



VSA Verein Schweizerischer Archivarinnen und Archivare
AAS Association des Archivistes suisses
AAS Associazione degli archivisti svizzeri
UAS Uniun da las archivarias e dals archivaris svizzers

GRUPE DE TRAVAIL ÉVALUATION

Échantillonnage

Terminologie et procédure

Le présent document définit les principaux termes relevant de l'échantillonnage (cf. schéma en annexe). Il a pour but de simplifier et d'unifier la terminologie utilisée afin de supprimer autant que faire se peut la confusion qui prévaut parfois dans le vocabulaire utilisé dans ce domaine. Conçu comme un guide pratique, il fournit des conseils et des astuces afin de faciliter le processus d'échantillonnage dans les archives suisses.

Dossiers sériels

Les dossiers sériels sont des séries de dossiers particuliers de même forme retraçant une ou plusieurs opérations précisément définies. Ils reflètent des procédures administratives analogues et contiennent différents dossiers particuliers qui se succèdent les uns aux autres, comme dossiers parallèles de même rang, sans lien interne entre les uns et les autres.¹ Des exemples typiques de dossiers sériels sont les dossiers fiscaux ou les dossiers par cas de l'aide sociale.

Dans le débat sur les procédures de sélection représentatives, les « dossiers sériels » renvoient à de grandes quantités de documents typiques de l'avalanche de papier de l'époque moderne. En Allemagne, on utilise désormais aussi la notion de « masse de dossiers de cas individuels homogènes » (« *massenhaft gleichförmige Einzelfallakten* »).

Évaluation des dossiers sériels

L'**élimination** est la destruction de documents que les Archives n'ont pas évalués comme étant dignes d'être archivés. Dès lors que l'on sélectionne, dans des dossiers sériels, un petit échantillon ou le plus petit échantillon pour en documenter la valeur d'évidence², on parle d'archivage à titre d'exemple.

Lorsque tous les documents proposés sont considérés comme dignes d'être archivés et conservés dans leur intégralité, il s'agit d'un **archivage complet**. Cela peut être le cas en

¹ Menne-Haritz, *Archivterminologie*, p. 78; Reimann, *Archivkunde*, p. 85.

² La valeur d'évidence relève de l'éloquence des méthodes, procédures et fonctionnements de l'administration cristallisés dans le document à archiver, de même que de sa matérialisation et son apparence graphique ; librement traduit de Menne-Haritz, *Archivterminologie*, pp. 63-64.

présence de dossiers sériels, lorsque les dossiers à évaluer présentent les caractéristiques de dossiers sériels mais que le nombre de dossiers est trop petit pour pouvoir opérer une sélection représentative. C'est souvent le cas par exemple pour de petites communes qui ne comptent qu'un petit nombre de dossiers de l'aide sociale. De même, une grande valeur informative justifie fréquemment l'archivage complet (par exemple dans le cas des dossiers de naturalisation).

L'archivage partiel consiste à ne conserver de manière pérenne qu'une partie des dossiers proposés. La sélection des documents à conserver est déterminée en priorité par la valeur informative des documents.³ Elle est souvent effectuée en prêtant une attention particulière d'une part aux éléments typiques de l'époque, d'autre part aux cas particuliers ou significatifs. Un échantillonnage représentatif est le meilleur moyen de rendre compte des éléments qui sont typiques d'une période. Les aspects « particuliers » ou « significatifs » doivent être déterminés selon des critères appropriés.⁴ La fiscalité est un domaine typique où intervient cette démarche combinée, puisqu'il faut pour les dossiers, d'une part conserver une sélection de pièces représentatives de la moyenne de la population et, d'autre part, conserver les dossiers de personnalités importantes.

L'archivage partiel peut toutefois aussi être effectué simplement via un échantillonnage, sans sélection préalable de dossiers précis sur la base de critères portant sur leur contenu. Pour de tels cas, il est particulièrement conseillé d'opérer selon une procédure représentative. Un exemple de ce type de démarche pourrait être l'archivage des dossiers des pensionnaires d'un EMS.

Une manière de réduire la quantité de pièces est l'élagage, méthode qui consiste à éliminer, dans des dossiers dignes d'être archivés, les pièces sans valeur archivistique. Il s'agirait par exemple d'éliminer des dossiers de tutelle les quittances et les documents d'assurance.

Échantillonnage (sampling)

L'échantillonnage désigne la sélection opérée pour l'archivage partiel de dossiers sériels, non pas selon des critères de contenu mais en fonction de critères quantitatifs ou statistiques.

Les critères qui déterminent le choix de l'échantillonnage adapté sont :

- le versement visé
- l'utilisation visée
- les caractéristiques structurelles du fonds

³ « La valeur d'information renvoie à la valeur significative de fonds d'archives sur des faits, des personnes, des lieux et des événements »; Menne-Haritz, Archivterminologie, pp. 72-73.

⁴ Cette forme d'archivage partiel est parfois appelée « sampling qualitatif ». Dans le présent document, la notion de « sampling » est pourtant utilisée exclusivement comme synonyme d'échantillonnage sans évaluation du contenu.

- la tradition archivistique / les échantillonnages pratiqués jusque-là
- la charge de travail / la faisabilité

Compte tenu de la « double structure »⁵ des dossiers sériels, en ce sens que les dossiers présentent une forme homogène tout en contenant des cas individuels ayant leur valeur propre, il convient d'opérer une sélection en plusieurs étapes ou selon plusieurs variantes, autrement dit de combiner plusieurs procédures. Si les échantillons aléatoires font partie d'une procédure combinée, il faudrait en règle générale les réaliser comme première étape de la procédure, quand l'ensemble est encore intact.

Échantillonnage aléatoire

L'échantillonnage aléatoire (random sample) est la procédure consistant à refléter de manière représentative la structure caractéristique de la totalité dans la sélection, ce qui signifie qu'à partir de l'échantillonnage, il est possible de déduire la totalité en tenant compte d'une marge d'erreur. La représentativité visée de la sélection permet des évaluations statistiques.

Dans l'*échantillonnage aléatoire simple*, l'on choisit en règle générale des pièces au sein d'un ensemble (une ou plusieurs entrées avec concordance des références de contenu et des organes de provenance). Le nombre de dossiers à choisir dans un ensemble N, autrement dit la taille de l'échantillon, se calcule selon Buchholz avec un intervalle de confiance usuel de 95 % et une marge d'erreur de 5 % en appliquant la formule suivante :⁶

$$\sqrt{\frac{N - 384}{N - 1}} \times 384$$

Un exemple d'application pourrait être l'échantillonnage des pensionnaires d'un établissement médico-social tel que nous l'avons mentionné ci-dessus.

Comme l'échantillonnage aléatoire simple ne reflète que la structure des caractéristiques de la totalité, il n'est pas possible d'évaluer de manière statistiquement correcte une caractéristique isolée de l'échantillon. L'*échantillonnage aléatoire stratifié* (stratified sample), consistant à commencer par structurer la totalité selon des caractéristiques précises, permet de combler cette lacune. L'on prélève ensuite un échantillon aléatoire simple de chaque strate, les différentes strates étant aussi hétérogènes que possible les unes par rapport aux autres et aussi homogènes que possible en elles-mêmes.⁷ Pour l'échantillonnage de dossiers fiscaux d'une commune importante, par exemple, on peut

⁵ Eder-Stein, Aktenstruktur, p. 562.

⁶ Sur la théorie et les notions de l'échantillonnage aléatoire, voir Buchholz, Stichprobenverfahren, pp. 134-151. On trouve des tableaux avec différentes tailles d'échantillonnage chez Buchholz, Stichprobenverfahren, p. 213ss.

⁷ Leesch, Sozialwissenschaften, p. 118.

commencer par catégoriser la totalité par quartiers ou districts, puis opérer une sélection aléatoire par groupe. Un processus de stratification fréquemment utilisé est le regroupement par année, décennie ou autres périodes temporelles.

L'*échantillonnage aléatoire par niveaux* (staged sample) consiste également à structurer la totalité par niveaux selon des caractéristiques précises et à prélever un échantillon aléatoire simple uniquement à partir de niveaux choisis au hasard. Cette démarche risque cependant d'engendrer des distorsions relativement importantes en raison d'erreurs de sélection.⁸ Une application possible est de classer par communes les autorisations cantonales pour les réservoirs et installations de combustion alimentées à l'huile, de choisir au hasard plusieurs communes et de soumettre les dossiers afférents à un échantillonnage aléatoire.

Hasard et représentativité

La formule ci-dessus indique que la taille de l'échantillon dépend de la taille de l'ensemble : plus un ensemble est petit, plus la proportion de dossiers sélectionnés doit être élevée pour prétendre à la représentativité. Par exemple, pour un ensemble de 800 dossiers, un intervalle de confiance de 95 % et une marge d'erreur de 5 %, il faut sélectionner près de 35 % des dossiers. Pour un ensemble de 2000 dossiers, la proportion de dossiers à sélectionner dépasse encore 17 %. Les Archives particulièrement petites ne constituent vraisemblablement pas des ensembles de cet ordre de grandeur, raison pour laquelle il faut soit renoncer à la représentativité, soit se résigner à un archivage complet.

Pour déterminer des chiffres aléatoires, il faudrait éviter d'utiliser des programmes informatiques car, du fait des algorithmes programmés, ils ne peuvent générer que des chiffres pseudo-aléatoires.⁹ Toutefois, selon la représentativité visée en fin de compte, il n'est pas inacceptable de recourir à des chiffres aléatoires générés par ordinateur. En effet, en raison notamment des erreurs de classement, des dossiers perdus etc., la totalité des dossiers ne constitue pas une base scientifique exacte pour une sélection aléatoire. Il importe uniquement que l'algorithme utilisé n'ait pas été programmé par la personne qui classe et archive les dossiers, car il faut exclure tout pilotage (in)conscient du choix.

Échantillonnage systématique

La méthode de l'échantillonnage systématique consiste à sélectionner un dossier tous les x dossiers au sein d'un ensemble. Le point de départ, donc le premier dossier sélectionné, peut être déterminé par un choix délibéré ou selon un chiffre pris au hasard. L'intervalle significatif x se calcule en divisant la taille de l'ensemble par la taille souhaitée

⁸ Thoma, Sampling, p. 20.

⁹ Amstutz, Bewertung, p. 1. Amstutz mentionne comme alternative à des programmes informatiques à utiliser en local le site internet www.random.org (consulté le 05.10.2009), lequel affirme fournir des chiffres au hasard non générés par ordinateur.

ou calculée de l'échantillon. Comme sa représentativité est contestée, l'échantillonnage systématique est particulièrement adapté à l'illustration d'actions administratives.¹⁰ De plus, l'ensemble ne devrait posséder qu'une structure simple et homogène en soi, par exemple une année de naissance classée par ordre chronologique, sans autre critère de classement tel que l'alphabet ou autre.

Échantillonnage en grappes

L'échantillonnage en grappes consiste à diviser la totalité en sous-groupes (ou grappes) définis selon des critères de contenu ou de forme. Cette procédure de sélection est également dénommée échantillonnage par la méthode des quotas.¹¹ Selon la structure de classement de l'ensemble, l'échantillonnage en grappes se caractérise par sa facilité d'application; la sélection n'est cependant pas considérée comme représentative.

Critères objectifs possibles pour la formation de grappes :

- Géographie : prise en compte de communes, de rue ou de districts précis
- Lettres : prise en compte de certains noms de familles. Est certes adéquat pour documenter les liens familiaux, mais les changements de nom et les mariages ne sont pas pris en compte. Le choix de la lettre dépend de ce que l'on veut transmettre. Il n'y a pas de lettres justes ou fausses.¹²
- Jours/mois de naissance : se traduit souvent par des distorsions de l'échantillon, en particulier pour les étrangers. Comme la date de naissance n'est souvent pas connue pour les réfugiés, elle est établie au 1^{er} janvier.
- Années de naissance
- Années des dossiers

Critères formels possibles :

- Volume (« fat files » épaisseur du dossier)

L'échantillonnage en grappes permet également de combiner plusieurs critères.

Les critères de sélection devraient être fixés surtout en tenant compte de la globalité et de l'objectif de versement. Il est par exemple peu économique de sélectionner certaines lettres dans des ensembles classés par ordre chronologique. Par des décisions prospectives d'évaluation, on peut cependant faire en sorte, avec l'organe qui produit les dossiers, que la structure de classement réponde aux exigences d'une procédure de sélection ultérieure.

¹⁰ Buchholz, Stichprobenverfahren, p. 119.

¹¹ Eder-Stein, Aktenstruktur, p. 563.

¹² Buchholz, Stichprobenverfahren, p. 194.

Bibliographie

- Amstutz, Irene, Bewertung massenhaft gleichförmiger Einzelfallakten: ein Leitfaden zur Anwendung tatsächlicher Zufallsauswahlen nach Zufallszahlen, Staatsarchiv des Kantons Basel-Stadt, 19.04.2009.
- Amstutz, Irene, Die Bildung repräsentativer Stichproben bei Massenakten. Zur Methode und Praxis der Bewertung mittels Zufallsauswahl, in: *arbido* 4 (2009), pp. 17-21.
- Buchholz, Matthias, Überlieferungsbildung bei massenhaft gleichförmigen Einzelfallakten im Spannungsverhältnis von Bewertungsdiskussion, Repräsentativität und Nutzungsperspektive: eine Fallstudie am Beispiel von Sozialhilfeakten der oberbergischen Gemeinde Lindlar (Archivhefte 35), Diss. Humboldt Universität Berlin, Köln 2001.
- Buchholz, Matthias, Stichprobenverfahren bei massenhaft gleichförmigen Einzelfallakten. Eine Fallstudie am Beispiel von Sozialhilfeakten, in: *Historical Social Research* 27 (2002), pp. 100-223.
- Couture, Carol et al., *Les fonctions de l'archivistique contemporaine*, Québec 1999.
- Couture, Carol, Rousseau, Jean-Yves et al., *Les fondements de la discipline archivistique*, Québec 1994.
- Couture, Carol, Evaluation d'archives et calendrier de conservation, dans: *arbido*, 7-8 (1998), pp. 10-13.
- Craig, Barbara L., *Archival Appraisal: Theory and Practice*, München 2004.
- Eder-Stein, Irmtraut, Aktenstruktur und Samplebildung. Überlegungen zur Archivierung von massenhaft anfallenden Einzelfallakten am Beispiel von Akten der Justiz, in: *Der Archivar* 45 (1992), pp. 562-571.
- Ernst, Albrecht et al., Überlieferungsbildung bei personenbezogenen Unterlagen, in: *Der Archivar* 61 (2008), pp. 275-278.
- Evaluation et formation des sources archivistiques, dans: *Revue suisse d'histoire* 51, (2001/4), pp. 413-534.
- Evaluation (L'), dans: *Archives* 28/1 (1996/1997), pp. 3-113.
- Guyotjeannin, Olivier, Tris et échantillonnages: empirisme et théorie, dans: *La Gazette des Archives* 124 (1984), pp. 5-26.
- Kluge, Arnd, Stichprobenverfahren zur archivischen Auswahl massenhaft gleichförmiger Einzelfallakten, in: *Der Archivar* 46 (1993), pp. 542-556.
- Leesch, Wolfgang, Sozialwissenschaften und Archive, in: *Der Archivar* 21 (1968), pp. 106-134.
- Menne-Haritz, Angelika, Schlüsselbegriffe der Archivterminologie (Veröffentlichungen der Archivschule Marburg 20, 2000).
- Thoma, Denise Diana, Sampling und Überlieferungsbildung. Bewertungskriterien am Beispiel von Bezirksgerichtsakten im Staatsarchiv Aargau, Masterarbeit HTW Chur, Chur 2009.
- Tiemann, Katharina, Bewertung und Übernahme von amtlichem Registraturgut, in: Reimann, Norbert (Hg.), *Praktische Archivkunde* (2004), pp. 85-87.

- Treffeisen, Jürgen, Archivische Überlieferungsbildung bei konventionellen Unterlagen im deutschsprachigen Raum – eine Auswahlbibliographie, in: Historical Social Research 29 (2004), pp. 227-269.
- Tri, sélection, conservation. Quel patrimoine pour l'avenir ? Ouvrage collectif réunissant les actes de la table ronde organisée sous l'égide de l'École nationale du patrimoine, les 23, 24 et 25 juin 1999, Paris 2001.

Karin Schleifer, Sandro Frefel, 19 mai 2010

Schéma : échantillonnage

