

**Zeitschrift:** Arbido  
**Herausgeber:** Verein Schweizerischer Archivarinnen und Archivare; Bibliothek Information Schweiz  
**Band:** - (2016)  
**Heft:** 3: Vernichten um zu bewahren? = Détruire pour conserver? = Distruggere per conservare?  
**Rubrik:** Wie viel darf Aufbewahren kosten? = Conserver à quel coût?

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. 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. [Siehe Rechtliche Hinweise.](#)

### **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. [Voir Informations légales.](#)

### **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. [See Legal notice.](#)

**Download PDF:** 02.02.2025

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

## II. Wie viel darf Aufbewahren kosten? / *Conserver à quel coût?*

### Können wir endlich alles behalten? Archivische Bewertung elektronischer Unterlagen

Thomas Neukom, Präsident der  
Arbeitsgruppe Bewertung des VSA

In der archivwissenschaftlichen Literatur wird die Bewertung als Kernkompetenz der Archivarinnen und Archivare, als archivische Schlüsseldisziplin, ja als «Königsdisziplin» unter den Kernaufgaben des Berufs bezeichnet.<sup>1</sup> Stimmt das wirklich? Ist die Bewertung nicht vielmehr eine logistische Lösung für ein logistisches Problem? Weil riesige Aktenmengen beschränkten Platzressourcen in den Archiven gegenüberstehen, kann nur ein Teil dieser Materialfülle übernommen und archiviert werden; der Rest muss vernichtet werden.

Würden wir, wenn wir genügend Platz und Ressourcen für die Bewirtschaftung der Unterlagen hätten, nicht lieber alles behalten, um der Forschung und allen weiteren interessierten Personen möglichst viele Auswertungs-

möglichkeiten offen zu halten? Bietet nicht gerade die elektronische Archivierung in Zeiten fast unbegrenzten Speicherplatzes die einmalige Chance einer vollständigen Überlieferung?<sup>2</sup>

#### Bewertung als Lösung logistischer Probleme?

Tatsächlich spielen im Archiv logistische und infrastrukturelle Aspekte eine bedeutende Rolle, sei dies nun beim Transport und bei der Einlagerung von Laufkilometern Papierakten, sei es bei der Übernahme und Archivierung elektronischer Unterlagen. Aufwand und Kosten für die Archivierung analoger Unterlagen sind hinlänglich bekannt. Mit der elektronischen Archivierung gibt es ebenfalls erste einschlägige Erfahrungen, und diese zeigen, dass die damit verbundenen Kosten vermutlich deutlich höher liegen als jene, die bei der Archivierung herkömmlicher Unterlagen anfallen.

Die Herausforderung dabei ist nicht die reine Speicherung der Information, denn die Frage nach der «mechanischen» Archivierung im Sinn von Lagerung oder Speicherung ist gelöst. Wie bereits Max Huber 2009 vermutet hat<sup>3</sup>, spielen anders als bei der Archivierung analoger Unterlagen die Kosten für zusätzlichen «digitalen Magazinraum» (also den nötigen Speicherplatz in einem Repository) eine eher untergeordnete Rolle. Immer mehr Speicherplatz ist zu immer günstigeren Konditionen zu haben. Das Problem ist der vom Menschen zu leistende Aufwand, um die Zugänglichkeit und Verständlichkeit der Informationen auch langfristig gewährleisten zu können, beispielsweise mittels Datenmigrationen. Damit kämpfen auch die Exponenten von Big Data.

Sehr viele der von den Aktenproduzenten eingesetzten Systeme und verwendeten Dateiformate eignen sich

nicht für die elektronische Archivierung. Dazu kommen die raschen Veränderungen bei Hard- und Software. Unter der Annahme, dass Organisationen im Durchschnitt ihre Technologie etwa alle drei Jahre ersetzen, muss ein Dossier, das 75 Jahre lang verfügbar gehalten werden soll, in mehr als 25 verschiedenen Kombinationen von Hard- und Software bewirtschaftet werden.<sup>4</sup> Um den mechanischen Prozess der Migration von Informationen verlustfrei zu gestalten, müsste die vollständige Interoperabilität der Systeme erreicht werden. Das würde allerdings strikte Vorgaben für die Beschreibung, die Speicherung und den Austausch

**Das Problem ist der vom Menschen zu leistende Aufwand, um Zugänglichkeit und Verständlichkeit zu gewährleisten.**

von Informationen bedingen. Entsprechende Versuche, Metadaten, Formate usw. zu kontrollieren, werden etwa in der Welt von Open Data oder Linked Data unternommen. Vor diesem Hintergrund gehe ich mit dem niederländischen Archivaren Ruud Yap einig, der findet, eine pragmatische, aber nichtsdestotrotz taugliche Alternative dazu sei die Kontrolle des Informationsflusses durch Bewertung, Auswahl und Vernichtung.<sup>5</sup>

Also stellt uns auch die elektronische Archivierung – zumindest bis dato – vor logistische und finanzielle Probleme, die uns dazu zwingen, die Menge an Informationen zu reduzieren. Wie immer gilt es, Aufwand und Ertrag in ein gutes Verhältnis zu bringen. Was aber wäre, wenn die Kosten und die geschilderten Herausforderungen der elektronischen Archivierung kein Problem mehr darstellen würden? Wenn die ganze Informatikinfrastruktur der-

- 1 Z.B. Maissen Anna Pia, «We are what we keep: die Bewertung als archivische Schlüsseldisziplin», *arbido*, Heft 4, 2009, S. 6.
- 2 Ich danke meinen Kolleginnen und Kollegen in der AG Bewertung des VSA für die zahlreichen Anregungen und Diskussionen rund um diesen Artikel.
- 3 Huber Max, «Archivische Bewertung: Aspekte, Probleme, Konjunktoren», *arbido*, Heft 4, 2009, S. 10–11.
- 4 DLM Forum Foundation, MoReq2010 @: Modular Requirements for Records Systems – Volume 1: Core Services & Plug-in Modules, 2011, published online (<http://moreq2010.eu/>), S. 23–24, zitiert bei Yap Ruud, «Some thoughts on appraisal in the digital age», *Archivpflege in Westfalen-Lippe*, Heft 80, 2014, S. 17.
- 5 Yap Ruud, «Some thoughts on appraisal in the digital age», *Archivpflege in Westfalen-Lippe*, Heft 80, 2014, S. 17.

einst sehr billig, die technische Umsetzung der Datenübernahme durch Standardisierung oder revolutionäre Neuerungen zum Kinderspiel würde? Vielleicht könnte man dann tatsächlich alles archivieren, und erst noch wohlstrukturiert?!

### **Informationsfülle oder Überblick durch Verdichtung?**

Verzichtet man auf die Bewertung, verschiebt sich die Selektion relevanter Informationen in den Lesesaal: Die Benutzenden wären mit einer ungleich grösseren Menge an Informationen konfrontiert, aus der sie selber auswählen dürften oder müssten, je nach Standpunkt. Um den dafür nötigen Überblick zu erhalten, müssten sie sich – im Fall eines staatlichen Archivs – sehr detaillierte Kenntnisse des Verwaltungsapparats aneignen: Bei welchen Stellen fallen welche Dokumente mit welchen Inhalten an? Welche Dokumente sind bei verschiedenen Organen mehrfach und wenn möglich in unterschiedlichen Versionen vorhanden? Welche Dokumente stellen nur eine Vorstufe zum finalen Dokument dar?

Eine unbegrenzte Fülle von Informationen bringt uns rasch an unsere kognitiven und zeitlichen Grenzen. Gefragt ist ja meist die effiziente Beantwortung von Fragen, das Erfassen des

Grundlegenden, der grossen Zusammenhänge. Sogar vertiefte wissenschaftliche Untersuchungen zu einem Thema müssen nicht auf jede jemals dazu verfügbar gewesene Information zurückgreifen. Wie beim hyperthymestischen Syndrom, das jemanden «befähigt», sich an jeden einzelnen Tag in seinem Leben in allen Details zu erinnern, wäre die vollständige Archivierung für die Forschenden wahrscheinlich mehr Hindernis als Chance.

Dazu kommt, dass sich alle Recherchierenden gewöhnt sind, vorhandene Quellen «gegen den Strich zu bürsten», Unterlagen für Fragestellungen auszu-

### **Verzichtet man auf die Bewertung, verschiebt sich die Selektion relevanter Informationen in den Lesesaal.**

werten, die fernab von deren ursprünglichem Verwendungszweck liegen. Es dürfte Archivarinnen und Archivaren noch kaum je gelungen sein – sofern sie dies überhaupt schon einmal versucht haben –, durch Bewertung die Erforschung eines Themas zu verunmöglichen. Auch für Zeiträume mit einer sehr dünnen Quellenbasis kann die Forschung (z.B. in der Archäologie oder in der Mediävistik) aussagekräftige Ergebnisse erzielen. Zugegebenermassen steigt bei einer sehr kleinen Quellenbasis – gleich wie bei einer überbordenden Fülle im Fall einer vollständigen Archivierung – der Aufwand zur Erlangung brauchbarer Resultate.

Relativ unbestritten – selbst bei allfälligen Anhängern einer vollständigen Archivierung – dürfte die rein formale Bewertung sein: Indem man beispielsweise Dossiers konsequent nur von den federführenden Stellen übernimmt und bei allen anderen beteiligten Stellen beseitigt, reduziert sich die Menge an Unterlagen, ohne dass relevante Informationen verloren gehen. Im Gegenteil: Bewertung schafft hier einen Überblick, führt zur Verdichtung von Informationen und zur Vermeidung von Redundanzen. Die Reduktion von Masse ist nicht das primäre Ziel der Bewertung, sondern die (willkommene) Folge davon.

Durch die Bewertung werden die Unterlagen in ihrem Zusammenhang

dargestellt und gewinnen an Informationswert, oder wie Ruud Yap es formuliert: «Appraisal can be described as updating. Appraisal ads context and thus updates the record and determines its value and meaning.»<sup>6</sup> Diese Arbeit muss für analoge Unterlagen, aber ganz besonders auch bei elektronischen Informationen geleistet werden, die in vielen Fällen ohne die Dokumentation des Entstehungszusammenhangs und des Ursprungssystems noch weniger verständlich wären als aus dem Zusammenhang gerissene Papierdokumente. Die Komplexität einer vollständigen Überlieferung würde erst recht nur zu Unsicherheiten führen. Mithilfe digitaler Instrumente lassen sich Informationen zwar oft genauer und gleichzeitig einfacher und rascher filtern, aber erst durch den menschlichen Eingriff im Rahmen der Bewertung erhalten diese ihren Wert und ihre eigentliche Bedeutung. Dieser Aufwand wird durch keine Technologie substanzial reduziert.

### **Alles ist nichts: Gerade das digitale Zeitalter braucht Bewertung**

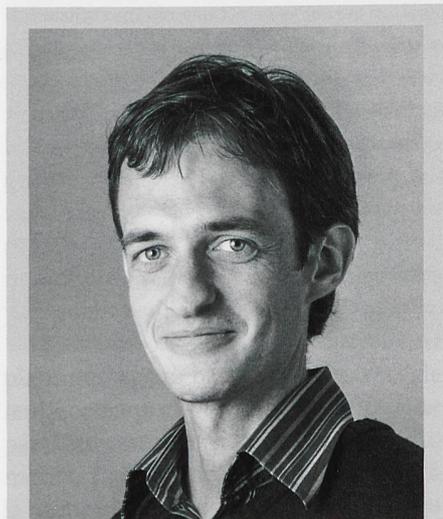
Bleibt noch die Frage, warum Archivarinnen und Archivare zusätzlich zur formalen Bewertung auch systematisch bzw. inhaltlich bewerten (müssen). Weshalb soll von einer wohlgeordneten,

### **Es dürfte noch kaum je gelungen sein, durch Bewertung die Erforschung eines Themas zu verunmöglichen.**

sauber registrierten Massenaktenserie nur ein kleiner Teil der Dossiers behalten werden? Ist beispielsweise das relevante Personendossier einmal vernichtet, kann eine bestimmte Frage zu dieser Person ja möglicherweise nicht mehr beantwortet werden. Das Gleiche gilt für die Suche von Betroffenen nach ihren Unterlagen.

Einmal abgesehen von den eingangs geschilderten logistischen Problemen, spielt bei dieser Frage der eigentliche Zweck der (insbesondere staatlichen) Archive die prominenteste

<sup>6</sup> Yap Ruud, op. cit., S. 17.



Thomas Neukom ist Abteilungsleiter Überlieferungsbildung im Staatsarchiv des Kantons Zürich und Präsident der Arbeitsgruppe Bewertung des Vereins Schweizerischer Archivarinnen und Archivare.

Rolle. Solange die Nachvollziehbarkeit jedes einzelnen Falls gewährleistet sein muss, ist es in der Regel Aufgabe der aktenproduzierenden Stelle, ihre Unterlagen während der vorgesehenen Aufbewahrungsfrist vollständig in ihren Registraturen oder elektronischen Ablagen zu behalten. Dies gilt – streng genommen – auch für alle Akten, auf die direkt von staatlichen Massnahmen

Die Reduktion von Masse ist nicht das primäre Ziel der Bewertung, sondern die (willkommene) Folge davon.

betroffene Personen Zugang haben sollen. Insbesondere bei langen Aufbewahrungsfristen kommt es des Öfteren zu einer Arbeitsteilung zwischen der aktenproduzierenden Stelle und dem zuständigen Archiv.

Sobald die Akten jedoch nicht mehr für ihren Primärzweck gebraucht werden, sind sie dem Archiv anzubieten und erfahren eine Zweckänderung. Im Archiv müssen die Unterlagen das

staatliche Handeln bzw. Nichthandeln im Allgemeinen dokumentieren und nachvollziehbar machen, um Rechtsstaatlichkeit, Rechtssicherheit und Transparenz zu gewährleisten, um der Gesellschaft ein «Gedächtnis» zu geben und um die bereits erwähnten Grundlagen für die Forschung bereitzustellen. Entsprechend nimmt das Interesse an einzelnen Vorgängen oder am Schicksal einer einzelnen Person auf der Zeitachse tendenziell ab. Es wäre für ein staatliches Archiv also schwierig zu rechtfertigen, dass es seinen Zweck nur mit einer vollständigen Übernahme aller Unterlagen erfüllen könnte, selbst wenn es dazu technisch, logistisch und finanziell in der Lage wäre.

Zusammenfassend lässt sich also sagen, dass die Archivierung elektronischer Unterlagen sowohl mit neuen Fragen und Herausforderungen wie auch mit neuen Lösungen und Chancen verbunden ist, dass sie die Grundsätze der archivischen Bewertung jedoch nicht infrage stellt. Robert Kretschmar betont sogar: «Die soweit entwickelten Vorgehensweisen, Strategien, Grundsätze und Methoden [der Bewertung] sind im digitalen Zeitalter noch sehr viel konsequenter anzuwenden.»<sup>7</sup> Eine These, die Ruud Yap noch zuspitzt: «The digital deluge demands appraisal and selection: everything is nothing.»<sup>8</sup>

Kontakt: thomas.neukom@ji.zh.ch

## ABSTRACT

*Peut-on finalement tout conserver? L'évaluation archivistique des documents électroniques*  
Est-ce que l'évaluation est vraiment la «discipline reine» de l'archivistique ou n'est-elle finalement qu'une solution logistique à un problème logistique? Comme des quantités de plus en plus grandes de dossiers sont confrontées à des ressources en place limitées dans les dépôts d'archives, seule une partie de cette masse matérielle peut être prise en charge. L'archivage électronique n'offre-t-il pas une chance unique de constituer un patrimoine archivistique complet?

Dans les faits, l'évaluation est une réponse à des enjeux logistiques, et cela ne change pas à l'ère du numérique. Le problème n'est pas ici les espaces de conservation nécessaires, mais plutôt les investissements consentis pour pouvoir garantir l'accès et la compréhensibilité de l'information sur le long terme. Pourtant, même si l'archivage électronique devenait plus facile et meilleur marché suite à une normalisation radicale ou à des nouveautés révolutionnaires, la question se poserait de ce que l'on gagnerait à tout conserver. Une masse illimitée d'informations nous amènerait rapidement à nos limites cognitives et temporelles. En général, ce qui est exigé est une réponse efficace aux questions, le nécessaire. L'évaluation crée un aperçu, elle mène à la densification des informations et à l'évitement des doubles. La réduction de la masse n'est pas le but premier de l'évaluation, mais une conséquence (bienvenue). Par ailleurs, les archives publiques ont le devoir de documenter les activités ou non-activités de leurs membres en général et de les rendre compréhensibles. Il serait paradoxal de devoir justifier pourquoi ce but ne pourrait pas être atteint si l'on constituait un patrimoine archivistique complet.

Les principes de l'évaluation sont donc toujours valables avec l'archivage électronique, peut-être même encore plus que pour la prise en charge de documents analogiques.

7 Kretschmar Robert, «Alles neu zu durchdenken? Archivistische Bewertung im digitalen Zeitalter», *Archivpflege in Westfalen-Lippe*, Heft 80, 2014, S. 14.

8 Yap Ruud, op. cit., S. 20. Die angelsächsische und frankophone Literatur zum Thema wird im Übrigen gut zusammengefasst bei: Laliberté Stéphanie, «Pour le maintien de la mémoire sociétale: la pertinence de l'évaluation à l'ère du numérique» (Semesterarbeit an der Ecole de bibliothéconomie et des sciences de l'information, Université de Montréal, trimestre d'hiver 2016) [online], [https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/13921/laliberte\\_stephanie\\_2016\\_pertinence\\_evaluation.pdf?sequence=1&isAllowed=y](https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/13921/laliberte_stephanie_2016_pertinence_evaluation.pdf?sequence=1&isAllowed=y) (Zugriff vom 29.6.2016).

# La croissance et la maîtrise des masses documentaires

Gilbert Coutaz, directeur des Archives cantonales vaudoises

**La surabondance informationnelle tend à devenir une composante fondamentale de la réflexion archivistique depuis quelques années, car elle agit sur les pratiques professionnelles et en modifie les processus. En outre, la croissance irréversible de l'informatique amplifie les phénomènes de masses, en en changeant le cadre d'expression et d'archivage. Celui-ci a toujours été lié à la notion de masse, indissociable du nombre de documents ou du volume d'informations.**

L'information est partout, elle est présente à tous les étages de la société et envahit l'ensemble des canaux de diffusion. Les masses documentaires se sont en conséquence intensifiées et diversifiées. Or cette surabondance informationnelle tend à devenir une composante fondamentale de la réflexion archivistique depuis quelques années, car elle agit sur les pratiques professionnelles et en modifie les processus. Elle est également un facteur discriminatoire entre les institutions chargées de l'intégrer à leur politique conservatoire, car elle dépasse souvent leurs possibilités d'accueil et met en cause les approches inadaptées aux exigences de son traitement. Enfin, la croissance irréversible de l'informatique amplifie les phénomènes de masses, en en changeant le cadre d'expression et d'archivage. Leur maîtrise ne peut pas se contenter de la simple accumulation au nom des capacités techniques offertes par les nouvelles technologies.

Les archivistes ont à l'évidence leur place dans ce débat, tant au travers de leurs concepts et outils de gestion que par leur commerce familier des documents. L'archivage a toujours été lié à la notion de masse, indissociable du nombre de documents ou du volume d'informations.

## Une terminologie parlante

De nombreuses expressions sont apparues, ces dernières années, pour caractériser le déluge des informations que

l'informatique génère, permet de partager et de diffuser sans limite de lieu ni de temps. Elles s'habillent des termes: «infobésité», «inforrhée», «big data», «déferlement», «inflation», «globalisation», «prolifération», «exaflood»; et de formules imagées: «augmentation effrénée et exponentielle des données», «boulimie d'informations», «gonflement hypertrophique de la mémoire», «flux grossissant et bouillonnant de données». Les raisons de l'accroisse-

**La numérisation des documents est conçue pour leur sécurité et leur diffusion, nullement pour leur substitution.**

ment des masses qui débutent dès la fin des années 1970, donc avant les nouvelles technologies, sont à chercher dans l'élargissement des domaines d'intervention de l'administration, de la complexification des dossiers traités, de la multiplication des procédures pape-rassières et de la dilatation des notions patrimoniales. Parallèlement, les de-

mandes sociétales et citoyennes ont crû fortement sous les effets de la transparence administrative, et avec le souci d'interpeller davantage les autorités sur leurs décisions passées et présentes.

## Le mètre linéaire, une mesure étalon emblématique dans une archivistique des flux

Selon le *Dictionnaire de terminologie archivistique des Archives de France* (2001), le mètre linéaire est l'«unité de mesure des archives correspondant à la quantité de documents rangés sur une tablette d'un mètre de longueur». Certes, il ne s'évalue pas par le volume, le poids, le support, le mode de rangement et le conditionnement des documents et, de ce fait, sa compréhension présente une certaine élasticité. Mais, par contre, il souligne que les archivistes ont pris très tôt l'habitude de recevoir des archives par lots, par fonds (cela induit le respect et l'unité du fonds). En comparaison, les professions voisines de bibliothécaire, de documentaliste et de muséographe sont attachées à compter leur accroissement par unité, par objet ou par individu, et



L'archiviste doit traiter des masses de documents qui arrivent souvent en vrac et sans description.

© Archives cantonales vaudoises, Olivier Rubin, photographe, 29 juin 2004

parlent de collections. En bibliothéconomie, on ne calcule pas le taux d'enrichissement des collections mais plutôt le taux d'accroissement ou le taux de renouvellement, ce qui permet de déterminer de combien de volumes s'accroît ou diminue la collection. Le désherbage fait partie de la gestion courante de la plupart des bibliothèques.

Par contraste, dans les lois y relatives, les archives des autorités sont inaliénables; elles ne peuvent pas être

acquises par prescription. Autrement dit, ce qui est conservé dans un dépôt d'archives ne peut qu'augmenter.

Les documents d'archives sont par définition uniques, ils n'ont pas leur correspondant dans une autre institution. Ils sont porteurs d'une valeur probatoire (ils servent de preuve de l'exécution d'une activité). Ils doivent être conservés dans leur forme originale et leur environnement originel. Les institutions d'archives doivent assumer

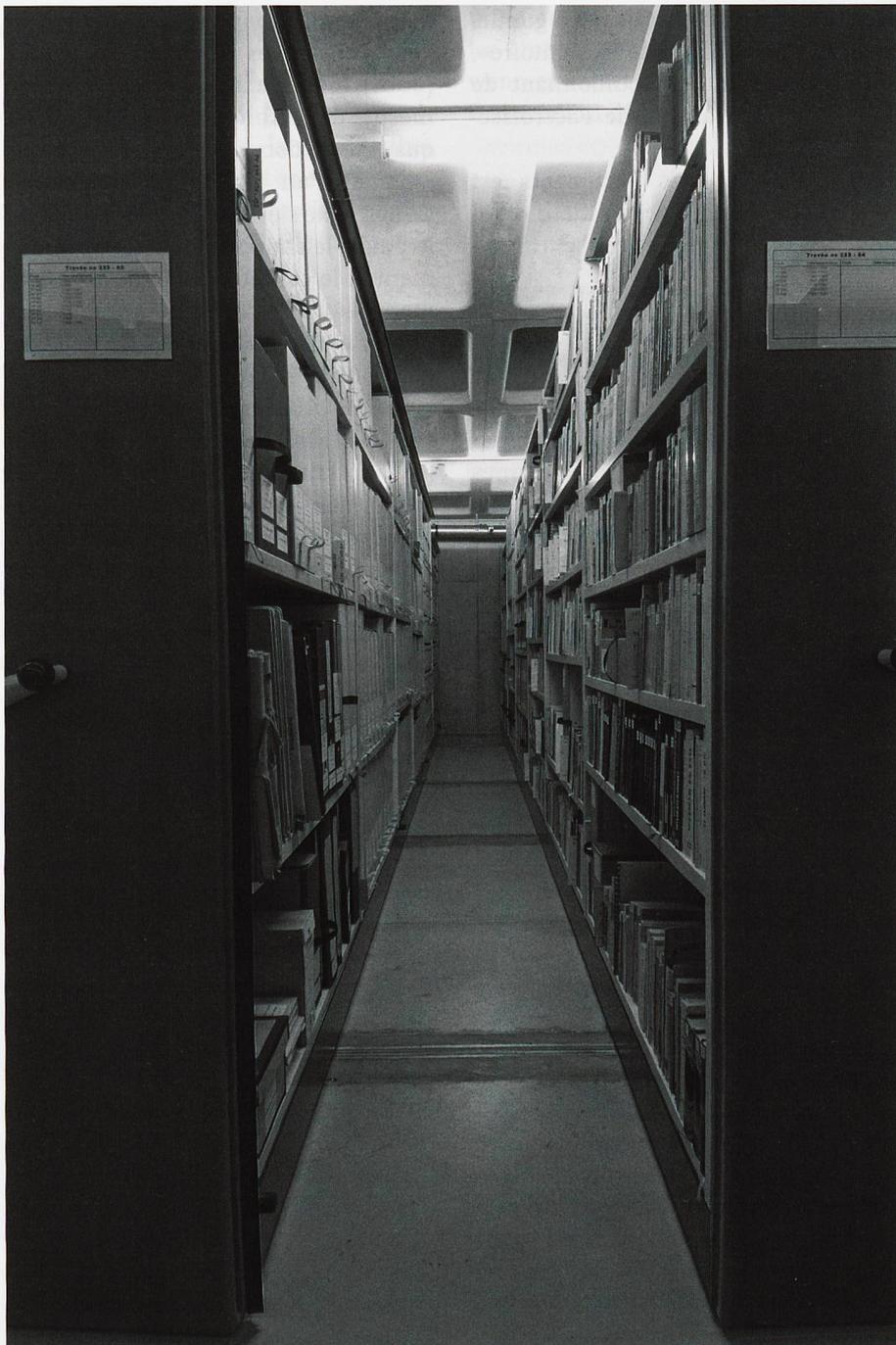
l'hybridité de la conservation. La numérisation des documents est conçue pour leur sécurité et leur diffusion, nullement pour leur substitution.

Avec la dématérialisation progressive des procédures, le métrage linéaire perd de sa pertinence au profit de la volumétrie qui s'exprime désormais en octets, accompagnés d'un multiple: kilo-, giga-, téra- ou penta-. La courbe du poids numérique, en raison de la lourdeur de certains formats, tend à prendre l'ascenseur et dépassera largement celle des documents analogiques. Ce qui est gagné en espaces physiques par la donnée virtuelle, est largement contrebalancé par les coûts de versements et de maintenance des données électroniques.

### Des chiffres exemplaires

Le 2 juin 2016, la Conférence des directrices et des directeurs d'Archives suisses a publié pour la première fois les statistiques des archives conservées et accessibles en Suisse: 370 kilomètres dont 31 kilomètres d'accroissement pour la seule année 2013. Les changements d'échelle quantitatifs les plus spectaculaires, ces dernières années, se vérifient en particulier pour les collec-

**L'archiviste est davantage qu'un spécialiste des rebuts, il doit être reconnu pour ses compétences d'évaluation.**



Des fonds d'archives traités, c'est-à-dire reconditionnés et décrits dans un inventaire, afin de pouvoir retrouver l'information recherchée.

© Archives cantonales vaudoises, Olivier Rubin, photographie, 21 février 2012

tions photographiques de presse. Un fonds ou une collection dépassant le millier de photographies constituait naguère l'exception, aujourd'hui, l'ordre de grandeur est passé en peu de temps à la centaine de milliers, au million, voire à plusieurs millions. C'est ainsi que les Archives du canton d'Argovie ont reçu, en 2009, du plus important groupe de presse de Suisse, Ringier, 7 millions de tirages photographiques, négatifs, etc. Depuis 2013, ce sont près de 2 millions d'enregistrements informatiques supplémentaires qui ont été créés. En août 2015, les Archives cantonales vaudoises ont acquis leur plus imposant fonds d'archives para-administratives et privées, la Fondation vaudoise du patrimoine scolaire, de 420 mètres linéaires. Ces quelques chiffres donnent l'ampleur des défis à

relever par les institutions patrimoniales dont les dépôts d'archives émergent.

### L'impact des masses sur les pratiques professionnelles

Le rôle des archives a longtemps commencé au moment où les secteurs d'activité ont fait appel à elles pour se défaire des dossiers qui encombraient leurs locaux. Archivage et élimination ont été des concepts antinomiques

**Il n'est pas possible de laisser la mémoire se définir par elle-même, surtout dans l'univers numérique.**

jusqu'au milieu du XX<sup>e</sup> siècle, les archivistes n'ayant été compris que comme des conservateurs tant dans leurs pratiques que dans leur approche intellectuelle. Dès les années 1950, l'autorisation d'éliminer des documents est confiée en Suisse aux seuls archivistes par les lois et les règlements. On commence à mesurer les limites des politiques de la conservation.

Depuis, la massification impose aux institutions patrimoniales de revoir leurs critères d'acquisition, leurs méthodes de gestion et d'examiner les possibilités de concertation et de mutualisation des engagements.

D'une attitude attentiste, voire passive, l'archiviste est désormais amené à intervenir sur le cycle de vie, dès l'amont de la vie de l'information jusqu'à son sort final: élimination (dans la grande majorité des cas) ou versement aux archives. Le modèle fonctionnel, Records Management, décrit toutes les étapes de la procédure, du temps de la production au temps de l'activité, et du temps de l'activité au temps historique. Si tous les documents naissent archives, ils ne sont pas appelés à être pour autant historiques. L'informatique a rendu à la fois caduque la notion d'original pour la remplacer par celle d'authentique, et suranné le «séquençage» des trois âges pour lui préférer le cycle de vie au continuum. Le plan de classement et le calendrier de conservation ont pris leur pleine place dans le dispositif de l'archivage, le premier désormais fondé sur les missions et les activités du producteur, le second décrivant les types de

documents produits et leur fixant les durées d'utilité administrative et d'utilité légale de chaque nature de documents, ainsi que leur sort final – ces analyses sont traduites en métadonnées qui constituent l'ADN du document concerné. Les normes archivistiques, à valeur universelle, permettent l'analyse hiérarchique et la description à plusieurs niveaux, modulaires. L'unité archivistique est adaptée selon la qualité et l'importance des documents à décrire. Il n'est pas étonnant que dans la norme de description générale et internationale de description archivistique (ISAD/G) le champ «Contenu» soit un champ facultatif, alors que les champs «Information sur le producteur» et «Information sur le traitement» doivent être renseignés. Cela traduit la primauté du contexte de production sur le contenu de l'information.

Dans le déplacement du curseur de l'intervention de l'archiviste dans la chaîne de l'archivage, il ne faut pas y voir seulement les effets de l'informatique; plusieurs pratiques sont antérieures à l'apparition des nouvelles technologies, le Records Management étant déjà appliqué aux documents analogiques. Il faut pouvoir planifier les versements, juger de la pertinence historique et de l'organisation interne de chaque fonds. Chaque étape de l'archivage est désormais encadrée par des normes qui sont autant d'obligations organisationnelles que techniques pour garantir le parcours sécurisé de l'information, mais surtout pour lui donner les contraintes de l'authenticité, de l'intégrité, de la fiabilité et de l'explo-

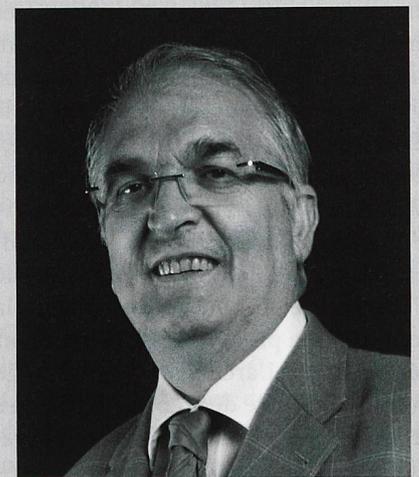
**Seule la rédaction d'inventaires permet de faire exister les masses documentaires.**

tabilité. A la différence de leurs prédécesseurs, les archivistes, tout en demeurant, au terme du cycle de vie, les dépositaires de la mémoire historique, sont présents dès les débuts de l'archivage. En guise de preuve, toutes les mises au concours des postes de records managers intervenues en Suisse dans les administrations publiques font dépendre cette fonction des archives, chargées de conduire et de contrôler la

politique de l'archivage au niveau de toute l'administration et, selon les cas, de formuler la politique de gouvernance de l'information.

### L'évaluation au cœur des politiques d'archivage

Lors de la double enquête systématique sur l'archivage dans l'administration cantonale vaudoise, entreprise entre 1996 et 1998, les chiffres suivants avaient été relevés: 48 000 mètres linéaires de documents attendaient dans les locaux d'être évalués: 30 000 mètres linéaires pour l'administration cantonale et 18 000 mètres linéaires pour l'Ordre judiciaire. Les Archives cantonales vaudoises conservaient alors dans leur bâtiment, inauguré en octobre 1985, 25 000 mètres linéaires d'ar-



Gilbert Coutaz est directeur des Archives cantonales vaudoises depuis 1995. Il a présidé l'Association des archivistes suisses entre 1997 et 2001, après avoir été membre du Comité directeur de la Section des Associations professionnelles d'archivistes du Conseil international des archives entre 1992 et 2000. Membre de plusieurs comités de sociétés d'histoire, il est à l'origine en 1998 de Réseau-PatrimoineS Association pour le patrimoine naturel et culturel du canton de Vaud, en 2004 de COSADDOCA (Consortium de sauvetage documentaire en cas de catastrophe), et en 2011 de Mnémo-Pôle. Entre 2006 et 2014, il a enseigné l'archivistique aux Universités de Berne et de Lausanne dans le cadre du Master of Advanced Studies in Archival, Library and Information Science (MAS ALIS) et est l'auteur de *Archives en Suisse. Conserver la mémoire à l'ère numérique* paru en 2016.

chives pour une disponibilité de locaux de 36 000 mètres linéaires. Face à de telles masses en souffrance, et au manque d'inventaires des masses conservées, que faire valoir auprès des autorités de tutelle, empêtrées dans des difficultés financières sans précédent? La construction d'un bâtiment nouveau? La conservation intégrale des documents, à ajouter progressivement aux fonds déjà existants et dans l'attente de nouvelles surfaces?

Pour ne pas risquer de reléguer les Archives cantonales à une institution subalterne et isolée, il a fallu proposer un plan d'action sur dix ans et des stratégies pour accomplir les objectifs du programme. Une des formulations fortes a été le recours à l'évaluation avec sélection. Près de 20 ans plus tard, le bâtiment des Archives cantonales vaudoises a été densifié jusqu'au maximum de ses capacités (56 000 mètres linéaires) et la couverture des inventaires atteint 98% des fonds existants. Une telle politique qui a nécessité la rédaction de deux Exposés des motifs et projets de décret adoptés par le Grand Conseil en mars 2006 et en juin 2015 n'aurait pas pu être appliquée sans l'affirmation que l'élimination fait partie d'une bonne gestion. Autrement dit, la conservation passe nécessairement par l'élimination. L'archiviste est davantage qu'un spécialiste des rebuts, il doit être reconnu pour ses compétences d'évaluation – celles-ci doivent être exercées dans la plus grande transparence, les choix étant consignés dans un bordereau d'élimination.

- 1 Coutaz, Gilbert, *Archives en Suisse. Conserver la mémoire à l'ère numérique*. Lausanne, Presses polytechniques et universitaires romandes, 2016, 131 p. (Opinion)
- 2 Coutaz, Gilbert, «Le paradigme de l'évaluation: l'élimination, inséparable de la conservation. Quelques réflexions autour des pratiques des Archives cantonales vaudoises en matière d'archives officielles», in: *Revue suisse d'histoire* 51, 2001, pp. 487–494.
- 3 Coutaz, Gilbert, «Enquête systématique sur l'archivage dans l'administration cantonale vaudoise», in: *Rapport d'activité des Archives cantonales vaudoises 1998*. Lausanne, Archives cantonales vaudoises, 1999, pp. 30–42.

De notre point de vue, il n'est pas justifiable, à l'ère numérique, de se priver de ce fondement de l'archivistique du fait que les capacités techniques permettent de réaliser la conservation intégrale. A quel prix, selon quelles certitudes et quelles contraintes? C'est opter pour une archivistique du mouvant et entrer dans une spirale explosive des coûts pérennes. Toute approche fondée sur la seule accumulation doit être combattue; elle est insoutenable sur la durée (c'est une fuite en avant) et inconsciente. Des parts sont appelées à périr, d'autres à être oubliées au nom de la protection des données. Le grand risque encouru par une telle posture est double: l'abandon de l'évaluation et le transfert de la responsabilité aux producteurs de données, très éloignés des besoins hors du temps administratif. Il n'est pas possible de laisser la mémoire se définir par elle-même, surtout dans l'univers numérique<sup>1</sup>, où l'information est souvent extrêmement éclatée et particulièrement volatile.

#### L'inventaire, l'indicateur de la maîtrise des masses

On a pris l'habitude dans le monde professionnel d'afficher les résultats selon

un certain nombre d'indicateurs: nombre de fonds d'archives, métrages linéaires, effectifs, ressources informatiques, pôle numérique, etc. De notre point de vue, un indicateur est peu avancé par les institutions patrimoniales, c'est celui du taux de couverture des inventaires. Seule la rédaction d'inventaires permet de faire exister les masses documentaires<sup>2</sup> et de leur donner une reconnaissance. Dans ce contexte, l'inventaire n'est pas seulement un outil archivistique, mais également un outil de gestion. Avec la norme ISAD/G, il peut être modulé selon les réalités institutionnelles et l'importance quantitative et qualitative des documents. Il doit rester un atout, dans le sens que selon les quantités, l'archiviste doit préférer un inventaire avec des niveaux de description généraux à un inventaire détaillé qui ne se terminera jamais<sup>3</sup>. C'est sans doute grâce à la flexibilité et la variété de ses méthodes de travail que l'archiviste peut le mieux répondre aux phénomènes de masses documentaires et qu'il est le plus habilité à les maîtriser.

Contact: gilbert.coutaz@vd.ch

## ABSTRACT

### *Dem Wachstum der Dokumentenmasse Herr werden*

Information ist überall, sie ist auf allen gesellschaftlichen Ebenen gegenwärtig und überflutet sämtliche Kanäle. Die Dokumentenmassen sind dementsprechend gross und vielfältig. Dieser Informationsüberfluss hat seit einigen Jahren die Tendenz, archivari-sche Überlegungen zu beeinflussen, denn er hat Auswirkungen auf die professionelle Praxis und verändert Abläufe. Er stellt eine grosse Herausforderung dar für Institutionen, die damit beauftragt sind, diesen Informationsüberfluss in ihre Konservierungspolitik zu integrieren, da dies oft ihre Kapazitäten übersteigt und Vorgehensweisen infrage stellt, die den Anforderungen nicht gewachsen sind. Schliesslich verstärkt das irreversible Wachstum von Information das Massenphänomen, da Ausdrucksformen, und damit auch deren Archivierung, sich ändern. Eine reine Akkumulation ist nicht ausreichend, auch wenn dank neuen Technologien mehr technische Kapazitäten zur Verfügung stehen, um diese Massen an Information bewältigen zu können.

Die Archive sind selbstverständlich Teil dieser Debatte, sowohl durch ihre Konzepte und Werkzeuge wie auch durch ihre Vertrautheit mit Dokumenten. Die Qualität der Verzeichnung (Inventar) ist daher für die Identifizierung von relevanten Informationen entscheidend. Die Archivierung war schon immer eng verknüpft mit dem Begriff Masse, sowohl mit der Dokumentenmenge wie mit dem Informationsvolumen.

# Les coûts d'une conservation électronique

Luc Schneider Tourn,  
co-fondateur de Practeo SA

**Alors que la conservation de documents papier consomme manifestement de la place dans les dépôts, la conservation de données électronique semble être illimitée avec les capacités de stockage de nos ordinateurs qui augmentent à des vitesses vertigineuses. Et pourtant, la conservation électronique a des coûts que l'on ignore souvent (ou feint d'ignorer). Le point avec un informaticien.**

## De la gestion de dossiers papier ...

La question de la gestion de la vie des documents a toujours été de celles que l'on écarte: tant qu'il reste de la place au sous-sol, on entasse les classeurs fédéraux d'archives ...

S'il est vrai que cette pratique a prévalu durant des décennies, s'il est vrai aussi qu'elle a permis par la suite à des chercheurs de retrouver des informations extrêmement intéressantes, toutes les entreprises ont dû se résoudre à mettre en place une gestion de la fin de vie des documents papier et prévoir la destruction d'une grande partie d'entre eux. Les dépôts en effet sont difficilement extensibles et tous les classeurs n'ont pas la même valeur au cours du temps.

## ... à la conservation de données électroniques

Or, à l'heure de la création des documents informatiques tout azimut, la place occupée sur des supports électroniques par les documents et les archives est globalement très faible. Passer du simple au double, au décuple même, de données ne change pas fondamentalement la donne: on a suffisamment de place à la cave pour quelques cassettes d'archives ou quelques disques durs.

Pour un volume donné, la densité de stockage augmente chaque année. On peut donc admettre que tout conserver n'occupera jamais davantage de place: à volume égal, un simple disque dur contenait 1 Go il y a vingt ans, et 10 To (soit 10 000 Go) aujourd'hui! Les cassettes d'archive doublent leur capa-

cité tous les deux ans environ. À elle seule, cette augmentation des capacités de stockage absorbe aisément, année après année, la somme totale des documents à conserver, anciens et nouveaux. Pour les documents bureautiques en tout cas ...

La quantité de données à conserver augmente année après année, mais la densité de stockage augmente encore plus vite. Le volume physique occupé par les archives n'augmente donc pas.

La question est donc évidente: pourquoi se préoccuper de ces données, si le coût induit par leur stockage physique n'augmente pas?

Pour tenter de répondre à cette question, il faut pouvoir la cerner.

Tout d'abord, il convient de différencier l'archive papier, dont la principale caractéristique est d'être directement lisible, de l'archive informatique qui a besoin d'un support et d'un outil pour être lue.

## Formats et supports des documents électroniques

L'outil nécessaire pour lire un document électronique, c'est le logiciel avec lequel on a créé le dit document. Un document informatique sous son format d'origine – Microsoft Excel, AutoCAD, etc. – ne pourra bien souvent plus être lu après quelques années de stockage: le format a évolué à un tel point que le logiciel actuel ne pourra tout simplement plus l'ouvrir. Les filtres d'importation vont gérer quelques versions antérieures, mais généralement pas les versions très anciennes. Il arrive également qu'un logiciel disparaisse du marché (Works par exemple). Faute d'outil, les documents créés dans ces formats sont devenus illisibles, ou très imparfaitement gérés par les filtres d'importation.

Le support, c'est la partie visible de la conservation des documents électroniques: disque, cassette, clé USB, etc. Tous les supports informatiques sont fragiles et sont sujets au vieillissement. Qu'ils soient optiques ou magnétiques, ils se dégradent inévitablement. Durant un certain temps, les redondances de



Luc Schneider Tourn est co-fondateur de Practeo SA, société active depuis près de 15 ans dans le soutien et l'accompagnement des PME romandes dans les différents domaines

touchant à l'informatique: gestion d'infrastructure, archivage des données, développement d'applications spécifiques et communication web.

données intégrées au support lui-même permettent de pallier cette dégradation en recomposant les portions perdues du message, mais finalement le contenu devient illisible, il est perdu. De plus, les supports statiques (clé USB, disque SSD) peuvent également tomber en panne. La difficulté réside dans le fait que, visuellement, rien n'indique que le support se dégrade.

## Contraintes des supports informatiques

Certaines catégories de supports sont dévolues à l'archivage et sont certifiées pour rester lisibles durant des décennies<sup>1</sup>. Toutefois, devant l'impossibilité de vérifier cet argument, dans les milieux de conservation des données, on recommande la recopie intégrale des documents tous les cinq ans au moins,

<sup>1</sup> Parmi ces tentatives de conception de supports voulus comme permanents, mentionnons le support sur verre développé par l'entreprise japonaise Hitachi: Schwan, Ben, «Speicher aus Glas», in: Technology Review, 26.10.2012, <http://www.heise.de/tr/artikel/Speicher-aus-Glas-1729182.html>  
Plus récemment, un projet de l'École polytechnique fédérale de Zurich (ETH) cherche à utiliser de l'ADN comme support d'archivage pérenne: «ETH-Forscher machen DNA zum Langzeit-Datenarchiv», in: ICT kommunikation, 16.02.2015, <http://ictk.ch/content/eth-forscher-machen-dna-zum-langzeit-datenarchiv>

et au minimum sur deux supports indépendants.

Une fois perdu, un support peut encore être sauvé par des entreprises spécialisées: la récupération en salle blanche est généralement possible, mais à un coût très élevé. Enfin, afin de parer la destruction par le feu, il convient de stocker des copies en plusieurs lieux.

Rappelons-nous finalement qu'il faudra encore posséder un lecteur en état de marche, capable de lire le disque ou la cassette. Qui possède encore un lecteur de disquettes? C'était pourtant le format standard il y a vingt ans. Combien d'archives comptables dorment sur ce support?

Alors, quel peut être le sens de conserver de tels documents sous cette forme, quand bien même leur stockage serait bon marché? On croit posséder des archives, mais on possède en fait de la poussière, en tout point comparable à un classeur dévoré par les souris.

### Solution à la préservation électronique?

Existe-t-il une solution pour préserver ces données volatiles? Oui, elle est composée de plusieurs éléments indissociables, que nous avons vus précédemment:

- Éviter pour l'archivage les formats de données propriétaires ou non pérennes, leur préférer les formats d'archives reconnus et standardisés: PDF archive, images standards, bases de données non verrouillées<sup>2</sup>.
- Renouveler périodiquement les supports. Cela permet de profiter de l'augmentation de densité des dernières générations de produits, mais il faudra tout remplacer: un ancien lecteur de bandes ne tirera aucun profit des bandes à haute capacité. Il va donc falloir remplacer le lecteur ou le

robot de gestion des cassettes. Il en va de même pour les disques durs.

- Contrôler le contenu des supports: les systèmes de stockage et d'archivage doivent périodiquement lire et comparer les données enregistrées, et ce automatiquement, afin de vérifier qu'elles sont encore lisibles.
- S'assurer de disposer d'une copie de sécurité, toujours située à l'extérieur du bâtiment principal, indispensable pour gérer les problématiques d'inondations, de feu et de vol.

Le stockage brut de données informatiques est donc totalement inutile si l'on souhaite pouvoir relire les données à l'avenir.

### Gérer la masse des documents et des données

L'augmentation de la densité des capacités de stockage permet d'absorber au fur et à mesure les nouvelles données, pour autant qu'il s'agisse de données de bureautique dont la quantité augmente

La quantité de données à conserver augmente, mais la densité de stockage augmente encore plus vite.

linéairement chaque année. Or il se trouve que ce type de données représente une part de plus en plus congrue des documents essentiels au fonctionnement de l'entreprise.

À l'inverse, les documents de travail des systèmes métier actuels représentent de très loin la plus grosse part

de l'information d'une entreprise et leur masse augmente de manière extrêmement rapide et massive: dossiers systématiquement scannés et stockés dans une application, prises de vue par drone, relevés topographiques, rendus 3D, photographies en haute résolution, films, radiographies médicales, IRM, scanners de terrain, scanners 360°, etc. Ces données, pour un simple dossier client, peuvent aisément représenter plusieurs dizaines de gigaoctets de données, voire bien davantage.

L'augmentation de la densité du stockage des données n'est de loin pas aussi rapide et les systèmes de stockage à plusieurs disques durs sont devenus la norme.

### Augmentation des données plus grande que la baisse des prix

Prenons pour exemple un bureau technique de 20 personnes. Aujourd'hui, ce bureau aura à gérer environ 15 à 20 To de données. Le stockage d'une telle quantité d'information et leur réplification dans un centre de données (ou datacenter) sécurisé en Suisse va représenter à lui seul un coût minimum de CHF 1500.– par mois.

Certes, ce coût va baisser, probablement de moitié tous les trois ans<sup>3</sup>, mais en revanche, durant ce même laps de temps, la quantité de données à stocker va quintupler<sup>4</sup>!

La raison de cette augmentation est simple, les appareils qui créent ces données deviennent de plus en plus gourmands en stockage: photos et films à résolution en hausse constante, utilisation

2 On trouvera un «Catalogue de formats de données d'archivage (Cfa)» sur le site du Centre de coordination pour l'archivage à long terme de documents électroniques (CECO), une entreprise commune des Archives suisses: <http://kost-ceco.ch/wiki/whelp/Cfa/>

3 C'est en tout cas le rythme constaté jusqu'à maintenant.

4 Là encore, c'est l'évolution passée qui indique cette tendance.

## ABSTRACT

### Die Kosten der elektronischen Aufbewahrung

Während die Erhaltung von Papierunterlagen offensichtlichen Platz in Magazinen beansprucht, scheinen die Möglichkeiten zur Aufbewahrung und Erhaltung von digitalen Daten mit den Speichermöglichkeiten unserer Computer, die rasant zunehmen, unbegrenzt zu sein. Und dennoch, die Erhaltung digitaler Daten verursacht Kosten, die häufig ignoriert werden (oder bei denen man so tut, als ob man sie nicht kennt). Die Speicher und die Formate für die Erhaltung sind nicht ewig haltbar, ebenso wenig wie die Lesegeräte: Man muss deshalb Entwicklungen und Veränderungen verfolgen und die digitalen Daten regelmässig auf neue Speicher und in neue Formate migrieren. Zusätzlich gilt, dass die Datenmenge rascher zunimmt als die Preisabnahme für den dafür benötigten Speicher. Es kommt deshalb teurer, alles aufbewahren zu wollen. Vielmehr sollte man diejenigen ausgewählten Daten bewältigen können, die man erhalten möchte, und dies ab ihrem Entstehungszeitpunkt.

tion systématique des relevés par drone et des scanners 3D, systématisation de la numérisation de documents (dossiers de candidature, courrier, factures), etc.

Le résultat est que la masse de données à stocker croît de manière bien plus rapide que la baisse de prix des espaces de stockage. Cette dernière est relativement linéaire, alors que l'augmentation de la quantité de données a tendance à suivre une courbe exponentielle.

La question de l'archivage de ces données devient donc très rapidement une nécessité économique: il n'est pas du tout raisonnable ni viable à long terme de conserver toutes ces données, ce d'autant plus que la majorité de ces données ne sera bientôt plus lisible si aucune politique de préservation n'est mise en œuvre. Ces données devien-

dront donc parfaitement inutiles, mais leur stockage continuera de coûter!

Ainsi, sans prendre de mesures de gestion de la vie des documents, cette entreprise va donc dépenser des fortunes pour stocker ... du vent. Si elle veut proposer des prestations meilleur marché et rester concurrentielle, elle doit impérativement maîtriser ce type de charges!

### **Gérer la vie des documents dès maintenant pour maîtriser les charges**

Que faire alors? Un bon plan d'archivage repose sur une gestion de la vie des documents! Et c'est un informaticien qui le martèle!

Il est indispensable de penser sur le long terme et de prévoir une organisation des documents qui permette leur archivage intelligent: par exemple le

passage d'un dossier actif de client en mode archive figée qui nécessite moins de ressources de stockage (car moins de duplication des données), la suppression des documents devenus inutiles dans le cadre de tel contrat, la conservation des documents essentiels sous un format PDF/Archives stable ou encore la conservation d'un minimum de sources brutes.

Nous sommes conscients que la gestion de cette problématique n'est pas aisée, elle est de surcroît chronophage. Se faire aider d'un professionnel en gestion documentaire allège singulièrement ce travail et évite les écueils habituels. C'est maintenant qu'il faut y penser, pas dans dix ans!

Contact: [luc.schneider@practeo.ch](mailto:luc.schneider@practeo.ch)

## **Le plaisir de tout conserver sans modération: une question de taille?**

Arnaud Gaudinat,  
adjoint scientifique, HEG Genève

**Pourquoi jeter lorsqu'on peut conserver? Exit le papier physique et les mètres linéaires. L'ère du numérique a tout chamboulé. L'espace dans les nuages est infini, c'est la promesse de la loi Kryder qui prédit empiriquement le doublement de la densité de stockage tous les ans depuis 60 ans. Mais conserver l'information c'est bien, la retrouver c'est encore mieux et indispensable. Google trouve plutôt bien son chemin parmi plus de 1000 milliards de documents décentralisés. Alors pourquoi devrions-nous perdre du temps à trier, archiver, sélectionner, effacer nos centaines d'emails, de photos et autres documents? Ici sont présentés quelques idées, repères et exemples relatifs à la problématique de la conservation de toute l'information numérique plutôt que de son élimination.**

### **Tous archivistes**

Dans nos sociétés modernes, nous sommes tous des archivistes ... Des

archivistes plus ou moins efficaces. Nous sommes bien entendu tous confrontés à la gestion de nos nombreux documents administratifs. Et tout comme des archivistes chevronnés, nous nous devons de faire des classements et, de temps en temps, de décider de nous séparer de certains documents pour retrouver un peu d'espace dans nos rangements. Nous sommes encore des archivistes lorsque nous décidons de mettre nos photos numériques sur un support optique, de les étiqueter, d'en imprimer certaines ou carrément d'en effacer d'autres, car elles sont légèrement floues. Bien entendu, l'analogie s'arrête ici, car l'archivistique professionnelle s'occupera au sein d'une entreprise ou d'une institution, entre autres, du cycle de vie du document, mais aussi d'archiver les documents de façon pérenne sur du très long terme. Cependant, cette question de l'élimination qui touche l'archiviste professionnel est universelle dans notre monde moderne. Elle se doit d'être mise en perspective, par rapport

à l'état de l'art du stockage et du traitement des données. À noter que l'élimination de données pour des aspects légaux, de confidentialité et de droit à l'oubli, ainsi que la conservation à long terme des documents numériques ne seront pas considérées ici. Cependant, en ce qui concerne la conservation à long terme, les principes de bases présentés dans cette revue semblent aussi valables (pour des coûts supérieurs). Et certaines solutions flexibles liées au cloud computing semblent aussi prometteuses<sup>1,2</sup> pour traiter ce problème.

- 1 Steven. C. Horii, «Archiving, Chapter 10: Future Storage Trends and Technologies» [en ligne], <[http://siim.org/?page=archiving\\_chapter10](http://siim.org/?page=archiving_chapter10)> (consulté le 22.07.2016)
- 2 Rosenthal. David, «The Future of Storage» [en ligne], 2016, <<http://blog.dshr.org/2016/05/the-future-of-storage.html>> (consulté le 22.07.2016)

## Taille de l'information, de quoi parle-t-on?

Dans le monde numérique, l'espace se compte en bits, la plus petite unité d'information. Avec un simple bit par exemple, on peut indiquer que l'on rend visible des millions de documents dans une interface ou pas, plutôt que de les effacer définitivement. L'octet, l'unité de base de l'informatique, qui représente 8 bits, permet lui de coder 256 informations. Mais que représente par exemple 1 To (téraoctet) de données, taille moyenne en 2016 des disques durs dans nos ordinateurs portables? Selon le Tableau 1, il est possible actuellement de stocker dans 1 To de 1 à 4 millions de livres, 250 DVD, approximativement 10% des ouvrages présents à la Bibliothèque nationale de France (BNF), ou 0,6% de Common Crawl. Et ainsi de conclure qu'un ordinateur actuel permettrait de stocker au format numérique la plupart de nos bibliothèques, sauf les plus grandes. Et qu'internet une fois de plus pose de nouveaux défis.

Pour arriver à stocker autant de données dans nos ordinateurs, il s'est passé 60 ans d'évolution et de défis techniques<sup>8</sup> (voir Tableau 2) qui ont donné naissance à la loi de Kryder<sup>9</sup> (du nom

- 3 Valeski, Jud, «Handling High-Volume, Realtime, Big Social Data», 2011, <<https://blog.gnip.com/tag/data-collection/>> (consulté le 22.07.2016)
- 4 Gurrin, Cathal, Smeaton, Alan F., & Doherty, Aiden R.. «Lifelogging: Personal big data». Foundations and trends in information retrieval, 8(1), 1-125. A. R. 2014.
- 5 Merity, Stephen. «Common Crawl» [en ligne], 2016, <<http://commoncrawl.org/>> (consulté le 22.07.2016)
- 6 «PetaBox» [en ligne]. 2014, <<https://archive.org/web/petabox.php>> (consulté le 22.07.2016)
- 7 «Internet lives stats» [en ligne]. 2016, <<http://www.internetlivesstats.com/>> (consulté le 22.07.2016)
- 8 Cocolova, Alex, «The astounding evolution of the hard drive» [en ligne]. 2013, <<http://www.pcworld.com/article/2048232/the-astounding-evolution-of-the-hard-drive.html>> (consulté le 22.07.2016)
- 9 Walter, Chip. «Kryder's law». Scientific American, 293(2), 32-33. 2001.

| Type de données  | Approximation de la taille | Par rapport à 1 To                       |
|--|----------------------------|--|
| Tweet <sup>3</sup>   | 2,5 Ko                     | 400 millions (moins d'un jour de tweets) |
| Livre  | Entre 256 Ko et 1 Mo       | 1 à 4 millions de livres                 |
| DVD  | 4 Go                       | 250 DVDs                                 |
| Encyclopédie Web, Wikipedia en anglais, en 2016  | 50 Go                      | 20 encyclopédies Web                     |
| 11 millions d'ouvrages présents à la BNF en 2015   | 10 To                      | 10% de la BNF                            |
| Capture vidéo pendant un an (selon Gurrin 2014) <sup>4</sup>                                   | 32,8 To                    | 3%                                       |
| Capacité du plus gros serveur de l'équipe fouille de données serveur à la HEG en 2013          | 100 To                     | 1%                                       |
| CommonCrawl <sup>5</sup> (moissonnage publique du Web en 2015)                                 | 168 To                     | 0,6% de Common Crawl                     |
| Archive.org <sup>6</sup> , site web, images, et vidéo en 2014 selon archive.org                | 50000 To                   | 0,002% de archive.org                    |
| Estimation du trafic Internet en 1 jour, en juillet 2016 selon internetlivesstats <sup>7</sup> | 1 556 000 To (1,5 Eo)      | 0,00006% du trafic mondial               |

Tableau 1: La taille de la donnée en 2016 et comparée à 1 To.

| Année                      | 1 Go              | Capacité pour 100 CHF |
|----------------------------|-------------------|-----------------------|
| 1956                       | 26 000 000.00 CHF | 3,8 Ko                |
| 1973                       | 3 000 000.00 CHF  | 33 Ko                 |
| 1980                       | 100 000.00 CHF    | 1 Mo                  |
| 1987                       | 40 000.00 CHF     | 2,5 Mo                |
| 1995                       | 800.00 CHF        | 125 Mo                |
| 2002                       | 2.00 CHF          | 50 Go                 |
| 2007                       | 0.28 CHF          | 357 Go                |
| 2015                       | 0.05 CHF          | 2 To                  |
| 2020 (selon loi de Kryder) |                   | (estimation) 64 To    |
| 2025 (selon loi de Kryder) |                   | (estimation) 2 Po     |

Tableau 2: 60 ans d'évolution du stockage et capacité pour CHF 100.- (source pcworld.com).

de l'ingénieur chez Seagate qui identifie cette loi en 2005). Cette loi empirique (similaire à la fameuse loi de Moore) indique que la densité de stockage est multipliée par deux tous les ans pour un coût deux fois moindre. Et même si un certain ralentissement de la croissance de stockage dans les zones de stockage personnelles est constaté

ces dernières années, le développement des zones de stockage dans le cloud devrait continuer dans les prochaines années à garder la même croissance. Cette croissance devrait permettre donc théoriquement de pouvoir stocker les ouvrages de la BNF en 2015 dans nos zones de stockage personnelles dans cinq ans.

La dernière colonne du tableau donne aussi la capacité qu'il est possible d'acheter pour CHF 100.-. Ainsi, si en 2015 il est possible d'obtenir 2 To pour CHF 100.-, pour le même prix en 1973 nous obtenions 33 Ko seulement.

### Pourquoi désherber?

Si on reprend les six objectifs de la méthode de désherbage CREW<sup>10</sup>, on trouve

- 1 Gagner de l'espace,
- 2 Gagner du temps,
- 3 Rendre la collection plus attractive,
- 4 Améliorer la réputation de la bibliothèque,
- 5 Adapter la collection aux besoins,
- 6 Avoir un retour sur les forces et faiblesses de la collection.

L'espace et le temps semblent en effet essentiels. Cependant, éliminer les données permet aussi de diminuer le bruit lors de la recherche.

Ainsi, l'objectif numéro 1 du désherbage est de récupérer de l'espace. Qu'il soit physique ou numérique, l'espace est forcément fini (en tout cas avec nos connaissances scientifiques actuelles) et a de facto un coût proportionnel à sa taille. Si le coût de l'espace physique a tendance à croître dans le temps, inversement le coût de l'espace numérique a tendance à diminuer et ce de manière importante (voir Tableau 2). C'est pourquoi on peut se poser la question sérieusement: dans le monde numérique, devrait-on tout conserver plutôt que de passer du temps (objectif 2) à sélectionner ce que l'on doit éliminer?

Le gain d'espace n'est certainement pas le seul objectif de l'élimination. Il est aussi d'éviter de se retrouver submergé par l'information lors de la recherche d'information. Car, comme

**Qu'il soit physique ou numérique, l'espace est forcément fini et a de facto un coût proportionnel à sa taille.**

mentionné précédemment, conserver l'information c'est bien, la retrouver c'est encore mieux et surtout indispensable. Google nous montre la voie en indexant plus de 1000 milliards de documents déjà en 2008. Bien entendu, nous ne sommes pas Google, mais la bonne nouvelle est que la plupart des solutions efficaces de traitement de

données massives sont des logiciels «Open Source». Par exemple, l'indexation des 26 millions de documents de Medline (citations de la littérature scientifique médicale) dans un logiciel tel que Elasticsearch<sup>11</sup> prend moins de dix heures sur un ordinateur portable datant de 2012.

Les objectifs 3 à 6 peuvent être réalisés aussi sans avoir besoin d'éliminer définitivement les documents, mais en les catégorisant comme tels pour les rendre moins visibles (mais encore accessibles) et aussi en les enrichissant automatiquement d'informations d'usages (objectif 6).

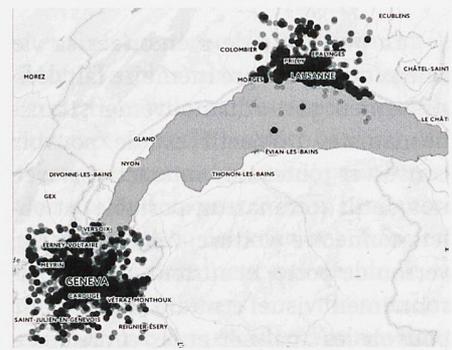
### La sélection, une certaine forme d'élimination déléguée?

La sélection permet le gain de place, mais ce gain de place ne se fait-il pas au détriment de la qualité de l'information lorsque le but du stockage de données est de procéder à des analyses? Cette question s'avère très importante lorsque l'on commence à vouloir faire parler les données éphémères. Prenons l'exemple

**Dans le monde numérique, devrait-on tout conserver plutôt que de passer du temps à sélectionner ce qu'on élimine?**

récent du projet GeoTweet<sup>12</sup>. Ce projet avait comme objectif premier la cartographie des tweets géolocalisés dans le seul grand Genève pendant une période de sept mois (voir Illustration 1).

La collecte des tweets, limitée par l'API (Interface de programmation) offerte par Twitter, se faisait en définissant une fenêtre de capture rectangulaire. Dans GeoTweet, la fenêtre de capture utilisée pour couvrir le grand Genève alla arbitrairement de Culoz en France à Fribourg en Suisse. Le grand Genève s'arrête bien entendu avant cette zone, mais l'avantage de ce choix a permis d'étudier de façon pertinente les différences de tweets entre Genève et Lausanne, alors que ce n'était pas prévu au départ. À noter que dans ce cas, les informations en dehors du rectangle de capture sont actuellement perdues, à moins de disposer d'un budget conséquent pour utiliser les services de GNIP (entreprise permettant d'accéder à la totalité des tweets émise depuis



Ill. 1: Comparaison des zones de points chauds des tweets en 2015 entre Genève et Lausanne (sur [geotweet.hesge.ch](http://geotweet.hesge.ch)).

les débuts de Twitter). Ainsi, une collecte de données plus large permettra de répondre à plus de questions, surtout si des questions additionnelles intéressantes apparaissent en cours d'analyse de données.

### Le mouvement *lifelogging* et *quantified-self*

Pour certains, la question de tout conserver l'information ou non ne se pose déjà plus. Bien que singulier, le mouvement du *lifelogging* – le fait d'enregistrer sa vie de manière la plus continue possible – est déjà lancé et trouve de plus en plus d'adeptes grâce aux nouvelles technologies<sup>4</sup>. Son origine date de 1945 par Vannevar Bush et la vision Memex (une sorte de bureau qui capture l'activité de son utilisateur). Aujourd'hui, ce mouvement est représenté par Gordon Bell et son projet MyLifeBits<sup>13</sup>. Quant au *quantified-self*, il est déjà en partie démocratisé grâce aux capteurs de nos téléphones portables, aux balances connectées et surtout aux bracelets permettant entre autres de mesurer son activité physique, sa position GPS ou son sommeil.

10 Larson, Jeanette. «Crew: A weeding Manual for Modern Libraries». [en ligne]. 2008, <<https://www.tsl.texas.gov/sites/default/files/public/tslac/ld/pubs/crew/crew-methodo08.pdf>> (consulté le 22.07.2016)

11 «Elasticsearch» [en ligne], 2016, <<https://www.elastic.co/>> (consulté le 22.07.2016)

12 Banfi, E., Béguelin, F., & Gaudinat, A. «GEOTweet» (No. TRMASID 7). Haute école de gestion de Genève. 2016.

13 Bell, C. G., & Gemmill, J. «Total recall: How the e-memory revolution will change everything». Dutton.

La promesse de mémoriser sa vie mieux que sa propre mémoire fait définitivement partie du mouvement transhumaniste. L'objectif est de pouvoir conserver toutes nos interactions avec nos outils (ordinateur, portable) et objets connectés (voiture, vélo, frigidaire, verrou de porte) et surtout, tout l'environnement visuel et audio proche pour pouvoir les analyser et les consulter a posteriori. Les valeurs du Tableau 3 représentent la taille réelle des captures typiques du *lifelogging* suivant les différentes sources<sup>14</sup>. Bien entendu, elles

**Un critère d'élimination qui vaut aujourd'hui ne vaut pas forcément demain.**

sont dépendantes de la personne et sont ici données à titre indicatif. Par ce biais, il est possible d'enregistrer un an de données audio sur une carte mémoire SD de 2016 de 512 Go. D'ici cinq ans, il sera possible en théorie d'enregistrer les données de la durée d'une vie sur le disque dur de son ordinateur portable. La capture totale des données nécessite la sauvegarde de toutes ces sources simultanément et, mise à part les données vidéo qui sont beaucoup plus lourdes, le stockage ne semble pas être une limitation. Dans le cas du stockage des données vidéo, le Tableau 2 indique que ce sera probablement possible dans une dizaine d'années.

Pour avoir une illustration du *lifelogging*, je ne peux que conseiller de voir l'épisode d'anticipation «The entire history of you» de la série *Black Mirror* de Charlie Brooker où les dérives d'un tel dispositif sont mises en exergue.

D'un point de vue technique, tous ces capteurs vont générer énormément de données qu'il faudra stocker, analyser, archiver, indexer afin de pouvoir les

rendre utiles pour l'utilisateur final. Ceci offre de nouveaux défis en termes de traitement de données hétérogènes, de stockage, de recherche et de visualisation. Dans le cadre du *lifelogging*, aucune donnée n'est éliminée, car même si elle n'est jamais utilisée, toute donnée est potentiellement utile.

### Les big data et la valeur de la donnée

Aujourd'hui, tout le monde parle des big data ou données massives et de la valeur de la donnée. Les grandes entreprises d'internet (par exemple Google et Facebook) l'ont bien compris en nous offrant des services gratuits dans le cadre desquels chacun d'entre nous offre ses données «idiotes» en surfant sur internet, utilisant telle ou telle application, etc. Ces données mises bout à bout et multipliées par le nombre d'utilisateurs ont une valeur considérable pour qui sait les faire parler. Faut-il éliminer de l'information? Ces entreprises ont déjà choisi et répondu clairement que non.

Si on prend comme exemple Twitter, application dans laquelle les utilisateurs publient des messages de 140 ca-

**Même si conserver le document, c'est bien, conserver la vie du document, c'est mieux!**

ractères maximum (les tweets), ceux-ci contiennent en réalité 20 fois plus d'informations (2,5 ko par tweet en moyenne<sup>15</sup>). Chaque tweet qui est échangé dans le monde contient à chaque fois, en plus du message proprement dit (les 140 caractères), la description de l'émetteur (pseudo, langue, origine géographique, etc.), le contexte (date,

géolocalisation si activée) et l'historique des retweets.

Ceci a l'avantage d'offrir une grande transparence, mais montre clairement qu'on essaie de garder toutes les informations disponibles, y compris le contexte (si cher aux archivistes). Aucune information n'est éliminée a priori. Pourtant, avec 500 millions de tweets par jour, soit 1,25 To, on pourrait faire pas mal d'économies d'espace de stockage en évitant la redondance de l'information.

Pour continuer avec ce fameux micro-blog, la Bibliothèque du Congrès des États-Unis semble avoir compris l'intérêt d'archiver des données sans faire de désherbage. Et ce malgré le fait que le contenu de beaucoup de tweets peut paraître a priori peu intéressant et peu pertinent (par exemple «il fait bo ce matin»). En effet, ils ont décidé en avril 2010 de s'associer avec Twitter et d'archiver la totalité des tweets<sup>14</sup> émis à ce jour et d'en offrir l'accès gratuit. Cependant, après six ans de bonnes intentions, les tweets ne sont toujours pas disponibles et Zimmer<sup>15</sup> semble indiquer aussi bien des problèmes techniques et juridiques que commerciaux.

### Le critère d'élimination diachronique

Les critères d'élimination utilisés de nos jours sont adaptés à nos connaissances et usages d'aujourd'hui. Un critère d'élimination qui vaut aujourd'hui ne vaut pas forcément demain. Pour faire un parallèle avec un domaine très différent, la sauvegarde de la biodiversité a aussi son intérêt et ces critères. L'exemple du rat-taupe nu du Kenya est emblématique. Ainsi, si le seul critère de la sauvegarde de l'animal était la beauté (voir Illustration 2), le pauvre n'aurait aucune chance d'être sauve-

| Contenu capturé    | Volume/an                  | En un an | En une vie (85 ans) |
|--------------------|----------------------------|----------|---------------------|
| Localisation       | 3,9 millions de points GPS | 0,01 Go  | 0,85 Go             |
| Accélération       | 21 millions d'entrées      | 0,05 Go  | 4,25 To             |
| Log de lecture     | Dépend de l'utilisateur    | 1 Go     | 85 Go               |
| Photo              | 1,65 millions              | 30,2Go   | 2,6 To              |
| Audio (mono 22khz) | 5840 heures                | 227,8 Go | 19,4 To             |
| Vidéo              | 5,840 heures               | 32,8 To  | 2,65 Po             |

Tableau 3: La taille des données du *lifelogging* selon Gurrin 2014.

<sup>14</sup> Matt, Raymond. «How tweet it is! Library acquires entire Twitter archive» [en ligne], 2010, Library of Congress Blog, <<http://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/>> (consulté le 22.07.2016)

<sup>15</sup> Zimmer, M. «The Twitter Archive at the Library of Congress: Challenges for information practice and information policy». First Monday, 20(7). 2015.



Ill. 2: Le rat-taupe nu (*Heterocephalus glaber*) est un petit rongeur remarquable non seulement pour son apparence physique. (Roman Klementschtz, Wien – travail personnel, CC BY-SA 3.0)

gardé en tant qu'espèce. Par contre, des recherches de 2005 ont montré que l'animal pouvait vivre jusqu'à 50 ans, qu'il était insensible à la douleur et surtout qu'il avait une forte résistance aux maladies cancéreuses. Cela l'a remis de facto sur le podium des animaux suscitant l'intérêt des humains. Par analogie, effacer une donnée aujourd'hui ne veut pas dire qu'elle n'aura pas d'importance à l'aulne des critères du futur. Et si les coûts sont inférieurs ou équivalents, pourquoi éliminer l'information?

### Conserver la vie numérique de la donnée

Une première tendance en informatique est de garder un historique le plus précis possible des applications et de la vie du document. La seconde tendance est la redondance et la distribution de

l'information. Ainsi, même si conserver le document, c'est bien, conserver la vie du document, c'est mieux! Pouvoir avoir la trace de la naissance du document, des premiers mots de son

**Ne pas éliminer l'information ne veut pas dire ne pas filtrer l'information.**

auteur. Pouvoir voir qui a contribué à tel ajout ou à telle élimination (mais toujours réversible) de manière non ambiguë. Pouvoir voir que plusieurs copies du document ont évolué de leur côté, alors que le document original de l'auteur reste inchangé.

Les historiens l'ont rêvé, les informaticiens l'ont créé: un système de gestion des versions pour le développe-

ment informatique (dit «versionning»). L'outil le plus abouti et populaire à l'heure actuelle est le dépôt GitHub, interface web collaborative basé sur Git. Il est principalement utilisé pour la gestion des codes source, mais l'est aussi pour du texte ou des sites web. Ici, c'est l'auteur du document qui décide de la granularité des versions, mais toutes ces informations et bien d'autres sont conservées et donnent une grande valeur à ce dépôt. De plus, la pérennité de ces données semble assurée, car l'INRIA vient d'annoncer l'initiative «software heritage»<sup>16</sup> qui a pour objectif d'être l'archive universelle de l'open source.

Un autre exemple contemporain et intéressant est celui du bitcoin, cryptomonnaie la plus populaire et controversée. Il est actuellement utilisé pour des échanges commerciaux non contrôlés par un établissement bancaire ou étatique. Le bitcoin est basé sur une base de données distribuée qui se nomme la blockchain (ou chaîne de blocs) et qui a pour particularité de conserver toutes les transactions financières effectuées depuis sa création. Elle fait actuellement plus de 70 Go<sup>17</sup> (juillet 2016), existe en copie sur les ordinateurs de plus de 100 000 «mineurs» – ceux qui minent les bitcoins (comme on mineait de l'or) – et gère de facto les transactions. Dans le blockchain, qui est une archive des transactions, tout est conservé (et partagé) et à valeur de preuve.

### Conclusion

À l'heure du numérique et des données massives, il est, dans la plupart des cas, inutile d'éliminer l'information, car d'une part le coût du stockage continue de baisser énormément et d'autre part les algorithmes de recherche de données structurées et non structurées sont taillés pour gérer plusieurs milliards de documents plus ou moins hétérogènes. *Lifelogging*, *quantified-self*, *blockchain* et *big data* sont autant

## ABSTRACT

*Vom Vergnügen, alles zu behalten: eine Frage des Volumens?*

Wieso etwas wegwerfen, wenn man es aufbewahren kann? Weg mit dem physischen Papier und den Laufmetern; die digitale Ära hat sowieso alles auf den Kopf gestellt. Der Raum in den «Wolken» ist unendlich, das verspricht Kryders Gesetz, das eine jährliche Verdoppelung der Speicherkapazität seit 60 Jahren voraussagt. Information aufzubewahren ist gut, aber sie wiederzufinden ist noch besser und unverzichtbar. Google findet seinen Weg durch die mehr als 1000 Milliarden dezentralisierten Dokumente ziemlich gut. Wieso also sollen wir Zeit verlieren mit Triage, Archivierung, Auswahl, dem Löschen von Hunderten von E-Mails, Fotos und anderen Dokumenten? In diesem Beitrag werden einige Ideen und Beispiele bezüglich der Problematik der Aufbewahrung sämtlicher digitaler Informationen anstelle ihrer Vernichtung präsentiert.

16 «Software Heritage» [en ligne], 2016, <<https://www.softwareheritage.org/>> (consulté le 22.07.2016)

17 «Taille de la blockchain» [en ligne], 2016, <<https://blockchain.info/fr/charts/blocksize>> consulté le 22.07.2016)

d'exemples où le choix de la conservation de toutes les données a déjà été effectué et ce malgré la masse considérable d'informations.

La tendance en informatique est à la conservation des données pour des raisons de traçabilité, de transparence et pour pouvoir reconstruire l'histoire du document. Néanmoins, ne pas éliminer l'information ne veut pas dire ne pas filtrer l'information. Ainsi, sans faire disparaître l'information définitivement, l'intérêt de classer, catégoriser, voire d'enrichir, a encore tout son sens.

L'espace physique est fini. L'espace numérique, qui est in fine physique, l'est tout autant. Cependant, le numérique est, sans aucune comparaison ou mesure, le champion de la compression d'espace physique. Les exemples et idées développés ici montrent que les limites actuelles et surtout futures sont

plutôt le fait d'utilisation extrême tel l'archivage du Web ou des expériences du CERN avec leurs tailles de stockage de plus de 200 Po (soit 200 000 To).

Paradoxalement, si la conservation et la diffusion sont facilitées par les technologies, il est urgent de trouver des moyens de mieux collecter l'information publique, voire l'information privée (courriels, réseaux sociaux fermés, photographies et films produits en masse comme autant de souvenirs qui ne dureront pourtant guère, faute de réflexion), car des millions de documents disparaissent tous les jours. Sur le web par exemple, la durée de vie d'une page est estimée entre 50 et 100 jours en moyenne selon Brewster Kahle, fondateur principal de l'«Internet Archive».

Mais la bonne nouvelle, c'est que la page Wikipédia du rat-taupe nu du

Kenya est bien à l'abri parmi les Peta données d'archive.org et ne «souffre» d'aucun risque d'élimination!

*L'auteur remercie Esther Bettiol pour sa relecture.*

Contact: [arnaud.gaudinat@hesge.ch](mailto:arnaud.gaudinat@hesge.ch)

Twitter: @AGaudinat

#### L'auteur

Arnaud Gaudinat est ingénieur en informatique de formation. Après 15 ans d'expérience en recherche appliquée, il intègre en 2010 la filière information documentaire de la HEG. Il est actuellement adjoint scientifique et enseigne les sciences de l'information, la programmation et la gestion de contenu sur le Web. Il effectue principalement sa recherche dans le domaine de la fouille de données sur le Web et de la webométrie.

## Big Data: Den digitalen Wandel aktiv gestalten

Beat Estermann, Wissenschaftlicher Mitarbeiter, Berner Fachhochschule

**Vor dem Hintergrund der zunehmenden Digitalisierung von Wirtschaft und Gesellschaft und als Antwort auf Entwicklungen, welche gemeinhin unter dem Oberbegriff «Big Data» zusammengefasst werden, hat der Bundesrat im April 2016 die Strategie «Digitale Schweiz» verabschiedet. Der vorliegende Artikel gibt einen Überblick über die Implikationen für Gedächtnisinstitutionen und gibt Denkanstösse, wie diese den digitalen Wandel aktiv gestalten können. Ein aktuelles Forschungsprojekt zeigt auf, dass die Chancen für Gedächtnisinstitutionen vor allem bei der verbesserten Sichtbarkeit der Institutionen und der Bestände liegen sowie bei der zunehmenden Vernetzung und der Neugestaltung der Beziehung zu den Nutzerinnen und Nutzern. Allerdings legen die Forschungsergebnisse auch nahe, dass die Schweizer Institutionen im internationalen Vergleich Aufholbedarf haben.**

Vor dem Hintergrund der zunehmenden Digitalisierung von Wirtschaft und Gesellschaft hat der Bundesrat im April 2016 die Strategie «Digitale Schweiz» verabschiedet. Sie dient dem Zweck, die digitale Entfaltung der Gesellschaft zu fördern, indem die digitale Transformation und der damit verbundene Strukturwandel proaktiv und vernetzt gestaltet werden. Der Bundesrat reagiert mit der neuen Strategie unter anderem auch auf Entwicklungen, welche gemeinhin unter dem Oberbegriff «Big Data» zusammengefasst werden.

Der vorliegende Artikel gibt einen Überblick darüber, was es mit Big Data auf sich hat, welche Chancen und Risiken für Gedächtnisinstitutionen damit verbunden sind und wie sich die Stossrichtungen der bundesrätlichen Strategie auf den Bereich der Bibliotheken und Archive übertragen lassen. Anhand der Ergebnisse einer Umfrage unter Gedächtnisinstitutionen in zehn Ländern wird aufgezeigt, wo die Schweizer Institutionen mit Blick auf die digitalen Transformationsprozesse

im internationalen Vergleich stehen. Den Abschluss bilden einige Denkanstösse und Handlungsempfehlungen, um in der Debatte über Bewahren versus Ausscheiden den Blick aufs Wesentliche nicht zu verlieren: die aktive Mitgestaltung des digitalen Wandels.

#### Chancen und Risiken von Big Data

Im Auftrag des Bundesamts für Kommunikation hat die Berner Fachhochschule 2015 eine Studie durchgeführt, um die Chancen und Risiken von Big Data zu erörtern und den kollektiven Handlungsbedarf aufzuzeigen (Jarchow/Estermann 2015). Grundlage der Studie waren qualitative Interviews sowie eine quantitative Erhebung bei über 800 Personen in der Schweiz, welche mit Daten zu tun haben – sei es privat, beruflich oder im Rahmen ihres politischen oder zivilgesellschaftlichen Engagements.

Wie aus der Studie hervorgeht, ist Big Data – definiert als «Praktiken, welche sich dadurch kennzeichnen, dass grosse Datenmengen aus sehr unter-

schiedlichen Quellen zusammengeführt und – oftmals sehr zeitnah – analysiert und ausgewertet werden» – sowohl mit Chancen als auch mit Risiken verbunden.

Was die positiven Aspekte von Big Data angeht, ergab die Studie, dass rund 80% der Befragten davon ausgehen, dass Big Data es ermöglicht, bisher unbekannte Zusammenhänge aufzudecken und neue Erkenntnisse zu gewinnen. Konkret wurde Big Data von rund zwei Dritteln der Befragten als Mittel gesehen, um genauere Prognosen und Risikoabschätzungen zu ermöglichen, um vermehrt daten- und faktenbasiert zu handeln und um operative und logistische Abläufe zu optimieren. Bei den Risiken standen die grundsätzliche Bedrohung der Privatsphäre und die missbräuchliche Verwendung von Daten, inklusive Datendiebstahl, im Vordergrund. Rund 80% der Befragten sahen darin ein ernst zu nehmendes Risiko von Big Data.

Interessanterweise liessen sich die Befragten nicht in zwei Lager von Big-Data-Pessimisten und Big-Data-Optimisten einteilen, sondern sie unterschieden sich primär darin, wie hoch sie das Potenzial von Big Data einschätzten – im Positiven wie auch im Negativen.

### Kollektiver Handlungsbedarf

Ist das Potenzial von Big Data erkannt, so stellt sich die grundsätzliche Her-

ausforderung, die Entwicklungen im Bereich Big Data so zu kanalisieren, dass sie zum Nutzen der Menschen reichen. Hier ist aus Sicht der Befragten insbesondere die öffentliche Hand gefragt, wobei die Studienergebnisse eine Drei-Punkte-Strategie nahelegen: den Aufbau und die Pflege einer nationalen Dateninfrastruktur, Massnahmen zum Schutz vor Missbräuchen und Datenschutzverletzungen sowie die Umsetzung des Prinzips der persönlichen Datenhoheit, mit dem langfristigen Ziel, die Menschen zu befähigen, die Kontrolle über ihre Daten besser auszuüben.

Diese drei Stossrichtungen finden sich in der Strategie des Bundesrates für eine «Digitale Schweiz» wieder, in der unter anderem die drei folgenden Ziele vorgegeben werden (Schweizerische Eidgenossenschaft 2016):

- **Die Schweiz verfügt über eine kohärente und zukunftsorientierte Datenpolitik.**
- **Die Schweiz verfügt über eine nationale Dateninfrastruktur:** «Im Sinne einer «Good Data Governance» wird die Bereitstellung von Daten als Infrastrukturressourcen zur freien Weiterverwendung verstärkt. Durch den Aufbau einer nationalen Dateninfrastruktur, analog zur bereits bestehenden Geodateninfrastruktur, soll dem Open-Data-Prinzip über die gesamte Verwaltung hinweg, in der Forschung und teilweise auch in der Privatwirt-

schaft zum Durchbruch verholfen werden» (ibid., S. 6).

- **Die Einwohnerinnen und Einwohner der Schweiz können die Kontrolle über ihre eigenen Daten ausüben:** «Damit das Grundrecht jeder Person auf informationelle Selbstbestimmung umgesetzt werden kann und um Missbräuchen und Ungleichgewichten entgegenzuwirken, müssen die Datenschutzregeln neu gestaltet werden. Zudem gilt es, in enger Kooperation mit allen Mitarbeiterinnen und Mitarbeitern von Daten Mechanismen zu schaffen und Dienste bereitzustellen, die es den Einzelnen erlauben, sie betreffende Personen- und Sachdaten – unter grösstmöglicher Kontrolle – für die Weiterverwendung durch Dritte freizugeben oder deren Weiterverwendung zu unterbinden» (ibid., S.6).

Eine Sondierung der Berner Fachhochschule bei Schlüsselakteuren mit Blick auf den Aufbau und die Pflege einer nationalen Dateninfrastruktur ergab folgendes Bild (Estermann et al. 2016):

Eine nationale Dateninfrastruktur ist eine schweizweite (verteilte) technische Infrastruktur (Portale, Plattformen, Services etc.), die den Zugang zu und den Austausch von Daten auf Basis von definierten Regeln ermöglicht. Ihr Zweck besteht darin, die datenbasierte Wertschöpfung zu unterstützen und den potenziellen Mehrwert bestehender Daten freizusetzen.

Eine nationale Dateninfrastruktur ist aus vier Perspektiven zu betrachten (siehe Abb. 1):

- Die **Open-Data-Perspektive** stellt die Sekundärnutzung von Daten durch Dritte in den Fokus. Um diese zu vereinfachen, ist in einem ersten Schritt die rechtliche und technische Offenheit der Daten und Inhalte (freie Lizenzen, freie Datenformate, offener Zugang) zentral; während in einem weiteren Maturitätsschritt die Herstellung der semantischen Interoperabilität zwischen Datensätzen unterschiedlichen Ursprungs (Linked Data) im Vordergrund steht.
- Die **Perspektive der Basis-Register** stellt die gemeinsame Verwendung von Referenzdaten ins Zentrum. Im Vordergrund stehen Authentizität, Vollständigkeit und Interoperabilität

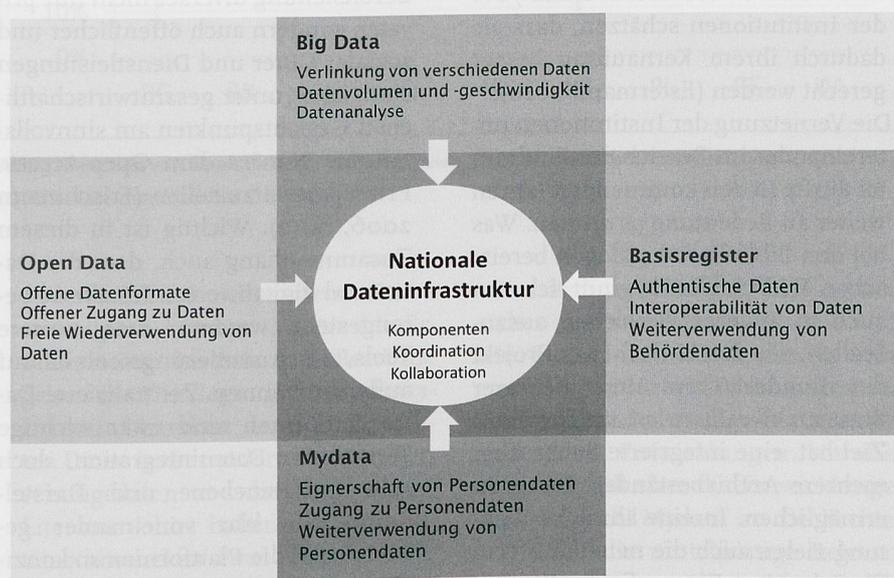


Abb. 1: Vier Perspektiven einer nationalen Dateninfrastruktur.

(Quelle: Estermann et al. 2016, CC BY-SA 4.0)

der Daten, aber auch die Bereitschaft verschiedener Akteure, einen gemeinsamen Pool von Basisdaten (z.B. ein nationales Unternehmensverzeichnis, ein zentrales Adressverzeichnis, ein Gemeindeverzeichnis oder Geo-Basisdaten) zu pflegen und zu nutzen.

- Die **Big-Data-Perspektive** fokussiert auf neue Methoden der Datenanalyse und das Gewinnen von neuen Erkenntnissen dank dem Aufbrechen von Datensilos und der Analyse von immer grösseren Mengen von Daten, welche oftmals in immer höherer Kadenz vorliegen.
- Und schliesslich stellt die **«Mydata»-Perspektive** das Empowerment des Einzelnen ins Zentrum, was die Verfügungsgewalt über die einen betreffenden Personendaten anbelangt, und damit einhergehend die Befähigung zu ihrer gezielten Freigabe für die Weiterverwendung zu ausgewählten Zwecken (z.B. für die medizinische Forschung).

Eine nationale Dateninfrastruktur ist kein monolithischer Block und soll auch nicht als Schweizer Insellösung funktionieren, sondern muss international vernetzt sein. Der Staat ist mitverantwortlich für die Bereitstellung der Daten und der technischen Infrastruktur, wobei es zum Aufbau einer nationalen Dateninfrastruktur eines schritt-

80% der Befragten gehen davon aus, dass Big Data es ermöglicht, neue Erkenntnisse zu gewinnen.

weisen Prozesses bedarf, der sowohl einer Bottom-up- als auch einer Top-down-Logik folgt: Die Bottom-up-Logik steht im Vordergrund, wenn einzelne Akteure ihre Daten veröffentlichen, Bürger und Bürgerinnen sowie die Privatwirtschaft mit einbeziehen und Anwendungsfälle aufzeigen. Nach der Top-Down-Logik sollten der strategische Rahmen abgesteckt und politische Aufträge formuliert werden. Für einen erfolgreichen Aufbau einer nationalen Dateninfrastruktur müssen diese beiden Entwicklungslogiken miteinander korrespondieren, wobei der Zivilgesellschaft eine Vermittlerrolle zukommt.

Neben diesen Grundpfeilern, welche bisher unumstritten sind, hat die Sondierung auch einiges an weiterem Klärungs- und Koordinationsbedarf zutage gefördert, was die Governance einer solchen Dateninfrastruktur anbelangt.

### Relevanz für Bibliotheken und Archive

Aus der Perspektive der Bibliotheken und Archive stellt sich die Frage der Relevanz dieser Entwicklungen für ihre

Die «Mydata»-Perspektive stellt das Empowerment des Einzelnen ins Zentrum.

eigenen Wandlungsprozesse angesichts der zunehmenden Digitalisierung der Gesellschaft. Diese soll im Folgenden in Form von kurzen Thesen skizziert werden, wobei diese als Diskussionsbeitrag in einer laufenden Debatte zu verstehen sind und nicht als abschliessende Bewertung:

- Wie aus einer internationalen Befragung von Gedächtnisinstitutionen hervorgeht, können Gedächtnisinstitutionen dank Open Data und Open Content die Sichtbarkeit der Institutionen und ihrer Bestände verbessern, die Vernetzung zwischen den Institutionen erhöhen, die Inhalte für Benutzerinnen und Benutzer einfacher zugänglich machen und die Interaktion mit ihren Nutzerinnen und Nutzern verbessern. Rund 70% der Institutionen schätzen, dass sie dadurch ihrem Kernauftrag besser gerecht werden (Estermann 2016).
- Die Vernetzung der Institutionen untereinander im Bereich der Findmittel dürfte in den kommenden Jahren weiter an Bedeutung gewinnen. Was bei den Bibliothekskatalogen bereits schon Tradition hat, beginnt sich nun auch in anderen Bereichen auszubreiten, wie das Linked-Data-Projekt des Bundesarchivs und mehrerer Staatsarchive illustriert, welches zum Ziel hat, eine integrierte Suche über mehrere Archivbestände hinweg zu ermöglichen. In eine ähnliche Richtung zielen auch die neue Plattform für Schweizer Fotografie des Fotobüros Bern und die geplante Plattform der darstellenden Künste der Schwei-

zerischen Theatersammlung und des Schweizer Tanzarchivs. Wir haben es mit einer Entwicklung zu tun, bei der im Rahmen von thematischen Plattformen die Metadaten in diversen Bereichen unter anderem dank Linked Data immer weiter integriert werden, was nicht nur die Suche, sondern auch die Verknüpfung mit weiteren Daten vereinfacht und im Endeffekt auch neuen Forschungserkenntnissen den Weg bereiten dürfte.

- Die Gedächtnisinstitutionen sollten sich an der Diskussion zu den Basisregistern beteiligen: Einerseits geht es darum, bestehende Register aus anderen Bereichen zu identifizieren, welche bei der Anreicherung und Vernetzung der Metadaten eine zentrale Rolle spielen könnten; andererseits gibt es auch im Bereich der Gedächtnisinstitutionen selber Verzeichnisse, die am besten gemeinsam gepflegt und allen Institutionen als Linked-Data-Dienste zur Mitbenutzung bereitgestellt werden. Der Betrieb eines Registers aller Schweizer Gedächtnisinstitutionen wäre ein naheliegendes Vorhaben.
- Zentral ist die Erkenntnis, dass Daten den Charakter von Infrastruktur-Ressourcen haben. Diese zeichnen sich dadurch aus, dass sie durch ihre Nutzung nicht aufgebraucht werden und gleichzeitig als Inputs für eine breite Palette von nachgelagerten Wertschöpfungsprozessen dienen. Insofern sie Voraussetzung für die Bereitstellung diverser nicht nur privater, sondern auch öffentlicher und sozialer Güter und Dienstleistungen sind, ist es unter gesamtwirtschaftlichen Gesichtspunkten am sinnvollsten, sie gemäss dem Open-Access-Prinzip bereitzustellen (Frischmann 2006, 2012). Wichtig ist in diesem Zusammenhang auch, dass die Daten und digitalisierten Inhalte so bereitgestellt werden, dass weitere Tools, z. B. Annotierungstools darauf aufbauen können. Zentralisierte Datenplattformen sind zwar wichtige Treiber der Datenintegration, doch sollten Datenebene und Darstellungsebene klar voneinander getrennt und die Plattformen so konzipiert werden, dass Dritte die Daten oder Teile davon auch in anderen Kontexten nutzen und im Rahmen

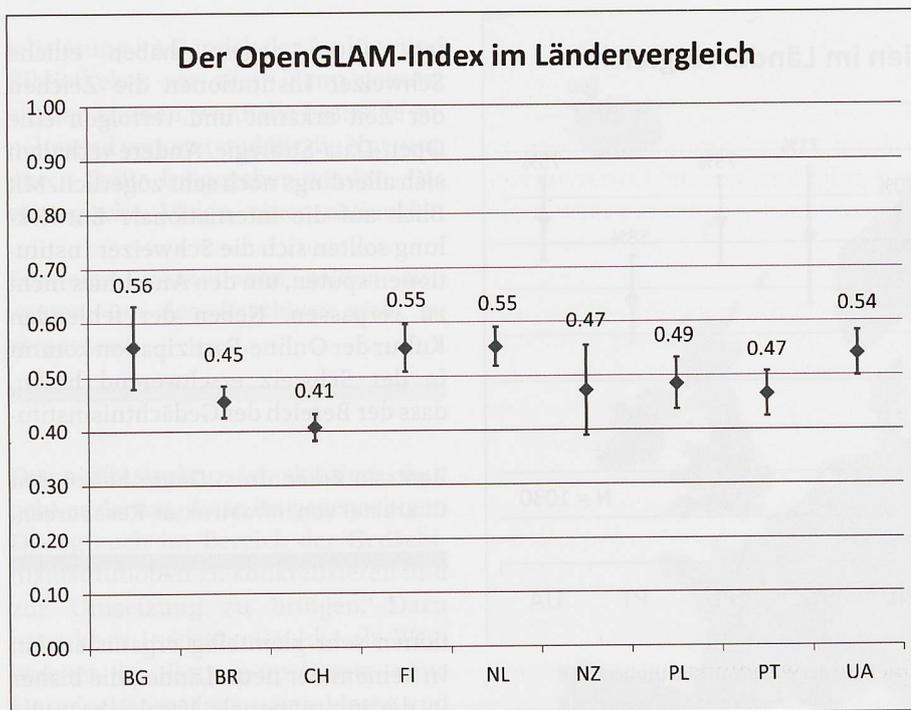


Abb. 2: Der OpenGLAM-Index im Ländervergleich. (Quelle: Estermann 2016, CC BY 4.0)

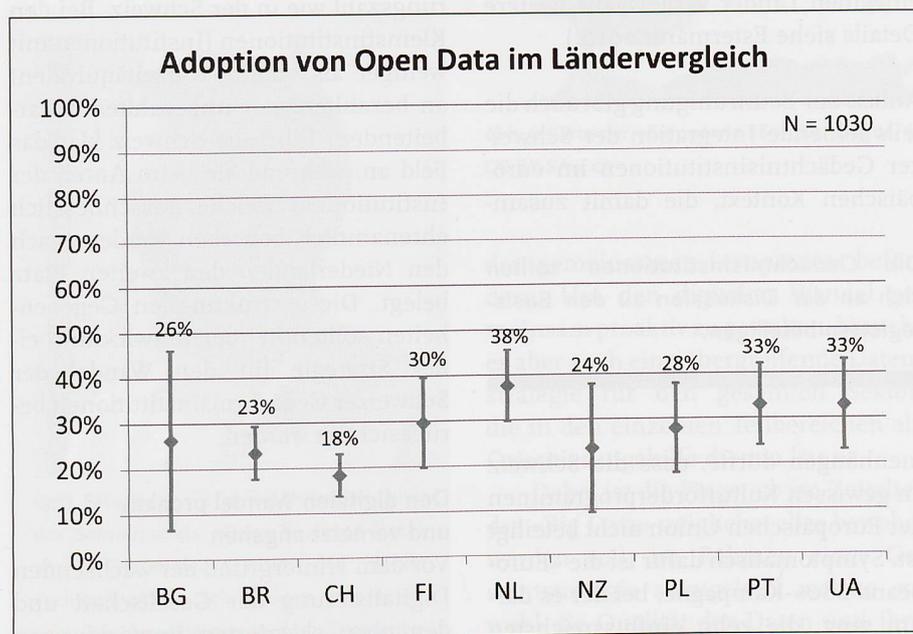


Abb. 3: Adoption von Open Data im Ländervergleich, in % der Gedächtnisinstitutionen. (Quelle: Estermann 2016, CC BY 4.0)

anderer Dienste mit weiteren Daten verknüpfen können. Neben der Erschliessung historischer Bestände gelten die Überlegungen zum Infrastruktur-Charakter der bereitgestellten Daten insbesondere auch für Forschungsdaten und die akademische Literatur – hier haben die Universitätsbibliotheken eine zentrale Rolle zu spielen – sowie für Datenbanken der Verwaltungsgeschäfte, deren Archivierung und Langzeitbetrieb im

Zuständigkeitsbereich der staatlichen Archive liegt.

- In Sachen informationelle Selbstbestimmung dürften Bibliotheken und Archive nicht unmittelbar von den aktuellen Entwicklungen betroffen sein. Es gilt aber, die Entwicklungen im Bereich Mydata zu beobachten und auf die mittelfristige Relevanz für Archive zu überprüfen, welche personenbezogene Daten in ihren Beständen haben, die für weiterfüh-

rende Auswertungen von Interesse sein könnten. Des Weiteren sollten die Vor- und Nachteile sowie die ethischen Implikationen im Zusammenhang mit der Personalisierung von Angeboten und dem Tracking der Nutzung von Beständen mit Blick auf den Datenschutz sorgfältig abgewogen werden. Wie die Big-Data-Studie der Berner Fachhochschule zeigt, stehen die Nutzerinnen und Nutzer der zunehmenden Personalisierung von Angeboten eher zwiespältig gegenüber: Einerseits kann daraus eine deutliche Verbesserung von Dienstleistungen resultieren, andererseits ruft sie bisweilen auch Unbehagen hervor und wird als invasiv empfunden.

- Im Zusammenhang mit Big Data dürfte der Fokus bei den Bibliotheken und Archiven vor allem auf dem Aspekt der Verknüpfung von heterogenen Datenbeständen liegen. Was die Verarbeitung grosser Datenmengen anbelangt, so werden bei den klassischen Daten eher geringe Volumina erreicht. Eine Ausnahme dürften die Entwicklungen im Bereich Bildanalyse und Bilderkennung bilden, wo beim eigentlichen Content durchaus mit grossen Datenmengen zu rechnen ist. Von einer hohen Aktualisierungskadenz dürfte bei der Datenanalyse hingegen kaum etwas zu spüren sein. Im Zusammenhang mit der Integration von heterogenen Datenbeständen dürfte es lohnenswert sein, ergänzend zum Linked-Data-Ansatz auch das Potenzial von Wikidata auszuloten, welches sich immer stärker in Richtung einer globalen Datendrehscheibe entwickelt und dessen Daten mittlerweile auch über einen SPARQL-Endpoint abrufbar sind, womit sie sich nahtlos ins Semantic Web (Linked Data) einfügen.

### Schweizer Institutionen international im Hintertreffen

Besondere Aufmerksamkeit ist auch dem internationalen Kontext zu widmen. Im Ländervergleich spielt die Schweiz, was die technische Infrastruktur anbelangt, in der besten Liga. Dies könnte mit Blick auf die aktuelle Digitalisierung der Gesellschaft durchaus zuversichtlich stimmen. Was die Kultur der Online-Partizipation (gemessen am

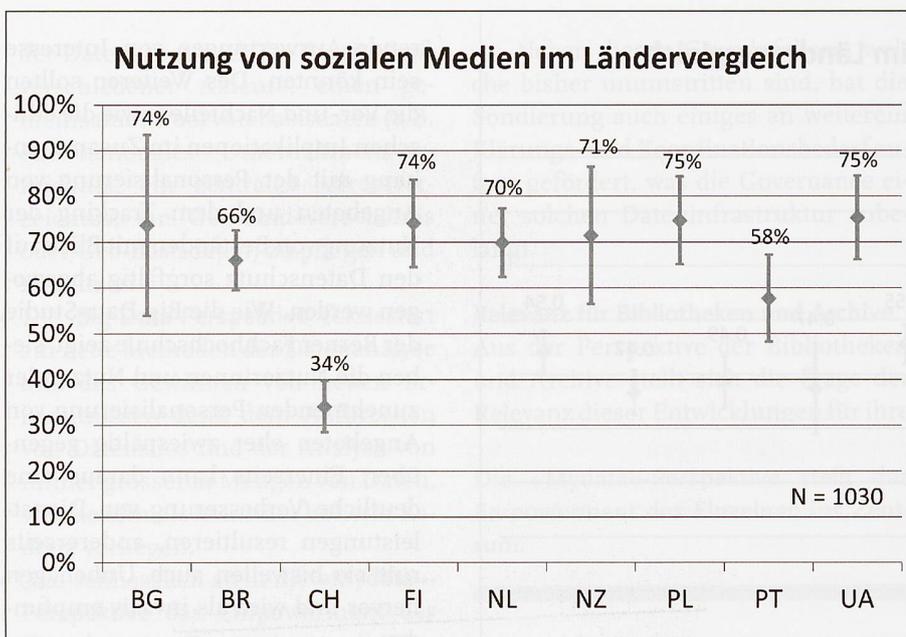


Abb. 4: Nutzung von sozialen Medien im Ländervergleich, in % der Gedächtnisinstitutionen. (Quelle: Estermann 2016, CC BY 4.0)

UNPACS-E-Participation-Index) be-  
trifft, rangiert die Schweiz allerdings  
im hinteren Mittelfeld. Im Ergebnis  
sind die Schweizer Gedächtnisinstitu-  
tionen, was die Öffnung der Daten und  
der Inhalte angeht, wie auch in Be-  
zug auf die Nutzung der Möglichkeiten  
der Online-Partizipation im europä-  
ischen Vergleich auf den hinteren Rän-  
gen zu finden. Abbildung 2 zeigt den  
OpenGLAM-Index im internationalen  
Vergleich. Der Index stellt eine Mess-  
grösse für die drei Dimensionen «Ein-  
bezug von Nutzern», «Open Data» und  
«Open Content» dar, womit er einen  
guten Anhaltspunkt gibt, wo die Ge-  
dächtnisinstitutionen eines Landes be-  
züglich des digitalen Wandels stehen.

**Eine nationale Dateninfrastruktur soll  
nicht als Schweizer Insellösung funk-  
tionieren, sondern international ver-  
netzt sein.**

Die Abbildungen 3 und 4 zeigen die  
Anteile der Gedächtnisinstitutionen in  
den jeweiligen Ländern, bei denen das  
Bereitstellen von Daten als «Open Da-  
ta» bzw. die Nutzung von sozialen Me-  
dien im Kontakt mit den Nutzerinnen  
und Nutzern gängige Praxis ist. (Die  
Balken zeigen 95%-Konfidenzintervalle  
an, deren Länge aufgrund der unter-  
schiedlichen Stichprobengrößen der

einzelnen Länder variiert; für weitere  
Details siehe Estermann 2016.)

Anlass zur Beunruhigung gibt auch die  
teils fehlende Integration der Schwei-  
zer Gedächtnisinstitutionen im euro-  
päischen Kontext, die damit zusam-

**Die Gedächtnisinstitutionen sollten  
sich an der Diskussion zu den Basis-  
registern beteiligen.**

menhängen dürfte, dass die Schweiz  
an gewissen Kulturförderprogrammen  
der Europäischen Union nicht beteiligt  
ist. Symptomatisch dafür ist die «Euro-  
peana 280»-Kampagne, bei der es dar-  
um ging, die zehn einflussreichsten  
Kunstwerke jedes der europäischen  
Länder zu identifizieren und online  
verfügbar zu machen. Wie in Abbil-  
dung 5 ersichtlich ist, erscheint die  
Schweiz dort als einsame Insel inmit-  
ten von Europa; ein Phänomen, das im  
kulturellen Bereich auch schon bei frü-  
heren Aktionen zu beobachten war. Vor  
diesem Hintergrund stellt sich die Fra-  
ge, was es für Schweizer Institutionen  
längerfristig bedeutet, bei europä-  
ischen Initiativen aussen vor zu bleiben.

Wie die Erfahrungen im Zusammen-  
hang mit der schweizerischen  
OpenGLAM-Arbeitsgruppe in den letz-

ten Jahren zeigen, haben etliche  
Schweizer Institutionen die Zeichen  
der Zeit erkannt und verfolgen eine  
Open-Data-Strategie. Andere verhalten  
sich allerdings noch sehr zögerlich. Mit  
Blick auf die internationale Entwick-  
lung sollten sich die Schweizer Institu-  
tionen spüren, um den Anschluss nicht  
zu verpassen. Neben der fehlenden  
Kultur der Online-Partizipation kommt  
in der Schweiz erschwerend hinzu,  
dass der Bereich der Gedächtnisinstitu-

**Zentrale Erkenntnis: Daten haben den  
Charakter von Infrastruktur-Ressourcen.**

tionen sehr kleinteilig organisiert ist:  
In keinem der neun Länder, die bisher  
in die internationale Vergleichsstudie  
einbezogen wurden, gibt es so viele In-  
stitutionen im Verhältnis zur Bevölke-  
rungszahl wie in der Schweiz. Bei den  
Kleinstinstitutionen (Institutionen mit  
weniger als einem Vollzeitäquivalent  
an bezahlten und unbezahlten Mitar-  
beitenden) führt die Schweiz klar das  
Feld an, während sie beim Anteil der  
Institutionen, welche ausschliesslich  
ehrenamtlich betrieben werden, nach  
den Niederlanden den zweiten Platz  
belegt. Diese strukturellen Gegeben-  
heiten sollten bei der Entwicklung einer  
Strategie für den Wandel der  
Schweizer Gedächtnisinstitutionen be-  
rücksichtigt werden.

**Den digitalen Wandel proaktiv  
und vernetzt angehen**

Vor dem Hintergrund der wachsenden  
Digitalisierung der Gesellschaft und  
den oben skizzierten Entwicklungen  
liegt es nahe, dass die Schweizer Ge-  
dächtnisinstitutionen den digitalen  
Wandel proaktiv und vernetzt angehen  
sollten: Proaktiv, weil sich der Wandel  
nicht automatisch aus der komfortab-

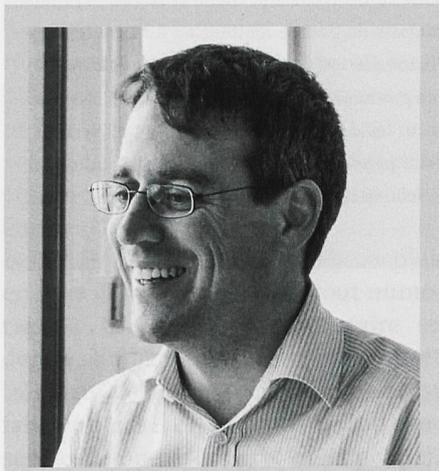
**Was die Kultur der Online-Partizipation  
betrifft, rangiert die Schweiz im hinteren  
Mittelfeld.**

len Situation hinsichtlich der techni-  
schen Infrastruktur und der finanziel-  
len Ausstattung ergibt. Vernetzt, weil  
die Chancen der fortschreitenden Digi-

talisierung im Bereich der Archive und Bibliotheken vor allem dann genutzt werden können, wenn die Bestände zunehmend vernetzt und für die Nutzung durch Dritte freigegeben werden. Die strategische Vision einer nationalen

Schweizer Institutionen sollten sich sputen, um den Anschluss nicht zu verpassen.

Dateninfrastruktur ist skizziert; nun geht es darum, diese im gegenseitigen Dialog auch im Bereich der Gedächtnisinstitutionen zu konkretisieren und zur Umsetzung zu bringen. Dazu braucht es Institutionen, die als Wegbereiter im Bottom-up-Ansatz neue Herangehensweisen ausprobieren und



Beat Estermann ist stellvertretender Leiter des Schwerpunkts «Open & Linked Data» am E-Government-Institut der Berner Fachhochschule. Seit einigen Jahren gilt sein Forschungsinteresse dem Themenbereich «Open Data» und im speziellen «OpenGLAM», d.h. Open Data und Crowdsourcing unter Gedächtnisinstitutionen. Er koordiniert die schweizerische OpenGLAM-Arbeitsgruppe des Vereins «opendata.ch», welche sich für Open Data im Bereich der Gedächtnisinstitutionen einsetzt, sowie die eCH-Fachgruppe «Open Government Data», welche sich Standardisierungsfragen rund um Open Government Data widmet. Er erbringt Consulting-Dienstleistungen für Bundesämter und Gedächtnisinstitutionen und hat 2015 im Auftrag des Bakom die Studie «Big Data: Chancen, Risiken und Handlungsbedarf des Bundes» mitverfasst.

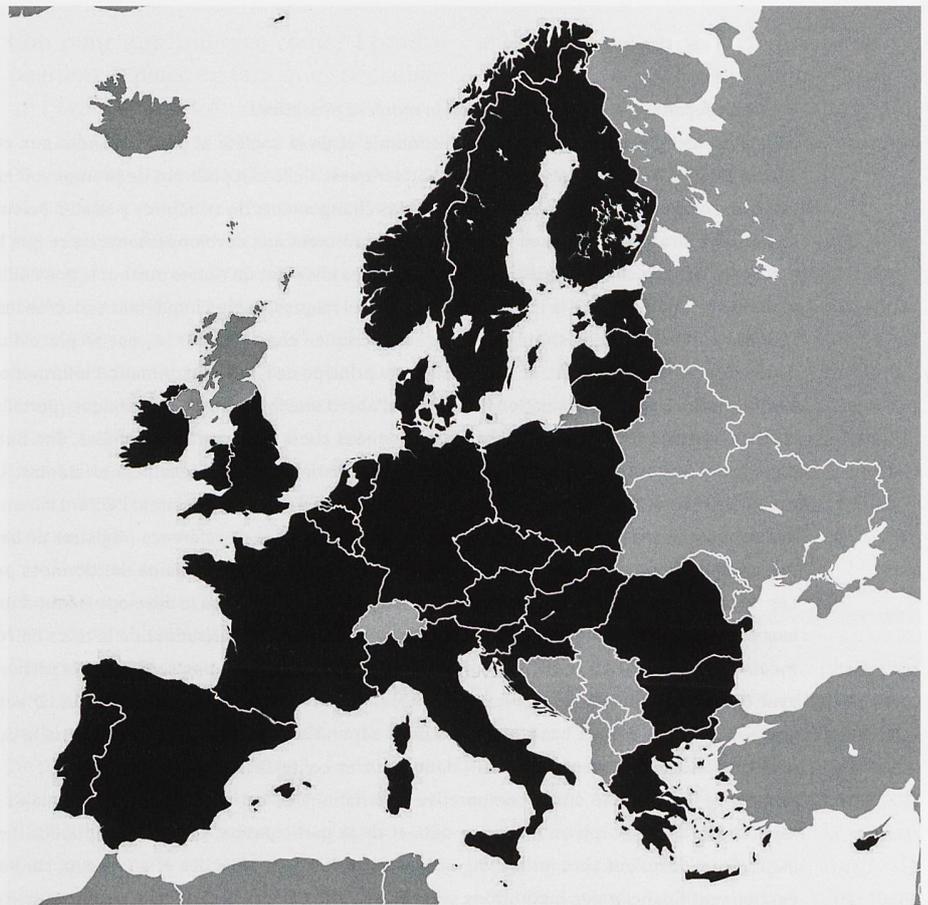


Abb. 5: Karte der «Europeana 280»-Kampagne. (Quelle: Llywelyn2000, Wikimedia Commons, CC BY-SA 4.0)

den gemeinsamen Lernprozess befördern. Um den digitalen Wandel gemeinsam proaktiv zu gestalten, braucht es aber auch eine übergreifende Datenstrategie für den gesamten Sektor, die in den einzelnen Teilbereichen als Orientierungshilfe dienen kann.

Dabei ist die Frage, ob im Zeitalter des «Big Data» möglichst alles bewahrt werden soll, im Rahmen einer Gesamtstrategie anzugehen, welche sowohl die Qualität der Daten und ihre Vernetzung als auch datenschützerische Belange berücksichtigt. In die entsprechenden Debatten sollten sich auch die Fachleute I+D einbringen.

Kontakt: [beat.estermann@bfh.ch](mailto:beat.estermann@bfh.ch)  
Twitter: @beatestermann

#### Bibliographische Angaben

- Estermann Beat, OpenGLAM Benchmark Survey – Measuring the Advancement of Open Data/Open Content in the Heritage Sector. Paper presented at the International Symposium on the Measurement of Digital Cultural Products, 9-11 May 2016, Montreal, Canada.
- Estermann Beat, Fraefel Marianne, Neuron Alessia, Conceptualizing a National Data Infrastructure for Switzerland. Paper Presented at the EGPA 2016 Conference, 24-26 August 2016, Utrecht, The Netherlands.
- Frischmann Brett M., An Economic Theory of Infrastructure and Commons Management. *Minnesota Law Review*, 89, 2006, pp. 917-1030.
- Frischmann Brett M., Infrastructure: The Social Value of Shared Resources. *Oxford Scholarship Online*, 2012.
- Jarchow Thomas und Estermann Beat, Big Data: Chancen, Risiken und Handlungsbedarf des Bundes. Studie im Auftrag des Bundesamts für Kommunikation, Berner Fachhochschule, Oktober 2015.
- Schweizerische Eidgenossenschaft, Strategie «Digitale Schweiz», April 2016.

## *Big data: façonner le virage numérique de manière proactive*

Face à la numérisation croissante de l'économie et de la société et pour répondre aux évolutions liées au «big data», le Conseil fédéral a adopté en avril 2016 la stratégie «Suisse numérique». Celle-ci a pour but de promouvoir et d'encadrer la transformation numérique de la société, de sorte que la mutation numérique et les changements de structures y relatifs soient façonnés de manière proactive et en réseau. Avec sa nouvelle stratégie, le Conseil fédéral réagit notamment aux développements de ce que l'on appelle communément les «big data». Ceux-ci recèlent autant de chances que de risques: parmi les chances, on notera surtout la possibilité de gagner de nouvelles connaissances en découvrant des relations jusque-là inconnues; du côté des risques, le plus inquiétant concerne les possibles violations de la protection des données. Une des parties-clés de la stratégie «Suisse numérique» consiste dans la mise en place d'une infrastructure de données nationale fondée sur le principe de l'open data et sur l'application du principe de l'autodétermination informationnelle dans le domaine des données personnelles. L'infrastructure de données nationale est tout d'abord une infrastructure technique (portails, plateformes, services, etc.) répartie à l'échelle de la Suisse qui permet l'accès et l'échange de données sur la base de règles définies. Son but est de soutenir la création de richesses à partir de données et surtout de libérer la valeur ajoutée potentielle des données déjà existantes. Une infrastructure de données nationale peut être examinée sous quatre perspectives différentes, chacune d'entre elles mettant l'accent sur un aspect particulier, comme l'exploitation secondaire des données (open data), l'utilisation en commun de données de référence (registres de base), de nouvelles méthodes d'analyse des données (big data) ou encore l'autodétermination informationnelle dans le domaine des données personnelles (mydata).

Les institutions patrimoniales devraient s'engager activement dans le développement d'une infrastructure de données nationale, car celle-ci leur permet d'être mieux adaptées à leur mission. Ainsi, le renforcement de la mise en réseau entre les institutions dans le cas des instruments de recherche est particulièrement significatif, car ce n'est pas seulement la recherche que cela simplifie, mais également la relation avec d'autres données, et par conséquent l'acquisition de nouvelles connaissances. D'autres opportunités découlent de l'exploitation collaborative des registres de base, ainsi que de la séparation claire entre les niveaux de la donnée et de sa représentation, ce qui permet à un tiers d'utiliser les données également dans d'autres contextes.

Comme le montre une étude comparative internationale, les institutions patrimoniales suisses se trouvent à la traîne d'autres pays par rapport à la mise en œuvre de l'open data et de la participation en ligne des publics. Les résultats de l'étude mettent en lumière que les institutions devraient aborder le virage numérique de façon proactive et en réseau, car l'excellence des infrastructures techniques et le bon équipement financier des institutions suisses ne suffisent pas. Pour que le virage numérique se produise avec succès, il faut aussi bien des institutions innovatrices qui se lancent dans de nouvelles approches, qu'une stratégie globale sur les données pour l'ensemble du secteur. Ainsi, on le voit, la problématique du «tout-conserver» chère aux partisans des «big data» est à penser dans une stratégie globale tenant compte à la fois de la qualité des données et de leur mise en réseau que de l'évaluation des risques sur la protection des données. Et les professionnels I+D sont tout désignés pour y apporter leurs compétences.

## Le coût environnemental de la gestion documentaire

Sophie Roy, responsable du blog *Convergence* de l'Association des archivistes du Québec

**De plus en plus, l'environnement est une préoccupation de nos sociétés et cela se reflète dans notre volonté de réduire la consommation de papier. Pourtant, est-ce que vraiment la production, l'utilisation et la conservation de données analogiques polluent moins que leurs consœurs numériques? Rien n'est moins sûr d'après plusieurs études.**

**L'environnement a la cote actuellement**  
Depuis l'avènement de l'ère industrielle, la santé de notre planète soulève

des questions. Le contexte de croissance économique fait s'accélérer la destruction de nos ressources naturelles et pousse l'humanité à agir. En 2015, la Conférence internationale de Paris sur le climat (COP21) semble avoir eu des résultats encourageants pour l'avenir de la planète. Les décideurs, gouvernements et entreprises, ont désormais la responsabilité d'agir. De plus, l'environnement a la cote dans notre quotidien avec les tendances zéro déchets, légumes moches, mobilité douce et voitures électriques, compostage et autres réutilisations des ressources.

L'industrialisation passe aujourd'hui à une étape de dématérialisa-

tion et de robotisation dans un but d'efficacité, de mobilité, mais aussi de réduction de la consommation de papier. Mais, dans le contexte de croissance mondiale, qu'en est-il du coût environnemental de la conservation et de la duplication sous toutes ses formes de la masse documentaire numérique? Et comment le spécialiste de la documentation peut-il intervenir, réagir face à ces enjeux de taille?

### **Quel coût environnemental des TI?**

Le calcul du coût environnemental des technologies de l'information n'est pas encore courant dans les organisations. Les réseaux et les serveurs consomment

raient environ 2% de la consommation énergétique mondiale<sup>1</sup>. Pouvez-vous imaginer, selon Google, qu'une recherche effectuée sur son moteur émet 0,2 g de CO<sub>2</sub>? Il y a 3,3 milliards de recherches effectuées chaque jour, ce qui équivaut à 660 tonnes de CO<sub>2</sub> émises par jour, ou 241 000 tonnes par an<sup>2</sup>. Un autre exemple, un livre papier représente 1,3 kg d'émission de carbone, alors qu'un livre numérique pour un iPad représente 135 kg<sup>3</sup>. Ou encore, l'envoi de courriels par 100 personnes d'une même entreprise équivaut à 14 allers-retours Paris New York ou 13,6 tonnes de CO<sub>2</sub> par année<sup>4</sup>.

Les méthodes de calcul existantes des coûts environnementaux des technologies de l'information ne font pas l'unanimité<sup>5</sup>, mais les gens de l'industrie s'entendent pour dire que le numérique pollue plus à long terme que l'utilisation du papier. Surtout par l'effet de duplication de l'information et du stockage infini des données. Des géants du numérique, tels que Google et Facebook, ont déclaré vouloir s'attaquer au problème en investissant dans les nouvelles technologies propres à faible émission de carbone et consommation énergétique.

Les entreprises deviennent de plus en plus des organisations «tout numérique». La gestion documentaire se déplace du papier vers le numérique et très peu d'élagage documentaire y est fait. On stocke, on stocke sans trop penser à cette masse qui grossit (et pollue). Il y a cependant des initiatives et des prises de conscience qui émergent tranquillement dans le monde documentaire. L'argument environnemental fait son chemin dans les organisations et cela peut représenter une opportunité pour la gestion documentaire. Surtout qu'il y a un enjeu monétaire pas très loin, car en réduisant la masse documentaire numérique, on réduit l'achat de matériel de stockage et le temps de gestion de tout ce matériel.

### Détruire ou acheter du stockage supplémentaire?

Cependant, la question de la diminution du stockage numérique ne ressort pas beaucoup dans les discours de Facebook, Twitter et autres. Les spécialistes de l'information documentaire ne pourraient-ils pas apporter une solu-

tion pour diminuer les coûts? Prendre position et place en tant que spécialiste de l'évaluation et du désherbage?

De mon expérience, la destruction de données numériques dans les organisations n'est pas une priorité. Acheter plus de stockage et tout conserver reste la norme. La question du quoi détruire est toujours épineuse et la masse documentaire, qui se chiffre en millions de documents et de fichiers, reste toujours trop grande pour les ressources humaines disponibles, car oui, quand on parle de tri, il y a normalement un humain derrière qui réfléchit et qui pose une action.

Il semble y avoir deux tendances chez les spécialistes. D'un côté, l'évolution des technologies peut nous aider. Programmer des super logiciels, voire des robots, à faire le tri selon les critères établis par les spécialistes de l'information peut être une solution. Structurer les données en amont est aussi un champ d'intervention à développer davantage par les spécialistes. D'un autre côté, à quoi bon faire le tri? Pourquoi ne pas laisser l'oubli et l'obsolescence faire le travail? Selon la professeure en Information Studies Anne Gilliland, «l'énergie mise pour la sélection des documents nés numériques devrait être transférée dans une autre activité»<sup>6</sup>, ici le «développement de stockage plus écologique, de moyens plus so-

**L'argument environnemental peut représenter une opportunité pour la gestion documentaire.**

phistiqués et réfléchis pour décrire et repérer les documents, ainsi que sur la compartimentation de la quantité massive de matériel préservé pour soutenir les besoins d'utilisation et de réutilisation toujours plus vastes des utilisateurs<sup>7</sup>. Gilliland souligne, avec raison, «qu'il est impensable de pouvoir éliminer toutes traces des documents numériques que l'on retrouve sur les réseaux»<sup>8</sup>. Les ressources devraient être engagées à centraliser, retenir et sécuriser «la preuve numérique et que les professionnels deviennent les agents du contrôle intellectuel et physique. La seule élimination sélective envisageable est celle basée sur l'oubli. Pour

elle, l'évaluation a fait son temps»<sup>9</sup>. L'archéologue et l'historien du futur se feront un plaisir de reconstituer et réinterpréter les traces que nous aurons laissées. Il en a toujours été ainsi d'ailleurs.

### Quels trous dans notre mémoire?

Dans ce cas de figure, les efforts devraient être mis sur la description et la catégorisation des données essentielles. Le reste pourrait être détruit après un délai établi par les parties prenantes.

**Le document numérique tel que nous le connaissons aujourd'hui n'existera plus dans quelques années.**

Les technologies évoluent vite et le coût du maintien des anciennes versions est important. Le big data peut apporter des pistes de solution, mais l'information se complexifie toujours plus. Le document numérique tel que nous le connaissons aujourd'hui n'existera

- 1 Sciences-tech, «Internet engloutit près de 8% de la consommation électrique suisse» [En ligne], 2015, <<http://www.rts.ch/info/sciences-tech/7286072-internet-engloutit-pres-de-8-de-la-consommation-electrique-suisse.html>> (consulté le 2016-06-20).
- 2 Manenti, Boris, «Les vidéos de chat polluent plus que les avions» [En ligne], 2015, <<http://tempsreel.nouvelobs.com/les-internets/20150930.OBS6808/les-vidéos-de-chat-polluent-plus-que-les-avions.html>> (consulté le 2016-06-21).
- 3 Jost, Clément, «COP21: nos pratiques numériques sont-elles écologiques?» [En ligne], 2015, <<http://www.archimag.com/vie-numerique/2015/11/30/cop21-pratiques-numeriques-ecologiques>> (consulté le 2016-06-15).
- 4 Jost, idem.
- 5 Manenti, idem.
- 6 Compte-rendu du livre Wathley, Patricia et Caroline Brown, Archives and Recordkeeping: Theory Into Practice, London: Facet Publishing, 2013. Fait par Alexandra Buthieux paru dans la revue Archives, volume 16, numéro 1, 2015-2016, p. 104.
- 7 Idem.
- 8 Idem.
- 9 Idem.

plus dans quelques années. La donnée va devenir reine et les systèmes de lecture plus sophistiqués nous la retourneront avec intelligibilité. Mais établir ces critères de priorisation des données essentielles et mettre le paquet dans leur préservation et leur accessibilité au détriment d'une masse moins impor-

tante qui peut disparaître, cela ne nous renvoie-t-il pas aux principes de l'évaluation archivistique? Ainsi, coûte que coûte, même sans destructions volontaires systématiques, nous devons vivre avec une mémoire électronique perforée. Que fait-on avec une belle robe de mariée en dentelle qui appartenait à notre arrière-grand-mère et que nous retrouvons dans une vieille malle avec des milliers de petites tâches de rouille, héritage d'anciennes mites? Nous en gardons le souvenir, mais le reste va aux ordures. Il en va de même pour les documents qui dorment sur nos réseaux, on les retrouve avec des trous, car les logiciels ne sont plus tout

à fait interopérables. On les supprime alors et la planète souffle un peu mieux. Toutefois, est-ce vraiment une réponse durable aux problèmes environnementaux? Si nous laissons aux hasards du temps le soin de nous concocter cette sélection, nous perdrons la maîtrise de la gestion documentaire. Est-ce vraiment le patrimoine que nous souhaitons léguer à nos descendants? Une mémoire trouée aléatoirement et une société déshumanisée gouvernée par la mémoire imbécile des ordinateurs? À réfléchir ...

Contact: [soroy541@gmail.com](mailto:soroy541@gmail.com)

Twitter: @So\_Roy



Sophie Roy a une maîtrise en sciences de l'information, profil recherche, de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal au Canada. Le sujet de son mémoire de maîtrise touchait la diffusion des archives avec les technologies du Web 2.0 dans un contexte muséal. Elle a d'abord travaillé en gestion documentaire et veille informationnelle dans le domaine de l'ingénierie civile, puis maintenant dans le domaine de la finance à Montréal.

## ABSTRACT

### *Auch Dokumentenmanagement generiert Umweltkosten*

Unsere Gesellschaft kümmert sich mehr und mehr um die Umwelt, und diese Haltung zeigt sich auch in unserem Willen, unseren Papierverbrauch zu reduzieren. Indes stellt sich die Frage: Ist die Produktion, Verwendung und Erhaltung von analogen Unterlagen allenfalls weniger belastend für die Umwelt als ihre digitalen Pendanten? Auch nach mehreren Studien gibt es darüber noch keine Klarheit. Die Entmaterialisierung der Information bringt eine Zunahme von technischen Geräten mit sich, was einen hohen Energieverbrauch verursacht. Welche Lösungen können Informationsspezialisten für die Reduktion dieser Masse von Dokumenten bieten? Vernichten oder zusätzlichen Speicher kaufen? Hier sind mehrere Tendenzen auszumachen: Auf der einen Seite die technologische Entwicklung und die Automatisierung, die uns helfen können, eine sorgfältige Auswahl zu treffen, auf der anderen Seite das fortschreitende Vergessen und das Veralten von Technologien, die unser virtuelles Gedächtnis allmählich auslöschen.

## Impressum arbido print arbido newsletter arbido website

print: © arbido ISSN 1420-102X  
website + newsletter: © arbido newsl. ISSN 1661-786X  
[www.arbido.ch](http://www.arbido.ch)

### Offizielle Publikationsorgane – *Organes officiels de publication*

Bibliothek Information Schweiz BIS ([www.bis.info](http://www.bis.info))  
Verein Schweizerischer Archivarinnen und Archivare  
VSA-AAS ([www.vsa-aas.org](http://www.vsa-aas.org))  
Bibliothèque Information Suisse BIS  
Association des archivistes suisses AAS-VSA

### Chefredaktor – *Rédacteur en chef*

Sara Marty (sm) [sm@arbido.ch](mailto:sm@arbido.ch)

### Redaktion BIS – *Rédaction BIS*

Katja Böspflug (kb), Myriam Erwin (me),  
Daniela Rüegg (dr)

### Redaktion VSA-AAS – *Rédaction AAS-VSA*

Marguérite Bos (mb), Gilliane Kern (gk)

### Layout, Druck – *Graphisme, Impression*

Stämpfli AG, Wölflistrasse 1, Postfach 8326, 3001 Bern,  
[www.staempfli.com](http://www.staempfli.com)

### Inserate – *Annonces*

Tel. 031 300 63 41, Fax 031 300 63 90  
E-Mail [inserate@staempfli.com](mailto:inserate@staempfli.com)

### Abonnemente – *Abonnements*

Tel. 031 300 63 44, Fax 031 300 63 90  
E-Mail [abonnemente@staempfli.com](mailto:abonnemente@staempfli.com)

4 Hefte/Jahr: Fr. 115.– (im Mitgliederbeitrag inbegriffen).

Auslandsabonnemente: Fr. 135.– (inkl. Porto).

Einzelnummer: Fr. 30.– (plus Porto und Verpackung).

4 cahiers/année: Fr. 115.– (compris pour les adhérents).

Abonnements de l'étranger: Fr. 135.–

(affranchissement compris).

Vente au numéro: Fr. 30.–  
(plus affranchissement et emballage).

### Termine – *Dates*

Die nächste Ausgabe erscheint am 1. Dezember 2016.  
*Le prochain numéro paraît le 1<sup>er</sup> décembre 2016.*

### Thema – *Thèmes*

arbido Nr. 4: Sichtbar werden, sichtbar sein / *Devenir visible, être visible*

Inserateschluss Nr. 4 – *Délai d'insertion N° 4*: 19.10.2016

### Newsletter

Pro Jahr erscheinen 8 elektronische Newsletters.

Siehe [www.arbido.ch](http://www.arbido.ch)

*Parution de 8 newsletters électroniques par an.*

Voir [www.arbido.ch](http://www.arbido.ch)

Publiziert mit Unterstützung der Schweizerischen Akademie der Geistes- und Sozialwissenschaften SAGW  
*Publié avec le soutien de l'Académie suisse des sciences humaines et sociales ASSH*

# Gérer le cycle de vie de l'information pour maîtriser les risques juridiques et financiers: le *discovery*

Lourdes Fuentes Hashimoto,  
chef du service des archives  
électroniques, Groupe Total

**Face aux procédures judiciaires américaines de *discovery*, l'évaluation archivistique se pose en termes différents pour les organisations qui risquent d'y être confrontées. La surconservation des données qui ne sont plus nécessaires aux activités de l'organisation présente un double risque: risque de communiquer des informations stratégiques qui peuvent porter atteinte à son action et à son image et risque de divulguer massivement des données susceptibles de porter atteinte aux droits des personnes.**

## Respect du cycle de vie et réglementations

L'identification des documents produits par une organisation – publique, parapublique ou privée – dans l'exercice de ses activités, l'attribution d'une durée de conservation à chaque typologie documentaire ainsi que l'association d'un sort final – destruction ou conservation définitive – lorsque la durée de conservation arrive à échéance constituent le pilier de la gestion documentaire (ou records management) et de l'archivage.

Cette pratique que les records managers ou archivistes appellent communément «gestion du cycle de vie de l'information» est au cœur d'une bonne politique de gouvernance de l'information: elle permet de maîtriser dans la durée l'ensemble des documents utiles à une organisation pour des raisons juridiques, stratégiques et/ou patrimoniales. La gestion du cycle de vie de l'information, quel que soit son support, est fondamentale pour éviter à la fois la destruction inopinée de documents ayant encore une utilité, la dispersion de l'information confidentielle, la surconservation génératrice de coûts et de pertes de productivité et pour garantir la constitution du patrimoine informationnel. L'attribution des durées de conservation s'opère en fonction du cadre réglementaire et juridique

applicable aux activités de l'organisation. Ce cadre peut être plus ou moins contraignant et complexe: il est impératif de respecter de nombreuses règles propres au pays où l'activité est exercée ainsi que la réglementation européenne, le cas échéant.

Toutes les organisations exerçant leur activité aux États-Unis, par exemple via des filiales implantées sur le territoire américain, ou ayant des liens commerciaux avec ce pays, sont soumises également aux injonctions de la justice américaine, en particulier à la procédure dite de «*discovery*». L'obligation de répondre à ces injonctions renforce la nécessité de contrôler davantage l'information. La gestion de son cycle de vie est incontournable. En quoi consiste exactement le *discovery*? Et quels sont les enjeux en termes de gouvernance de l'information?

## Les procédures de *discovery*

Le *discovery* est une procédure en droit civil fédéral américain (*federal rules of civil procedure*). Elle s'applique en amont d'un procès pour permettre la constitution du dossier de chacune des parties qui a donc le droit de demander des dossiers/documents à l'autre, voire d'intervenir directement à la recherche de tout élément pouvant conduire «raisonnablement à une preuve ou évidence recevable par la justice» selon les termes de la loi américaine. Ainsi, chaque partie est libre de réunir tous les éléments qui permettent de constituer le dossier qui sera présenté au tribunal pour attaquer l'autre partie dans le cadre d'un contentieux.

La recherche d'éléments pouvant constituer une preuve ouvre la voie à une recherche documentaire très élargie qui peut être qualifiée d'intrusive en raison de son caractère soudain et contraignant. Il convient de noter que si une partie prend la décision de ne pas communiquer des informations, son refus pourra être utilisé contre elle. Elle encourt alors des sanctions. Toutes les parties ont donc intérêt à jouer le jeu en autorisant la recherche documentaire

et en livrant tout élément réclamé par la partie adverse. Toute information dissimulée qui serait découverte a posteriori risque de porter atteinte à la partie, voire de lui faire perdre le procès.

Le terme d'«*e-discovery*» est utilisé lorsque l'on se réfère à des documents mobilisés dans les procédures de *discovery* qui sont sous forme électronique (données applicatives, logs, documents



Lourdes Fuentes Hashimoto est une archiviste franco-mexicaine spécialisée dans la conception, l'audit et le pilotage des systèmes d'information pour la gestion des archives, tous supports confondus, et la conduite de projets d'archivage électronique dont les interfaçages entre les applications métier et les systèmes d'archivage électronique. Elle a conduit plusieurs projets informatiques depuis plus de sept ans. Elle travaille, depuis 2013, au sein du Groupe Total où elle a dirigé le projet e-TRACES de refonte et de migration du système d'archivage électronique. Le système est en production depuis le mois d'avril 2016 et sera interfacé fin 2016 avec plusieurs applications. Il est ouvert aux collaborateurs du groupe dans le monde (130 pays) et permet de conserver les documents dès leur validation et pendant tout leur cycle de vie, y compris pour conservation définitive. Auparavant, elle a exercé des missions similaires au ministère français des Affaires étrangères et au conseil général de la Seine-Saint-Denis.

électroniques sous différents formats, etc.).

Certains types de document sont exclus de la procédure car ils sont considérés comme étant «protégés». Les données personnelles peuvent dans certains cas être concernées par cette protection. Cependant, celle-ci dépend entièrement du type de procès. Par conséquent, des données personnelles sont susceptibles d'être sollicitées. Cela implique une vigilance accrue pour ne pas porter atteinte aux droits des personnes.

Dans la justice américaine, les procédures de *discovery* permettant de constituer les dossiers en amont du procès conduisent généralement à des accords entre les parties qui sont libres de décider ensuite si elles vont, ou non, au tribunal. Ainsi, bon nombre de contentieux sont réglés par des accords et ne donnent pas lieu à des procès.

La procédure de *discovery* peut se traduire par des demandes formelles d'information (*request for production of documents*), des recherches directes d'information par l'intermédiaire de

**La recherche d'éléments ouvre la voie à une recherche documentaire très élargie qui peut être qualifiée d'intrusive.**

tiers (comme un moniteur par exemple), des dépositions, des interrogatoires, etc. La plupart des États américains ont adhéré au *Uniform Interstate Depositions and Discovery Act* et appliquent donc la loi fédérale en matière de *discovery* avec toutefois quelques exceptions. Au Royaume-Uni, la procédure de *discovery* est plutôt connue sous le nom de «*disclosure*».

Le *discovery* existe depuis la fin des années 1940, mais il a été renforcé de manière significative en 2006 lorsqu'il a été manifestement élargi à toutes les informations électroniques incluant notamment les logs de connexion et les

mails du personnel. Par conséquent, les procédures de *discovery* et d'*e-discovery* conduisent à d'importants transferts de données vers les États-Unis à la demande de juridictions américaines. Pour cette raison, en France, la Commission Nationale Informatiques et Libertés (CNIL) s'est saisie de la problématique<sup>1</sup> en 2007-2009 et le G29 en 2009. La CNIL a émis la délibération n°2009-474 «portant recommandation en matière de transfert de données à caractère personnel dans le cadre de procédures judiciaires américaines dite de *Discovery* ». Elle rappelle la nécessité de s'assurer de la légitimité et de la finalité des traitements d'information opérés dans le cadre du *discovery*.

### **Discovery et respect des droits des personnes**

Le respect des conventions internationales et des dispositions nationales applicables, telles que la Convention de La Haye et la loi du 26 juillet 1968, est nécessaire afin de protéger les droits fondamentaux des personnes concernées. La CNIL a rappelé que la législation américaine diffère de celle appliquée en Europe en matière de protection des données personnelles. Les transferts de données doivent être déclarés à la CNIL.

Le G29, groupe de travail européen, a demandé le respect d'un principe de proportionnalité dans le cadre de «procédures civiles transfrontalières» afin de ne pas divulguer des informations non requises pour le procès en question. Il préconise une coordination internationale menée par les gouvernements des États concernés, éventuelle-

**Les liens entre le *discovery* et la protection des données personnelles sont épineux.**

ment par l'adoption d'un traité ou d'une convention (avis du 11 février 2009)<sup>2</sup>. Les entreprises américaines sont tenues de certifier le respect de conventions européennes afin d'obtenir le transfert de données personnelles. Les États-Unis ont mis en place un cadre spécifique appelé *Safe Harbor*. En concertation avec la Suisse, il existe un «US-Swiss *Safe Harbor Framework*». Toutefois, les accords du *Safe Harbor* ont été invalidés en 2015 par la Cour de justice de l'Union européenne, sujet qui a fait la une des journaux spécialisés. Les liens entre le *discovery* et la protection des données personnelles sont donc épineux.

## ABSTRACT

*Der Dokumentenzyklus will gemanagt sein, um juristische und finanzielle Risiken zu bewältigen: Discovery-Verfahren*

Immer komplexere Regelungen, die dazu tendieren, den lokalen, europäischen und amerikanischen Kontext zu überdecken, machen für Organisationen die Beherrschung der produzierten Informationen für die Ausübung ihrer Aktivitäten erforderlich. Ein umfassendes Life Cycle Management von Information, das heisst das Identifizieren der verschiedenen Arten von Dokumenten, die produziert werden – Festlegung der Aufbewahrungsfristen von Fall zu Fall sowie endgültige Sortierung (Vernichtung oder Aufbewahrung) – ist unumgänglich für Organisationen, die mit zahlreichen Anträgen von verschiedenen Gerichtsbarkeiten, einschliesslich amerikanischer, konfrontiert werden. Im Fall einer Rechtsstreitigkeit kann die amerikanische Justiz aufgrund sogenannter *Discovery-Verfahren* von einer Organisation die Lieferung zahlreicher physischer oder elektronischer Informationen verlangen. Bei elektronischen Daten spricht man deshalb auch von *E-Discovery*. Jede Organisation, die in den oder mit den Vereinigten Staaten tätig ist, kann jederzeit in Verfahren solcher Art involviert werden. Durch die Offenlegung von Informationen kann ihre Tätigkeit durch die Konkurrenz beeinträchtigt werden. Oder sie könnte im Fall einer Verbreitung von persönlichen Daten die Rechte von Personen kompromittieren. Eine Organisation ist daher verpflichtet, ihre Bemühungen für die Beherrschung des Informationsmanagements sowie der rechtlichen und finanziellen Risiken zu erhöhen.

1 <<https://www.cnil.fr/fr/quel-cadre-juridique-pour-les-procedures-de-discovery-o>> En ligne. Consulté le 30 juin 2016.

2 <[http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2009/wp158\\_fr.pdf](http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2009/wp158_fr.pdf)> En ligne. Consulté le 30 juin 2016.

## Nécessité de maîtriser l'information dès sa production

Le *discovery* et l'*e-discovery* sont des procédures intrusives, coûteuses en temps et en argent et difficiles à maîtriser s'il n'y a pas de réflexion en amont pour être en mesure de répondre efficacement, sans compromettre les droits des personnes, aux injonctions de la justice américaine.

En France, toute organisation soumise à ces procédures doit s'assurer par ailleurs du respect des recommandations de la CNIL. Elle doit répondre à toutes les demandes d'information dans le cadre d'un contentieux en lien avec la justice américaine, permettre la recherche directe en ouvrant l'accès à l'information à son opposant, recherche qui se traduit dans bien des cas par l'accueil d'un moniteur qui investit sur place. Les informations conservées par une partie qui sont exigées par l'autre partie peuvent être soumises à un gel à la destruction et/ou à la communication. Autrement dit,

- 1) il est demandé d'interrompre l'application du cycle de vie tel que défini dans la politique de gestion documentaire de l'organisation lorsqu'elle existe et
- 2) il est exigé que certains documents ne soient plus consultés par d'autres interlocuteurs pendant une période qui peut être plus ou moins longue en fonction des contentieux.

La fourniture d'informations et la gestion des gels nécessitent une organisation appropriée qui se traduit notamment par la mise en place des procédures et des modes opératoires clairs et validés par les responsables juridiques (la direction juridique lorsqu'elle existe). En effet, le temps nécessaire pour mettre la main sur les informations demandées peut être assez long si les informations sont éparpillées ou bien si on ignore où se trouve cette information. Dès lors, les impacts en termes de gestion de l'information sont importants.

Toute organisation susceptible d'être engagée dans une procédure de *discovery* se doit de maîtriser l'information qu'elle produit, elle doit être en mesure de prouver que la destruction de documents, si elle a eu lieu, a été

**Le *discovery* est une procédure intrusive, coûteuse et difficile à maîtriser s'il n'y a pas de réflexion en amont.**

effectuée conformément à une politique interne reconnue pouvant être audité. C'est ainsi que des politiques de gouvernance de l'information se sont développées dans la plupart des entreprises multinationales dans le monde. Elles se sont dotées de politiques internes permettant de garantir la conformité aux règles de protection des données personnelles et la mise en

œuvre de politiques d'archivage appropriées au contexte local, européen et américain.

## Trop conserver, une pratique à risque

La gestion du cycle de vie de l'information devient donc un enjeu majeur pour la maîtrise des risques financiers et juridiques: il s'agit de maîtriser l'information qu'on produit, de bien gérer en conséquence son cycle de vie, c'est-à-dire de conserver et de détruire ce qui doit être conservé ou détruit en fonction de la réglementation en vigueur.

Ce contexte favorise également la mise en place de politiques d'archivage électronique qui permettent de bien contrôler l'information sous forme électronique: l'identifier, la répartir en catégories, lui attribuer des durées de conservation et un sort final. Toute organisation doit savoir où se trouvent ses données et qui peut y avoir accès.

La surconservation des données électroniques qui ne sont plus nécessaires aux activités de l'organisation présente un risque qui est double: le risque de communiquer de manière élargie des informations stratégiques à un secteur d'activité qui peuvent porter atteinte à son action et à son image et le risque de divulguer massivement des données susceptibles de porter atteinte aux droits des personnes.

Contact: [lourdes.fuentes-hashimoto@total.com](mailto:lourdes.fuentes-hashimoto@total.com)

Twitter: @e\_archiviste

[ a [ r [ b | i ] d ] o ]

Newsletter arbido → [www.arbido.ch](http://www.arbido.ch)