



Internationaler Archivkongress 2004
23.-29. August - Wien - Österreich

Archive, Gedächtnis und Wissen



Appraisal: Process, Selection Criteria, Sampling

Stephen Twigge
Cassandra Findlay
Vincent Doom
Markku Leppanen

Sampling - An Obsolete Method for the Appraisal and Disposal of Modern Records ?

Markku Leppänen

Member of the Committee on Appraisal (ICA/CAP)

Senior Research Officer

National Archives of Finland

1. Introduction

Main themes of my presentation are: sampling in general, different sampling methods, practical experiences of sampling with regard to traditional records and sampling in electronic records management. First, I would like to raise some provocative questions:

- Is the use of sampling an obsolete method for the appraisal and disposal of modern records (i.e. electronic records)?
- Why should we take samples of records/files, if we are able to preserve all of them?
- Why should we use sampling techniques, if the costs of preservation are the same whether you keep only the sample or the total information, the base, and if the researchers always want to get the whole material for their own purposes?
- Is the use of sampling a relevant option any more?

I will not give you here any final answers to those questions; I rather let them be hanging on the screen for your further consideration. These questions will also be discussed by the members of the Committee on Appraisal (ICA/CAP), as we are working with the draft of guidelines concerning the appraisal and selection of records.

I have not found any recent comprehensive study on sampling practice on international level. I wonder, is this not a matter of general interest any more? For the background I will use my own study from 1987. It was about the theory and practice of sampling in Finland and some other European countries. I will also give you some examples of the practice of sampling in Finland. But let me first explain some basic terms and facts about sampling in the framework of appraisal and disposal of records.

2. Sampling in general

The concept of *sampling* is a loan word from statistics. In Finland the first decision on sampling of records was made more than 60 years ago.

Sampling is a form of appraisal and disposal, a sort of compromise that is used especially in connection with large, in some respect homogeneous series or other groups of records. The goal of sampling is to preserve a sample of such quantity and quality that, with regard to the whole population (the whole totality of records), enables inferences about the records creator, its functions and operational setting as well as about the form, functions, and content of records.

By the concept of *sample* is meant that on the basis of some sampling technique a part is separated for preservation from the totality of records designated for destruction. A sample is meant for permanent retention and it is supposed to serve as a basis for inferences about the characteristics and content of the totality of records; it is judged to be appropriate as regards the objectives and methods of future research and to be generally interesting from the research point of view.

Sampling is an option when:

- the information content of records is relatively thin and/or the amount of unique information is small
- the volume of records is great and/or
- the records are relatively uniform as regards form and partly as regards content, too.

The filing order of the records has to be taken into consideration when we are deciding on sampling methods. Sometimes the sampling decision may be the reason for the change of the filing order.

Provided that sampling methods have been appropriate, researchers can rely on the samples and at the same time one has saved a lot of money due to the fact that it has not been necessary to preserve all of the records in archives rooms to occupy space. The size of the sample must be determined individually case by case. In most cases the size of the sample has varied between 5 and 20 percent.

The sampling of records in the framework of different appraisal strategies

I will first emphasize that sampling decisions have been and will be exceptions. The appraisal of records should not give anyone an opportunity to perform useless tricks. The decisions of sampling should be planned carefully and be coordinated with other appraisal decisions concerning the connected material elsewhere.

There are several alternative strategies for the appraisal of records. It seems to me that the macro appraisal strategy with the functional approach will be the main alternative for most national archives and other organisations responsible for the appraisal of public records. This is also the case in Finland.

Let me give you two examples that will show the impacts of the appraisal strategy to the practice of sampling of records.

1. If the principal idea of appraisal is to preserve information that gives you documentation (evidence) of the social functions, policy, procedures and activities of a records creator, and if the statistics of the results of the organisation seem to be comprehensive enough, you will probably not preserve the related mass of records showing the impacts of the policy on the micro level (on the level of citizens) at all. This kind of appraisal decisions will however preserve enough evidence for the researchers of the public administration.
2. But if you pay more attention to the daily life and social problems of citizens as well as the impacts of the public administration on people, you have to preserve personal records/files created by public organizations, either in its totality or in samples.

In Finland we have decided to walk on the middle of the road, which means that we are using the macro appraisal strategy with the functional approach and considering both the evidential value and informational value. We try to preserve for researchers as many-sided and representative archival materials as possible that certainly will contain also records and data of the activities and everyday conditions of citizens. We could not have implemented our appraisal strategy for the records of public administration without using sampling methods.

3. Different sampling methods and practical experiences

What kind of sampling methods there are and how can we use them in order to reach the goals of our appraisal policy? Let me give you a list of sampling methods that have been used in the appraisal and disposal of records:

1. *systematic sampling method* (based on the internal order of pieces; e.g. every 10th piece will be preserved permanently)
2. *periodic sampling method* (for chronological series of records)
3. *time-series sampling method* (for chronological series of records)
4. *topographical (geographical) sampling methods*
5. *alphabetical sampling method* (used e.g. in Germany)
6. *qualitative (purposive) sampling methods*
7. *birthday sampling method* (for personal files; it is a sort of “Grand Solution” in the Nordic countries)
8. *multi sampling methods* (combined and successive sampling methods)

Most of these sampling methods are widely used and there is, I think, no need to explain them here. Let me, however, explain the *birthday sampling method* that is perhaps unfamiliar for most of you, but is widely used in

the Nordic countries (Finland, Sweden, Norway and Denmark), nowadays even in Estonia.

In Finland the National Archive Service has made sampling decisions on the disposal of certain personal records/files of e.g. labour, social welfare, tax and prison administrations. Let's take one example: as regards the hospitals records, only those relating to people who were born on the 18th and 28th days of each month are retained permanently. As the same dates have been used also in other governmental and municipal offices and institutes, the National Archive Service can guarantee an extensive pool of information on the same persons, which can be studied throughout their life with the preserved archival information. In cases involving very large bases (e.g. if we have records or files of the whole population of Finland, five million citizens) the samples, or at least a great number of their variables, can be fairly representative. So, by using the *birthday sampling method* we have a possibility to draw general conclusions of the everyday life of citizens in Finland.

Practical experiences of sampling with regard to traditional records

In public administration a lot of space and costs have been saved by using sampling methods. In Finland the researchers have not made complaints because of the sampling decisions, yet. Some sampling decisions made by the National Archive Service have been too labour intensive, some too complicated and the size of some samples have been changed by making reappraisal decisions (reduction decisions).

4. Sampling in electronic records management

A short analysis of modern records creation and sampling

The records management of the public and private sector has changed rapidly into the electronic form. The role of paper records in the administration is decreasing, but it will not disappear for years. The electronic records management will cause a lot of problems. We still have to make appraisal decisions and dispose a lot of records and files in order to decrease the costs of the preservation and at least to increase the accessibility of the rest of the information. What kind of impacts these changes have caused to the practice of sampling of records?

Electronic records in records management systems - is sampling irrelevant?

The sampling seems to be - at least in Finland - irrelevant when we are appraising electronic records in records management systems. These electronic systems, which are extensively maintained in the public administration, contain both reference data (meta-data) and electronic records, either digitized or born in electronic form. Such material needs to be appraised, but there is, I think, no need for sampling.

Databases - sampling as an option

There are a great variety of databases or registers maintained in public administration. In Finland most of them have been appraised without using any sampling method. There are several alternatives for that:

1. preservation of all information,
2. disposal of all information and
3. preservation of some information of all cases/objects (the rest will be destroyed).

Another example of the practice of sampling in Finland: the National Archive Service has decided to use the *time-series sampling method* combined with the *birthday sampling method* on the nation wide database maintained by the taxation administration. Here the decision made in 1980 has been applied unchanged. For permanent preservation will be sampled the data only from the years ending with zero and number five and only the data concerning individuals who are born on 8th, 18th and 28th days of each month in such years. So we are using the same principles and criteria of appraisal and the same sampling methods regardless of the form of the information.

Unfortunately the design of some databases have been planned and implemented without taking into consideration the preservation and disposal of information. I am sure that you will face the same kind of problems in your

country. In cases like this the sampling is not an option. The question is then only, whether to keep the database as a whole or let it be vanished.

My final example of the sampling practice is as follows: if the amount of information in a public register, that has minor research value, is small, and there are a lot of textual records connected to the register, we may decide to preserve the register and only a sample of textual records permanently. Sometimes the register seems to contain enough information to satisfy the information needs of the researchers; then only the register is preserved permanently.

5. Sampling - not an obsolete method

The public administration will produce, receive and maintain paper records for years and years ahead. Therefore, we will have to find ways to manage *hybrid archives*, where we have data both in paper and electronic forms, even in microform. The *hybrid archives* will cause challenges for the appraisal and disposal projects and may have impacts also for the practice of sampling.

If we consider sampling as an option, as I do, we have to avoid too labour intensive sampling decisions. We should take into consideration the different branches of research when we are appraising electronic data, too. When a sampling decision has been made and a method fixed we should not change it lightly; certain conservatism is a virtue here.

As regards the appraisal and disposal of personal files, the "*Grand Solution*" (the wide use of *birthday sampling method*) in the Nordic countries seems to offer possibilities to a wide spectrum of research, including research that aims at following individual life histories from beginning to end.

Finally, in case you are interested in the practice of sampling techniques and e.g. of the *birthday sampling method*, I refer to my study in English in the following web site: <http://www.narc.fi/parnu/3.pdf>