TECHNICAL LEAFLET SERIES

Number 8:

Archival Sampling: A Method of Appraisal and A Means of Retention

James Gregory Bradsher
and
Bruce I. Ambacher

Mid-Atlantic Regional
Archives Conference

Mid-Atlantic Regional Archives Conference

Reviewers
David Kepley
Karen D. Paul
Cynthia Requardt

Editor's Note: The idea for a technical leaflet series originated with the New York Caucus of MARAC. The series, as developed by the Caucus, is intended to provide brief, practical information about selected archival topics suitable especially for beginners in the profession.

The series editor welcomes proposals for future publications and comments from readers.

Acknowledgements: The editor would like to acknowledge the invaluable contributions of the reviewers and of Bruce Ambacher toward the preparation of this leaflet for publication.

The authors wish to acknowledge the assistance of Joyce Ray, Charles Dollar and Lida Churchville in the preparation of this leaflet.

Technical Leaflet Series

Kathryn Allamong Jacob, *Series Editor*

James Gregory Bradsher
Chair, MARAC Publications Committee

Karl J. Niederer
Chair, MARAC Steering Committee

ARCHIVAL SAMPLING: A METHOD OF APPRAISAL AND A MEANS OF RETENTION

by James Gregory Bradsher and Bruce I. Ambacher

INTRODUCTION

Appraising records is the most important aspect of archival administration. Appraisal decisions result in the destruction of most records. In deciding what to retain and what to destroy, archivists evaluate the research value of the records knowing that all records have some value. With limited resources, archivists also must consider the costs associated with processing, storing, and preserving those records accessioned into their archival repositories.

The appraisal decision has become more difficult due to the growing volume of records and changes in research methodologies. As governments, institutions, organizations, and other entities expand their activities and operations, and technological advances facilitate faster records creation, the result has been an explosion in volume. Although bulk is the bane of archivists, it is often that same bulk that many scholars desire to use in their research. Increasingly researchers are supplementing their traditional research methods with quantitative analysis. This enhances the research potential of many series of case-type files. Thus, it is these series which present the major problems of costs, particularly when the large series is a continuing one.

When appraising relatively large series of case-type records, especially those that are homogeneous in format and content, archivists are confronted with the task of delicately balancing researcher needs and archival interests. Frequently these large

series contain information in each case file with varying archival values, as well as information with little or no archival value. In the past such series often were kept in toto, preserving the valuable information along with the superfluous. Just as often such series were destroyed. Archivists assumed that the large volume did not warrant the cost of processing, preserving, and storing the entire series just for the few exceptional case files it contained. They also were uncertain whether the series would support quantitative analysis.

Thus, deciding what to do with large series containing records of mixed archival values is perhaps the most challenging task facing the archival appraiser. Archivists have been criticized for not being more objective in their appraisals, especially those relating to large, mixed-value series. To answer such criticisms, and to address the interests of researchers and the reality of bulk, archivists have begun sampling as an appraisal method, that is deciding the archival value of a body of records based on an examination of a scientifically created sample.

Four large sampling projects, completed within the past fifteen years, employed sampling both for appraisal and for retention. They are the appraisals of the records of the Massachusetts courts, the Federal Bureau of Investigation, the Department of Justice litigation case files, and the Quebec courts.

Sampling is taking any portion of a population as representative of that population or universe. To be truly representative, the portion of the population selected must be selected randomly. Sampling leads to more objective appraisals of the whole series, and points out what should be retained both quantitatively and qualitatively.

Quantitative sampling permits statistical analysis, but may not provide for the retention of files with exceptional research potential - those pertaining to important or interesting persons and subjects. Some archivists, therefore, have attempted qualitative selection to provide for the retention of the "exceptional" case files - those that generally would be destroyed if only some form of quantitative sample was retained.

Like other appraisal methods and decisions, there are advantages and disadvantages to using sampling both as a means of appraisal and as a method of selection and retention. This technical leaflet addresses some of the advantages and disadvantages of sampling and suggests when it is appropriate to sample and, if used, how it is accomplished.

SAMPLING AS AN APPRAISAL METHODOLOGY

When appraising records, archivists generally supplement or confirm the written descriptions of the records by looking at the records. They may examine the first, last, and several files from the middle of a series and make a subjective decision about the series' archival value. In most instances this is the easiest, quickest, and, as often as not, a satisfactory method of appraisal. In the past however, appraisals have been criticized for being too subjective. Past decisions to retain quantitative samples and/or qualitatively selected files also have been criticized, both because they were done in the first place, and because of the method used to select what was and was not retained.

A more objective appraisal can overcome these criticisms. Scientifically sampling the records to be appraised provides an objective appraisal. This method involves sampling a statistically valid number of files, collecting uniform information about each file, assigning a value to each file, analyzing the information, and making retention decisions about which files, if any, to retain.

The major advantage of sampling for appraisal is that it provides an objective, unbiased basis for appraisal decisions. The major disadvantages are that it can require more time to determine the appropriate sample, to actually sample the records, and to document the process. But often archivists fail to take the time to ensure that their appraisals are the best they can be.

Quantitative Sampling as an Appraisal Method

Sampling for appraisal involves five distinct steps. The process begins with the appraiser examining the records (the universe) to determine the recordkeeping practices, procedures, objectives, and content of the records (Step one). This review will help determine if one sample is required or if the records actually represent more than one "strata" for the purposes of sampling for appraisal. In sampling court cases, for example, separate strata would be used for criminal cases, civil cases, and bankruptcy cases. If more than one strata is present, each sample drawn should represent the same percentage of all records sampled as that strata represents of the total universe of records to be appraised. Thus if strata 1 contains the criminal cases and they represent 10% of all the records, then that sample should represent 10% of all the records sampled. Each sample is treated as a separate universe.

Step two is to determine the proper number of files to review. The number should be sufficient to ensure that the sample will reflect the characteristics of the entire series and that valid inferences about the universe can be drawn. Unfortunately, there is no preset formula. The larger the sample, the smaller the error. This is due to the increased probability of randomness in larger samples. Larger samples, however, require more time and expense but add little to the accuracy of the results. Too small a sample causes ambiguity about the representativeness of the results. Thus sample size is a "compromise" dependent upon the level of confidence required, the amount of error that can be tolerated, the homogeneity of the population being sampled, and the resources (time, staff, and money) available.

Archivists who are uncertain about determining sample size can consult an expert survey methodologist or utilize a computer based expert system software to assist in determining the sample. One such system is EX-SAMPLE, developed by The Idea Works.

In creating the samples for its 1980-1981 appraisal of the records of the Federal Bureau of Investigation, the largest appraisal project to employ sampling for appraisal, the National Archives developed the ratios listed in Table I to reflect the minimum number of cases to be sampled based on the number of cases in a universe.

The sample sizes suggested actually oversample universes of less than 1500 cases in order to compensate for the greater variation from the norm and to ensure that the sample will reflect the entire universe.

TABLE I
RATIO OF UNIVERSE TO SAMPLE SIZE

NUMBER OF CASES	SIZE OF SAMPLE
<40	5
40 - 300	10
301 - 750	15
751 -1500	20
>1500	35

The selection of records to be examined should be unbiased. Developing and implementing a proportionate systematic sample with a random start will eliminate any prejudgment which might be introduced by the appraisers. In a proportionate systematic sample the records selected for examination will be spaced evenly throughout the universe, that is every nth case is selected without regard to any other criteria such as size, subject, or completeness.

Implementing a random number for the first case further ensures impartiality and allows specific levels of confidence to be assigned the results based on the size of the sample. Use a table of random numbers, found in the back of most statistics text books, to

determine the number of the first case to be reviewed. The number for each subsequent case is determined by the interval between cases.

For example, you are examining the state motor vehicle violation case files. The universe consists of 350,000 case files dated between 1936 and 1988. You have determined that a sample size of 165 (every 2,122nd case) is required to achieve a "confidence level" of 0.95. Using a table of random numbers you pick 1652 as the first case to review, the next case would be 3774 (1652 + 2122), the third would be 5896 (3774 + 2111), etc.

The size of the sample should be sufficient to achieve a confidence level of 0.95. This widely accepted measure of accuracy means that if you created 100 different samples employing a proportionate systematic random sample each time - but with a different starting case number each time - 95 of the 100 samples would produce the exact statistical profile.

STANDARDIZING INFORMATION COLLECTION

Step three is to develop a "data collection sheet" on which to record similar information about each file reviewed. The first step in this process starts with an examination of the various files to determine what information they contain that should be collected in order to ensure an impartial evaluation. In creating the data collection sheet, the goal should be to create a form that is simple yet collects the basic and necessary information on which to base an appraisal. The data collection sheet should provide spaces for collecting routine information, such as the size of the file, certain

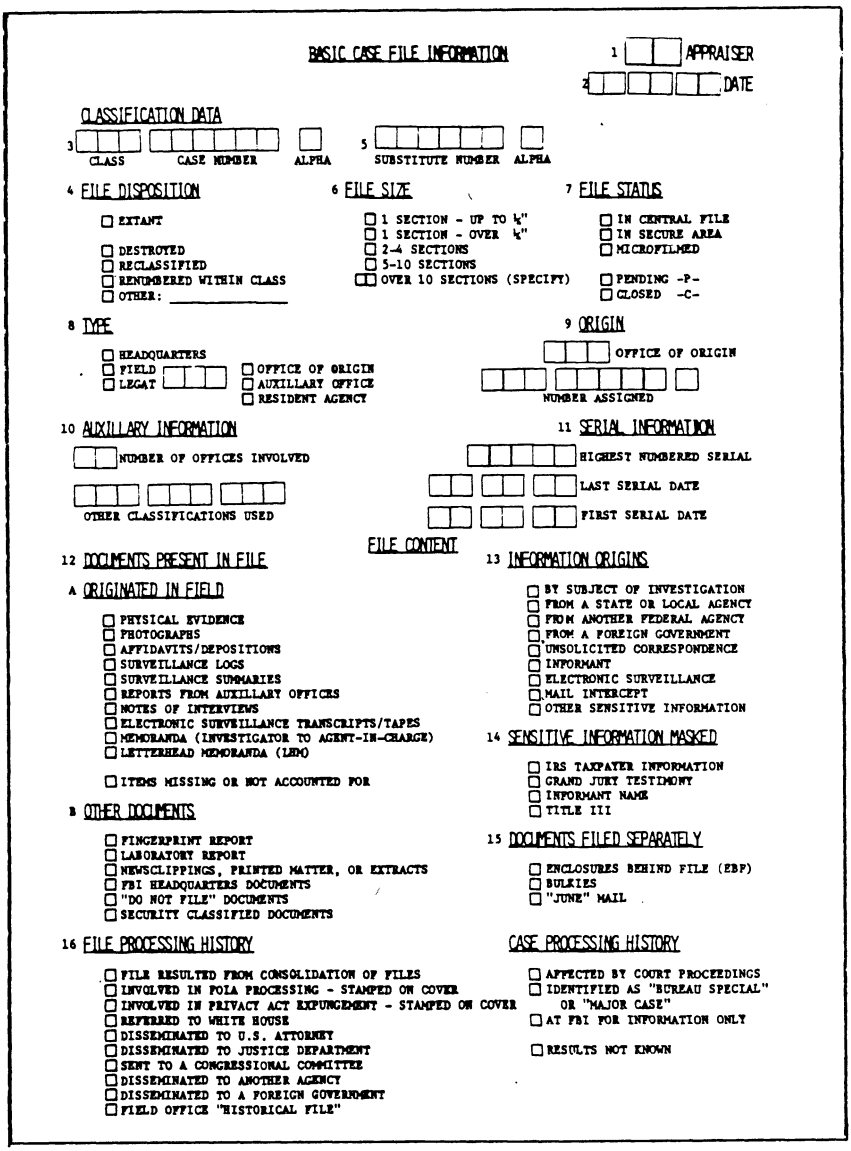
types of documents, the result of the case, and a rating of the file reviewed. It also should provide space for remarks. Figure 1 is an example of a data collection sheet used by the National Archives in its 1981 appraisal of FBI investigative case files. Every effort should be made to minimize the amount of information to be collected on the data sheets. Interactive computer based expert systems software can assist in determining which information is essential.

The most important piece of information on the data collection sheet is the reviewer's rating of the research potential of the case file. The rating of the files can easily be defined as high, medium, low, and none. The definitions of these ratings should be clearly understood by all who will complete the data collection sheets so that the aggregate data will be comparable and can be standardized. Four possible ratings and definitions could be:

High: evidence or information that is unique and of such substantive detail and richness that the case file could stand alone as a primary historical source.

Medium: evidence or information that is sufficiently rich that the case file significantly complements other historical sources.

Low: evidence or information so lacking in detail and richness that it is only a modest supplement to other historical sources and the case file has significance only in the context of other case files.



BASIC CASE FILE INFORMATION

1 ☐☐ APPRAISER

2 ☐☐☐☐☐☐ DATE

CLASSIFICATION DATA

3 ☐☐☐ ☐☐☐☐☐☐ ☐ 5 ☐☐☐☐☐☐ ☐

CLASS CASE NUMBER ALPHA SUBSTITUTE NUMBER ALPHA

4 FILE DISPOSITION

☐ EXTANT

☐ DESTROYED
☐ RECLASSIFIED
☐ RENUMBERED WITHIN CLASS
☐ OTHER: _____

6 FILE SIZE

☐ 1 SECTION - UP TO ½"
☐ 1 SECTION - OVER ½"
☐ 2-4 SECTIONS
☐ 5-10 SECTIONS
☐☐ OVER 10 SECTIONS (SPECIFY)

7 FILE STATUS

☐ IN CENTRAL FILE
☐ IN SECURE AREA
☐ MICROFILMED

☐ PENDING -P-
☐ CLOSED -C-

8 TYPE

☐ HEADQUARTERS
☐ FIELD
☐ LEGAT ☐☐☐

☐ OFFICE OF ORIGIN
☐ AUXILLARY OFFICE
☐ RESIDENT AGENCY

9 ORIGIN

☐☐☐ OFFICE OF ORIGIN

☐☐☐ ☐☐☐☐☐☐ ☐ NUMBER ASSIGNED

10 AUXILLARY INFORMATION

☐☐ NUMBER OF OFFICES INVOLVED

☐☐☐ ☐☐☐ ☐☐☐ OTHER CLASSIFICATIONS USED

11 SERIAL INFORMATION

☐☐☐☐☐ HIGHEST NUMBERED SERIAL

☐☐ ☐☐ ☐☐ LAST SERIAL DATE

☐☐ ☐☐ ☐☐ FIRST SERIAL DATE

FILE CONTENT

12 DOCUMENTS PRESENT IN FILE

A ORIGINATED IN FIELD

☐ PHYSICAL EVIDENCE
☐ PHOTOGRAPHS
☐ AFFIDAVITS/DEPOSITIONS
☐ SURVEILLANCE LOGS
☐ SURVEILLANCE SUMMARIES
☐ REPORTS FROM AUXILLARY OFFICES
☐ NOTES OF INTERVIEWS
☐ ELECTRONIC SURVEILLANCE TRANSCRIPTS/TAPES
☐ MEMORANDA (INVESTIGATOR TO AGENT-IN-CHARGE)
☐ LETTERHEAD MEMORANDA (LHM)

☐ ITEMS MISSING OR NOT ACCOUNTED FOR

B OTHER DOCUMENTS

☐ FINGERPRINT REPORT
☐ LABORATORY REPORT
☐ NEWSCLIPPINGS, PRINTED MATTER, OR EXTRACTS
☐ FBI HEADQUARTERS DOCUMENTS
☐ "DO NOT FILE" DOCUMENTS
☐ SECURITY CLASSIFIED DOCUMENTS

13 INFORMATION ORIGINS

☐ BY SUBJECT OF INVESTIGATION
☐ FROM A STATE OR LOCAL AGENCY
☐ FROM ANOTHER FEDERAL AGENCY
☐ FROM A FOREIGN GOVERNMENT
☐ UNSOLICITED CORRESPONDENCE
☐ INFORMANT
☐ ELECTRONIC SURVEILLANCE
☐ MAIL INTERCEPT
☐ OTHER SENSITIVE INFORMATION

14 SENSITIVE INFORMATION MASKED

☐ IRS TAXPAYER INFORMATION
☐ GRAND JURY TESTIMONY
☐ INFORMANT NAME
☐ TITLE III

15 DOCUMENTS FILED SEPARATELY

☐ ENCLOSURES BEHIND FILE (EBF)
☐ BULKIES
☐ "JUNE" MAIL

16 FILE PROCESSING HISTORY

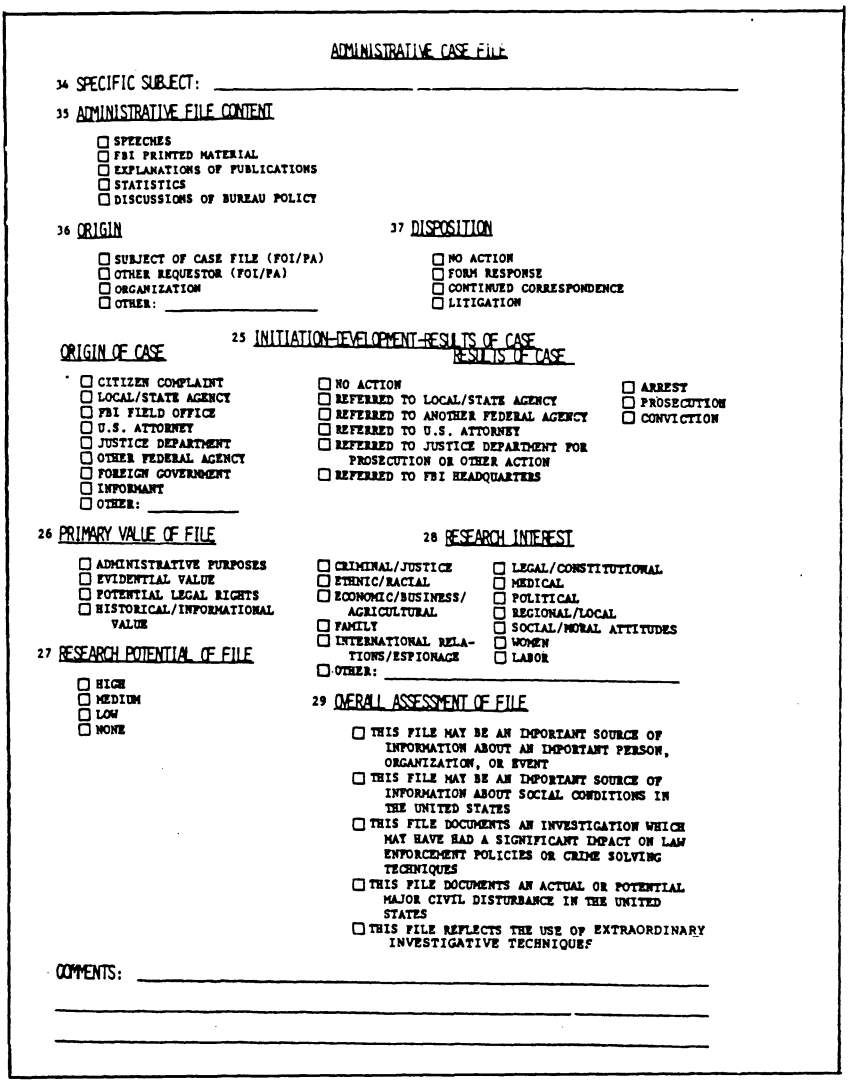
☐ FILE RESULTED FROM CONSOLIDATION OF FILES
☐ INVOLVED IN FOIA PROCESSING - STAMPED ON COVER
☐ INVOLVED IN PRIVACY ACT EXPUNGEMENT - STAMPED ON COVER
☐ REFERRED TO WHITE HOUSE
☐ DISSEMINATED TO U.S. ATTORNEY
☐ DISSEMINATED TO JUSTICE DEPARTMENT
☐ SENT TO A CONGRESSIONAL COMMITTEE
☐ DISSEMINATED TO ANOTHER AGENCY
☐ DISSEMINATED TO A FOREIGN GOVERNMENT
☐ FIELD OFFICE "HISTORICAL FILE"

CASE PROCESSING HISTORY

☐ AFFECTED BY COURT PROCEEDINGS
☐ IDENTIFIED AS "BUREAU SPECIAL" OR "MAJOR CASE"
☐ AT FBI FOR INFORMATION ONLY

☐ RESULTS NOT KNOWN

Figure 1. Sample Data Collection Sheet
Side 1



ADMINISTRATIVE CASE FILE

34 SPECIFIC SUBJECT: _____

35 ADMINISTRATIVE FILE CONTENT

- ☐ SPEECHES
- ☐ FBI PRINTED MATERIAL
- ☐ EXPLANATIONS OF PUBLICATIONS
- ☐ STATISTICS
- ☐ DISCUSSIONS OF BUREAU POLICY

36 ORIGIN

- ☐ SUBJECT OF CASE FILE (FOI/PA)
- ☐ OTHER REQUESTOR (FOI/PA)
- ☐ ORGANIZATION
- ☐ OTHER: _____

37 DISPOSITION

- ☐ NO ACTION
- ☐ FORM RESPONSE
- ☐ CONTINUED CORRESPONDENCE
- ☐ LITIGATION

25 INITIATION-DEVELOPMENT-RESULTS OF CASE

ORIGIN OF CASE

- ☐ CITIZEN COMPLAINT
- ☐ LOCAL/STATE AGENCY
- ☐ FBI FIELD OFFICE
- ☐ U.S. ATTORNEY
- ☐ JUSTICE DEPARTMENT
- ☐ OTHER FEDERAL AGENCY
- ☐ FOREIGN GOVERNMENT
- ☐ INFORMANT
- ☐ OTHER: _____

RESULTS OF CASE

- ☐ NO ACTION
- ☐ REFERRED TO LOCAL/STATE AGENCY
- ☐ REFERRED TO ANOTHER FEDERAL AGENCY
- ☐ REFERRED TO U.S. ATTORNEY
- ☐ REFERRED TO JUSTICE DEPARTMENT FOR PROSECUTION OR OTHER ACTION
- ☐ REFERRED TO FBI HEADQUARTERS
- ☐ ARREST
- ☐ PROSECUTION
- ☐ CONVICTION

26 PRIMARY VALUE OF FILE

- ☐ ADMINISTRATIVE PURPOSES
- ☐ EVIDENTIAL VALUE
- ☐ POTENTIAL LEGAL RIGHTS
- ☐ HISTORICAL/INFORMATIONAL VALUE

27 RESEARCH POTENTIAL OF FILE

- ☐ HIGH
- ☐ MEDIUM
- ☐ LOW
- ☐ NONE

28 RESEARCH INTEREST

- ☐ CRIMINAL/JUSTICE
- ☐ ETHNIC/RACIAL
- ☐ ECONOMIC/BUSINESS/ AGRICULTURAL
- ☐ FAMILY
- ☐ INTERNATIONAL RELATIONS/ESPIONAGE
- ☐ OTHER: _____
- ☐ LEGAL/CONSTITUTIONAL
- ☐ MEDICAL
- ☐ POLITICAL
- ☐ REGIONAL/LOCAL
- ☐ SOCIAL/MORAL ATTITUDES
- ☐ WOMEN
- ☐ LABOR

29 OVERALL ASSESSMENT OF FILE

- ☐ THIS FILE MAY BE AN IMPORTANT SOURCE OF INFORMATION ABOUT AN IMPORTANT PERSON, ORGANIZATION, OR EVENT
- ☐ THIS FILE MAY BE AN IMPORTANT SOURCE OF INFORMATION ABOUT SOCIAL CONDITIONS IN THE UNITED STATES
- ☐ THIS FILE DOCUMENTS AN INVESTIGATION WHICH MAY HAVE HAD A SIGNIFICANT IMPACT ON LAW ENFORCEMENT POLICIES OR CRIME SOLVING TECHNIQUES
- ☐ THIS FILE DOCUMENTS AN ACTUAL OR POTENTIAL MAJOR CIVIL DISTURBANCE IN THE UNITED STATES
- ☐ THIS FILE REFLECTS THE USE OF EXTRAORDINARY INVESTIGATIVE TECHNIQUES

COMMENTS: _____

Figure 1. Sample Data Collection Sheet
Side 2

None: evidence or information so ordinary and routine that the case file has no significance as a historical source even in the context of other case files.

Gathering the Information and Recording it on the Data Collection Sheet

The fourth step is gathering the information on which to base an analysis of the series. The appraiser should examine each case file or record as thoroughly as possible. He/she should ascertain the presence or absence of each type of information sought, examine the record to determine the research value of that case file or record, assess its value as part of the universe, and complete the data collection sheet.

Whenever the size of the universe to be examined justifies it, more than one person should review the cases to ensure that certain prejudices and arcane interests don't skew the results. Multiple reviewers provide additional perspectives upon which to base the final appraisal judgment. This is especially useful when the sample is large.

The remarks section should contain information from that record which is special or unique. This information can be used during the decision-making process to complement or offset the purely statistical results. It also can be used in writing up the appraisal. Finally, it can be used to embellish the series descriptions in a finding aid.

Analyzing the Results and Developing Disposition Decisions

Once the data is collected, the fifth step is to analyze it to ascertain if the series has value and to find correlations between the various data collection elements and the rankings. If the information is entered into a computer system it can be easily manipulated; correlations between the data and the archival value can be easily ascertained even for voluminous bodies of records. The computer computations can make clearer possible relationships between the research value of a case file and size; the length of time it was open or active; the presence or absence of certain information, activities, or document types; the subject; the final disposition; or any other of a number of relevant aspects of the records. Numerous correlations can be made. But it can be time consuming and it may be beyond the resources of some archives. The analysis can be done manually if there is not a lot of information data on each collection sheet or a lot of sheets.

Analyzing the data may make it readily evident that the series should be kept or destroyed in its entirety, either because the files showed substantial or little or no research value. If this is the case, then the process stops. Sampling for appraisal was a worthwhile effort, because it provided a more objective appraisal method than simply looking at a few case files and making a subjective decision. And equally important you have a record to document your appraisal decision.

Frequently the analysis will indicate that some files within a series have great value while others have little or none. This was the case with the three major appraisal projects involving sampling. To enhance the statistical accuracy of the samples it retained from

the Federal Bureau of Investigation the National Archives contracted with Westat, Inc. to develop a retention table. Table II illustrates the relationship that exists between the number of historically interesting cases identified in the sample of cases appraised and the percentage of historical cases which should be present in the entire series being examined when a 0.95 level of confidence is achieved in selected cases or records to be appraised. The Table can be used with any statistically selected sample to help establish retention criteria. For example, if your sample of 35 cases had three cases which had research potential, twenty percent of all cases in that universe have similar research potential.

Of course this also may have been determined subjectively by a cursory review of the series. But if either method of appraisal determines that some files have value and that retention is warranted, then three disposition options present themselves: retaining a quantitative sample, retaining qualitatively selected files, and/or doing both.

In deciding the most appropriate disposition, consideration should be given to the anticipated use of the files in the series; the volume that may be retained; the cost of storage, preservation, and processing of the files; and whether the series is homogenous or heterogenous in nature. Consideration also should be given to whether the series is a continuing one with potential for substantial growth. All of these factors are equally important in determining whether to keep or destroy the series in its entirety or whether to retain some of it, and if so, what and how much.

In making the disposition decisions, it might be worthwhile to consult subject-matter experts. But in doing so, always remember

TABLE II
ARCHIVAL VALUE MATRIX

Decision rules which will preserve the 0.95 confidence level for retaining a universe with more than the specified percentage of records having historical or archival value.
RULE: Destroy if fewer than the listed number of sampled case files has historical or archival value.

	PERCENT						
	10	20	25	30	40	50	60
SAMPLE SIZE							
10	*	*	*	1	2	2	3
15	*	1	1	2	3	3	6
20	*	1	2	3	4	6	8
35	1	3	5	6	9	13	17
42	1	4	6	8	11	15	20
48	2	5	7	9	13	18	23
51	2	6	8	10	14	19	24
54	2	6	8	11	16	21	28
78	4	10	13	17	23	32	40
141	9	21	27	34	47	61	76
171	11	26	34	42	58	75	93
177	12	27	35	44	61	78	96
258	18	42	54	66	91	116	142
333	25	55	71	87	119	158	191
354	27	59	76	93	127	162	198

* Sample size is too small to preserve the 0.95 confidence level, even if no cases of historical interest are found.

that any researcher can think of a reason for keeping everything, especially when it pertains to records that relate to their field of study. Final disposition decisions always must be made by archivists. They have a better understanding of the uses to which records will be put as well as an understanding of the implications of the costs involved in retaining records.

QUANTITATIVE SAMPLING AS A RETENTION CRITERION

After the analysis and consultation you may decide to retain a quantitative sample to facilitate use of the records for statistical analysis. Generally this will be the case when most files reviewed are voluminous, homogeneous in nature, and seem to have some value, but not enough to warrant retaining the whole series.

It is important to remember that not all types of records are suitable for quantitative sampling. As a general rule, only those series which contain records that are essentially homogeneous in character can be sampled successfully. In such series each individual file contains similar records and information and the variability between the files will be small; thus the statistical precision will be high.

When the documentation and information within a series is heterogenous, sampling will produce a statistical bias. Thus quantitative sampling should not be done to such series. Because statistical sampling generally can be applied only to documents which contain mathematically quantifiable information, quantitative sampling normally cannot be applied to non-textual series of

records. This is because those files generally deal with unique subjects.

Quantitative sampling offers both advantages and disadvantages. The major advantage is that it is easy to do. Second, it will result in the retention of a representative sample containing both routine case files for evidential purposes and some significant case files. The disadvantage, of course, is that it may result in the destruction of valuable information contained in some case files. This can be minimized if qualitatively selected files also are retained.

Evidential and Informational Samples

There are two basic types of quantitative samples. The first is an evidential sample. This small sample is retained to show how transactions or routine activities were handled. The other is an informational sample. This much larger, statistically valid sample of 1500 or more cases provides an aggregate of information on the individuals or activities involved in the transaction. This latter sample can be useful to those who employ statistical techniques in their research.

Determining the size of a sample is not always easy. But it should be statistically sound, that is, large enough to contain all the key elements of evidence or information within the series; thereby allowing researchers to conduct valid statistical analysis.

As a general rule, the size of the sample depends upon the research potential, the degree of statistical "confidence" required, the intended purpose of the sample, and the volume of records. While

the size of the sample will vary according to the nature of the documentation, it is always true that a larger sample will be more representative of the whole series and provide a more satisfactory coverage of the whole and therefore will be more likely to meet researcher requirements.

Often a five per cent sample is a sufficient standard of retention. There may be cases, however, where, because of the size of the series and its content, a smaller sample of only one percent will be valid. Conversely, there might be examples where a ten per cent retention is justified. Because of the importance of determining the size of the sample, it might be worthwhile to consult a subject expert or a statistician or utilize a computer based expert system.

The exact nature of the sample will depend upon the disposition instructions established as part of the appraisal process. A simple systematic sample would be created by identifying and marking every nth file. Use a random number to identify the first case file designated for retention. If the selected numbered case file is missing in a numerically arranged records universe, then, alternating, substitute the next higher and then the preceding lower numbered case file.

QUALITATIVE SELECTION AS A RETENTION CRITERION

Even in a series of relative insignificance, which is suitable for quantitative sampling, there may be case files of great research potential. Thus, another disposition option is to retain qualitatively selected case files, whether or not quantitatively sampled case files

are retained. It should be noted that qualitative selection is really not sampling, since the intention is not to represent the totality, but only to keep the "exceptional" or "most important" to meet specific criteria.

Generally, decisions to retain certain qualitatively selected files result from a sampling for appraisal method which reveals that it is likely that case files containing certain characteristics have research value, and thus should be retained as a part of or independent from the sample. These case files may be those that relate to certain subjects, contain certain types of documentation, and/or result in certain transactions. They may meet specific retention criteria such as all files over one inch thick, all files open more than one year, all files forwarded to another agency, or all files opened on the fifteenth of each month.

Based on several major appraisal projects it is relatively safe to say that the most likely candidate for research potential correlation is the size of the file and its value. In other words, if most or all case files over a certain size have a rating of some value, then all files over a certain size should be retained.

Sampling for appraisal provides an objective basis for selective retention within a series. Such appraisal and disposition decisions, though time-consuming, result in the best appraisals. Quantitative selection can be supplemented with qualitative, i.e., subjectively-based selection criteria. These may include criteria such as case files containing information about cases or subjects mentioned in the annual reports of specific agencies or organizations; cases involving extensive litigation; cases setting major legal or policy precedents; cases receiving marked attention by the mass media;

cases involving significant legislative or administrative interest or investigation; and cases regarded by authorities in that particular field of study as having special importance or uniqueness. These include the type of criteria archivists have been including in disposition instructions for decades.

In developing disposition instructions for qualitative selections the selection criteria must be either very specific or as comprehensive as they can reasonably be made. Failure to do so will result in the right case files not being retained and/or the wrong ones being retained. Thus, specific objective criteria leave little room for interpretation and increase the likelihood that the right case files will be retained. Archivists may wish to consult with subject area specialists to develop criteria.

Qualitative selection is quite tempting; it reduces bulk and retains the "best." But as a general rule, the more homogeneous the series of records, the less appropriate are qualitative selections, because of their "exceptional" character and built-in bias. This does not preclude retaining the exceptional case files, but as another general rule, qualitative samples should not be taken of homogeneous series in lieu of a statistically valid sample. If qualitatively selected files within a homogeneous series are to be retained, they should be selected after the statistical sample is developed to ensure statistical validity. When qualitative samples are retained they should be labeled as such and kept separate from the quantitative samples.

Qualitative samples have both advantages and disadvantages. The major advantage of qualitative selections is that generally they ensure the retention of much, if not virtually all, of the valuable information. The disadvantages of qualitative selections is that they

are time consuming to implement, they do not result in the retention of all of the valuable information, and they will not result in the retention of the routine case files if not taken in conjunction with quantitative samples. Further, they provide researchers with a distorted perspective of the content and research potential of the series; they inflate the value of the series.

RETAINING BOTH QUANTITATIVE SAMPLES AND QUALITATIVE SELECTIONS

In some instances, with very large series of mixed archival values, the appraisal archivist may decide to retain both quantitative samples and qualitatively selected files. The quantitative sample is determined first. Then the qualitative selection is made from the remaining case files. The result will be the retention of both the routine case files and the most important historical information.

IMPLEMENTATION AND PERIODIC EVALUATION

All quantitative samples and qualitative selections should be physically separated and well-marked. They also should be labeled in a way that makes the distinction between what is acceptable for quantitative analysis and what is not, clearly apparent.

The appraisal report and accessioning dossier must contain full notes of every action taken and of the various elements of the sample if more than one has been taken. Subsequent finding aids should clearly explain the reasons for the sampling and the methodology used. They also should clearly describe any

qualitative selection criteria. Clear descriptions make users more fully aware of what was done, why it was done, how it was done and, when applicable, where they can find the other elements of the series.

After the initial sample has been accessioned, the archivist should determine whether researchers are using the records, how they are using them, and what complaints, if any, they have about what was and was not retained. If the records are from a continuing series, it is quite possible in time, to change the disposition and/or provide for the internal disposal of the quantitative and/or qualitative portions of the series.

CONCLUSION

Sampling records for appraisal is an excellent method to determine the value of a series and the files within it. It may be a time consuming process, but when used appropriately, it is well worth the effort.

Quantitative sampling and qualitative selections of records for retention are effective methods to determine the disposition of the great volume of some series of records. They also reflect a growing realization, because of costs associated with preserving large bodies of records, that not everything can be saved, and also that within some series, only some case files should be saved.

Archival purists dismiss sampling by saying that if part of a series is worth retaining then the whole series should be retained. They argue that selective retention within a series is "tampering" with the

integrity of the series. They base this on the belief that the integrity of the series always should be maintained and the files kept in their context. This argument also is based on past misuses of sampling such as vague or subjective qualitative criteria such as all "significant" cases. Frequently in the past, neither archivists nor researchers were satisfied with the disposition decisions, the selection criteria or the samples retained. The present, more fully developed sampling criteria should reduce their anxiety about the decisions that are made.

Disposition decisions involving quantitative samples and qualitative selections should be undertaken thoughtfully and should not be adopted unless there is no alternative solution. Sampling should be implemented infrequently. It should result from careful consideration of the disposition options and determination that the universe meets the criteria developed above for quantitative sampling.

Leonard Rapport, in his article "In the Valley of Decision: What to do about the Multitude of Files of Quasi Cases" in the Spring 1985 *American Archivist*, observed that when appraising records there is "one immutable law: there are no perfect appraisals and the best appraisal is the one that does the least harm." Thus in doing the appraisal as well as deciding what, if anything, should be retained, archivists always need to look at all the possibilities and then act and decide with the intent of doing the least harm to their collections or to the researchers they serve. Knowing full well that not every appraisal will be perfect, the archivist can nonetheless, be confident that the decisions were reached with a full knowledge of the appraisal and disposition options, which include quantitative sampling and qualitative selection.

SELECT BIBLIOGRAPHY

Archives National du Quebec. **Report of the Interministerial Committee on Court Records** (Montreal, Quebec: 1989. Abridged English-language edition 1991). ix, 113 pp.

Carvelho, Joseph. **Archival application of mathematical sampling techniques.** *Records Management Quarterly* **18** (January 1984): 60-62.

Cook, Terry. **"Many are called but few are chosen": appraisal guidelines for sampling and selecting case files.** *Archivaria* (Canada) no. 32 (Summer 1991): 25-50.

Dean, Jenny and Wendy Southern. **The practice of sampling in the disposal of Commonwealth Records.** *Archives and Manuscripts* (Australia) 18, no. 1 (May 1990): 53-62.

Hindus, Michael, Theodore M. Hammett and Barbara M. Hobson. **The Files of the Massachusetts Superior Court, 1859-1959.** Boston: G.K. Hall and Co., 1980. xii, 264pp.

Hull, Felix. **The use of sampling techniques in the retention of records.** RAMP Study PGI 81\WS\26. Paris: UNESCO, 1981. iv, 64 pp.

Kepley, David. **Sampling in archives: a review.** *American Archivist* 47 (Summer 1984): 237-42.

Lewinson, Paul. **Archival Sampling.** *American Archivist* 20, no. 4 (October 1957): 291-312.

National Archives and Records Service. **Appraisal of the Records of the Federal Bureau of Investigation.** Washington, DC: NARS, 1981. 2 vols. ca. 1500 pp.

Sly, Margery N. **Sampling in an archival framework: mathoms and manuscripts.** *Provenance* 5 (Spring 1987): 58-59.

Smith, K.J. **Sampling and selection: current policies.** In ***The records of the nation: the Public Record Office, 1838-1988, the British Record Society, 1888-1988***, edited by G.H. Martin and Peter Spufford, 48-59. Woodbridge, Suffolk, England: Boydell Press, 1990.

Stoddart, Mark. **Sampling - the much maligned archival choice - a New Zealand view.** *Archifacts* (New Zealand) (October 1989): 19-32.