Zeitschrift: Arbido

Herausgeber: Verein Schweizerischer Archivarinnen und Archivare; Bibliothek Information Schweiz

Band: 18 (2003)

Heft: 10

Artikel: La conservation et l'évaluation des banques de données aux Archives fédérales suisses

Autor: Zürcher Thrier, Thomas

DOI: https://doi.org/10.5169/seals-769956

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 13.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

La conservation et l'évaluation des banques de données aux Archives fédérales suisses

Thomas Zürcher Thrier
Archives fédérales suisses
Collaborateur scientifique
de la section ARELDA

Traduction française de Jean-Daniel Zeller et Thomas Zürcher Thrier, de l'article original allemand paru dans Arbido, n°3, 2003, pp. 19–21

Les bases de données au sein de l'administration fédérale

Quand M. X est-il devenu premier-lieutenant? Quel était le salaire du fonctionnaire Y en 1991? Est-ce que Z a réellement déposé une demande pour consulter sa fiche de protection d'Etat?

Au sein de l'administration fédérale les bases de données sont omniprésentes. Elles forment des outils de travail (gestion d'adresse, etc.) et sont utilisées à des fins statistiques, mais de plus en plus, elles sont également utilisées pour le contrôle du déroulement des affaires. Des bases de données sont à la base de chaque système de gestion de documents et elles sont sous-jacentes à chaque site web interactif. Les questions posées au début de cet article, et beaucoup d'autres questions de ce genre, peuvent être résolues à l'aide des bases de données déposé dans les magasins des Archives fédérales.

L'équipe projet ARELDA est en train d'établir un modèle de métadonnées et un catalogue de données pour la description préarchivistiques des bases de données. Cette description devrait garantir que les informations des bases de données ne seront pas seulement disponibles dans un demi-siècle, mais également compréhensibles. Pour atteindre cet objectif il ne faut pas seulement archiver les données elles-mêmes, mais aussi les informations sur ces données (métadonnées). Cela n'est possible que si les archives réclament déjà lors de la phase de conception des bases que la qualité archivistique des données soient strictement respectée. Seul ce qui est défini dès le début et qui est saisi en période d'exploitation peut être transféré ultérieurement comme métadonnées aux archives.

Les systèmes doivent être conçus de telle façon qu'ils permettent de reconstituer le cours des affaires ultérieurement. Ainsi, les nouvelles données doivent être rajoutées aux données préexistantes et non pas se superposer aux enregistrement existants (historicisation). Le système devrait également documenter, à travers des log-files, qui a modifié quels enregistrements et à quel moment. Finalement, les interface pour l'archivage devraient être définis à l'avance, pour permettre la prise en charge des données. Dans ce sens, les offices de l'administration fédérale sont invités à laisser participer les collaborateurs des archives déjà lors de la phase de conception d'un nouveau système informatique[1].

L'évaluation des bases de données: une typologie des systèmes informatiques

Depuis déjà quelques années les Archives fédérales ont peu à peu introduit le principe de l'évaluation prospective. Selon ce principe, les dossiers ne sont pas sélectionnés au moment où ils sont proposés aux archives, mais ce sont les positions du plan de classement qui sont préalablement sélectionnées selon la pertinence des documents qui seront produits ultérieurement. Le principe d'évaluation prospective est particulièrement indispensable dans un environnement de gestion informatisée des affaires. L'énorme investissement financier et organisationnel nécessaire à la sauvegarde et à l'archivage des bases de données ne se justifie que si l'on sait déjà au moment de sa conception si ces données auront une valeur archivistique.

Les base de données sont composées de deux parties: une collection de données sur un thème spécifique et l'application qui permet de les traiter. Dans notre contexte, on entend par données des informations structurées et codées en vue de leur traitement par des ordinateurs. La partie application du système de gestion des données facilite la saisie et permet un accès efficace selon différents critères de recherche.

Souvent les bases de données sont catégorisées selon des critères techniques (banques de données relationnelles, orientées objets, etc.) ou selon leur contenu (banques de données personnelles, systèmes d'information géographiques, etc.). Pour évaluer les bases de données d'un point de vue archivistique il est recommandé de les considérer comme des systèmes d'information développés pour un certain objectif. Dans ce sens, le groupe d'évaluation des Archives fédérales et l'équipe ARELDA proposent une typologie fondée sur les critères suivants:

- *Pertinence pour les affaires:* Est-ce que le système a pour objectif primaire, de documenter le déroulement des affaires ou sert-t-il en premier lieu de boîte à outils?
- *Originalité et provenance des données:* S'agit-il de données produites par l'office lui-même ou sur sa demande ou est-ce que les données sont des copies d'un autre système?
- *Structuration du document:* Est-ce que le système gère des documents entiers (fichiers texte, images, pages web) ou bien contient-il surtout des données structurées de manière homogène (Codes sous forme chiffrée, mots ou abréviations)?

A partir de ces trois critères principaux dérive une série de types de bases de données:

RMS (Records Management System ou système de gestion d'affaires): Ils servent au contrôle des affaires . Le système intègre un plan de classement et une composante workflow, qui permettent de documenter complètement le cours des affaires. Dans l'administration fédérale il existe des systèmes qui ne gèrent que les méta données des dossiers. Quelques offices utilisent des RMS avec une fonction de gestion des do-

[1] Instructions concernant l'obligation de proposer et le versement des documents aux Archives fédérales suisses, du 28 septembre 1999, Art. 4, § 3. (accessible seulement en allemand sur le site *http://www.bundesarchiv.ch*).

cuments. Les système de gestion des affaires ont normalement une grande valeur d'évidence, ce qui les rend particulièrement archivables[2].

DMS (Document Management System ou système de gestion électronique de document, GED): Au sein de l'administration des systèmes de gestion électroniques des documents sans gestion d'affaire sont également utilisés. Ils ne permettent pas de documenter explicitement le déroulement des affaires. Ils ont une valeur d'évidence seulement si les documents sont classés selon un plan de classement. Autrement, la valeur archivistique de ces documents est douteuse. Ils peuvent être éventuellement utilisés comme instruments de recherche pour retrouver efficacement certains documents (indexation).

Bases de données de mesure et d'enquête: A ce type appartiennent les premières bases de données transférées aux Archives fédérales. Elles servent a mesurer des processus quantifiables ou comptables pour une publication ou une exploitation interne. Normalement, leur valeur d'évidence est faible mais leur valeur d'information est élevée en particulier pour la recherche en sciences sociales grâce aux accès variés que les base de données permettent par opposition aux statistiques publiées sur papier[3]. Dans ce cas là c'est l'originalité des données qui décidera de leur valeur archivistique. Pour ce type de base de données les Archives fédérales ont statué une décision sur une base interdépartementale[4].

Données de travail: Il existe un grand nombre de bases de données petites ou moyennes qui sont censées faciliter certaines tâches (inventaires, carnet d'adresse, etc.). Pour la plupart elles n'ont pas de valeur archivistique, à l'exception des bases de données contenant des numéros de dossiers, qui peuvent éventuellement être utilisées comme instruments de recherche. Il peut arriver qu'une telle base puisse avoir une grande valeur d'évidence, lorsqu'elle prouve l'existence d'un dossier «disparu».

Les bases de données et leur documentation

Pour être compréhensibles également dans un futur lointain les bases de données archivées doivent être documentées avec soin. En plus des métadonnées utilisées normalement pour les fonds sous forme papier comme les informations sur la provenance ou la taille du versement, il faut également des métadonnées spécifiques aux bases de données. Avant tout les informations sur la structure doivent être complètes et la signification des codes et des abréviations utilisés doivent être minutieusement expliqués. Dans ce contexte il faut se méfier du fait que les définitions des champs et des structures des bases de données peuvent changer au fil du temps. Pour cette raison, les versements successifs d'une même base ne peuvent pas être réunis automatiquement.

Ces fondements correspondent à une pratique communément admise. Le National Digital Archive of Datasets (NDAD) anglais décrit ses bases de données selon une adaptation particulière de l'ISAD(G)[5]. Les descriptions peuvent être consultées en ligne et les données elle-mêmes peuvent être téléchargées moyennant payement[6]. Le Data-Archive de l'Université d'Essex, qui contient plus de 5000 banques de données dans le domaine des sciences sociales et humaines, dispose également d'une documentation exemplaire[7]. Elle se base sur un standard spécifique de la Data Description Initiative (DDI) pour les données du domaine des sciences sociales et politiques, qui est codé en XML. La DTD (Document Type Definition) correspondante est basée sur le Dublin Core et contient plus de 200 éléments de description[8].

Les pratiques de description pour les bases de données qui existent jusqu'à présent ont deux désavantages importants. Premièrement, elles sont pratiquées par des archives qui hébergent uniquement des bases de données. Par conséquent, les bases de données sont toutes traitées au même niveau de description et elles ne sont pas intégrées aux autres fonds. En plus, les descriptions existantes sont surtout orientées vers des bases de données de mesure. Pour cette raison les documentations contiennent des informations détaillées sur les enquêtes qui ont produit les données primaires (les instructions pour la saisie des données, des spécimens de questionnaires, etc.) Pour la description des systèmes de gestion d'affaire l'histoire de leur utilisation serait beaucoup plus importante. Pour la reconstitution du déroulement d'une affaire il sera très instructif de savoir quel fonctionnaire a pu saisir ou consulter quelles données.

Description à plusieurs niveaux des bases de données en fonction de leur pertinence

Une partie de métadonnées saisies dans la phase préarchivistique sera intégrée dans le système de recherche archivistique lors du transfert de la base de données aux archives. Le point d'accès primaire aux fonds ne peut pas être élaboré uniquement à partir du contexte de provenance, la structure de description doit en fournir la clé. Pour intégrer les bases de données dans cette structure de façon adéquate, ils faut les considérer en premier lieu en fonction de ces objectifs et pas de ces aspects techniques. Les propositions d'ARELDA pour la description et le classement des bases de données sont les suivantes:

ARELDA: Variantes de propositions de classement

Variante	Sous-fonds / Dossier		Dossier / sous-dossier	Document(s)
1	**base de données**			*tables*
2	**base de données**		*groupe(s) d'enregistrements*	*enregistrement*
3	**base de données**	*groupe(s) d'enregistrements*	*enregistrement*	*champs*

Tab. 1: Intégration des bases de données dans le classement des fonds selon les propositions d'ARELDA:
enregistrement = groupe de ligne d'une ou de plusieurs tables concernant la même affaire (par exemple une personne)
champs = champs unique d'une table (par exemple une lettre)
gras: description prioritaire, *italique:* description de deuxième priorité

[2] Sur la différence entre valeur d'évidence et valeur d'information voir: Schellenberg, Theodore R.: The Appraisal of Modern Records, Bulletins of the National Archives, number 8 (Washington, D.C., October 1956).

[3] Bütikofer, Niklaus: Archivierung von statistischen Daten. In: Etudes et Sources – Revue des Archives Fédérales Suisses1990–1991, 16/17: pp. 263–293.

[4] *http://www.bar.admin.ch/webserver-static/docs/d/Bewertung_Grundsatzentscheid_statistische_Daten_032003_d.pdf* [cons. 28.5.2003]

[5] Shepherd, Elizabeth et Smith, Charlotte: The application of ISAD(G) to the description of archival datasets. In: Journal of the Society of Archivists 2000, 21, (1): pp. 55–86.

[6] *http://ndad. ulcc.ac. uk/* [cons. 5.2.2003].

[7] Musgrave, Simon und Winstanley, Bridget: Archiving Statistical Data: The Data Archive at the University of Essex. In: RLG DigiNews 1998, 2, (6), *http:// www.rlg.org/preserv/diginews/diginews 2–6.html-essex* [cons. 5.2.2003].

[8] *http://www.icpsr.umich.edu/DDI/CODEBOOK/index.htrnl* [cons. 5.2.2003].

La première variante est prévue exclusivement pour des bases de données de mesure simples, qui ne consistent qu'en une seule ou quelques tables. Les bases de données relationnelles modernes sont composées de beaucoup de tables qui contiennent surtout des valeurs de références pour d'autres tables. Ces tables sont des artefacts purement techniques qui n'ont aucune pertinence pour les affaires.

La deuxième variante propose un groupe d'enregistrements qui concerne la même affaire, par exemple la même personne.

La troisième variante est prévue pour des bases de données qui gèrent des documents entiers. Les champs pourraient représenter les documents. Ils vont êtres saisis avec des métadonnées générées automatiquement, comme le titre du document, l'auteur, le destinataire, etc.

La décision de savoir si une base de données sera traitée comme sous-fonds ou un dossier dépend de sa position dans le service versant. Si une base de données traite un domaine qui correspond à la compétence du service dans son ensemble, elle sera décrite au niveau le plus haut possible. Dans ce cas elle n'a d'habitude aucune position attribuée dans le cadre de classement du service. Elle est considérée son comme propre système de classement et devrait être traitée par conséquent comme un sous-fonds. Si par contre la base de données ne traite que d'un aspect du service versant (par exemple les ressources humaines), sa description sera faite au niveau du dossier. Dans ce cas elle aura – dans l'idéal – une position dans le cadre de classement.

Musique d'avenir

Même si la majorité des métadonnées sera transmise automatiquement dans les systèmes de recherche des archives, la main de l'archiviste sera indispensable pour compléter et coordonner les données. Il ne s'agit pas seulement des questions de classement et de normalisation des noms et des abréviations. Ils devront aussi créer des index secondaires partout où les fonds seront mis à disposition pour des usagers qui ne sont pas familiers avec les coutumes de l'administration ou qui n'ont pas le temps d'étudier des documentations approfondies. Des descripteurs peuvent êtres extraits de certaines bases de données. Ce sont surtout des bases de données personnelles qui contiennent des noms de personnes et des numéros de dossiers d'où on peut extraire de tels index. La procédure technique de l'extraction ne représente pas un grand effort, mais plutôt la normalisation des descripteurs de façon à ce qu'ils permettent des requêtes à travers tous les fonds. De toute façon, l'usager devrait être sûr de recevoir absolument toutes les données concernant son Pierre Echantillon et que les entrées concernant des différentes personnes portant le même nom ne puissent pas être confondues.

Jusqu'à présent la majorité des consultations des archives se passe dans une salle de lecture. Il est prévu d'ouvrir l'accès à certains fonds par un portail Internet[9]. Si le traitement des fonds veut «rendre accessible des documents archivés et les informations qu'ils contiennent à un usage prévu et imprévu»[10], il doit aussi anticiper – ou pour le moins ne pas empêcher – cette variation dans l'usage. C'est surtout vrai pour les données relevant du domaine des sciences sociales qui pourraient être d'intérêt pour un public international étendu. Mais l'usager web du futur ne consultera pas une salle de lecture virtuelle après l'autre pour trouver l'information dont il aura besoin. Il consultera un portail central, qui sera capable de récolter des informations de plusieurs archives simultanément. L'European Social Science Tools and Resources (NESSTAR) offre un tel portail déjà maintenant, portail qui utilise le standard DDI pour normaliser les données[11]. La description archivistique encodée (Encoded Archival Description (EAD)) pourrait avoir une fonction comparable pour des données des affaires[12]. Les archives devront de toute façon tenir compte à l'avenir de tels jeux de métadonnées normalisées.

contact:
E-mail:
thomas.zuercher@bar.admin.ch

[9] Chiquet, Simone et al.: Zugang öffnen: Die Vermittlung im Schweizerischen Bundesarchiv [Ouvrir l'accès: la communication aux Archives Fédérales Suisses]. In: Etudes et Sources - Revue des Archives Fédérales Suisses, 2002, 28: pp. 351–376.

[10] Förster, Bärbel et Kellerhals, Andreas: Das Erschliessungskonzept des Schweizerischen Bundesarchivs: Vom Findmittel zum Findsystem [Le concept de description des fonds des Archives Fédérales Suisses: des instruments de recherche au système de recherche]. In: Etudes et Sources – Revue des Archives Fédérales Suisses, 1997, 23: pp. 335–353, p. 335.

[11] *http://wivw.nesstar.org/* [cons. 5.2.2003].

[12] *http://lcweb.loc.gov/ead/* [cons. 5.2.2003].

ARBIDO: PROCHAINS NUMÉROS/NÄCHSTE NUMMERN

	Clôture de rédaction/ Redaktionsschluss	*Clôture de publicité/* Inserateschluss	*Date de parution/* Erscheinungsdatum
11/03	8.10.03	15.10.03	10.11.03
12/03	10.11.03	15.11.03	10.12.03

Anzeige

Texte schnell und einfach kategorisieren und analysieren mit LexiQuest

// **LexiQuest** basiert auf linguistischer Technologie und findet Konzepte und Beziehungen zwischen diesen Konzepten in Dokumenten.

// **LexiQuest** entdeckt die wichtigen Informationen in der Textflut von Datenbanken, Internet und E-Mails.

// **LexiQuest** kann unterschiedlichste Datenquellen analysieren und dabei 250'000 Textseiten pro Stunde verarbeiten.

// **LexiQuest** ist weltweit erfolgreich im Einsatz und in mehreren Sprachen zuhause.

LexiQuest

LexiQuest ist ein Tool von SPSS - The industry leader in comprehensive data mining technology and analytic applications for enhanced decision making.

Gerne zeigen wir Ihnen die Möglichkeiten von LexiQuest persönlich, rufen Sie uns unverbindlich an unter:

Telefon 01 266 90 30

SPSS®

www.spss.ch
www.spss.com/lexiquest/