sampled of such a character that it justifies advanced sophisticated methods of retrieval - will it in fact ever be used or is it of such a nature that traditional roll film can be used with its consequent difficulties of usage and of comparison of exposures? Is the cost of filming, bearing in mind the preparatory work which will inevitably be involved, so high as to make the project unsound in economic terms?
- 5.20 Unfortunately, microphotography does not of itself eliminate the basic problem of sampling, which lies in the nature of the documentation to be so treated. If the records had had a high, perhaps even a moderate, evidential or informational content, unavailable elsewhere, then the proposal to sample would not have arisen. That proposal has been made because the records are vast in bulk, comparatively uniform in character, probably growing at an appreciable rate and, in general terms, are regarded as having a minimal research value, their administrative significance having already been determined in a negative sense. If, therefore, the creating agency has not considered microfilming to be a Justifiable exercise and one having a sound economic basis, it is highly improbable that the archivist with still more limited resources will find it to be justifiable as a process in lieu of sampling.
- 5.21 (d) The computerisation of significant data: essentially the question must arise as to whether advanced modern technology cannot be utilised to overcome the various dilemmas which face the archivist in the matter of sampling. Perhaps the best example so far examined is that of the process used in Canada for the records of Unemployment Insurance.(6)

The actual process has been considered earlier, but for the purposes of this present discussion the following points are relevant. (1) The twenty million benefit claim files are maintained in microfilm cassettes and are therefore much less storage consuming than would be the originals and that the information is also more easily accessible, and (2) the Unemployment Commission itself maintains its own continuous sample of benefit files. It is therefore relatively simple for the Canadian Archives to take its own sample of these computerized files because the information is already in a form which can be easily utilised for a variety of purposes. It is therefore noteworthy that not only can the basic random sample be compiled, certain other samples may also be taken of a specific nature once alphabetical distribution of names has been determined in respect of immigrants. Once more it is necessary to emphasize that the relative ease of adoption of this methodology is dependent to a degree upon business practice: the agency, in this case the Unemployment Commission, has found that filming and the computerisation of data is a viable economic exercise and as a result the archivist can adopt highly sophisticated means of retaining information in sample form. It is undoubtedly a splendid solution and a very adaptable one, but if its operation depended upon the retention and analysis by the archivist of a vast series of paper files until the adequate data base had been established, then the viability of the project in terms of staff time, storage and cost effectiveness would surely have militated against its application. If the resources are available at a high enough level there is no doubt that what is eminently suitable for non-traditional archives could be applied with advantage elsewhere but it would be a time consuming and very costly exercise since it would entail the preparation of computer in-pup from the original paper files and not the application of A.D.P. methods to material already in suitable form.

- 5.22 In one respect this study has been less than wholly effective for there has been little evidence of consumer reaction. That there is concern over what is done is evident in many countries and the present investigation into the control of records in the United Kingdom by the Wilson Committee stems, in part, from academic concern lest the present methods with their traditional emphasis on administrative needs are too restrictive and tend to the elimination of a far greater proportion of the records than should be. Nevertheless the speed of change of techniques in this field and the combined pressures of economic factors and of contemporary historical and sociological study present the archivist with a serious dilemma of choice and the most noteworthy development is that considered opinion tends to involve the archivist increasingly in appraisal methodology. The burden of argument throughout this study has also been that, except in rare instances, sampling is an archival activity, although on occasions academic consultation may be desirable if the outcome is to be generally acceptable. It may be that, in time, a more detailed study of user reaction would be possible simply because conclusions reached at the present time may prove, in the light of experience, to be less than perfect.
- 5.23 In summing up, therefore, it has been seen that research needs and archival interests do not of necessity coincide. Moreover, traditional interests in the historical field are now supplemented by social science and an approach to questions based on quantitative analysis. This fact carries the corollary that many categories of files, more particularly those relating to individuals and to individual cases, bear a higher research potential than was previously envisaged. It is, however, just these series which present the major problems of bulk, space for storage, continuing growth and costs in terms of maintenance

and processing. The archivist, faced with this problem sees the retention of samples as a practical alternative to wholesale destruction. In the report on 'Appraisal' produced for the I.C.A. in 1976 one reply was quoted which stated 'that the selection of samples for permanent preservation is the pre-requisite for the elimination of the bulk of the material'.(7) Unfortunately, no one system is universally regarded as satisfactory; local needs and local differences lead archivists to take independent decisions as to the methodology to apply and, although one can identify certain types of sample, even within each type different methods and techniques will be used. It is claimed that random statistical samples provide not only the essential objectively chosen material for analysis, but also offer the more traditional scholar what he seeks. At the same time, in archives, there are so many special considerations to be borne in mind that even with the most stereotyped and limited case file it would appear that far more than one factor will be found, and indeed, that each individual file may be of concern to a number of students in varying projects.

- 5.24 It becomes necessary, therefore, to consider records both from the quantitative and the qualitative view point and no one sampling method seems to be equally acceptable in both areas. Alternatives are possible by the use of combined methods or by the use of film or tape, but unless a prearchival decision has been taken by which the record itself has become machine readable, there remain serious economic doubts as to the validity of operating an alternative process with material which is already judged to be of low informational value. On the other hand, if such documentation has been mechanised as a management procedure, then it is possible for a data base to be created which will provide ample scope for sampling of significant information.
- 5.25 Processes and methods in 1981 depend upon the technological advance within the countries, agencies and archives concerned and therefore, despite misgivings, it seems likely that forms of purposive and systematic sampling will continue alongside the random statistical sample for the foreseeable future. Indeed, it may be questioned whether they will ever be eliminated entirely, or indeed, should be, so long as large series of paper files remain in common use. The problem therefore remains of attempting to find some common factors and procedures which can provide archivists with guidelines when attempting to put sampling projects into practice.

### Notes

1. This difference of approach has been highlighted in a letter from Dr. F.B. Evans to Dr. Eckhardt Franz of 15 April 1981. He writes regarding the confusion which may occur between 'appraisal' and 'sampling', 'that we all view the concept of "selective retention" differently, depending upon our training and experience, with different record-keeping systems. (To me, this is the basic problem).  
In the USA we "appraise" the total organic accumulation of an office, and select certain series (in their totality) for preservation. In Europe, it seems to me you select dossiers within certain classes (series) and that is your "appraisal". If we in the US or Canada go below the **series level to select** only certain files (folders, dossiers, etc.), we maintain that we are violating the organic character and archival integrity of the series. You don't see it that way in Europe, since it seems that you apply the concept of archival integrity (when you do apply it) to the individual dossier (which may still be 'weeded' in the pre-archival stage!) while we apply it to the entire series.

I feel we should get some basic agreement on this important matter. But for the present study, selective retention at whatever level (series or file unit) is not "sampling", since the intention is not to represent the totality, but only to keep the "most" important". To sample, by definition, is to select so that a part, represents the totality; unless that is the intent, we are not sampling, and I feel that is the use of the term that should be made in this study. Perhaps we should make very clear that 'purposive (selection) sampling" is not sampling, but is rather selective retention according to a criterion other than to have a part represent the physical form, informational content, etc. of the whole.'

2. In fact, although claimed to be 'random' the method used was a numerical series of every tenth box.
3. H. Jenkinson, op. cit. , P.149
4. The records of the Finnish Social Insurance Agency are treated in this manner, the final digit of the file numbers being utilised for this purpose.
5. La Gazette des Archives, vol. 59, No.3, 1967, p.221.
6. See pp, 32-3
7. A. Kromnow, op.cit., p.19

6. Conclusions and Guidelines

(i)

- 6.1 In his book Archives Administration, published in 1977, Michael Cook briefly examined sampling as a technique as part of his chapter on appraisal. Having identified three types - random, selective and representative(1) - he completed his examination by claiming that there is 'no such thing as a statistical sample of general utility' and then stated that in any event, sampling was inadvisable without some specialist advice.(2) After considering the conflicting evidence and views of archivists in Europe and America, one comes reluctantly to much the same conclusion and feels sympathy with one correspondent who wrote that 'the first thing about archival sampling . . . is that one should only do it if one really has to.'
- 6.2 Nevertheless the purpose of this study has been to try to identify the cases when sampling is an appropriate exercise and under what terms it can properly be carried out and it is therefore desirable to attempt some statement of general principles and to offer guidelines for those faced by sampling problems. In this respect, it is necessary to return to the basic distinction between appraisal and sampling and to attempt to define the terms in use. Appraisal, therefore, is the fundamental selection of documentation leading to preservation or to destruction, under whatever criteria are considered valid at the time the choice is made. Every archive class, at least in theory, has passed through an appraisal process, for a decision, conscious or unconscious, has been taken in respect of its retention. In contrast to this process, sampling can only occur in those cases, where despite the appraisal decision, some uncertainty remains because of the positive but limited informational content of the records and also because of their bulk. For good or ill the archivist is influenced by costs, for storage, maintenance and care are expensive commodities and potential, perhaps unknown, research values (that possibility of usage for secondary purposes), have to be set against the costs of retaining the records in their original state.
- 6.3 It seems abundantly clear, that sampling is a technique which, in whatever form, is subject to some criticism and uncertainty by custodian and searcher alike and that the archivist will be well advised to adopt sampling methods as infrequently as possible and only then after the most careful consideration of the methodology to be used and, perhaps - for the final decision may well depend upon circumstances beyond the archivist's control - with the advice of an expert in the field of study involved, and, or, in statistical method.
- 6.4 Appraisal, therefore, may determine the fate of classes, series or even unit pieces; sampling will only apply when some further reduction in bulk is deemed necessary over and above that first decision, or if, for some reason it is decided to keep an example of a type of record otherwise destroyed.
- 6.5 A second consideration will relate to those types of record where such a decision may be appropriate. Bulk has already been mentioned and, except for the 'example', there are few occasions when the factor of bulk coupled with that of continuing growth will not provide the key element in the final decision making. Yet not all types of record are equally suitable for sampling and it has been necessary to stress repeatedly, that a series of records which is suitable for this technique, will be homogeneous. If the documentation is wholly individual in content and if it provides significant evidential detail, then it is improbable that sampling can apply. One manual of advice which refers to statistical sampling states that it can only 'be applied to documents which contain mathe-

matically quantifiable information in standard form and in sufficient depth, either because they cover a long period of time, or because they are complete for the particular subject with which they deal', (3) a ruling which goes some way towards explaining why statistical methods are seldom wholly applicable to non-conventional archives.

6.6 As opposed to the somewhat cautious and qualified approach of European archivists and their tardy acceptance of fully random statistical methods, is the enthusiasm evinced by the American paper on 'Statistical Sampling of Archives' (k), which argues forcibly that since 'the record' is never complete we should 'be able to sample without undue anguish about the integrity of the records' . Certainly, as suggested earlier, the trans-Atlantic attitude is much less overburdened with fears that the 'significant' item will be lost and, although NARS still employs both selective (purposive) and systematic sampling on occasions, the movement of thought in the case of large homogeneous series of files and case papers is very much in line with that proclaimed in Canada.

6.7 A related matter which must be briefly mentioned is the size of the sample if, in fact, one is to be taken, though here again, it is difficult to find completely common ground. Eleanor McKay considered that in order to obtain a satisfactory and authoritative sample of Congressional papers a twenty per cent sample was required. (5) McReynolds argues from the statistical point of view that 'the greater the reduction the greater the imprecision of the resulting sample' and he, like Cook, recommends the seeking of advice. He also suggests that if a record group has a number of series of varying significance, the archivist can 'create a stratified sample of the whole record group' . (6) This is, in fact, essentially what is being attempted in Kent with the sampling of Social Service records, where the size of the sample taken from various series of records is dependent upon the relative size, importance or nature of the individual series. This concept suggests that the size of sample will vary according to circumstances, though it will remain true that the larger the sample the more truly representative it is likely to be. Once again the overriding factor is likely to be cost, for if a very large sample is to be retained may it not still be preferable and possible to retain the whole, which after all is the ideal solution?

## 6.8 Guidelines

We can conclude so far:

- (1) Sampling should only take place (a) when there is some doubt about the validity of retaining the whole class or series of conventional (paper or textual) records, but when automatic destruction is regarded as too drastic a course of action, or (b) when it is felt proper to retain some examples from an otherwise destructible category of records.
- (2) Where the material itself is appropriate for this kind of technique, i.e. where classes or series of files are homogeneous ('if the individual files contain similar records in each file, the variability will be small and the statistical significance or precision of the sample will be high')(7). Heterogeneous or highly variable records will produce a serious bias and should not be sampled for that reason unless there is a sub-series within the main series which has special characteristics indicating that retention of the sub-series is desirable.
- (3) In most instances these criteria indicate that sampling is not applicable for the selection of cartographic, audio-visual or machine-readable records.

- (4) While the size of sample will vary according to the nature of the documentation and the circumstances under which decisions have been made, a larger sample will provide a more satisfactory coverage of the whole and will therefore be more likely to provide researchers with their special requirements.
- (5) In all cases the methodology used and the reasons which led to sampling must be indicated in any finding aid which is prepared.

(ii)

6.9 It now becomes necessary to consider the various types of sampling in greater detail and to attempt to define areas of usage and significant variations in value in terms of methodology. In this section, therefore, it is most important that exact terminology should be adopted if at all possible. This is particularly so with words such as 'random' and 'statistical', for it is clearly apparent that not only has there been much confusion in the past, but that both these terms are still being used to describe significantly different processes. Although 'random' is still regarded as suitable as a description of various patterns of systematic sampling, its use here, so far as possible, will be limited to describe the precise statistical process described on pp. 24-5. The terminology, therefore, will be that adopted in Chapter 2, but wherever applicable known variants will also be given in the heading to any particular method.

#### 6.10 (a) The example

Although it would appear that examples are taken fairly widely in appropriate circumstances, it must be stressed that within the terms of definition 'an example' cannot be regarded as a true sample. It neither illustrates the qualities of the whole mass, nor does it provide a representative experience of the series. On the other hand, it does indicate that a certain class or series existed or even that a particular type of individual document was once in use. From time to time circumstances may arise when it seems desirable to retain an example of what would otherwise be destroyed.

#### 6.11 Guidelines

This is a valid appraisal decision, but -

- (6) Any description must indicate the provenance of the example and why the residue was destroyed.
- (7) The significance of the example rests in its nature as an example and in nothing else; it has virtually no research potential except as an indicator of what was formerly in existence, even though it may have value as a precedent for the agency concerned.

#### 6.12 (b) Purposive Sampling (Qualitative or Selective Sampling)

The most dramatic example of this kind of sampling and the one which led to considerable dispute was the French attempt some years ago at producing 'models'. This was attacked in 1953 by R.H. Bautier(8) and was further discussed in 1967 by Pierre Boisard in La Gazette des Archives.(9) As a method it is superficially very tempting to the archivist, who considers that a class or series must contain material of special or particular merit in respect of individual topics, areas or personalities. As a method

it attempts to answer the criticism that sampling removes the exceptional and thus loses what is of special significance. In considering this argument, one author has referred to this form of sampling being carried out under 'criteria of significance', but has then pointed out that such criteria are atypical simply because they are of special significance and that 'the resulting sample does not, in any way, reflect the whole group of records'. While he does, somewhat grudgingly, admit that selective sampling 'within series or record groups is a valid technique for archivists in some circumstances', it is also plain that this system is not very far from the taking of examples. It is nevertheless a method which has been used by many archivists and, almost as often, it has given rise to criticism because of its inevitable built-in bias and its unsuitability for statistical purposes. It must be admitted, however, that so long as doubts remain about the wholly satisfactory nature of systematic or true random sampling, there will be a tendency to continue to use this method. It is of interest to note that when the sampling of the papers of Congressmen in Wisconsin was attempted, the persons involved still considered it desirable to take a purposive sample from the eighty per cent residue in case the statistical sample did not prove to cover all contingencies.(11) Indeed, if purposive sampling is to be accepted at all, it is probably most appropriate in this kind of context where, after the taking of a statistically acceptable sample, it is used as a secondary system - a kind of safety-net. In other words, that once a statistical sample has been taken, it is permissible to extract from the residue, other files or papers on a qualitative basis.

#### 6.13 Guidelines:

In view of the above facts and of the continued use of this method, even if there remain many reservations, it is well to establish certain basic rules for its adoption and usage:

- .(8) Purposive sampling, because of its 'exceptional' character and built-in bias, is the less appropriate the more homogeneous the original series of records.
- (9) No purposive sample should be taken in the place of a statistically valid sample - it may be a supplementary process, but is not really acceptable as the primary method to be employed.
- (10) Criteria of selection must either be very specific (e.g. the case files of conscientious objectors referred to on pp.20-1) or must be as comprehensive as they can reasonably be made. Once again the problem arises that the attempt to be comprehensive may only be achieved by the retention of the whole series.
- (11) A very clear description of any criteria must preface any finding aid, so that the user is made fully aware of what was done, why it was done, how it was done and where he can find the other element of the sample which can be used for quantitative analysis.

#### 6.14 (c) Systematic Sampling (Representative, Quantitative or Statistical other than 'random', Time-Series or Chronological, Numerical

There is no doubt that of all sampling methods, systematic sampling is the method with the greatest number of devotees at the present time. The preference for this method is usually expressed in terms of the simplicity with which it can be carried out. Nevertheless there has been some criticism of methods in the past, and indeed, systematic sampling covers such a very



wide spectrum of techniques, varying from ones bordering on purposive methods to systems designed to provide a near 'random' sample that it is difficult to define in any one simple manner. It is worthy of note, however, that whereas European archivists still tend to regard systematic methods as the most suitable for general purposes, the pressure in North America is towards the greater use of truly random methodology. Each of the varieties of systematic sampling must be briefly considered and for each some rules of action must be established.

- 6.15 Only one remove from purposive sampling and still far from being what one usually regards as a systematic sample is the representative topographical sample. This occurs when, faced with a large number of similar groups—(fonds) within the archives of an agency, each based on a particular topographical area or regional office, it is decided to retain whole archive groups for a selected number of offices. It is immediately clear that this is not true sampling, rather it is more closely associated with basic appraisal in that it is not series which are being reduced in size but rather a determination to retain some units of archives as opposed to others. Moreover there is no homogeneity within the records chosen for retention - the group may contain many disparate classes - but each group will approximate to its fellows in the character of the records created and those selected for retention will therefore present a representative sample of the records of the central agency as they existed at the local level. This approach has been strongly criticised in France as creating a false sense of uniformity and for damaging the resources of local history within those areas where destruction took place. There is a sense, of course, in which this process is of the kind of arbitrary selection which the ravages of time have created and, indeed, such selected groups of records are only valid for research within the parameters of their own topographical area and cannot be cited as authoritative evidence for what took place elsewhere in areas for which the records have been destroyed. Finally, this is a method which underlines the archivist's dilemma very forcibly - the archives of the unit office or area can only be regarded as marginally worth preservation; can the cost of keeping all such records and therefore all such units be justified? If it cannot then possibly a topographical sample is to be considered as a comprehensive example of what once existed and took place in one area only.

6.16 Guidelines:

It should be understood therefore:

- (12) That topographical samples are only acceptable where a large central agency has many local offices and where the central agency's records are already retained in an adequate form.
- (13) The records of chosen units within the system should be retained in their entirety as group - this is a very special kind of example and its validity rests in the completeness of the archive groups of which it is composed.

- 6.17 The second form of systematic sampling frequently used is the retention of files based on letters of the alphabet. Alphabetical sampling is regarded as having a certain statistical justification and, in Canada, is combined with a still more 'random' numerical selection. The weakness of this system lies in the national and local variability of letter usage, so that although the choice of initial letter may appear representative of the whole series, sub-categories of individuals may be entirely missed. For example the use of this method in the German Federal Republic has concentrated on the letter H, but while this is satisfactory for names of Germanic origin, it omits those

of Romance origin; and a similar use in the United Kingdom would result in the exclusion from the sample of certain immigrant groups in the population. Nevertheless with large series found in alphabetical order it is a form of sampling very easy to put into practice.

#### 6.18 Guidelines:

It may be said to be appropriate when:

- (14) there is a large homogeneous series of personal files arranged alphabetically, so that the sample will be of a large enough size to provide reasonably accurate information (e.g. at Köln the use of the initial letter H provides an 8.5 per cent sample);
- (15) where an analysis has been carried out to establish which initial letter will be satisfactory according to the purpose of the exercise as a whole and will also effectively represent the records in question (e.g. it is useless to select a letter of very infrequent usage like Q or Z; equally to adopt M in Scotland might lead to an undesirably high percentage sample).

6.19 A more nearly 'random' and statistically sound sample is achieved by numerical selection. Numerical or Serial sampling can be simple, i.e. every tenth or twentieth box, files, etc., according to the format of the records; or it can be based on far more complex criteria, e.g. the Social Insurance Number selection in Canada, where files with terminal digit 5 alone are selected. It must be indicated that the 'tenth box' method can lead to some problems if the make-up of the records results in files which overlap the box arrangement. It is said that this method used in the Public Record Office for the records of the Registrar General of Shipping has resulted in a sample, which, while it provides a statistical base, is not particularly suitable for other research purposes. Nevertheless numerical sampling is one of the three most widely accepted methods which while not wholly random in the statistician's sense can provide an acceptable base for most purposes. one of the difficulties illustrated in the evidence submitted for this study rests once more in the loose use of terminology and one cannot always be certain when 'random' methods are mentioned whether that is really so or whether some form of numerical selection is not being practised. It is essential, however, that if the intention is to produce a sample which is valid for statistical purposes, then bias must be avoided and not all records will be equally suitable for this kind of sampling.

#### 6.20 Guidelines:

It can therefore be stated:

- (16) 'A serial sample may be acceptable for statistical study if the existing order of the whole body of the records is random (e.g. a series of returns filed in no systematic order). '
- (17) 'A serial sample is the only practicable method of sampling if the individual items cannot be separated and the assemblage has to be taken as the unit. ' (14)
- (18) This method should not be used if there is an undoubted alphabetical, topographical or chronological arrangement to the records.
- (19) The degree of acceptability depends upon every unit in the series having its unique individual number, a vital element if statistics are to be meaningful.

(20) The **numerical** series to be adopted must be established in advance and must be adhered to rigidly.

6.21 Chronological or Time-Series samples. This form of systematic sampling, which finds much favour, depends upon the chronological arrangement of the papers to be sampled and, in most cases, results in the survival of records for every fifth or tenth year, often using census years because of their association with other demographic material. The weakness of this form of sampling rests primarily in the fluctuations of human society and that the years thus selected may avoid vital changes of a political or economic or legislative character. This is a cause for concern and has made searchers suspicious of a method which tends to concentrate on the short term facts rather than the long term trends. It has been this consideration which has led in France to the somewhat complex pattern of sampling adopted for the records of Santé-Travail, where the papers for one year in thirteen are retained and those for one month, in rotation, kept for the intervening years. This system is further refined, however, by permitting the year for which there is a total retention to be determined not by an arbitrary series, but by the significance of events of that year. In the end therefore one is met in essence with a statistical sample based on the monthly series with its built-in variable in order to obtain a representative cover over a period of years, and then superimposed on that, what is in effect a purposive sample dependent upon 'criteria of significance'. This complication must be recognized in any statistical work carried out, for while some valid comparisons are possible with the monthly series, the chosen years will not be similarly comparable.

#### 6.22 Guidelines:

For a satisfactory time-series sample therefore:

- (21) the records must be homogeneous and arranged chronologically;
- (22) the time-series should be selected irrespective of political or other changes happening in between the retention years and this time-series should be adhered to at all times if the result is to provide statistical information;
- (23) the closer together the selected years are, the more likely it will be that sudden aberrations in society will be picked up, but since it is only a sample one cannot and should not regard special circumstances as reasons for special variation; if there is cause for doubt then it may be that a selective (purposive) sample should be taken in addition to and after the chronological sample.

6.23 It must be stressed again that it is the relative ease with which systematic samples can be taken which is their principal attraction. In the R.A.D. paper from the Public Record Office, quoted above, the comment is made that 'in practice it may be too expensive or time consuming to take 'a random sample, and that therefore 'the alternative of a "serial" or "systematic" sample' may have to be adopted. This is now very much the preferred method in European repositories, but it is gradually giving way to the true random sample in the United States and Canada.

#### 6.24 (d) Random Sampling

The essential problem in this technique is to establish the fully random nature of the sample; to apply a statistically sound method of selection with no element of bias; and to be satisfied that the needs of traditional

research are as adequately covered as those of quantitative analysis. In the view of the Canadian Public Archives these criteria are all met and this is also the opinion of R.M. McReynolds of the National Archives and Records Service of the U.S.A. on this side of the Atlantic there are still doubts and the much slower adoption of computer techniques in archives and lack of resources has limited the use of random methods. One comment received reads that 'calculations based on the (random) sample will not provide historical accuracy in the sense of tying the creating authority's operations to particular cases, but they should give an accurate overall view of the effect of policies or the extent of problems'. Material to be processed in this way must be essentially homogeneous, i.e. with a very low variability of content, and should 'contain mathematically quantifiable information in standard form or in sufficient depth, either because they cover a long period of time, or because they are complete for the particular subject with which they deal'. (17)

## 6.25 Guidelines

The choice and practice of this methodology depends upon:

- (24) a suitable series of homogeneous records;
- (25) the use of a random number table(18) or, possibly, of a highly sophisticated numerical series; (19)
- (26) the numerical individuality of all the pieces (units) in the file series so that bias is eliminated;
- (27) the careful determination of an appropriate size for the sample, bearing in mind that 'the greater the reduction the greater the imprecision of the resulting sample'. It should likewise be remembered that 'to double the accuracy of a sample it is necessary to quadruple its size'; (21)
- (28) that in this area in particular, the advice of a statistician and expert in historical quantitative research can be invaluable and can prevent serious error.

6.26 The random sample is taken by a precise scientific process, all other samples only approximate to a greater or lesser degree to that objective ideal and since they are easier to adopt, and are in some ways more natural in methodology to traditional archival thinking, they will tend to be used, especially where the technology associated with true random sampling is still difficult to acquire and the skills of the persons who must carry out the work limited. Nevertheless, as computer technology expands and becomes less expensive, it would appear that the random sample based on the random number table, or perhaps on some essentially random system like the Canadian S.I.N. numerals, will become increasingly the standard adopted for long homogeneous series of paper files. It will never be appropriate for records with a high variable factor, but it is questionable whether sampling of any kind should be advised in those circumstances.

(iii)

6.27 This chapter has not considered at length the question of non-traditional archives, but it may be recalled that the essential argument of Chapter 4 was that, in most instances, sampling was not a technique suitable for material

which must be selected on a unit basis. There were a few exceptions, depending upon the provenance of the material and in such cases it was almost invariable that time-series or chronological samples were taken, unless indeed the basic record had been microfilmed and a purposive sample appeared more appropriate. In all these instances the rules which would apply are those which have already been set out in the appropriate section of this chapter. There seems to be no room for quantitative, statistical samples in these areas, though of course the argument is confused by the availability of almost unlimited samples for research purposes, where the records themselves form a data base available for investigation.

- 6.28 In conclusion, therefore, sampling is a methodology forced on archivists by the sheer bulk of documentation and the cost of preservation. It should not be adopted unless there is no alternative solution, for it can seldom be wholly satisfactory. Random statistical sampling is appropriate for homogeneous series of paper files and can form a satisfactory base for quantitative research and, dependent upon that homogeneity, a reasonable base for traditional research also. The archivist, however, faced with costs, staff problems and many classes which are somewhat more variable than should ideally be the case, will often tend to adopt simpler methods, which can still be statistically based even though less completely satisfactory than random sampling and which provide more scope for the retention of the exceptional as well as the normal. In all cases, however, any sample which is intended to be statistically valid must be taken first and any other type of selection made subsequently. Full notes must always be retained of every action taken and of the various elements of the sample if more than one has been taken. The records themselves, too, must be stored in such a way that the distinction between what is acceptable for quantitative analysis and what is not, is clearly apparent.
- 6.29 In one sense, sampling is the worst of all worlds, but there is a growing opinion which sees in random sampling a set of criteria acceptable for all purposes provided the basic record is suitable for that kind of technology. Microphotography and the preparation of data bases may help to overcome the dilemma presented by archival sampling, but even apart from cost, the application of modern techniques depends upon the availability of such technology and it will be many years before the less sophisticated methods are rejected entirely. Unfortunately sampling will always leave the archivist and perhaps the scholar in a state of uncertainty, even though the methods used may be totally acceptable. In all cases, the residue must be destroyed and something may be lost thereby. Accepting that irreducible factor, forms of sampling will continue so long as very bulky series of essentially similar records have to be appraised and cost of storage prevents the retention of the whole.

#### Notes

1. These terms approximate to what have been identified in this study as random, statistical, sampling; purposive sampling; and systematic sampling. They reveal once-again the difficulties of determining the appropriate terminology in this area.
2. Cook, op.cit., P.89
3. 'Manual of practice', East Sussex County Council.

4. R.M. McReynolds, 'The Statistical Sampling of Archives', NARS, Nov. 1975
5. American Archivist, vol, 41, No.3, pp.181-9
6. McReynolds, op. cit.
7. Ibid.
8. R.H. Bautier, 'Triages et eliminations des documents d'archives', in Archives, Bibliothèques, Collections, Documentation, No.9, May-June 1953, P.250
9. P. Boisard, 'Pour une politique des eliminations: reflexions sur la pratique des Archives de la Seine' , La Gazette des archives, No.59, 4eme trimestre, 1967
10. McReynolds, op. cit.
11. American Archivist, vol. 41, No.3, pp.181-9
12. See p.105 for further notes on S.I.N. numbers
13. It has been pointed out that the major use so far made of the Registrar General of Shipping records has been for research into particular vessels or individuals, which would not be served by any sample however taken. The sample is also valueless for studying trends in types of vessel, particular ports or areas or particular trades.
14. RAD Occasional Papers, No.5, Public Record Office, Oct. 1975
15. This is similar to the method used for Finnish Social Insurance files
16. Cook, OP. cit., P.88
17. Manual of practice, East Sussex County Council
18. See P.106
19. The Canadian S.I.N. system is such and for that reason is described more fully in appendix 1, see P.105
20. McReynolds, op.cit.
21. RAD Occasional Papers, No.5

BIBLIOGRAPHY

Note: the books listed below are mainly of a general nature though containing some information of relevance either to appraisal or to sampling or to both topics; the pamphlets and articles are usually of a more specific character.

BOOKS AND CHAPTERS OF BOOKS.

- Baudot, M.: 'Selections and Eliminations'. In: Manuel d'Archivistique, Paris, S.E.V.P.E.N. , 1970, pp. 161-185
- Cook, M.: Archives Administration, Dawson, 1977, 258 p.
- Hindus, M.S. et al.: The Files of the Massachusetts Superior Court, 1859-1959: an analysis and a plan of action, G.K. Hall, 1980.
- Jenkinson, H.: Manual of Archive Administration, Percy Lund, Humphries & Co.Ltd., 1965 (2nd ed. rev. R. Ellis), 261 p.
- Lamb, W.K.: 'The Fine Art of Destruction'. In: Essays in Memory of Sir Hilary Jenkinson (ed. Hollaender), Soc. of Archivists, London, 1962, pp.50-56
- Langley, R.: Practical Statistics, Pan Books Ltd., 1970. 398 p.
- Mitchell, T.W. (ed.): Norton on Archives, Southern Illinois University Press, 1975, 288 p.
- Schellenberg, T.: Modern Archives, Principles and Techniques, University of Chicago Press, 1956

PAMPHLETS AND ARTICLES

- Anderson, R.J.: 'Public Welfare Case Records - a study of Archival practices', American Archivist, vo. 43, No.2, Spring 1980, pp.169-79
- Australian Archives: 'Administrative Instructions: Procedure for the Destruction/ Review of time-expired Accessions', 1977/3
- Bell, L.: 'The Professional Training of Archivists', UNESCO Bull.Lib., XXV, No.4 July-Aug. 1971, pp.191-97
- Boisard, P.: 'Pour une politique des eliminations: reflexions sur la pratique des archives de la Seine', La Gazette des Archives, No.59, 4th trimestre, 1967, pp.220-22
- Brichford, M.: Archives and Manuscripts: Appraisal and Accessioning, Society of American Archivists, 1977. 24 p.
- British Home Office: Public Records: Clerk of the Peace, 2nd Schedule, 1941 (832, 538/11); Clerk of the Peace - Public Records Acts, 1958, 1967 (38/1968): Coroners Records (250/1967).
- British Records Association: Memorandum - Rating Records
- Burckhardt, F.: 'Les eliminations', La Gazette des Archives, No.108, 1st trimestre, 1980.

- Darter, L.J. : 'Records Appraisal: A daunting task', Indian Archives, 18, 1969
- Dollar, C.M.: 'Appraising Machine-Readable Records', American Archivist, vol.41, No.4, Oct. 1978, pp.423-30
- Duchain, M.: 'Les Procédures de Tri', La Gazette des Archives, No.77, 2nd trimestre 1972, pp.82-3
- East Sussex County Council (U.K.): Modern Records - Manual of Practice
- Evans, F.B.: 'Modern concepts of archive administration and records management', UNESCO Bull. Lib. xxiv, No.5, Sept.-Oct. 1970, pp.242-7
- Modern Archives and Manuscripts: a Select Bibliography, Society of American Archivists, 1975, 209 p.
- Fishbein, M.: 'Viewpoints on Appraisal of National Records', American Archivist, vol. 33, No.2, April 1970, pp.175-87
- Grigg Committee: Report on the Public Records (Command 963), H.M.S.O., 1954
- Kent County Council (U.K.): 'Guide for Officers in charge of Modern County Council Records', 1963, rev. 1979, 38 p.
- Kromnow, A.: The Appraisal of Contemporary Records, VIII International Congress on Archives Papers, Washington D.C., 1976 26 p.
- Lewinson, P.: 'Archival Sampling', American Archivist, vol. 20, No.4, Oct. 1957, pp.291-312
- McKay, E.: 'Random Sampling techniques: a means of reducing large homogeneous series of Congressional papers', American Archivist, vol. 41, no.3, July 1978, pp. 181-9
- McReynolds, R.M.: Statistical Sampling of Archives, NARS, Washington D.C., 1975 19 p.
- Meteorological Office, Bracknell (U.K.): Data Review Board, Amex 7C, 'Retention of Technical Records', 1980
- National Archives and Records Service, Washington D.C.: 'Disposition of Federal Records', 1978
- Perotin-Dumont, A.: L'audiovisuel, nouveau territoire de la conservation', La Gazette des Archives, No.109, July 1980
- Public Record Office, London: 'New Perspectives on the Selection of Documents' and 'The Sampling of Records for Statistical Use', R.A.D. Papers, Nos.4 and 5, 1975
- Reiger, M.: 'Modern Records Retirement and Appraisal Practice', UJISLAA, vol.1 No.3, 1979, pp.200-9
- Rundell, W., Jun.: 'Photography as Historical Evidence - early Texan Oil', American Archivist, Vol.41, No.4, Oct. 1978, pp.373-98



- Schellenberg, T.R.: 'The Appraisal of Modern Public Records', NARS Bull., No.8, 1956
- Singapore National Archives: 'Guidelines for the Selection and Disposition of Public Records', 1977. 5 p.
- Smith, W.I.: 'Archival Selection, a Canadian View', Journal of the Society of Archivists, Vol.3, Oct. 1967, pp.275-80
- Steck, H. and Blouin, F.: 'Hannah Lay & Co.: sampling records of a century of timbering in Michigan', American Archivist, vol.39, No.1 Jan. 1976, PP.15-20
- Wilson Committee: Modern Public Records (Command 8204), H.M.S.O., 1981

## APPENDIX I

### Guidelines for Sampling Procedures

The various forms of sampling may be regarded, generically, as one of the tools available to the archivist involved in the process of selection. This process, termed appraisal, affects all records and may be applied at differing times in the life of a record and in respect of differing criteria, administrative or research.

Sampling comes into the picture when a class or series, having been appraised as a whole, is, nevertheless, not being retained in toto, but when an attempt is being made to retain a sufficient element to provide an adequate and representative part of that class or series. Ideally, sampling is an exact statistical technique, but the complication arises with records that the archivist is desirous both of providing a pure statistical sample for quantitative analysis and yet is also anxious lest the individuality of the pieces making up any class or series should be overlooked.

For that reason the guidelines which are enunciated in Chapter 6 of the study The Use of Sampling Techniques in the Retention of Records are given below in the order in which they occur, but without the argument leading to them. Starting with general principles, they move to those factors controlling the use of particular methods. It is hoped that this re-statement will provide a simple and uncomplicated guide to the application of sampling procedures, which vary from the selection of 'examples' with little research potential to the sophisticated 'random sample' of absolute statistical validity.

#### 1-5 General rules

- 1 Sampling should only take place (a) when there is some doubt about the validity of retaining the whole class or series of conventional (paper or textual) records, but when automatic destruction is regarded as too drastic a course of action, or (b) when it is felt proper to retain some examples from an otherwise destructible category of records.
- 2 Sampling should only take place where the material itself, is appropriate for this kind of technique, i.e. where classes or series of files are homogeneous ('if the individual files contain similar records in each file, the variability will be small and the statistical evidence or precision of the sample will be high'). Heterogeneous or highly variable records will produce a serious bias and should not be sampled for that reason, unless there is a sub-class within the main series which has specific characteristics indicating that retention of the sub-class is desirable.
- 3 In most instances these criteria indicate that sampling is not applicable for the selection of cartographic, audio-visual or machine-readable records.
- 4 While the size of the sample will vary according to the nature of the documentation and the circumstances under which decisions have been made, a larger sample will provide a more satisfactory coverage of the whole and will, therefore, be more likely to provide researchers with their special requirements.
- 5 In all cases the methodology used and the reasons which led to sampling must be indicated in any finding-aids which are prepared

#### 6-7 The taking of Examples

- 6 In the taking of examples any finding aid must indicate the provenance of the examples selected and why the residue of the class or series was destroyed.

7 The finding-aid should also indicate that the significance of the example rests in its nature as an example and in nothing else; it has virtually no research potential except as an indicator of what was formerly in existence, even though it may have value as a precedent for the agency concerned.

8-11 Purposive (quantitative or selective) Sampling

8 Purposive sampling, because of its exceptional character and built in bias, is the less appropriate the more homogeneous the original series of records.

9 No purposive sample should take the place of a statistically valid sample - it may be a supplementary process, but it is not really acceptable as the primary method to be employed.

10 Criteria of selection must either be very specific or must be as comprehensive as they can reasonably be made. The problem always arises that the attempt to be comprehensive may only be achieved satisfactorily by the retention of the whole series.

11 A very clear description of any criteria must preface any finding-aid, so that the user is made fully aware of what was done, why it was done, how it was done and where he can find the other element of the sample which can be used for quantitative analysis.

12-23 Systematic (representative or topographical, alphabetical, numerical or serial, chronological or time-series, and quantitative or statistical other than 'random' ) Sampling

12 Topographical samples are only acceptable where a large central agency has many local offices and where the records of that central agency are already retained in an adequate form.

13 The records of chosen units within the system should be retained in their entirety as groups -this is a very special kind of 'example' and its validity rests in the completeness of the archive groups of which it is composed.

14 Alphabetical sampling is appropriate where there is a large homogeneous series of personal files arranged alphabetically, so that the sample will be of a large enough size to provide reasonably accurate information of the whole.

15 It is appropriate for an analysis to be carried out to establish which initial letter will be satisfactory according to the purpose of the exercise as a whole and which will also effectively represent the records in question (e.g. it is useless to select a letter of very infrequent use like Q or ~~S~~ similarly to adopt M in Scotland might lead to an undesirably high percentage sample).

16 Numerical or Serial Sampling may be acceptable for statistical study if the existing order of the whole body of records is random (e.g. a series of returns filed in no systematic order)'. .

17 'It is the only practicable method of sampling if the individual items cannot be separated and the assemblage has to be taken as a unit'.

18 This method should not be used if there is an undoubted alphabetical, topographical or chronological arrangement of the records.

- 19 The degree of acceptability depends upon every unit in the series having its unique individual number, a vital element if statistics are to be meaningful.
- 20 The numerical series to be adopted must be established in advance and must be adhered to rigidly.
- 21 Chronological or Time-Series Sampling is appropriate when the records are homogeneous and are arranged chronologically.
- 22 The time-series chosen should be selected irrespective of political or other changes happening between the retention years or months and this time-series should be adhered to at all times if the result is to provide statistical information.
- 23 The closer together the selected years are, the more likely it will be that sudden aberrations in society will be picked up, but since it is only a sample one cannot and should not regard special circumstances as reasons for special variations: if there is cause for doubt then it may be that a selective (purposive) sample should be taken in addition to and after the chronological sample.
- 24-28 Random Sampling
- 24 Random statistical sampling depends upon the existence of a suitable series of homogeneous records.
- 25 It results from the use of a random number table, or, possibly, a highly sophisticated numerical series.
- 26 It is only valid if all the pieces (units) in the file series have numerical individuality so that bias is eliminated.
- 27 There must be a careful determination of the size of sample bearing in mind that 'the greater the reduction the greater the imprecision of the resulting sample'. It should likewise be remembered that to double the accuracy of a sample it is necessary to quadruple its size.
- 28 Random sampling is an area where the advice of a statistician and expert in historical quantitative research can be invaluable and can prevent serious error.

In the application of these guide-lines it is well to remember that there may often be a case for combined methods, provided always that the pure statistically valid sample is taken first and that in very many cases expert advice into methods, size of sample and research potential may be highly desirable.

APPENDIX 2

Reference has been made many times to the method adopted in Canada for sampling the records of the Unemployment Commission and to the use of the final digit of Social Insurance Numbers as a method of producing a satisfactory random sample. The following note on the make-up of SIN numbers is adapted from a report to the National Archives by Jake Knoppers in 1979. The present system resulted from a revision of practice in 1964 when new numbers were issued. These consisted of eight digits, the first being an indicator of topographical province. The ninth digit is a check digit and it is this alone that is used in the sampling exercise.

The first or provincial digits are as follows:

- 1 = Atlantic Province
- 2 & 3 = Quebec
- 4 & 5 = Ontario
- 6 = Prairies
- 7 = Pacific Coast, Yukon and Northern Territory
- 8 = Non-citizens requiring work permits

The check digit is established on a MOD 10 base as follows: it is considered statistically sound and that there is an even distribution of terminal numbers.

	Specimen number	214	232	68	
(1)	Take each odd position number and double the value . . .	2	4	4	16
(2)	If the resultant value is greater than 9 (i.e. one digit) add the digits together . . .	2+4+4+	7	=	17
(3)	Take the value of the even positions . . .	2+4	+3+6	=	<u>15</u>
(4)	Total . . . . .				32
(5)	Go to the next MOD 10 value and then subtract: i.e. 40-32 = 8. 8 is therefore the check number and the whole SIN number will read . . . . .	214	232	688	

In the actual sampling exercise only those numbers with final check digit 5 are selected.

APPENDIX 3

Part of a table of Random Numbers

---

03 47 43 73 86	36 96 47 36 61	46 98 63 71 62	33 26 16 80 45
97 74 24 67 62	42 81 14 57 20	42 53 32 37 32	27 07 36 07 51
16 76 62 27 66	56 50 26 71 07	32 90 79 78 53	13 55 38 58 59
12 56 85 99 26	96 96 68 27 31	05 03 72 93 15	57 12 10 14 21
55 59 56 35 64	38 54 82 46 22	31 62 43 09 90	06 18 44 32 53
16 22 77 94 39	49 54 43 54 82	17 37 93 23 78	87 35 20 96 43
84 42 17 53 31	57 24 55 06 88	77 04 74 47 67	21 76 33 50 25
63 01 63 78 59	16 95 55 67 19	98 10 50 71 75	12 86 73 58 07
33 21 12 34 29	78 64 56 07 82	52 42 07 44 38	15 51 00 13 42
57 60 86 32 44	09 47 27 96 54	49 17 46 09 62	90 52 84 77 27
18 18 07 92 46	44 17 16 59 09	79 83 86 19 62	06 76 50 03 10
26 62 38 97 75	84 16 07 44 99	83 11 46 32 24	20 14 85 88 45
23 42 40 64 74	82 97 77 77 81	07 45 32 14 08	32 98 94 07 72
52 36 28 19 95	50 92 26 11 97	00 56 76 31 38	80 22 02 53 53
37 85 94 35 12	83 39 50 08 30	42 34 07 96 88	54 42 06 87 98
70 29 17 12 13	40 33 20 38 26	13 89 51 03 74	17 76 37 13 04
56 62 18 37 35	96 83 50 87 75	97 12 25 93 47	70 33 24 03 54
99 49 57 22 77	88 42 95 45 72	16 64 36 16 00	04 43 18 66 79
16 08 15 04 72	33 27 14 34 09	45 59 34 68 49	12 72 07 34 45
31 16 93 32 43	50 27 89 87 19	20 15 37 00 49	52 85 66 60 44
68 34 30 13 70	55 74 30 77 40	44 22 78 84 26	04 33 46 09 52
74 57 25 65 76	59 29 97 68 60	71 91 38 67 54	13 58 18 24 76
27 42 37 86 53	48 55 90 65 72	96 57 69 36 10	96 46 92 42 45
00 39 68 29 61	66 37 32 20 30	77 84 57 03 29	10 45 65 04 26
29 94 98 94 24	68 49 69 10 82	53 75 91 93 30	34 25 20 57 27
16 90 82 66 59	83 62 64 11 12	67 19 00 71 74	60 47 21 29 68
11 27 94 75 06	06 09 19 74 66	02 94 37 34 02	76 70 90 30 86
35 24 10 16 20	33 32 51 26 38	79 78 45 04 91	16 92 53 56 16
38 23 16 86 38	42 38 97 01 50	87 75 66 81 41	40 01 74 91 62
31 96 25 91 47	96 44 33 49 13	34 86 82 53 91	00 52 43 48 85
66 67 40 67 14	64 05 71 95 86	11 05 65 09 68	76 83 20 37 90
14 90 84 45 11	75 73 88 05 90	27 41 14 96 52	22 98 12 22 08
68 05 51 18 00	33 96 02 75 19	07 60 62 93 55	59 33 82 43 90
20 46 78 73 90	97 51 40 14 02	04 02 33 31 08	39 54 16 49 36
64 19 58 97 79	15 06 15 93 20	01 90 10 75 06	40 78 78 89 62
05 26 93 70 60	22 35 85 15 13	92 03 51 59 77	59 56 78 06 83
07 97 10 88 23	09 98 42 99 64	61 71 62 99 15	06 51 29 16 93
68 71 86 85 85	54 87 66 47 54	73 32 08 11 12	44 95 92 63 16
26 99 61 65 53	58 37 78 80 70	42 10 50 67 42	32 17 55 85 74
14 65 52 68 75	87 59 36 22 41	26 78 63 06 55	13 08 27 01 50