

**Charles Kecskeméti
et Lajos Körmendy**

Les écrits s'envolent

**Le défi de la conservation des archives
dans le long terme**

« Les peuples ont une mémoire, à l'image des personnes. L'humanité possède elle aussi une mémoire commune. La mémoire des peuples ne fonctionne pas comme un ordinateur, mais comme un cœur. Pour une famille ou un peuple, c'est le noyau de la vie. » (Pape François, 21 février 2014.)

TABLE DES MATIÈRES

Avant-propos	6
ARCHIVES ET HISTOIRE	7
1 Regard sur l'histoire de l'archivistique	8
1.1 Prologue	8
1.2 La révolution érudite et la révolution libérale	9
1.3 Les archives, laboratoires de l'histoire	12
1.4 Le début tardif de la coopération internationale des archivistes	15
2 L'apogée de l'archivistique papier	17
2.1 Vers le renouveau de la science des archives	17
2.2 Maîtriser la masse	19
2.3 Préarchivage et records management	21
2.4 Conservation	22
2.5 Communication	25
2.6 Développement	26
3 Scripta debent manere	27
3.1 Contestations	27
3.2 Protection de la vie privée vs intégrité du patrimoine	31
3.3 L'horizon de 500 ans	34
3.4 Garantie des droits, mémoire institutionnelle, sécurité de l'environnement	37
4 Certitudes et conjectures en 2014	39
4.1 Acquis et pertes	39
4.2 Une lacune à combler	40
4.3 La restauration du droit international : un impératif enterré	42
4.4 Regards sur l'avenir	43
4.5 En guise d'épilogue de ces regards sur le passé et l'avenir	45

LES ARCHIVES DANS LE MONDE NUMÉRIQUE	47
5 Les fonctions des archives et les valeurs des documents	48
5.1 Les fonctions des archives	48
5.2 Les valeurs des documents	49
5.2.1 La valeur administrative/pratique	49
5.2.2 La valeur probante	50
5.2.3 La valeur historique	51
6 Défis d'archives dans le monde analogue et numérique	53
6.1 L'informatisation, la communication et les documents électroniques	53
6.2 La numérisation de la mémoire humaine et l'explosion des données	55
6.3 Les difficultés de la conservation pérenne des documents électroniques	56
7 Le document d'archives et ses éléments constitutifs	58
7.1 Le support	58
7.2 La forme	59
7.3 Le contenu	60
7.4 L'identifiant	60
7.5 La structure	61
7.5.1 La structure interne	62
7.5.2 La structure externe	62
7.6 Le contexte	64
7.7 La séparation des éléments constitutifs des documents électroniques	65
7.8 Le caractère singulier des documents d'archives	65
8 Les métadonnées	66
9 L'intégrité et les caractéristiques essentielles des documents	70
10 L'authenticité des documents	74
11 Les formats de sauvegarde des fichiers	77

12	Les normes	82
12.1	Norme de langage : XML	84
12.2	Norme pour les archives courantes : MOREQ	85
12.3	Norme pour la conservation des données : OAIS	86
12.4	Norme pour la description des archives 1 : ISADg et ISAAR(CPF)	86
12.5	Norme pour la description des archives 2 : EAD et EAC	87
13	Stratégies pour la conservation pérenne des documents électroniques	89
13.1	Migration	92
13.2	Émulation	93
13.3	Conservation en forme analogique et hybride	95
13.4	Documents spéciaux	97
13.4.1	E-mails	97
13.4.2	Bases de données	99
13.4.3	Pages web	101
14	La préservation des supports physiques des documents électroniques	103
15	La transformation de l'archivistique à l'ère des documents électroniques	106
16	L'enjeu de la conservation pérenne des documents électroniques	110
	BIBLIOGRAPHIE SÉLECTIVE	113

Avant-propos

Ce livre est né de la volonté de lancer un cri d'alarme sur le danger de voir disparaître les archives de l'ère de l'informatique. Le grand public, y compris les honnêtes hommes du XXI^e siècle, ignore ce danger. Nous avons écrit cet essai pour sensibiliser les décideurs et tous ceux qui s'intéressent au passé et à l'histoire en train de se faire. Il leur appartient d'agir pour conjurer le danger. Nous l'avons écrit aussi à l'intention des conservateurs d'archives qui ne se sentent pas chez eux dans le monde numérique. C'est pour les réconcilier avec la réalité qui les entoure que nous insistons, tout au long du texte, sur la continuité qui transcende les tournants dont est jalonnée l'histoire de l'archivistique.

La menace qui pèse sur la préservation des sources dont devra se nourrir la recherche historique de demain ne préoccupe guère l'opinion publique. La principale raison en est, sans doute, que les archives ne font pas partie de l'environnement culturel du grand nombre. Si on fréquente les bibliothèques et on visite les musées dès les années de l'école primaire, puis tout au long de la vie, les archives n'accueillent que quelques centaines ou milliers de lecteurs poursuivant des recherches et un public d'amateurs de l'histoire lors de leurs expositions.

Il nous a fallu donc présenter la pensée, la pratique, les soucis et les priorités changeantes du monde professionnel des archives pour que le danger soit perçu imminent et palpable par les lecteurs de cet essai. Nous nous sommes efforcés de mettre en relief la logique propre de l'histoire des archives pour faire ressortir l'immensité du défi que doit relever l'humanité de notre temps pour sauvegarder la mémoire des générations à venir. L'ère numérique remet en cause la pertinence du vieil adage *verba volant, scripta manent*. Les écrits dématérialisés s'envolent, eux aussi. Stockés près de leur lieu de création ou dans un *nuage* lointain, leur préservation ne sera assurée que pendant une durée limitée, au bout de laquelle ils s'envoleront définitivement.

Le livre est composé de deux parties. Charles Kecskeméti est l'auteur des chapitres sur les archives et l'archivistique d'avant l'ère de l'informatique. La deuxième partie sur les archives dans le monde numérique est l'œuvre de Lajos Körmendy.

Nous tenons à exprimer nos remerciements à nos collègues Jens Boel, Marie-Anne Chabin, Jacques Schouller et André Vanrie pour leur aide amicale à la préparation de ce livre.

ARCHIVES ET HISTOIRE

1 Regard sur l'histoire de l'archivistique

1.1 Prologue

La révolution informatique, enclenchée il y a un demi-siècle environ, a effectivement changé les conditions de vie et de travail à l'échelle mondiale au cours des quinze dernières années. En ce court laps de temps, la certitude s'est imposée à l'opinion que nous vivions une période de mutation, que nous venions d'entrer dans l'ère de l'information.

Le terme information a pris, implicitement, la signification d'information actuelle, immédiate. Il n'est donc pas surprenant que les Archives, institutions créées pour préserver la continuité de la société dans le long terme, aient du mal à s'intégrer dans ce nouvel univers, d'autant plus de mal qu'elles doivent ajouter à la conservation de documents pérennes leur fonction originelle : la pérennisation d'images volatiles.

La révolution informatique s'est produite au moment où la profession s'enorgueillissait d'avoir bâti un édifice doctrinal et de s'être dotée d'un arsenal méthodologique et technique qui lui permettait d'exercer convenablement sa quadruple mission d'engranger, de préserver, de mettre en valeur et de communiquer les archives dont elle avait la charge. Cette archivistique sereine et ambitieuse du milieu du XX^e siècle, qui concluait l'histoire millénaire des archives européennes, était le fruit de quatre siècles de réflexion et d'expérience pratique. La longue maturation de la pensée et du savoir-faire archivistiques était ponctuée de découvertes intellectuelles, de révolutions provoquées par des mutations culturelles d'ampleur variable et de virages brusques, imposés tantôt par des réformes du système institutionnel et des changements intervenus dans la production des documents, tantôt résultant de l'étude de l'archive par ceux qui en assuraient la garde, tantôt encore pour répondre aux demandes et aux espoirs des utilisateurs. Or, à la différence des tournants du passé, celui de la dématérialisation a bouleversé d'emblée l'ensemble du champ archivistique. Il y a lieu donc de s'interroger sur ses conséquences dans la longue durée.

Une remarque préliminaire s'impose. La partie introductive de cet essai ne porte pas sur l'histoire des archives de l'Antiquité, dont les débuts remontent aux royaumes du Moyen-Orient d'il y a cinq mille ans. Appartiennent également à cette histoire générale des archives celles des empires et royaumes d'Asie et d'Afrique, comme la Chine et l'Empire ottoman. Mais ces pays, possesseurs de riches patrimoines écrits qu'ils ont su préserver à travers des siècles, ne participent à la réflexion professionnelle et à l'activité de la communauté

archivistique que depuis quelques dizaines d'années. Jusqu'à un passé récent, l'histoire de l'archivistique se déroulait bel et bien dans le monde européen.

1.2 La révolution érudite et la révolution libérale

Des manuels sur l'art de lire, conserver, classer et répertorier chartes, comptes et autres documents furent publiés à partir de la fin du XVI^e siècle en Allemagne, en Italie, en France et en Espagne. Puis la parution du traité *De Re Diplomatica* du bénédictin Jean Mabillon en 1681 posa les fondations intellectuelles de la science des archives en inventant une grille de règles et de méthodes de la critique des sources, servant à vérifier l'authenticité des actes. Car « les faux abondent à toute époque et dans tous les pays »¹ et cette affirmation, datant de 1929, reçut une actualité dramatique avec l'informatisation généralisée de la production des documents. À trois cent vingt ans de distance, Luciana Duranti de l'Université de la Colombie-Britannique, renouant avec la diplomatie des bénédictins, a lancé les recherches sur les critères permettant la détection des faux et des manipulations des documents électroniques². Le groupe InterPARES, qu'elle a fondé en 1999, s'attache à élaborer les techniques et les normes visant à pérenniser l'authenticité et la consultabilité de ces documents.

Au cours du siècle et demi allant du temps de la Grande Encyclopédie au suicide de l'Europe en 1914, les archives du continent passèrent par deux révolutions successives. La révolution érudite des Lumières se déclencha simultanément sur deux terrains.

D'une part, l'Europe émerveillée découvrit sa richesse fabuleuse de sources portant témoignage de son passé. De Lisbonne à Moscou, des conservateurs et des érudits entreprirent d'ambitieux programmes de publication de chroniques, de diaires, de chartes et de lettres. Le nom de l'un de ces éditeurs de sources, celui de Ludovico Antonio Muratori (1671-1750), archiviste-bibliothécaire ducal à Modène, est devenu légendaire.

D'autre part, l'extension et la diversification des compétences de l'État conduisirent, de Florence à Varsovie, à un vaste effort d'organisation et de concentration d'archives, conçues pour servir d'instrument de gouvernement. Sous les règnes de Marie-Thérèse et de Joseph II,

¹ A. de BOÛARD, *Manuel de diplomatie française et pontificale. I. Diplomatie générale*. Auguste Picard, Paris, 1929, p. 14.

² Voir Marie-Anne CHABIN, *Peut-on parler de diplomatie numérique*.
<http://www.marieannechabin.fr/diplomatique-numerique/>

la Monarchie des Habsbourg fut la puissance pionnière de la nouvelle politique. Après la création à Vienne, en 1749, du *Geheimes Hausarchiv* (Archives Secrètes de la Maison), appelé, depuis le début du XIX^e siècle, *Haus-, Hof- und Staatsarchiv* (Archives de la Maison, de la Cour et de l'État), les royaumes et grandes provinces de la Monarchie (Lombardie, Pays-Bas, Bohême, Hongrie, Galicie) furent dotés de services d'archives regroupant les fonds de leurs autorités particulières.

En France, l'Ancien Régime finissant n'établit pas d'Archives centrales comme celles de Vienne, mais un Bureau littéraire, un Comité des travaux historiques et un Bureau des Archives furent créés. En 1770, une enquête recensa cinq mille sept cents dépôts d'archives de toute nature. « Dans ce nombre se trouvaient, outre les archives des corps souverains et des corps administratifs, des archevêchés et évêchés : mille sept cent soixante abbayes, prieurés et couvents ; trois cent quatre-vingt-quatorze chapitres et collégiales ; mille sept cents maisons seigneuriales et huit cent cinquante-cinq hôtels de ville.³ » Il en allait de même partout en Europe, car ces archives « étaient essentiellement formées des titres sur lesquels pouvaient se fonder les droits et les obligations de leurs possesseurs. »⁴

La révolution libérale, commencée en France en 1789, a transféré à la Nation la propriété des archives provenant de l'ensemble des services publics. Il semblait évident que la vaste gamme de règles de classement alphabétique, chronologique, géographique ou méthodique, qui avait permis aux dicastères, cours de justice, évêchés et villes de contrôler leurs propres archives, pouvait servir pour organiser les nouveaux ensembles composites. Ce fut l'échec. Au bout de quelques dizaines d'années d'expérimentation, il s'est avéré que l'application des méthodes héritées des archivaires et feudistes de jadis aboutissait à produire des agrégats certes ordonnés, mais souffrant d'un handicap : en sortant les pièces de leur contexte d'origine, elle en réduisit la valeur informative. Les faits consignés restaient accessibles, mais les séries ainsi organisées ne suivaient plus l'enchaînement des faits, le déroulement de l'action qui avait donné lieu à l'établissement des documents.

La solution vint dans les années 1840 avec la découverte que l'entité d'archive fondamentale n'était pas le document individuel, la pièce, mais le fonds, c'est-à-dire l'ensemble des documents réunis par une institution ou une personne dans l'exercice de ses fonctions ou activités. Le *principe du respect des fonds* fut formulé en 1841 par Natalis de

³ A. CHAMPOLLION-FIGEAC, *Notice sur les Archives départementales en France*, Grenoble 1887, pp. 8-9.

⁴ Direction des Archives de France, *Manuel d'archivistique. Théorie et pratique des Archives publiques en France*. S.E.V.P.E.N., Paris, 1970, p. 21.

Wailly, chef de section aux Archives royales (nom des Archives nationales sous la Monarchie de Juillet), dans l'instruction du ministère de l'Intérieur fixant le cadre de classement des archives départementales françaises. L'instruction prescrivit de laisser groupées les archives émanant d'un même producteur et de ne pas les mélanger avec des archives d'autres provenances.

Le principe, qui est à la base de l'authenticité globale des archives, allait être développé par des archivistes allemands et par les auteurs du *Manuel néerlandais*⁵ en principe du respect de l'ordre primitif selon lequel, pour conserver intact le pouvoir informatif des archives, le classement établi par l'organisme d'origine devait être maintenu, ou reconstitué s'il se trouvait dégradé. Depuis le dernier tiers du XIX^e siècle, la théorie, la pratique et la méthodologie des archives publiques des pays européens, ouvertes à la recherche, reposent sur ce principe. Dans quelle mesure la structure des archives couvre-t-elle la structure interne de l'organisme producteur ? Cela dépend, avant tout, des pratiques et traditions administratives et archivistiques du pays. Dans les pays qui suivent le système de *Registratur* d'origine allemande, les deux structures se recourent, ailleurs la correspondance est moins nette⁶.

Le Congrès international des Archivistes et des Bibliothécaires de Bruxelles, tenu en 1910, apporta l'adhésion formelle de la profession au principe de provenance qui est « le meilleur système à adopter pour classer et inventorier un fonds d'archives, non seulement au point de vue du classement logique des pièces, mais aussi dans l'intérêt bien compris des études historiques. »⁷ Les restructurations au sein de l'appareil d'État, de plus en plus fréquentes dans nombre de pays, mettent souvent à mal l'application du principe du respect des fonds. Des compétences de l'État peuvent aller d'office en office, telle ou telle fonction, comme les affaires d'immigration ou la protection de l'environnement, peut changer de ministère 5 ou 6 fois en un quart de siècle. Des changements administratifs à des intervalles

⁵ S. MULLER, J. A. FEITH, R. FRUIN, *Handleiding voor het ordenen en beschrijven van archieven*. Erven B. van der Kamp, Groningen, 1898, iv-156 p.

⁶ Voir Michel DUCHEIN, "Le respect des fonds en archivistique: principes théoriques et problèmes pratiques" in *La Gazette des archives*, 1977, pp. 90-91 et Michel DUCHEIN, "Le principe de provenance et la pratique du tri, du classement et de la description en archivistique contemporaine" in *Janus* 1998.1, pp. 96-97, ainsi que Peter HORSMAN, „The Last Dance of the Phoenix or the De-discovery of the Archival Fonds" in *Archivaria* No. 54, Fall 2002, p. 12.

⁷ *Congrès international des Archivistes et des Bibliothécaires 1910. Actes* publiés par J. CUVELIER et L. STAINIER. Commission permanente des Congrès internationaux des Archivistes et des Bibliothécaires, Bruxelles, 1912, LXIII-812 p. Les Actes de ce Congrès ont été réédités sur CD-Rom par les Archives générales du Royaume à Bruxelles en 2010.

aussi courts compliquent le travail des archives et rendent difficile celui des chercheurs. Pour pallier ces inconvénients, les archivistes australiens ont élaboré un nouveau concept de série dit *Record Series*⁸ qui englobe les dossiers et pièces créés dans l'exercice de la même fonction par les services producteurs successifs.

L'applicabilité du principe de provenance aux archives électroniques est au centre de la réflexion archivistique depuis les années 1990⁹. Les partisans de la table rase que réclamerait l'entrée dans l'âge informatisé en recommandent l'abandon.

1.3 Les archives, laboratoires de l'histoire

Des historiens nés aux alentours du tournant du XIX^e siècle, tels Leopold von Ranke (1795-1886), Jules Michelet (1798-1874) ou Thomas Babington Macaulay (1800-1859), se fixèrent l'objectif de présenter comment les choses se sont vraiment produites, de reconstituer l'histoire comme elle s'est effectivement déroulée. Leur ambition d'obtenir pour l'art de l'historien le rang de science était fondée sur l'existence de services publics, les Archives, chargés de mettre à la disposition des chercheurs les produits finis de l'action des administrations, transformés en matière première de la recherche. Les documents qui arrivent aux Archives, vingt, trente ou cinquante ans après leur création, n'avaient pas été créés pour alimenter la culture historique de la postérité, mais comme preuves de l'expédition d'affaires concrètes. Leur valeur informative est infinie, elle dépend de ce que le chercheur y découvre en les confrontant et les combinant avec d'autres sources.

Le regard de l'Europe du XIX^e siècle restait vissé sur la Révolution française. D'innombrables récits, pamphlets et livres lui furent consacrés d'un bout à l'autre du continent, à commencer par les *Réflexions* d'Edmund Burke, parues en 1790. Entre 1823 et 1853, année du séjour d'Alexis de Tocqueville à Tours, quatre historiens, parmi les plus influents de leur temps – Adolphe Thiers (dix volumes), François Mignet (deux volumes), Thomas Carlyle (trois volumes) et Jules Michelet (sept volumes) – écrivirent leur synthèse de l'histoire de la Révolution. Dans la préface de l'édition de 1868 de son ouvrage, ce dernier se souvient : « Un mot sur la manière dont ce livre se fit. Il est né du sein des Archives. Je l'écrivis six ans (1845-1850) dans ce dépôt central, où j'étais chef de la section historique.

⁸ Voir Peter SCOTT, „The Record Group Concept: a Case of Abandonment” in *American Archivist* 29, 1966, pp. 493-505.

⁹ Voir Terry COOK, „What is Past is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift” in *Archivaria* 43, pp. 17-63.

Après le 2 Décembre, j'y mis deux ans encore et l'achevai aux Archives de Nantes, tout près de la Vendée dont j'exploitais aussi les précieuses collections » et d'énumérer les groupes et catégories des sources dépouillées. S'il cite rarement ses sources, c'est qu'il ne travaillait pas avec des pièces détachées, mais avec des grandes collections « où tout se suit dans l'ordre chronologique ».

Dès les années 1830, Tocqueville évoquait les « liens invisibles » qui « attachent les idées d'un siècle à celles du siècle qui l'a précédé » et s'interrogeait sur ce que les Français de son temps ont conservé de leur État ancien. Il s'attaquera à ce double problème quelque quinze ans plus tard, après avoir été écarté de la vie publique par le coup d'État du 2 décembre 1851. Il avait la conviction que la société nouvelle qui sortait de la Révolution, malgré tous les efforts déployés après 1789 pour faire table rase du passé, qu'ont décrite avec « trop d'éclat » les historiens de la Révolution, se rattachait par de multiples liens à celle de l'Ancien Régime, qu'elle continuait tout en s'étant entourée d'institutions remodelées. Pour vérifier son hypothèse, il choisit d'interroger l'Ancien Régime à travers ses archives.

Dans l'avant-propos de *L'Ancien Régime et la Révolution*, fruit de ses recherches publié en 1856, Tocqueville expose avec clarté et précision ce que signifient les archives :

« Dans les pays où l'administration publique est déjà puissante, il naît peu d'idées, de désirs, de douleurs, il se rencontre peu d'intérêts et de passions qui ne viennent tôt ou tard se montrer à nu devant elle. En visitant ses archives on n'acquiert pas seulement une notion très exacte de ses procédés, le pays tout entier s'y révèle. Un étranger auquel on livrerait aujourd'hui toutes les correspondances confidentielles qui remplissent les cartons du ministère de l'Intérieur et des préfectures en saurait bientôt plus sur nous que nous-mêmes. Au XVIII^e siècle, l'administration publique était déjà, ainsi qu'on le verra en lisant ce livre, très centralisée, très puissante, prodigieusement active. On la voyait sans cesse aider, empêcher, permettre. Elle avait beaucoup à promettre, beaucoup à donner. Elle influait déjà de mille manières, non seulement sur la conduite générale des affaires, mais sur le sort des familles et sur la vie privée de chaque homme. De plus, elle était sans publicité, ce qui faisait qu'on ne craignait pas de venir exposer à ses yeux jusqu'aux infirmités les plus secrètes. J'ai passé un temps très long à étudier ce qui nous reste d'elle, soit à Paris, soit dans plusieurs provinces. »

Tocqueville ajoute en note : « Je me suis particulièrement servi des archives de quelques grandes intendances, surtout de celles de Tours, qui sont très complètes [...] D'autres

généralités, entre autres celles de l'Île-de-France, m'ont fait voir que les choses se passaient de la même manière dans la plus grande partie du royaume. »¹⁰

L'adjectif *secret/geheim* accolé, sous l'Ancien Régime, au nom des archives centrales de différents pays, signifiait que celles-ci étaient à la disposition du souverain qui seul était habilité à donner l'autorisation d'y accéder. Les réseaux d'archives publiques, mis en place progressivement dans l'Europe libérale du XIX^e siècle, s'ouvrirent à la recherche, mais les nations souveraines, héritières des monarques, se devaient de les entourer de règles d'accès prudentes, d'autant plus prudentes que l'esprit du nationalisme imposait partout une écriture patriotique de l'histoire. Un épisode archivistique du conflit austro-prussien des années 1850 concernant l'avenir de l'Allemagne (devait-elle devenir « grande » avec l'Autriche ou « petite », l'Autriche en étant exclue) fournit un exemple éloquent.

Heinrich von Sybel, qui allait être nommé en 1875 directeur du *Geheimes Staatsarchiv* de Prusse, publia en 1853 sa monographie monumentale sur l'Histoire de la période révolutionnaire, 1789-1795¹¹. À partir de sources d'archives prussiennes, hollandaises, anglaises et françaises, il prouva que la responsabilité pour la conclusion par la Prusse de la paix séparée avec la France, à Bâle en 1795, revenait au cabinet intrigant de Vienne. Alfred von Vivenot répliqua à cette accusation en démontrant, dans la biographie du Prince Albert de Saxe-Teschen publiée en 1859, la responsabilité de la Prusse. Vivenot, qui fit ses recherches dans les Archives de Vienne où Sybel ne fut pas admis, ne put consulter les sources conservées à Berlin et à Paris.

Depuis lors, dans toute controverse entre historiens, avec ou sans enjeu patriotique ou politique, les parties se battent à coups d'archives, chacune exhibant les pièces opposables à la position adverse. Dans la quête de la vérité historique, objective parce qu'étayée par des documents authentiques, chaque auteur suit la voie que lui imposent ses convictions intimes. Neutres, les archives ? Certes, et cette neutralité confère à tout utilisateur la liberté de les interpréter.

¹⁰ Alexis de TOCQUEVILLE, *L'Ancien Régime et la Révolution* dans *Œuvres* III, pp. 44-45, Gallimard, 2004 (Bibliothèque de la Pléiade).

¹¹ Traduit en français en 1868, sous le titre *Histoire de l'Europe pendant la Révolution française*, 6 volumes, Paris, 1869-1888.

1.4 Le début tardif de la coopération internationale des archivistes

Dans l'Europe de l'ère moderne, le progrès des sciences et des techniques passe par la circulation des idées et des hypothèses. Au moyen traditionnel des correspondances et rencontres privées entre les membres de la République des lettres s'ajoutent, depuis deux siècles, des structures institutionnelles de plus en plus étoffées.

Le Congrès international d'astronomie, réuni à Gotha en 1798, premier événement de ce type, ne fut suivi d'initiatives semblables que des dizaines d'années plus tard. Mais il ne fut pas oublié. En 1817, l'abbé Grégoire (1750-1831), député à partir de 1789, militant infatigable de l'abolition de la traite et de l'esclavage, rédigea à la demande de Sir John Sinclair (1754-1835), homme politique écossais, un *Plan d'association générale entre les savants, gens de lettres et artistes pour accélérer les progrès des bonnes mœurs et des lumières*. L'association tiendrait des congrès périodiques où chaque science serait examinée et où « chacun entre en conférence avec ceux qui cultivent la partie à laquelle il est voué ». L'ère de la Sainte Alliance et de la Restauration ne se prêtait guère à la vie associative internationale de l'élite lettrée. Cependant, l'intérêt de celle-ci pour le progrès scientifique ne faiblissait pas, ce qui encourage le baron de Férussac (1786-1836) à lancer, en 1823, son *Bulletin général et universel des annonces et des nouvelles scientifiques* dont les huit sections couvraient l'ensemble des sciences exactes et humaines. Cent soixante-dix volumes parurent pendant les huit années de l'existence du Bulletin.

La glaciation prit fin en 1830. La Belgique, née cette année-là, prit sur elle d'offrir un foyer à la collaboration intellectuelle internationale qu'elle allait animer et piloter des années 1840 jusqu'aux années 1870. Elle accueillit une bonne partie des trente-neuf congrès internationaux tenus entre 1840 et 1859. Comme il n'était pas seyant de se réclamer de l'abbé Grégoire, réprouvé par l'Église pour avoir été évêque constitutionnel et accusé d'avoir voté la mort de Louis XVI, nul ne fit allusion à son projet lorsque l'état-major libéral belge le réalisa en organisant à Bruxelles, en 1862, le premier congrès de *l'Association internationale pour la promotion des sciences sociales*. Trois autres congrès suivirent à Gand, Amsterdam et Berne. L'idée de confier à la Belgique la convocation du premier congrès vint de Lord Brougham, homme politique libéral britannique, qui fonda, en 1857, la *National Association for the Promotion of Social Sciences*.

Les congrès travaillaient en cinq sections : législation comparée, instruction et éducation, art et littérature, bienfaisance et hygiène et économie politique. L'Association ne survécut pas aux dissensions provoquées dans les pays européens, et plus particulièrement au sein de la

classe politique belge, par l'encyclique *Quanta Cura* et le *Syllabus* de Pie IX, qui dénonçaient avec vigueur les erreurs rationalistes et libérales ainsi que les thèses contestant le pouvoir temporel et le monopole spirituel de l'Église. En rejetant en bloc les acquis du progressisme européen du XIX^e siècle – la liberté d'association, le mariage civil, l'école laïque etc. – et en désignant le libéralisme comme ennemi principal, ces textes pontificaux interdirent aux catholiques, intellectuels et hommes politiques, de participer à des événements pilotés par le parti libéral belge.

Mais l'Association avait accompli sa mission. La formule de l'association générale, regroupant le monde des sciences, des lettres et des arts, fut abandonnée, mais grâce à l'élan donné par la Belgique, la coopération intellectuelle connut une période de progression spectaculaire pendant cinquante ans. Les unes après les autres, les professions organisèrent leurs associations et leurs congrès : les anthropologues en 1865, les architectes en 1867, les orientalistes en 1873... Le mouvement allait en s'accéléralant. En 1912, il existait cinq cents organisations internationales de droit privé. Le nombre des congrès internationaux doublait de décennie en décennie, pour atteindre 1070 entre 1900 et 1910.

Les archivistes ont rejoint le mouvement assez tardivement¹². La première initiative internationale fut prise conjointement par les trois métiers du patrimoine culturel : la *Revue internationale des Archives, des Bibliothèques et des Musées*, éditée à Paris en 1895-1896, s'arrêta après la publication du neuvième numéro. Au cours de la première décennie du XX^e siècle, les communautés professionnelles des archivistes et des bibliothécaires sont arrivées à la conclusion que le cloisonnement devait céder la place à la confrontation transfrontière des idées et des expériences. Le *Congrès international des Archivistes et des Bibliothécaires*, qui se réunit à Bruxelles en août 1910, répondait à un besoin évident ; cinq cents participants venant de vingt-six pays y assistèrent.

Le projet de programme de la Section des Archives du Congrès proposait la mise à l'étude de vingt-cinq thèmes. Cette liste, composée par la génération de 1910, donnait un avant-goût de l'archivistique du XX^e siècle. Après 1950, nombre de ces thèmes seront repris par la revue *Archivum* (archives notariales, registres paroissiaux), traités par des comités ou sections du Conseil international des Archives (formation professionnelle, archives économiques, préservation et restauration) ou feront l'objet de manuels ou de directives (construction de bâtiments d'archives, tri et élimination, instruments de recherche) et

¹² Voir l'excellent résumé sur les débuts de la coopération archivistique internationale dans Elio LODOLINI, *Archivistica. Principi e problemi*, Franco Angeli, Milano, 1987, pp. 292-298.

figureront au programme de congrès et de conférences. Deux vœux adoptés en 1910 méritent d'être signalés, car ils resteront d'actualité tout au long du siècle. L'un demandait que « toutes les difficultés diplomatiques soient supprimées pour l'accès des Archives de l'État aux étrangers », l'autre que les administrations qui versent leurs documents aux archives publiques utilisent du papier normal et de l'encre indélébile. Le Congrès se sépara en confiant à une Commission permanente siégeant à Bruxelles la convocation des congrès futurs et choisit Milan comme lieu du congrès de 1915.

La guerre stoppa net l'action qui venait d'être entreprise. Reprendre le fil après la destruction du « monde d'hier » exigeait du temps. Le *Comité international des Sciences historiques* établi en 1928, à Washington, une commission d'archives qui reçut pour tâche d'œuvrer pour améliorer les conditions d'accès aux sources. Elle ne put obtenir de résultat concret et disparut au bout de quelques années. L'Institut international de coopération intellectuelle de la Société des Nations chargea en 1931 un comité d'experts archivistes d'élaborer et de mettre en œuvre un programme d'activités. Le *Comité consultatif permanent d'experts*, présidé par Eugenio Casanova (1867-1951), qui avait joué un rôle éminent au Congrès de Bruxelles, adopta un plan d'échange d'informations et publia, en 1934, le *Guide international des Archives* présentant les archives publiques et privées des quarante-deux États européens. Mais il ne put poursuivre son travail alors que la moitié du continent s'enfonçait dans la démence. Finalement, c'est sous l'égide de l'UNESCO, à la fin des années 1940, que la profession put renouer avec l'esprit du Congrès de 1910.

2 L'apogée de l'archivistique papier

2.1 Vers le renouveau de la science des archives

L'idée de créer des Archives d'État, pour accueillir les documents dont les services d'origine n'avaient plus besoin, fit son chemin entre les deux guerres. La Russie, devenue URSS, entama dès 1918 l'organisation d'un système d'archives publiques. Le modèle soviétique du contrôle par l'État-Parti de la totalité des documents d'archives existant sur le territoire de l'Union, appelé « Fonds d'Archives de l'État », fut imposé/suivi, avec plus ou moins de rigueur, dans les pays appartenant au bloc de l'Est. Son démontage se poursuit depuis un quart de siècle. Quant à l'autre superpuissance du XX^e siècle, les États-Unis en sont venus à la fondation de leurs Archives nationales cent soixante ans après la Déclaration d'Indépendance.

À partir de la fin de la Seconde Guerre mondiale, les États-Unis exercent une influence grandissante sur l'évolution de la théorie et de la pratique archivistiques. C'est grâce à leur initiative, prise en 1946, de doter enfin la profession d'une organisation internationale, que l'archivistique a pu s'affirmer comme une discipline de plein droit, liée à la science historique et à ses sciences auxiliaires, à la bibliothéconomie et aux sciences politiques et administratives.

Comme indiqué plus haut, les fondements intellectuels du traitement des archives datent d'une part du XVII^e et d'autre part du XIX^e siècle. Or, bien que les deux concepts de la critique diplomatique et paléographique des sources et du respect des fonds aient présidé à peu près partout en Europe à la réflexion et à la pratique professionnelles, et que des modèles étrangers aient pu inspirer des emprunts, chaque pays élaborait ses propres lois et règlements s'appliquant à l'organisation des services, à la production des documents, à leur versement aux Archives et à leur communication au public. De même, la formation professionnelle et le corpus des méthodes de tri, de classement et de description et des mesures techniques de préservation, en d'autres termes l'armature intellectuelle de l'exercice du métier d'archiviste, variaient de pays à pays.

En 1950, lorsque se réunit à Paris le 1^{er} Congrès international des Archives, il était entendu que tout document à conserver existait physiquement et que le papier constituait la matière par excellence à entreposer sur les rayonnages des magasins. L'acte constitutif du Conseil international des Archives (ICA), adopté par le Congrès, fixait à l'organisation naissante les objectifs généraux suivants :

- Établir des relations entre archivistes et organismes professionnels de tout pays et organiser la coopération internationale dans le domaine des archives ;
- Promouvoir la protection du patrimoine archivistique de l'humanité et contribuer au progrès de la théorie et de la pratique archivistiques ;
- Faciliter l'utilisation des archives et l'étude impartiale des documents au moyen d'instruments de recherche, de production de copies et en s'efforçant de rendre l'accès aux archives plus aisé ;
- Coopérer avec d'autres organismes s'occupant de la conservation et de l'utilisation du patrimoine historique de l'humanité.

Ce document de 1950 ouvrit une nouvelle ère dans l'histoire des archives. Dans le monde métamorphosé qui sortait de la guerre, la pensée archivistique se métamorphosait elle aussi.

Au cours des décennies de l'après-guerre, la hiérarchie des priorités ne cessait d'évoluer, le domaine archivistique de s'élargir et les ressources humaines et financières de croître à l'échelle mondiale, jusqu'au début de notre siècle.

En 1910, comme en 1950, le domaine était tenu par les Archives publiques nationales et régionales. Tout au plus y étaient admises, outre les archives municipales, quelques bibliothèques américaines, des archives historiques des églises, et reconnaissait-on l'existence d'archives économiques. Archiviste était synonyme de conservateur érudit d'archives publiques, maîtrisant la paléographie, la diplomatique, la géographie historique ainsi que l'histoire des institutions et des méthodes d'archivage ayant été appliquées dans son pays. Correspondaient à cette norme les archivistes-paléographes formés à l'*École des Chartes* de Paris (fondée en 1821), les diplômés de l'*Institut für Österreichische Geschichtsforschung* de Vienne (fondé en 1854) ou encore des Écoles d'archives italiennes (de Naples, fondée en 1811, de Florence, fondée en 1817, etc.) et les stagiaires de l'administration prussienne des archives.

Au fil des ans, la domination des Archives de l'État, d'une part, et le monopole des archivistes érudits, d'autre part, s'érodaient sous l'effet conjugué de plusieurs facteurs. Le traitement de la masse croissante des documents contemporains, de même que celui des documents sonores et visuels, n'exigeaient pas de qualification de médiéviste. Les municipalités, banques et autres entreprises, universités, partis politiques, fondations, etc., soucieux de préserver leur mémoire, entendaient recruter des professionnels bien formés, mais pas nécessairement de type chartiste. Enfin, l'univers professionnel naguère encore limité à l'Europe allait en s'élargissant aux autres continents. Il fallait y bâtir des systèmes d'archives dotés de moyens de formation professionnelle appropriés.

2.2 Maîtriser la masse

À partir des années 1914-1918, la production documentaire des administrations n'a cessé de croître. Les guerres et les après-guerres, les périodes de récession et celles d'essor économique, les progrès de l'éducation et de la protection sociale ont été accompagnés de l'élargissement du champ d'intervention de l'État et de la croissance de l'appareil administratif. Les services et commissions, tous producteurs de documents publics, se sont multipliés partout, chez les belligérants comme chez les neutres, dans les pays ayant subi des dictatures idéologiques, où la bureaucratisation atteignit des sommets, comme dans les pays

qui échappèrent au fléau du totalitarisme. L'existence même des Archives dépendait de leur capacité à faire face à ce défi, de leur inventivité, car les crédits alloués à la fonction archivistique de l'État ne pouvaient tenir le rythme de l'augmentation de la masse documentaire.

La réduction du volume des archives à prendre en charge devint une priorité absolue vers 1940. Même si leurs recommandations pratiques différaient, les théoriciens les plus influents de l'époque, Sir Hilary Jenkinson (1882-1961) et Theodore Schellenberg (1903-1970) estimèrent indispensable l'élimination d'un pourcentage significatif des documents lors de la préparation du versement. Le thème du tri entre documents à conserver et ceux à détruire a servi d'axe principal à la réflexion sur le concept et la gestion des archives jusqu'à l'émergence de la priorité informatique. Un corpus d'analyses et de textes normatifs publié sous l'égide de l'UNESCO¹³ et de l'ICA¹⁴, auquel s'ajoutent d'innombrables études, articles, recommandations et directives, parus dans tous les pays producteurs de littérature archivistique, traitent du thème et du réaménagement du domaine des archives induit par la priorité de la maîtrise de la masse documentaire.

Quelles que soient les modalités pratiques retenues, la maîtrise de la masse passe par la coopération organique entre les Archives et le service producteur. La décision en matière d'élimination se prend en deux étapes. Il appartient au service producteur de désigner les groupes/catégories destinés à rejoindre les Archives, le moment venu, pour y être conservés en raison de leur valeur juridique ou pratique, et d'établir la liste des éliminables avec mention du délai de conservation. C'est ensuite à l'instance compétente de l'administration des Archives de donner l'autorisation de destruction en retirant de la liste les éléments dont la conservation définitive s'impose en raison de leur valeur pour la recherche.

La responsabilité pour le contrôle et la gestion de l'accroissement continu du patrimoine revient donc aux archivistes. Il s'agit sans doute de la plus délicate et de la plus complexe de leurs fonctions. On rencontre parfois la suggestion qu'il faudrait tout conserver, car même une contravention routière ou une facture de supermarché recèle des informations exploitables par la recherche. Le vœu, au demeurant irréalisable en raison de son coût, relève des *fallacies* brocardées il y a près de deux cents ans par Jeremy Bentham. Submergés par le flux de

¹³ Série des études RAMP comprenant plus de cent titres.

¹⁴ Études et manuels du CIA.

quantités immenses de documents, les chercheurs ne pourraient identifier les sources susceptibles de contenir des informations relatives au thème qu'ils étudient.

Les choix qu'opèrent les archivistes ne leur sont pas dictés par leur capacité à prévoir les curiosités des historiens de l'avenir. Nul ne possède pareille prescience. La décision de préserver des dossiers jugés inutiles par le service producteur est fondée sur la connaissance des fonds du dépôt d'archives auxquels ils viendront s'ajouter, fournissant ainsi des informations complémentaires aux chercheurs. Elle prend aussi en considération les fonds et collections existant dans les autres dépôts d'archives du pays.

La conservation partielle par échantillonnage permet de concilier l'impératif de réduire le volume des versements à prendre en charge avec l'intérêt de la recherche historique. Le critère de l'échantillonnage est déterminé par le type des dossiers concernés. Il peut être chronologique (année de la création du dossier), topographique (lieu de création) ou alphabétique (nom de personnes commençant par telle ou telle lettre), mais toujours mécanique, ne faisant pas la distinction entre les affaires « banales » à éliminer et les affaires « intéressantes » à conserver. Une politique nationale d'échantillonnage coordonné, répartissant les tâches entre services d'archives territoriales, peut apporter une aide efficace aux historiens.

2.3 Préarchivage et records management

Le gonflement rapide de la « paperasse » menaçant d'envahir bureaux, couloirs, caves et greniers, les administrations devaient repenser et réorganiser le traitement de leurs archives courantes. L'insertion de l'étape du préarchivage dans le cycle de vie des documents servait à désengorger les bureaux avant de pouvoir procéder, en raison des délais de prescription légale à respecter, aux éliminations et au versement dans le dépôt d'archives. Les documents n'ayant plus d'utilité administrative pouvaient être transférés dans des centres de préarchivage, appelés en Allemagne *Verwaltungsarchiv* (archives administratives) ou *Zwischenarchiv* (archives intermédiaires), gérés conjointement par les Archives et les services producteurs. Cet expédient pratique d'une utilité évidente, imaginé aux États-Unis, puis appliqué un peu partout dans le monde, a été théorisé par le concept du cycle de vie en trois âges des documents. Pendant le premier âge *actif* ou *administratif*, les documents restent dans le service d'origine, passent le deuxième âge *intermédiaire* dans un dépôt de préarchivage, pour entrer ensuite, une fois versés aux Archives, dans le troisième âge *historique*.

Une digression s'impose ici sur l'extrême pauvreté de la terminologie, qui rend inévitables ambiguïtés et obscurités. Ainsi, le même mot sert à désigner les fonds et collections des Archives nationales françaises entreposés sur 366 000 mètres de rayonnages, les papiers d'une famille ou d'une association occupant une étagère, une armoire ou un local. De plus, il n'existe pas en français de terme correspondant à la *Registratur* allemande et aux *records* anglais, d'où le recours au terme *archives courantes*. À ce premier handicap s'ajoute la difficulté de traduire les concepts et les termes techniques liés à l'histoire et à la pratique des archives d'un pays donné, comme les subdivisions des fonds, les types de documents et d'instruments de recherche, etc., qui n'ont pas d'équivalents dans d'autres langues.

La définition du mot « archives » n'est pas gravée dans le marbre. Jusqu'au milieu du dernier siècle, elle se limitait aux documents d'intérêt historique de valeur permanente conservés dans les archives publiques. La théorie des trois âges a étendu le sens du mot à « l'ensemble des *documents*, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale, et par tout service ou organisme public ou privé, dans l'exercice de leur activité, *documents* soit conservés par leur créateur ou leurs successeurs pour leurs besoins propres, soit transmis à l'institution d'*archives* compétente en raison de leur valeur *archivistique*. »¹⁵

Records management fait partie des termes sans équivalent français. Il s'agit de la fonction, exercée dans le cadre de la gestion administrative, de gérer avec économie et efficacité la création, la conservation, le tri et l'utilisation des documents. L'innovation majeure de cette pratique née pour faire face à l'inflation documentaire, l'intervention en amont pour simplifier et normaliser les procédures, ont fait naître la profession des *records managers* (gestionnaires de documents), distincte de celle des archivistes. Grâce à l'adjonction du contrôle de la création des documents, le *records management* allait faciliter le passage à l'ère des archives dématérialisées. Il est appelé maintenant *information management*, terme plus approprié dans le nouveau contexte.

2.4 Conservation

Une fois la vie normale rétablie après la Seconde Guerre mondiale, la capacité des locaux, hérités de l'avant-guerre, s'est avérée dramatiquement insuffisante en Europe face aux

¹⁵ Peter WALNE (ed.), *Dictionary of Archival Terminology, English and French with Equivalents in Dutch, German, Italian, Russian and Spanish*. K.G. Saur, München–New York–London–Paris, 1984, 226 p. (ICA Handbooks Series. Volume 3).

versements à prendre en charge. L'impulsion à entreprendre des programmes de construction à grande échelle vint de Charles Braibant (1889-1976), directeur des Archives de France de 1948 à 1959. Des chantiers s'ouvrirent les uns après les autres. Furent inaugurés les bâtiments des Archives départementales de Toulouse en 1956, de Mézières en 1957, de Tours en 1958, et depuis lors l'effort ne s'est pas affaibli. En l'espace de cinquante ans, les Archives de tous les départements français ont été relogées au moins une fois. Bien d'autres pays européens (Allemagne, Autriche, Belgique, Italie, Slovaquie, Suède, etc.) se sont engagés sur la même voie avec, peut-être, des résultats moins spectaculaires.

Au cours des quarante dernières années, les Archives centrales (nationales, fédérales) de dizaines de pays ont reçu de nouveaux sièges : Algérie, Allemagne, Australie, Autriche, Belgique, Brésil, Burkina Faso, Canada, Costa Rica, États-Unis, France, Hongrie, Indonésie, Malaisie, Mexique, Norvège, Nouvelle-Zélande, Pakistan, Pays-Bas, Portugal, Royaume-Uni, Russie, Singapour, Slovaquie, Sri Lanka, Suède, République tchèque, Venezuela et sans doute bien d'autres.

Lors du démarrage des premiers chantiers en France et en Allemagne, faute de littérature spécialisée, les architectes qui se lançaient dans l'aventure ne pouvaient compter que sur leur inventivité pour concevoir le plan et diriger la construction du bâtiment à réaliser. Il fallait donc identifier et exposer, de façon ordonnée, les spécificités et servitudes dont les architectes devaient tenir compte, depuis le choix du terrain jusqu'à l'équipement des magasins et des ateliers de restauration, en passant par les options structurales, les normes de sécurité, la régulation du microclimat, la répartition de l'espace et les circuits de déplacement à l'intérieur du bâtiment. Le manuel de Michel Duchein, *Les Bâtiments et équipements d'archives*¹⁶, publié dix ans après le début des chantiers français, comblait enfin cette lacune. Les grands chantiers se multipliant, l'architecture appliquée au domaine archivistique, née avec le manuel de Michel Duchein, progressait rapidement, pour dégager un ensemble de normes (telle la double paroi assurant l'isolation thermique sans climatisation) qui sont entrées dans les habitudes. Les responsables des projets de construction, architectes et archivistes, ont maintenant à leur disposition une littérature riche, quelle que soit la zone climatique d'implantation ou le support matériel des archives à préserver.

¹⁶ Paris, Conseil international des Archives, 1966, 314 p. Le manuel révisé *Les Bâtiments d'archives, construction et équipements*, Paris, Direction des archives nationales, la Documentation française, 1985, 256 p., est toujours en vente.

Les anciennes techniques minutieuses de la réparation des sceaux, des parchemins, des pièces papier précieuses et des reliures continuent à servir, mais face à la dégradation de la qualité du papier et des encres, combinée avec la pollution de l'air, la tâche de la préservation a changé de dimension et d'acuité. Les ateliers de restauration se sont étoffés et des campagnes de recherche ont été entreprises par des laboratoires et des chimistes individuels dans divers pays. Les méthodes de désacidification du papier, le laminage thermoplastique et manuel et le comblement par pulpe de papier ont augmenté la productivité des ateliers, élargi leurs possibilités d'intervention et amélioré la qualité de leur travail. Par ailleurs, on pouvait faire appel au microfilmage pour sauver le texte de documents irrémédiablement détériorés.

Dès avant la fin du XIX^e siècle, des archives et des bibliothèques avaient eu recours à la production de copies photographiques pour satisfaire les demandes d'institutions partenaires et de lecteurs. Le microfilmage d'envergure pour produire des copies de sécurité à conserver sur un autre site et des copies de lecture pour ménager les originaux prit son essor à la fin des années 1940. Les ateliers purent s'équiper d'appareils de prise de vue spécialement conçus pour le microfilmage de documents et de livres, les fabricants de pellicule eurent à cœur de répondre aux besoins des archives et des bibliothèques et une gamme d'appareils de lecture de différentes marques était disponible pour les salles de recherche. Le microfilmage servait aussi à entreprendre des programmes d'échange entre pays ainsi que la fourniture de copies aux lecteurs.

En Amérique latine dans les années 1970, en Europe un peu plus tôt, on lançait la rumeur que, grâce au microfilmage, l'État pouvait faire de substantielles économies. Au lieu de financer la construction de dépôts d'archives de centaines ou de milliers de mètres carrés de surface, pour y conserver des cartons d'archives sur des centaines ou des milliers de mètres linéaires de rayonnages, il suffisait d'aménager des locaux de quelques dizaines de mètres carrés pour y entreposer la totalité des documents versés sous forme de microfilms. Cette solution plaisante pour les firmes commercialisant des pellicules est à classer dans la catégorie des *fallacies* déjà évoquée. Les travaux préparatoires, classement des pièces et liste précise sur chaque bobine, demandent du temps. Les prises de vue exigent du réglage, car les formats des pièces qui se suivent varient, de même que le contraste entre le support et le texte ou dessin qui y figure. Il en résulte que la construction qui permet de conserver les originaux coûte moins cher que le microfilmage dit de substitution.

2.5 Communication

Une évidence d'abord. L'accès aux sources dépend de l'état de classement des fonds et de leur description, donc du travail professionnel des archivistes. Dans chaque dépôt, les chercheurs trouvent, dans ou près de la salle de lecture, l'ensemble des instruments de recherche manuscrits et imprimés : guides, inventaires, répertoires, catalogues, etc. En raison du passage à l'informatique, il a fallu repenser les méthodes de description. La norme générale et internationale de description archivistique ISADg, élaborée au cours des années 1990 sur l'initiative des Archives du Canada, est disponible maintenant en quatorze langues. Les nouveaux instruments de recherche, inventaires et bases de données, se consultent sur écran en ligne. Pour faciliter le travail des chercheurs, les anciens inventaires sont progressivement mis en ligne dans un nombre croissant de pays.

Avec le programme de Séville, réalisé pour célébrer le cinquième centenaire de la traversée de l'Atlantique par Christophe Colomb, l'Espagne ouvrit la voie à la numérisation de grands ensembles documentaires. Le programme, inauguré en 1992, consistait à lier la numérisation des fonds de l'*Archivo General de Indias* à la construction d'une base de données intégrant l'ensemble des informations contenues dans les inventaires, pour permettre l'étude des documents numérisés et aussi préparer des recherches dans la salle de lecture. Environ 10 % des documents ont été numérisés et mis en ligne. Les logiciels du programme de Séville, adaptés au besoin du projet, ont servi ensuite à l'informatisation des archives du Komintern, réalisée à Moscou entre 1996 et 2004. En réalisant le programme de Séville, l'Espagne a partagé, avec les pays qui s'en sont détachés il y a près de deux cents ans, sa mémoire américaine. Quant à la réalisation du projet de Moscou, elle a radicalement amélioré les conditions de la recherche de l'histoire de l'entre-deux-guerres en Europe et hors d'Europe.

Les débats au sujet de l'accès aux archives se cristallisent autour des restrictions auxquelles se heurtent les lecteurs. Les pouvoirs publics aussi bien que la société civile estiment indispensable que la législation protège, pendant des périodes déterminées :

- les intérêts publics prépondérants dignes de protection, tels que la défense nationale, la politique étrangère et l'ordre public ;
- les particuliers contre la divulgation de données relatives à leur vie privée.

Or, les règles de communication héritées de l'avant-guerre contenaient des restrictions devenues incompatibles, après la Seconde Guerre mondiale, avec les exigences de la science

historique en plein essor, et avec l'intérêt croissant pour l'étude du passé récent. L'action internationale, entreprise par la communauté des archivistes dans les années 1960, s'est assigné comme buts d'en finir avec les restrictions obsolètes, comme la longueur de période de clôture générale fixée à cinquante ans dans la plupart des pays européens, et avec celles dont l'application est devenue inutile comme la discrimination entre nationaux et étrangers. De nombreux pays ont suivi les recommandations en ce sens des Congrès internationaux de 1966 et de 1968 et ont libéralisé leur réglementation.

Puis, après le tournant de 1989, se mettait en route sous l'égide du Conseil de l'Europe le démontage du régime du secret, hérité de l'ère de l'État-Parti, dans les nouvelles démocraties européennes. L'action a abouti à l'adoption de la Recommandation n° R(2000)13 du Conseil de l'Europe sur une politique européenne en matière de communication des archives. Ajoutons que les seize principes éthiques, procéduraux et techniques recommandés dans ce texte restent valides dans le monde des archives dématérialisées.

2.6 Développement

La majorité des dirigeants des États devenus indépendants lors du démantèlement du système colonial était convaincue que l'organisation des archives était un luxe que pouvaient se permettre les riches pays industrialisés, mais une dépense inutile là où toutes les ressources devaient être consacrées à la lutte contre le sous-développement. Le rôle que jouent les services d'archives pour assurer l'efficacité et la continuité de l'activité administrative n'était tout simplement pas perçu. Facteur aggravant, la pratique observée pendant trois cents ans d'inclure des clauses sur la dévolution des archives dans les traités portant sur la succession d'États fut interrompue lors de la décolonisation, occultant ainsi le rôle qui revient aux archives dans la gestion des relations internationales. La non-organisation des archives figure parmi les symptômes du gaspillage et de la mauvaise gouvernance à déplorer dans maints pays du Sud.

Le défi fut relevé par l'UNESCO qui savait pouvoir compter sur l'appui de la communauté professionnelle. La mission confiée au Dr. F. R. J. Verhoeven d'organiser les Archives nationales de Malaisie, accomplie entre 1962 et 1966, a marqué le début d'une nouvelle étape dans l'histoire des archives du tiers-monde. En 1968 fut créée, à Kuala Lumpur, la première organisation régionale des archives : la SARBICA (Southeast Asian Regional Branch of the ICA). L'idée de compter sur la coopération régionale pour

promouvoir le développement des archives fit le tour du monde : Afrique de l'Est et du Sud (1969), pays arabes (1972), Amérique latine (1973), Caraïbes (1976), Asie du Sud et de l'Ouest (1976), Afrique de l'Ouest (1977), Pacifique (1981), Afrique centrale (1982), puis, onze ans plus tard, Asie orientale. L'UNESCO a apporté son soutien à la fondation et aux activités de toutes ces branches régionales de l'ICA. Elle mit sur pied les deux premières écoles d'archives africaines, à l'Université de Dakar en 1971 et à l'Université Legon du Ghana en 1974.

Pendant quelque trente ans, le soutien des programmes de développement occupait la première place dans la hiérarchie des priorités de l'action internationale. Des ressources purent être mobilisées pour financer des bourses d'études, des cours et des séminaires de formation, ainsi que des missions de consultants pour préparer des projets législatifs, des plans de construction et d'adaptation de bâtiments, l'équipement d'ateliers de restauration ou encore le recensement des fonds dispersés dans les ministères et l'organisation des versements. Certes, il reste encore des États où les Archives nationales, tout en figurant dans l'organigramme de l'appareil administratif et possédant quelques fonds et collections historiques, n'ont qu'une existence nominale, mais l'action internationale a changé l'image des archives à travers les continents du Sud : leur utilité administrative n'est plus contestée.

3 Scripta debent manere

3.1 Contestations

« L'histoire est le témoin des temps, la lumière de la vérité, la vie de la mémoire, l'instructrice de la vie, la messagère de l'antiquité.¹⁷ » Formulée dans son traité sur l'art oratoire, cette phrase de Cicéron a sa place dans toute introduction (optimiste) aux études historiques. Les historiens d'aujourd'hui parlent de leur art avec plus de retenue et s'interrogent sur deux des qualificatifs cicéroniens, intimement liés, « lumière de la vérité » et « instructrice de la vie ».

La vérité historique change d'auteur en auteur et de génération en génération. La recherche s'étend sans cesse à des sources non encore utilisées, et la liberté d'interpréter ou de réinterpréter les sources appartient à tout historien. Ranke, qui cherchait *la lumière de la vérité*, ne croyait pas que l'historien eût une tâche pratique de pédagogue. Il devait se

¹⁷ « Historia vero testis temporum, lux veritatis, vita memoriae, magistra vitae, nuntia vetustatis. »

contenter de cultiver sa discipline sans prétendre à en enseigner les leçons. Ces leçons sont tirées par d'autres, les philosophes, les politologues, les sociologues, les économistes, les théoriciens et stratèges militaires et aussi les hommes politiques et les journalistes. Quant au citoyen, il est libre de choisir la leçon qui lui semble le plus crédible, car confortant la mémoire collective du moment.

On aurait tort de croire que le statut de l'archive, comme le socle par excellence sur lequel s'édifie le savoir historique, rencontre l'approbation unanime de la corporation des historiens et des lettrés en général. Socle par excellence ne signifie pas d'ailleurs monopole. Il signifie que grâce à leur structure basée sur les fonds, ensembles issus du fonctionnement d'entités publiques ou privées, les archives accompagnent le déroulement de l'histoire. Elles ne consignent pas des faits, mais des enchaînements de faits, comme l'avaient observé Michelet et Tocqueville, cités plus haut. L'histoire s'alimente aussi d'une vaste gamme de sources matérielles et immatérielles, sonores et visuelles, conservées dans des bibliothèques, des musées, chez des particuliers, auxquelles s'ajoutent des monuments, des sites naturels et architecturaux. La liste est longue.

Le dossier de la mise en doute de la valeur de l'érudition, servante humble des philosophes qui écrivent l'histoire, est ouvert depuis deux cent cinquante ans. Des modes surgissent, de temps à autre, contestant l'utilité des documents d'archives, en particulier des archives publiques, pour la recherche. Les arguments varient, mais tous dérivent soit de la condescendance envers le travail d'érudition soit de la non-connaissance de ce que sont, en fait, les archives.

Le « culte » de l'archive source authentique s'est donc attiré beaucoup de foudres. Dans une série de conférences tenues à Cambridge en 1961, Edward Hallett Carr s'en prend passionnément au respect « fétichiste » des sources, à la prétention de reconstruire le passé à l'aide des documents d'archives¹⁸. La littérature anti-archives publiques ne manque pas d'inventivité. On oppose les archives « d'en bas » vivantes et sincères aux archives fabriquées par l'élite de l'État¹⁹, dont la valeur informative est douteuse du fait de l'immensité de la distance qui sépare l'exercice des fonctions administratives des préoccupations réelles de la société civile. Selon Karl Raimund Popper, les faits historiques à notre disposition « ont été

¹⁸ E.H. CARR, *What is History?* The George Macaulay Trevelyan Lectures Delivered in the University of Cambridge, January-March 1961. London, Penguin Books, 1973, 169 p. (1ère édition: Macmillan, 1961).

¹⁹ Voir: Petra HEINE und Jürgen BACIA, "Plädoyer für die Bewahrung des Geschichte von unten" in *Der Archivar*, 50, 1997, 2, col. 299-326.

collectés conformément à un point de vue préconçu ; les soi-disant "sources" de l'histoire consignent seulement des faits qui ont paru suffisamment intéressants pour être consignés, en sorte que les sources ne contiendront souvent que des faits qui s'accordent avec la théorie préconçue. »²⁰ Ironique et dubitatif, le discours postmoderne sur les archives partage cette conviction, comme l'observe Terry Cook²¹.

Il n'est peut-être pas inutile de rappeler une phrase de Marc Bloch : « Contrairement aux sources narratives, chroniques, livres d'histoire et mémoires, délibérément voués à l'information des lecteurs, le souci d'instruire l'opinion, soit des contemporains, soit des historiens futurs n'avait aucune part dans la création des documents d'archives »²². Dans le même essai, écrit dans la clandestinité, il rend un vibrant hommage au travail des érudits et aux outils dont dispose grâce à eux la recherche historique.

Une mode lancée il y a une bonne trentaine d'années oppose la *mémoire* « affective et magique » à *l'histoire* qui délégitime le passé vécu par la reconstruction problématique et incomplète de ce qui n'est plus²³. La raison et la finalité de ce débat sont difficiles à saisir. Les *mémoires individuelles* recueillies font partie des sources au même titre que les documents écrits qu'elles peuvent compléter, car éclairant différemment les *faits* analysés par l'historien. Par contre, les *mémoires collectives* ne constituent pas des sources, mais des objets de l'étude historique, d'autant plus passionnants qu'elles sont plurielles, qu'elles évoluent en prenant parfois des virages abrupts, et leurs oublis (les souvenirs rayés) ont la même importance pour comprendre l'histoire que les faits retenus. Le psychologue Simon-Daniel Kipman, militant convaincu du droit à l'oubli, affirme avec raison que « dans l'écriture, comme dans l'histoire, comme dans la vie l'oubli est omniprésent » et que « les oublis collectifs s'inscrivent dans un continuum, une logique qu'il s'agit de décrypter. »²⁴

L'œuvre d'historien naît du « respect fétichiste des sources ». J'ai choisi trois monographies, l'une suisse, l'autre hongroise, la troisième française, pour illustrer ce que représentent dans le travail d'historien les sources, c'est-à-dire la combinaison des

²⁰ Karl Raimund POPPER, *The Open Society and Its Enemies*, Routledge, London 1962 (première édition 1945), I, p. 265.

²¹ Terry COOK, „Archival Science and Postmodernism: New Formulations for Old Concepts”, <http://www.mybestdocs.com/cook-t-postmod-p1-00.htm>

²² Marc BLOCH, *Apologie pour l'Histoire ou Métier d'Historien*, pp. 37-38

²³ Voir: Enzo TRAVERSO, *Le passé modes d'emploi histoire, mémoire, politique*. La fabrique, Paris, 2005.

²⁴ Simon-Daniel KIPMAN, *L'oubli et ses vertus*. Albin Michel, Paris, 2013, p. 14 et p. 127.

informations tirées d'une multiplicité de fonds, conservés dans différentes Archives et bibliothèques.

Pour dresser le bilan du mouvement puis du gouvernement libéral au pouvoir de 1830 à 1845, dans le canton de Vaud²⁵, Gérard Arlettaz a poursuivi des recherches dans les Archives cantonales vaudoises, à la Chancellerie d'État du canton de Vaud, aux Archives fédérales, aux Bibliothèques cantonales et universitaires de Lausanne et de Fribourg, à la Bibliothèque nationale suisse, au Musée de l'Élysée à Lausanne, auxquelles s'ajoutaient deux fonds détenus par des familles. Il a ainsi pu suivre l'évolution des idées politiques et religieuses, le fonctionnement des institutions, la politique économique et sociale, l'éducation, l'exercice du pouvoir libéral et la montée de la gauche radicale. Son enquête couvrait également des archives imprimées, la presse, des essais, rapports, discours, mémoires et correspondances de l'époque et quelque deux cents ouvrages historiques.

L'ouvrage hongrois et d'une autre nature. Le procès en crime rituel de Tiszaeszlár défraya la chronique en 1882-1883. L'histoire de l'affaire, la première offensive de l'antisémitisme en Hongrie, échouée sur la résistance de l'État de droit, est connue. Dans son ouvrage monumental²⁶, György Kövér procède à une dissection minutieuse de la genèse et du déroulement de l'affaire. Son analyse englobe le village menacé en permanence par le fleuve, la population appartenant à quatre cultes, sa pauvreté, les tensions à l'intérieur de la commune et les acteurs du drame : la jeune fille disparue, les inventeurs de l'accusation, les accusés, les témoins et les familles de ceux-ci, les gendarmes, les juges, les avocats et les journalistes qui assistaient au procès. Pour réaliser cette fresque, Kövér a dépouillé de multiples fonds et collections conservés aux Archives nationales et dans une douzaine de dépôts d'archives départementales, dans les archives des églises catholique et calviniste, dans la Bibliothèque nationale et la Bibliothèque de l'Académie des Science, au Musée national et au Musée juif. Il a consulté des manuscrits de conférences et de thèses inédites et la presse de l'époque. La bibliographie énumère cinq cents titres d'ouvrages historiques, sociologiques et psychologiques.

²⁵ Gérald ARLETTAZ, *Libéralisme et société dans le canton de Vaud, 1814-1845*. Lausanne 1980, 742 p. (Bibliothèque historique vaudoise, N° 67).

²⁶ KÖVÉR György, *A tiszaeszlári dráma* [Le drame de Tiszaeszlár]. Osiris, Budapest, 2011, 749 p.

Pour écrire son ouvrage pionnier²⁷ sur le thème à peine effleuré des camps d'internement français fonctionnant de 1938 à 1946, Denis Peschanski a conduit, en France et hors de France, une enquête d'ampleur exceptionnelle : vingt et une Archives départementales (où étaient situés les camps), Archives nationales (onze séries dépouillées), Archives nationales d'Outre-mer, Services historiques de l'Armée de Terre et de la Gendarmerie, Musée et archives de la Préfecture de police de Paris, Centre de documentation juive contemporaine, Bundesarchiv Coblenz, Politisches Archiv des Auswärtigen Amtes, Bonn, Centre russe qui conserve les archives du Komintern (RCHIDNI devenu RGASPI) et divers fonds privés dont les photographies prises par l'inspecteur Philiponeau. En tout, 600 000 internés passèrent des périodes plus ou moins longues dans les deux cents camps : républicains espagnols et brigadistes, militants communistes, étrangers indésirables, Juifs, Tsiganes, proxénètes, prostituées, trafiquants du marché noir, puis après 1944 des collaborateurs. Peschanski examine les quatre politiques successives d'internement, le rôle du gouvernement de Vichy et celui des autorités d'occupation, la gestion et la vie à l'intérieur des camps, les personnels de garde, les déportations, les structures d'entraide et de sauvetage et l'utilisation de l'internement administratif dans le cadre de l'épuration.

À la lumière de ce qu'est l'œuvre de l'historien, les discours sur le caractère idolâtre du crédit accordé aux archives, sur l'origine impure et la nature intrinsèquement trompeuse des documents issus des organismes de l'État, rappellent les disputations théologiques de jadis. Mais tant qu'on ne brûle pas les archives pour incitation à l'hérésie...

3.2 Protection de la vie privée vs intégrité du patrimoine

Diverses menaces pèsent sur la pérennité de la mémoire de l'État et de la société, fondée sur la consultabilité des sources primaires produites par le fonctionnement des institutions publiques et privées. Ainsi, les règles conçues pour protéger les données à caractère personnel, qui figurent dans les premiers instruments intergouvernementaux, semblent s'infléchir dans une direction insoupçonnée lors de leur adoption au début des années 1980.

Les restrictions traditionnelles à la communication de documents contenant des données personnelles répondaient à la nécessité de « protéger l'honneur des personnes vivantes, ou des familles proches de personnes décédées, contre les indiscretions que pourrait entraîner la divulgation des renseignements contenus dans les documents d'archives récents », « de

²⁷ Denis PESCHANSKI, *La France des camps. L'internement 1938-1946*. Gallimard, Paris, 2002, 555 p.

protéger les intérêts légitimes des citoyens contre la rupture du secret administratif, notamment en matière fiscale » et « d'assurer la liberté d'action de l'Administration face aux curiosités injustifiées durant un délai raisonnable »²⁸. La réglementation antérieure aux années 1970 se fondait donc essentiellement sur des critères assez flous, en particulier une série d'adjectifs, en sorte que l'autorisation d'accès à des documents contenant des données personnelles se faisait sur l'appréciation de la demande et du degré de confidentialité des données concernées. L'essor des recherches sur le « temps présent » (Première Guerre mondiale, entre-deux-guerres, Seconde Guerre mondiale et après-guerre) multipliant les demandes, des règles plus précises devaient être introduites.

Les réglementations détaillées de la communication des documents, approuvées à partir des années 1970, devaient être coordonnées avec les instruments nationaux et intergouvernementaux relatifs à la protection des données à caractère personnel. Au départ, les instruments, lois ou directives, répondaient à l'inquiétude des organisations de la société civile, des forums judiciaires et des Parlements devant le progrès de l'informatisation dans les services publics, qui rendait possible l'interconnexion des fichiers contenant des données personnelles produits dans les diverses branches de l'Administration. On craignait de voir les États disposer d'outils les mettant à même de tout savoir sur chacun de leurs citoyens.

En France, dès 1978, la Commission Nationale de l'Informatique et des Libertés (CNIL) reçut la mission de veiller à ce que les techniques informatiques ne pussent être utilisées pour porter atteinte aux droits, aux libertés et à la vie privée des citoyens. La loi française du 3 janvier 1979 sur les archives prescrivit des délais de communicabilité de soixante à cent cinquante ans pour les divers types de renseignements individuels, à compter de la date de naissance de la personne concernée. Ailleurs en Europe, les durées de la période protégée s'échelonnent de trente à cent vingt ans.

Le premier texte intergouvernemental, les *Lignes directrices sur la protection de la vie privée et les flux transfrontières de données à caractère personnel*, fut approuvé par l'OCDE en 1980. Cet instrument ainsi que la *Convention du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel*, adoptée l'année suivante, étaient censés faire barrage à l'abus informatique. Ainsi, les lignes directrices de l'OCDE ne s'appliquent pas à des « données de caractère personnel qui, manifestement, ne présentent aucun risque pour la vie privée et les libertés individuelles ».

²⁸ *Manuel d'archivistique*, p. 296.

Quant à l'article 6 de la Convention du Conseil de l'Europe, où sont énumérées les données considérées particulièrement sensibles, il est toujours libellé dans l'esprit de la préoccupation d'origine : « les données à caractère personnel révélant l'origine raciale, les opinions politiques, les convictions religieuses ou autres convictions, ainsi que les données à caractère personnel relatives à la santé ou à la vie sexuelle, ne peuvent être traitées automatiquement à moins que le droit interne ne prévoise des garanties appropriées. Il en est de même des données à caractère personnel concernant des condamnations pénales. »

La *Directive 95/46/CE du Parlement européen et du Conseil du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données* amplifie dans plusieurs directions les principes contenus dans la Convention de 1981. Elle en étend l'application à tout ensemble structuré de données à caractère personnel, sons et images compris, à tout traitement, automatique ou non, de ces données, et interdit la collecte et le traitement des données sensibles identifiées dans l'article 6 de la Convention. Ces interdictions, fondées sur le vécu européen du XX^e siècle, sont destinées à prévenir le glissement des États de droit démocratiques vers l'autocratie.

Les études en cours pour la révision de la Directive vont jusqu'à envisager des mesures plus radicales, à savoir l'anonymisation voire la destruction des ensembles de données personnelles dès lors qu'ils n'ont plus d'utilité pratique. Il faut espérer que cette idée sera abandonnée lors de l'adoption de la Directive révisée.

L'anonymisation des données indiffère le sociologue et, en ce qui concerne l'épidémiologie, elle ne gêne que les enquêtes sur les maladies héréditaires. En revanche, les murs de protection de plus en plus hermétiques, érigés autour des données à caractère personnel (« sensibles » ou non) figurant dans les documents d'archives, gênent toute recherche d'histoire politique, sociale et culturelle, et font barrage devant les études dont les sujets sont des individus ou des familles. Il ne s'agit pas seulement des travaux biographiques et généalogiques, mais aussi de trois branches particulièrement dynamiques de la recherche : la micro-histoire, les études de genre et la prosopographie.

La *micro-histoire* a reçu ses lettres de noblesse grâce au livre de Carlo Ginzburg sur un meunier du Frioul du XVI^e siècle²⁹. Aujourd'hui, c'est le terrain privilégié des historiens non

²⁹ Carlo GINZBURG, *Le fromage et les vers. L'univers d'un meunier du XVI^e siècle*. Flammarion, Paris, 1980, 220 p.

professionnels qui se penchent sur le passé de leur commune ou sur l'histoire de la firme où ils avaient travaillé. Dans les années 1980, les *études de genre* se fixèrent comme objectif de s'affranchir de la vision traditionnelle masculinisée de l'histoire pour y intégrer les femmes. La *prosopographie* fait partie des mots savants dont la signification est façonnée par l'évolution de la recherche. Dans leur nouvelle acception qui date des années 1980, les études prosopographiques portent sur l'examen individuel des membres d'un groupe déterminé (association, parlement, unité militaire, loge maçonnique, etc.) selon une grille de critères, comprenant notamment la généalogie et la parenté, les études, le parcours professionnel, les biens, les revenus, l'appartenance religieuse et la carrière à l'intérieur du groupe³⁰. Les données individuelles ainsi ordonnées et combinées donnent la clé pour comprendre le poids social et le rôle historique de l'institution ou de l'organisation étudiée.

L'histoire ne se réduit pas aux gestes des grands hommes, grands commis, grands capitaines et génies. Y interviennent aussi, pour ne citer que les trois ouvrages évoqués ci-dessus, les citoyens, fonctionnaires et pasteurs vaudois, les villageois de Tiszaeszlár et les internés, les directeurs, les inspecteurs, les médecins et les gardiens des camps de France. Comme, dans un environnement informatisé, il suffit d'appuyer sur quelques boutons pour anonymiser ou supprimer des données et que toute cause a ses protecteurs zélés qui n'hésitent pas si leur conscience leur dicte d'agir, des pertes de données à caractère personnel se produiront certainement, et l'on peut même s'attendre à des hécatombes.

3.3 L'horizon de cinq cents ans

Dans leurs entretiens menés par Jean-Philippe de Tonnac³¹, Umberto Eco et Jean-Claude Carrière affirment leur certitude, avec arguments convaincants à l'appui, que l'e-book ne détrônera pas le livre. Interviewé pour Télérama en octobre 2009, Umberto Eco reprenait l'essentiel de ces arguments, dont celui-ci : « Nous pouvons encore aujourd'hui lire des livres vieux de cinq cents ans. En revanche, nous n'avons aucune preuve scientifique que le livre électronique puisse durer au-delà de trois ou quatre ans. En tout cas, il est raisonnable de douter, compte tenu de la nature de ses matériaux, qu'il conserve la même intensité

³⁰ À voir, par exemple, les excellents ouvrages prosopographiques d'Eric ANCEAU, *Dictionnaire des députés du Second Empire*. Presses Universitaires de Rennes, 1999, 421 p. et *Les députés du Second Empire. Prosopographie d'une élite du XIX^e siècle*. H. Champion, Paris, 2000, 1018 p.

³¹ Jean-Claude CARRIÈRE, Umberto ECO, *N'espérez pas vous débarrasser des livres*. Grasset, Paris 2009, 283 p.

magnétique pendant cinq cents ans. Le livre, c'est une invention aussi indépassable que la roue, le marteau ou la cuiller. »³²

Puisque les livres ne sont pas condamnés à disparaître, les bibliothèques continueront à les conserver, tout en s'occupant également de la préservation d'e-books et d'autres documents électroniques comme les pages web de l'Internet. Pareille certitude n'est pas de mise en ce qui concerne les archives. C'est la disparition en cours et à venir des dossiers traditionnels qu'analyse la partie *Archives dans le monde numérique* du présent livre.

La disparition programmée du papier ne s'est pas encore réalisée. Les documents authentiques analogiques servent toujours aux personnes morales et physiques pour prouver leurs droits, à l'État pour contrôler le respect des dispositions légales et réglementaires et à la justice pour garantir la régularité des procédures. Dans notre monde où coexistent documents analogiques et électroniques, si perquisition il y a dans le cadre de poursuites pénales, la justice saisit aussi bien des ordinateurs, des disques durs externes et des clés USB que des lettres, comptes, reçus, factures, plans, etc., sur papier. Dans les reportages télévisés sur des administrations, la caméra s'arrête toujours sur des cartons d'archives. Mais la métamorphose est en cours et toute l'énergie intellectuelle de la communauté professionnelle des archivistes s'investit dans la recherche théorique et technique qu'exige le passage de l'ère des archives papier à celle des archives électroniques dématérialisées.

L'indifférence que l'on constate un peu partout quant au sort des documents s'explique en partie par le syllogisme dévastateur répandu à travers cinq continents : la Bureaucratie est Papier, l'Archive est Papier donc l'Archive est Bureaucratie. D'où le succès du slogan « bureau sans papier », lancé dès le début de l'ère informatique. En fait la bureaucratie est le système (et la bureaucratie informatisée en est la variante sophistiquée) destiné à camoufler les responsabilités pour éviter de devoir rendre compte de ses actes. L'archive n'est pas papier, mais le corpus organisé de preuves concernant des faits, des décisions et des transactions³³. Aussi est-elle probablement le seul outil efficace pour établir qui était responsable pour quoi. Théoriquement, l'archive dématérialisée peut rendre le même service en empêchant camouflage et fuite. Mais, du fait de la volatilité et de la destructibilité des documents électroniques, elle risque de s'avérer moins efficace que l'archive papier.

³² Propos recueillis par Catherine PORTEVIN – Télérama n° 3117, <http://www.telerama.fr/livre/umberto-eco-internet-encourage-la-lecture-de-livres-parce-qu-il-augmente-la-curiosite,47983.php>

³³ Le rôle des archives de garantir l'*accountability* de l'organisme producteur est affirmé avec force dans la littérature archivistique australienne.

Le chemin frayé par une technologie qui maîtrise la production, la gestion, la mise à jour et la communication de documents électroniques dématérialisés, sans pouvoir en garantir la lisibilité à l'horizon plusieurs fois séculaire, pointe précisément dans la direction d'un monde débarrassé des archives. Le livre récent de Christian Duverger³⁴ vient d'illustrer avec virtuosité la légitimité du souci de la conservation de longue durée, en l'occurrence de cinq cents ans.

L'Histoire véridique de la conquête de la Nouvelle Espagne, mentionnée dans la correspondance administrative et des ouvrages historiques à partir de 1575, paraît pour la première fois en 1632, sous le nom de Bernal Díaz del Castillo, modeste homme de troupe de Cortés, qui l'aurait composée quarante-cinq ans après la prise de Mexico. Sa paternité du texte n'a été mise en doute ni dans les multiples éditions de *l'Histoire véridique* en espagnol et en d'autres langues, ni dans la littérature surabondante sur ce chef-d'œuvre du Siècle d'Or. Cette conviction bien enracinée est résumée en quatre lignes dans l'essai sur l'historiographie de John Burrow : « Un autre historien [de la Conquête, à côté de Las Casas] est un compagnon de Cortés dans la marche de la côte à la capitale aztèque et dans le dur combat qui aboutit à la prise de la ville, Bernal Díaz. Sa relation, qu'il écrivit dans sa vieillesse, est un récit alerte et vigoureux, exempt de tropes conventionnels et de fioritures rhétoriques.³⁵ »

Christian Duverger prouve, au-delà de toute possibilité de doute, que le soldat conquistador de base, Bernal Díaz, installé à Santiago de Guatemala au début des années 1540, ne put être l'auteur de *l'Histoire véridique*, écrite par un lettré possédant une culture classique solide. Au temps de Charles Quint et de Philippe II, en métropole comme dans les possessions américaines, l'administration espagnole opérait à tous les niveaux au moyen de l'établissement de documents et de leur archivage. Or, il ressort des documents émanant de Bernal Díaz, reçus par lui ou le concernant, des suppliques adressées au souverain et des cédules royales de réponse, des témoignages au sujet d'affaires d'héritage et des pièces familiales, conservés à l'*Archivo general de Indias* de Séville, à l'*Archivo histórico nacional* de Madrid et à l'*Archivo general de Centroamérica* de Guatemala, que le prétendu auteur d'un chef-d'œuvre des lettres espagnoles savait à peine écrire. À cela s'ajoute la preuve négative, non moins importante, que le nom de Bernal Díaz qui dans *l'Histoire véridique* suit

³⁴ Christian DUVERGER, *Cortés et son double. Enquête sur une mystification*. Seuil, Paris, 2013, 317 p. Édition espagnole: *Crónica de la eternidad. ¿Quién escribió la Historia verdadera de la conquista de la Nueva España?* Taurus. Madrid, 2013. 384 p.

³⁵ BURROW, *op. cit.* p. 424.

Cortés comme son ombre, n'apparaît dans aucun des documents créés au cours de la conquête de la Nouvelle Espagne, ni dans les listes des noms de l'entourage de Cortés.

À une distance de cinq cents ans, grâce aux archives, une mystification magistrale a pu être dévoilée. Quant à l'autre hypothèse de Duverger, attribuant à Cortés lui-même la paternité de l'*Histoire véridique*, hypothèse qui n'est étayée que par un faisceau de preuves indirectes, elle fait l'objet d'un débat entre américanistes depuis la parution de la *Crónica de la eternidad*.

Grâce aux procédés d'authentification, la production de faux électroniques est devenue, semble-t-il, aussi difficile que la falsification des manuscrits traditionnels. En revanche, malgré les recherches en cours dans de nombreux pays, la question prioritaire au point de vue des Archives, à savoir comment peut-on garantir la préservation des e-documents dans la longue durée (cent ans et plus), reste ouverte. Elle reste ouverte par définition, puisque la réponse ne viendra que dans cent, deux cents et cinq cents ans. Consensus tacite des plus inquiétants, selon la littérature professionnelle la réponse au souci de la longue durée ne relève pas de la réflexion archivistique mais de la recherche technologique. Il faut faire confiance à la technique qui apportera la solution. En attendant, quitte à accepter des pertes d'éléments, la conservation pérenne des documents électroniques est possible. Elle exige le respect des conditions exposées ci-après dans les chapitres 8-14. Mais intacts ou amoindris, ils ne pourront être lus qu'à l'aide d'équipements et de courant électriques.

3.4 Garantie des droits, mémoire institutionnelle, sécurité de l'environnement

Alors que l'informatisation générale de la gestion des affaires publiques et de la vie économique est un fait acquis, certains impératifs imposent la production d'originaux sur papier. La protection des droits fondamentaux des citoyens ne peut dépendre d'une coupure de courant ou d'une panne de serveur. Il s'agit en premier lieu de l'état civil. La législation de nombreux pays prescrit que du papier permanent et de l'encre indélébile soient utilisés pour la tenue des registres de naissances, de mariages et de décès. Des exigences similaires s'appliquent aux cadastres, croquis fonciers et autres titres de propriété immobilière pour garantir le droit de propriété.

Restent sur papier les originaux des textes législatifs et réglementaires, des accords diplomatiques et commerciaux et des adhésions aux conventions intergouvernementales, bref les instruments et pièces censés documenter et garantir la continuité de l'État. Dans le même

esprit, les délibérations des conseils municipaux et les arrêtés et décisions des maires sont inscrits dans des registres en papier.

Le concept du programme de protection des documents essentiels (*vital records program*) apparut aux États-Unis, au début des années 1950, face au danger d'une guerre nucléaire. Tous les services fédéraux reçurent l'instruction de prendre des mesures en vue d'assurer la continuité gouvernementale en cas d'état d'urgence. Le choix des documents essentiels à protéger, parce qu'indispensables pour que l'organisme puisse reprendre ses activités, défendre ses droits et accomplir ses devoirs, faisait partie de ces mesures. Dans la suite, ces programmes se sont généralisés et étendus au secteur non gouvernemental, notamment aux États-Unis, au Canada et en Australie, non plus pour faire face à la menace de guerre, mais pour préserver la continuité de l'organisme en cas de destruction des archives par un sinistre naturel ou provoqué (séisme, inondation, incendie, etc.). Tous les textes portant sur les règles de la gestion des documents jugés essentiels comportent une bizarrerie. Ils évoquent la possibilité et les avantages de la préservation sur papier et sur microfilm, tout en admettant, malgré le coût et les incertitudes, le support numérique. En fin de compte, ils s'en remettent au bon sens des gestionnaires des programmes.

Dans les pays du notariat latin, les preuves authentiques d'actes relatifs au droit de la famille (contrats de mariage, donations, testaments, tutorat, etc.) et d'actes de nature économique et financière (contrats, ventes, fusions, etc.) sont établis par des notaires. Des études sont en cours, notamment en Allemagne et en Autriche, sur la conservation à long terme des actes notariés informatisés. En Espagne, en France et en Italie, pays d'une tradition notariale remontant au Moyen Âge, la législation impose aux notaires de conserver sur papier les minutes des actes délivrés et régit le traitement informatique des archives notariales.

Produits pour répondre à l'impératif de la conservation pérenne, les documents analogiques sont immédiatement numérisés et intégrés dans le système informatique d'archivage. Aussi, les actes d'état civil et les extraits de casier judiciaire, pour ne citer que les exemples les plus courants, peuvent-ils être fournis instantanément aux demandeurs.

La conversion peut s'opérer en sens inverse. Tout organisme privé ou public produit des documents qu'il estime devoir préserver en permanence sous forme directement lisible. C'est le cas, comme le rapporte Bruno Delmas³⁶, de l'Agence nationale chargée de la gestion des

³⁶ Bruno DELMAS, *La société sans mémoire. Propos dissidents sur la politique des archives en France*. Bourin, Paris, 2006, pp. 139-140.

sites de stockage des déchets radioactifs de France (ANDRA) et des organismes similaires opérant dans le monde. Comme les données de ces sites doivent être préservées pendant des siècles, parfois des millénaires, une fonction « mémoire à très long terme » a été créée, qui signifie que « les archives numériques qui servent à l'archivage courant, sont converties en documents papier permanents. »

Cet échantillonnage rapide ne fait qu'effleurer un aspect, essentiel malgré le peu d'intérêt qu'il suscite, de la conservation de la mémoire du XXI^e siècle. Le moment est sans doute venu d'entreprendre des enquêtes aux échelles nationale, régionale voire internationale, pour avoir une idée plus précise de la place que garde le document lisible à l'œil nu dans les sociétés où la gestion de toutes les affaires administratives et économiques intéressant la collectivité passe par l'informatique.

4 Certitudes et conjectures en 2014

4.1 Acquis et pertes

Les Archives se sont informatisées de par le monde. Toutes leurs activités pour préserver, décrire et communiquer les milliers de kilomètres linéaires de cartons et de liasses confiés à leur garde passent par des logiciels et des équipements informatiques, depuis la gestion de l'espace-magasin jusqu'à la reconstitution, par nettoyage des images numérisées, de textes devenus illisibles. Cette mutation leur impose une veille permanente pour être à jour du progrès continu de la technologie.

Au cours des vingt dernières années du siècle passé, la réflexion archivistique a connu une période faste. Elle est devenue mondiale, avec des foyers particulièrement actifs et hardis en Australie, au Canada et aux États-Unis. Cette archivistique globale repensée qui se bâtit sur le socle de l'informatique, qui a élaboré son cadre conceptuel, sa terminologie et ses normes présentées ci-après dans les chapitres 7-14, est remise constamment en question. On s'interroge sur sa légitimité en tant que discipline, sur son appartenance à la science historique ou aux sciences de l'information. Les réponses à ces questions varient en fonction de la préparation intellectuelle et du profil professionnel du répondant. Science à part entière, science auxiliaire de l'histoire, technique professionnelle, bibliothéconomie spécialisée ou branche des sciences de l'information, ce débat abstrait de nature déductive est condamné à la stérilité, le désaccord étant d'ordre idéologique.

Il vaut donc mieux opter pour une approche inductive, prendre pour point de départ la réalité. Gérer les archives, les préserver, les structurer, les communiquer, quelles qu'elles soient, exigent un corpus de connaissances théoriques et pratiques. Ce corpus est suffisamment riche pour répondre aux besoins d'une profession diversifiée. Bâtir une base de données sur les chartes des Hohenstaufen ne fait pas appel aux mêmes savoirs que la description des archives (analogiques et électroniques) d'une banque de notre temps. Le brassage continu des idées depuis trente ans, dans des universités, des associations professionnelles, des revues et des colloques, a abouti à la conclusion que les archives sur support numérique et les archives sur support analogique appartiennent à un même ensemble et relèvent, sur le plan intellectuel, de la même discipline : l'archivistique.

L'ère de l'Internet a révolutionné l'image et l'accessibilité des Archives. Des milliers d'institutions d'archives de tous les continents font connaître sur leur site web leurs activités, leurs bâtiments, leurs instruments de recherche, leurs données statistiques et mettent en ligne des sources numérisées pour faciliter la recherche. L'acquis est spectaculaire. Il y a vingt ans, seul le cercle restreint des visiteurs des salles de lecture connaissait la finalité et les activités des Archives.

La croissance gigantesque de la masse d'informations générées dans le monde s'accompagne de pertes qui seront certainement ressenties par la postérité. Le courrier électronique et les messages SMS ont fait disparaître les correspondances privées, déjà malmenées par la concurrence du téléphone. Or, elles permettaient de saisir, avec plus de sûreté que toute autre source, « l'esprit du temps », autrement dit les soucis, les goûts, les idées et les émotions de la société civile.

La survie des premières versions des manuscrits littéraires et savants dépend dès maintenant du hasard puisque les corrections se font sur écran. Il en va de même des brouillons, projets et autres textes provisoires, sauf dans les institutions où un règlement intérieur en prescrit la conservation. Certes, les disques durs gardent toutes les traces, mais où seront-ils dix, vingt ou trente ans après la rédaction du texte ?

4.2 Une lacune à combler

L'archive objet tangible et intangible sert de base aux principes et doctrines évoqués tout au long des chapitres précédents. Toutes les approches théoriques ont en commun de pouvoir faire et, parfois, de faire l'impasse sur l'institution d'archives. Celle-ci ne participe pas de la

théorie, elle relève de la pratique archivistique propre à chaque pays. En matière d'organisation des archives, il n'existe pas de norme prescriptive. Le Conseil international des Archives a publié, en 2011, deux textes normalisant 1) la description des fonctions des collectivités associées à la production et à la gestion des archives³⁷ et 2) la description des institutions de conservation des archives³⁸. Ces deux textes descriptifs pourraient servir d'arrière-plan à une directive ou à une recommandation sur les responsabilités et les pouvoirs que doivent exercer les archives publiques en tant que services de l'État. En attendant l'élaboration d'un tel texte, seule la norme OAIS³⁹ formule quelques exigences auxquelles doit répondre une institution d'archives publiques.

L'absence s'explique sans doute par le rôle dévolu aux Archives en tant qu'attributs de la souveraineté des États. Nulle autorité internationale ou supranationale n'est habilitée à dicter à un État souverain comment il doit prendre soin des documents produits par le fonctionnement de ses services. Les juristes de l'Union européenne estiment que cette matière relève de l'autorité exclusive des États membres. Une « subsidiarité » semblable est observée aux États-Unis. Chacun des cinquante États de l'Union est libre de gérer ses archives comme il l'entend. Il peut établir un service d'archives, mais il peut tout aussi bien les confier à une bibliothèque universitaire ou à une association d'historiens. Le législateur fédéral n'exerce aucun pouvoir en la matière.

Le passage à la dématérialisation des documents publics affecte nécessairement les relations entre les Archives et l'appareil administratif. Subsidiarité ou pas, une recommandation spécifiant les responsabilités qui reviennent à l'autorité archivistique comblerait une lacune. Il va du fonctionnement de l'Union européenne que les archives soient gérées convenablement partout et que les Archives des États membres puissent coopérer conformément à la Résolution du 6 mai 2003 et de la Recommandation du 14 novembre 2005 du Conseil de l'Union.

³⁷ ISDF - Norme internationale pour la description des fonctions.

³⁸ ISDIAH - Norme internationale pour la description des institutions de conservation des archives. Les deux textes sont téléchargeables à partir du site ica.org.³⁸ Voir ci-dessous chapitre 12.3

³⁹ Voir ci-dessous chapitre 12.3

4.3 La restauration du droit international : un impératif enterré

Pendant trois cents ans environ, jusqu'à la veille de la Seconde Guerre mondiale, le droit international d'archives observait trois principes :

- 1) en cas de succession d'États, l'État prédécesseur remet à l'État successeur les documents nécessaires à exercer la souveraineté et à assurer la continuité administrative ;
- 2) les archives déplacées d'un État devenu dépendant sont restituées lorsque cet État redevient indépendant ;
- 3) les archives saisies et déplacées pendant une guerre sont restituées après la fin des hostilités à la puissance contre qui la guerre a été menée.

La destruction du droit international d'archives s'est accomplie, entre 1939 et 1960, en deux étapes. Dans la première étape de la marche vers le non-droit, le troisième principe fut aboli de fait. Des immenses masses d'archives publiques et privées, saisies dans les pays occupés et transférées dans le 3^e Reich, seuls les lots récupérés par les Alliés occidentaux furent restitués aux propriétaires. Ceux récupérés par l'armée soviétique furent transférés secrètement à Moscou avec les archives des pays ennemis saisies⁴⁰.

L'existence de ces archives spéciales contenant 27 000 mètres linéaires de documents étrangers fut révélée en février 1990. Dans un premier temps, les autorités russes opposèrent aux revendications des pays spoliés leur droit de conserver les « archives trophée ». Puis, à partir de 1994, les pays alliés ont recouvré progressivement leurs archives perdues. Mais la Russie n'envisage toujours pas de rendre « les archives trophée » aux pays ennemis (Allemagne, Hongrie, Roumanie).

Il n'a pas encore été possible de recenser les archives dispersées pendant la guerre dans divers pays, car les institutions détentrices s'abstiennent de les signaler. Les armes s'étaient tues en 1945. Mais le chapitre de la Seconde Guerre mondiale n'est toujours pas clos dans l'histoire des archives.

L'abandon des deux premiers principes relatifs aux successions d'États intervint dans la deuxième étape, après la fin de la guerre. Le droit international coutumier étant tombé en désuétude, l'accession à l'indépendance d'environ quatre-vingt-dix anciennes colonies,

⁴⁰ Grâce aux travaux de Patricia K. GRIMSTED, nous connaissons l'ampleur des pillages et le destin extraordinaire de ces masses d'archives. Voir Alexandre SUMPFF, Vincent LANIOL (dir.). *Saisies, spoliations et restitutions. Archives et bibliothèques au XX^e siècle*. Presses Universitaires de Rennes, Rennes, 2012, 383 p. La liste des livres et articles de Patricia K. Grimsted se trouve à la page 373.

entre 1950 et 1990, n'a pas donné lieu à la conclusion d'accords sur la dévolution des archives.

Les Nations Unies prirent l'initiative de combler le vide juridique en élaborant le projet d'une *Convention internationale sur les Successions d'États en matière de biens, de dettes et d'archives*. La préparation du projet a pris 16 ans. Le texte, exposant sous forme d'articles d'une convention les revendications des anciennes colonies, fut soumis à la Conférence intergouvernementale de Vienne en avril 1983. Une majorité d'États de l'Est et du Sud vota le texte pour soutenir l'argumentaire opposé aux anciennes métropoles tout en sachant qu'il ne deviendra jamais un instrument international opérationnel. En effet la Convention contient des clauses incompatibles avec les législations archivistiques, puisque prescrivant aux États signataires de transférer des catégories d'archives sans accord particulier entre l'État prédécesseur et l'État successeur.⁴¹

La Conférence de Vienne eut une conséquence désastreuse. Pour les Nations Unies, le problème est réglé. Peu importe que la Convention ne puisse jamais entrer en vigueur, faute d'États signataires en nombre suffisant⁴² elle existe dans les dossiers des Nations Unies, il n'y a donc pas lieu de se remettre à l'ouvrage. Un texte conçu dans l'esprit de la guerre froide empêche, en 2014, la mise en chantier d'une convention appelée à restaurer le droit international. Le vide juridique persistant, ni la dislocation de l'URSS ni celle de la Yougoslavie n'ont donné lieu à la conclusion d'accords sur la dévolution des archives. C'est d'autant plus regrettable que, grâce aux technologies de numérisation, nombre de contentieux pourraient être abordés dans un climat détendu.

4.4 Regards sur l'avenir

Pour scruter l'avenir des archives, on est automatiquement amené à chercher les constants dans leur histoire au cours des derniers siècles. La première constatation qui ressort de cet examen incite à la modestie. Quelle que soit la politique suivie, quelles que soient les ressources investies, il n'en résultera pas un système parfait, donnant satisfaction à tout le monde, au législateur, à l'autorité de tutelle, aux services versants, au personnel professionnel et aux lecteurs. Car les Archives doivent répondre à des exigences contradictoires : accorder

⁴¹ Voir Charles KECSKEMÉTI, „Disputed archival claims. Activities of UNESCO and ICA since 1976” in *Sovereignty Disputed Claims Professional Cultures. Essays on Achival Policies*, Brussels, 2000 pp. 247-266.

⁴² De 1983 à 1993, dix États seulement ont adhéré à la Convention.

aux lecteurs un accès libéral aux sources et protéger les secrets ; éliminer la masse pour pouvoir préserver ce qui deviendra la matière première de la recherche ; mettre en place une politique nationale cohérente et respecter les compétences des collectivités régionales et locales, pour ne citer que trois exemples.

Les défis à relever dépassent parfois la capacité de riposte des Archives, comme dans le cas du papier qui se décompose et des encres qui pâlisent puis disparaissent. Elles sont confrontées actuellement au défi global de l'entrée dans l'ère des archives dématérialisées. Au-delà des recherches technologiques, élément central de la riposte, on peut s'attendre à une gamme de conséquences affectant l'avenir de l'archive.

Une des options possibles, le maintien des fonds dématérialisés en état actif par le service d'origine, sans date butoir, pointe dans la direction du morcellement de la mémoire électronique de l'État. L'option de confier la responsabilité de la préservation aux autorités archivistiques, qui semble avoir la préférence des pays pionniers, soulèvera à plus ou moins brève échéance le problème du financement. L'État pourra-t-il couvrir, décennie après décennie, le coût de la préservation d'une masse croissante d'archives dématérialisées ?

Les versements à l'ancienne iront en décroissant mais, comme indiqué plus haut, ne tariront pas complètement. Nous savons par ailleurs, puisque les Archives en conservent les preuves, que les acteurs de la vie publique, grands commis, élus et hommes des cabinets, constituent leurs propres collections de copies et de courrier qu'ils emmènent chez eux en quittant leur poste. Ces collections, utilisées ou non pour la rédaction de mémoires, restent dans la famille. Elles peuvent reparaître et rejoindre les Archives, une ou plusieurs générations plus tard. L'habitude peut survivre dans le monde politique de demain pour le plus grand bien des Archives.

L'accès en ligne à des sources d'archives et des livres numérisés ainsi qu'aux données bibliographiques a d'ores et déjà transformé les conditions et les pratiques de la recherche. Le travail sur écran, qui sera la règle à mesure que les restrictions seront levées sur les fonds dématérialisés, s'est généralisé. Comme ne seront mis en ligne que des agrégats intéressant un public suffisamment large, les Archives continueront à fournir des copies (numériques ou analogiques) à leurs lecteurs.

Avec la possibilité du repérage automatique des mots dans les bases de données et les documents numérisés, le souhait maintes fois exprimé des professionnels d'offrir aux lecteurs des *user-friendly finding aids* s'est réalisé.

L'espionnage dans le monde d'avant se faisait par transmission de documents et de microfilms de la main à la main. Aujourd'hui, les services de renseignements procèdent par pénétration dans les systèmes informatiques. La vulnérabilité des systèmes a été prouvée de façon spectaculaire par *WikiLeaks*, avec la publication de 250 000 télégrammes diplomatiques américains. Il est de l'intérêt vital aussi bien des gouvernements que du secteur privé que soient sécurisés les systèmes contre les fuites et les pénétrations. Le souci de sécurité promet de rester longtemps d'actualité à cause de l'omniprésence de la volonté de connaître les secrets d'autrui, politiques et économiques.

Enfin, nous sommes sortis de l'époque où la mémoire collective se nourrissait de l'information distillée par la galaxie des institutions publiques et privées ambitionnant de servir ou d'influencer la société civile. Grâce aux nouveaux outils, le web et ses blogs, les ordinateurs, les tablettes, les smartphones (et ceux qui s'y ajouteront année après année), des centaines de millions d'individus produisent et échangent des informations écrites et des images au moyen des réseaux sociaux. L'océan des informations disponibles par tous ne cessera de s'étendre.

4.5 En guise d'épilogue de ces regards sur le passé et l'avenir

Dans les romans policiers, les enquêteurs, qu'ils soient policiers, détectives privés, avocats, journalistes ou amis de la victime, cherchent souvent la clé du mystère dans les archives. L'auteur place parfois les archives au centre de l'énigme. Dans *Fault Line* de Robert Goddard⁴³, le narrateur est chargé de retrouver les six cartons d'archives des années 1960 d'une compagnie minière, pour que l'historienne engagée par la compagnie puisse en écrire l'histoire. Le roman se déroule en 1968, 1969, 1984 et 2010, dans les Cornouailles et à Capri. À côté de l'intrigue principale, la trahison que documentent les cartons recherchés, à l'origine de six décès, une intrigue accessoire a pour objet une autre trahison, une lettre dénonçant en 1944 à la Gestapo un résistant italien.

Dans ce roman, tous les documents existent physiquement. Les archives recherchées sont annotées en marge et une expertise en écriture permet d'identifier l'auteur de la dénonciation. N'en doutons pas, les commissaires et les privés de la littérature policière future sauront exploiter les archives électroniques pour démasquer les délinquants. Mais les documents

⁴³ Corgi Books, 2012.

disparus devront être récupérés sur les disques durs et l'expertise en écriture ne leur sera d'aucun secours.

LES ARCHIVES DANS LE MONDE NUMÉRIQUE

5 Les fonctions des archives et les valeurs des documents

5.1 Les fonctions des archives

Les documents furent toujours rassemblés et conservés pour des besoins précis et nous pouvons distinguer, à partir de ces besoins, les quatre fonctions essentielles des archives, qu'elles soient courantes ou historiques : le service de l'administration, la garantie des droits, le service de la recherche et l'action culturelle.

- *La fonction administrative.* Au-dessus d'un certain niveau de développement, la société humaine ne peut fonctionner sans recours à l'écriture parce que, sans enregistrer les informations, on ne pourrait pas administrer, organiser et évaluer la coopération des individus, des groupes et des bureaux. Les organisations, les administrations et même les individus créent leurs archives dans le but de servir à leur propre fonctionnement.
- *La fonction de garantie des droits.* Dans les sociétés développées, les droits acquis (par exemple titres de noblesse, donations) furent consignés et les documents produits conservés par des institutions compétentes, le notariat en particulier, afin de pouvoir prouver ou infirmer les droits. Dans la Rome antique, le *Tabularium* fut le dépôt des documents privés ; aux Archives de la Chancellerie royale de Hongrie, il y a la série des registres portant le titre de *Libri regii* où on enregistra les titres de noblesse et les donations royales. Dans les sociétés de l'ère moderne, la sécurité juridique est devenue une exigence fondamentale et, pour ces besoins, l'établissement de nouvelles catégories documentaires, tels les registres paroissiaux puis d'état civil et les cadastres, s'est généralisé. Dans un État de droit, un document produit dans des conditions assurant son authenticité et conservé dans un établissement qui la garantit, par exemple les Archives de l'État, peut bien servir de preuve.
- *Service de la recherche.* Les documents d'archives peuvent être utilisés non seulement à des fins administratives et de garantie de droits, mais aussi pour servir la recherche. L'exploitation la plus connue est l'utilisation par la recherche historique des documents ne servant plus aux besoins pratiques de l'organisme producteur et versés aux Archives. Les grands ensembles de données, comme les bases de données géologiques ou astronomiques, serviront de la même manière les travaux scientifiques dans l'avenir.

- *La fonction culturelle* est apparue dans la vie des archives au XX^e siècle seulement. L'objectif des expositions, des visites, des conférences, des publications sur les archives elles-mêmes ou sur les documents qu'elles conservent, est de faire connaître le travail et les fonctions des Archives d'une part et d'attirer l'attention sur le trésor que représentent les documents d'autre part.

Dans toute société, il y a un lien étroit entre les fonctions des archives et les valeurs intrinsèques des documents conservés. Comme on le verra dans le chapitre suivant, les documents d'archives peuvent passer, même en peu de temps, d'une valeur à l'autre au cours de l'histoire, entraînant ainsi des modifications dans la hiérarchie des fonctions exercées.

5.2 Les valeurs des documents⁴⁴

Les documents, au cours de leur cycle de vie⁴⁵ – depuis la production jusqu'à l'archivage définitif ou l'élimination – portent plusieurs sortes de valeurs. Ces valeurs coexistent mais leur importance varie au fil du temps.

5.2.1 *La valeur administrative/pratique*

La valeur administrative/pratique est l'avantage ou le bénéfice (potentiel) que retire ou retirera de l'information contenue dans le document l'utilisateur, en exerçant ses activités. Pour l'organisme producteur, c'est la valeur la plus importante, puisqu'il a produit le document afin de lui être utile dans son activité. Dès que la transaction est close, la valeur administrative/pratique des documents commence à diminuer et la diminution se poursuivra avec l'écoulement du temps. La valeur pratique des documents est fort variable tout comme sa durée. Le dossier d'une petite commande faite à un magasin perdra rapidement sa valeur administrative/pratique après la livraison et sera probablement éliminé au bout d'une année ou deux. Mais la valeur de la recette secrète du concentré de Coca-Cola – d'après la presse seulement quelques membres du personnel ont accès à ce document vieux de cent vingt ans – a augmenté avec le temps parce que la compagnie ne cessait de se développer et son chiffre d'affaires d'augmenter. Mais une telle hausse de la valeur administrative/pratique ne se

⁴⁴ Sur les valeurs des documents voir Lajos KÖRMENDY, „Schriftgutbewertung. Anmerkungen zur aktuellen Diskussion aus ungarischer Sicht” in *Archivalische Zeitschrift*, 79. Band, Böhlau Verlag 1996, pp. 65-78.

⁴⁵ Voir ci-dessus chapitre 2.3 et la *Terminologie archivistique multilingue de l'ICA*. Accessible: <http://www.ciscra.org/mat/termdb/term/1028>

produit qu'exceptionnellement ; la tendance typique est celle de la diminution rapide. Toutefois, on ne peut jamais dire que cette valeur est nulle, même s'il s'agit de documents inutilisés depuis des siècles, car à tout moment peut survenir un événement qui leur rend leur valeur pratique. Nous connaissons plusieurs cas de plans, conservés dans les archives et « oubliés » depuis longtemps, qui furent utilisés pour la reconstruction de bâtiments⁴⁶.

5.2.2 *La valeur probante*

La valeur probante est l'avantage ou le bénéfice (potentiel) de ce que retire ou retirera l'utilisateur de l'information contenue dans le document pour apporter la preuve de son statut juridique ou du bien-fondé de sa position dans une action légale. Bien que cette valeur ait son importance pour l'organisme producteur, relativement peu de documents sont créés à cette fin, parce qu'un document peut servir de preuve contre son producteur aussi. Dans ce cas sa valeur probante devient négative pour le producteur (s'il l'avait su, il ne l'aurait pas produit).

Conformément à l'ordre légal, certains droits peuvent se périmier ou être abrogés, les uns plus tôt, les autres plus tard. Il y a des documents qui perdent leur valeur probante (ou une partie de celle-ci) en quelques minutes : ainsi nous protesterions certainement en vain en présentant la facture du supermarché établie un quart d'heure plus tôt, pour réclamer la monnaie que le caissier ne nous a pas rendue⁴⁷. Par contre, il y a des documents qui ont une valeur pérenne, par exemple les cadastres. Dans la plupart des cas, la loi stipule le délai de prescription pour chaque type d'affaires (par exemple les affaires fiscales). Il s'ensuit que des documents de même type perdent leur valeur en masse et en même temps. S'agissant de l'ensemble des documents, leur valeur probante diminue avec le temps mais ne sera jamais nulle, comme c'est le cas avec la valeur administrative/pratique, car ils peuvent toujours servir de preuve dans une procédure judiciaire. Il y a des cas exceptionnels comme le procès entre Lübeck et Mecklenburg en Allemagne, il y a quelques décennies, où le tribunal accepta comme preuve une charte du XIII^e siècle⁴⁸. Un meilleur exemple est fourni par ce qui s'est passé en Europe de l'Est dans la seconde moitié du XX^e siècle : il s'agit des nationalisations massives sans indemnisation qui eurent lieu après 1945. Le jour de l'entrée en vigueur des lois et des décrets d'expropriation, des milliers et des milliers de certificats de propriété et de

⁴⁶ Voir par exemple *Budapest enciklopédia*, Budapest, 1981, p. 70.

⁴⁷ Mais la même facture peut servir de preuve dans une procédure criminelle, même après plusieurs années.

⁴⁸ Carl HAASE, „Studien zum Kassationsproblem” in *Der Archivar*, November 1975, p. 408.

contrats de vente perdirent leur valeur probante. Après le tournant politique de 1989, les gouvernements démocratiques rétablirent légalement la propriété privée⁴⁹, et celui qui put prouver la confiscation de ses biens dans l'ère communiste eut droit à une certaine indemnisation. C'est dire que les certificats de propriété et les contrats de vente mentionnés ci-dessus sont redevenus, après plusieurs décennies, des documents probants.

5.2.3 La valeur historique

La valeur historique est l'importance que l'utilisateur attribue à l'information contenue dans le document par rapport à la connaissance du passé. Cette valeur indiffère le producteur : il produit les documents à des fins pratiques ou juridiques. À mesure que le temps passe, la quantité des documents d'une période donnée diminue à cause des éliminations programmées et des destructions accidentelles. Parallèlement à cette diminution, la mémoire individuelle et collective s'estompe, en sorte que la masse des sources d'information disponible sur toute époque ne cesse de se réduire. La rareté augmente la valeur historique au sens intellectuel et la valeur du document au sens matériel. Par conséquent, contrairement aux deux valeurs précédentes, la valeur historique augmente avec temps et les documents qui subsistent voient leur importance croître.

Mais cette valeur a beaucoup plus d'éléments subjectifs que les deux autres. En effet, la valeur probante est réglée par la loi et la valeur administrative/pratique, attachée à une activité, ne se maintient que pendant une période courte dont la durée est prévisible. Il existe toutefois des facteurs extérieurs, notamment au niveau de la société, qui déterminent la valeur historique du moment.

- *L'échelle de valeurs de la société.* Un exemple : dans les années 1960 un mouvement spontané s'est engagé simultanément en Amérique du Nord et en Europe pour la recherche des ancêtres. Au cours des décennies suivantes, des dizaines de milliers de personnes quelconques, sans ascendance nobiliaire ou patricienne, découvrirent les sources d'archives, en particulier les registres paroissiaux, les recensements et les minutes notariales. Au tournant du millénaire, ces chercheurs constituèrent déjà la majorité des

⁴⁹ Par exemple, en Hongrie la loi XXV/1991.

utilisateurs des salles de lecture dans plusieurs pays⁵⁰. Tout cela résultait d'un changement sociétal, d'une recherche d'identité et d'une démocratisation des esprits qui – et il faut insister sur ce point – ne dépendait pas de la politique, puisque la tendance fut similaire dans les pays de l'Europe de l'Est longtemps soumis à des régimes dictatoriaux.

- *La politique.* Dans la seconde moitié du XX^e siècle, la vie intellectuelle des pays de l'Europe de l'Est était soumise au marxisme vulgaire devenu idéologie officielle. Sur ordre du pouvoir central, de nombreux programmes de recherche historique furent lancés sur les thèmes préférés de cette idéologie, à savoir l'histoire économique, les révoltes paysannes et le mouvement ouvrier. Il suffit de parcourir les listes des publications historiques parues dans ces pays au cours années 1950 et 1960 pour s'en rendre compte. Abstraction faite de la qualité de ces travaux – il y en avait des bons et des mauvais –, il est incontestable que ces sujets historiques négligés auparavant vinrent au premier plan sous la pression du pouvoir politique. En tout état de cause, la valeur des sources relatives à ces thèmes a considérablement augmenté pour les chercheurs des époques concernées.
- *La science.* Il ressort de l'étude de l'historiographie que la recherche historique ne cesse de s'étendre à de nouveaux terrains. Dans ses débuts, elle ne s'occupait que des actes glorieux des rois et des chefs de guerre, pour s'emparer progressivement des domaines les plus divers, la politique, l'économie, la société et la culture. Qui parlait, il y a cent ans, de l'histoire des mentalités ou du climat, ou bien, il y a cinquante ans, de l'histoire des femmes ? Des modes peuvent surgir qui stimulent des recherches intensives sur tel ou tel thème pendant une période relativement courte, puisque c'est la nature de la mode. Mais l'extension de la recherche à de nouveaux terrains implique la découverte de sources nouvelles dont la valeur augmente en raison de l'intérêt qu'elles suscitent. Dès lors que la recherche s'en désintéresse, leur valeur diminue. Nous ne pouvons guère prévoir dans quelles directions s'étendra et comment se diversifiera la science historique, mais nul doute qu'elle saura, dans l'avenir, découvrir des parcelles inexplorées.

On peut se demander quelles valeurs il faut attribuer aux grands ensembles de données scientifiques, par exemple aux bases de données astronomiques, géologiques, etc. En fait,

⁵⁰ Voir par exemple Gérard ERMISSE, "La communication" in *Gazette des Archives*, 1988 No. 141, p. 200 et Gérard ERMISSE, "L'étude sur les publics des Archives de France" in *Comma*, 2003, 2-3, p. 71.

mettre l'accent sur leur valeur administrative/pratique ou sur leur valeur historique n'est qu'une question d'approche.

À la fin de notre analyse, nous pouvons tirer quelques conclusions. La première est que les trois valeurs (administrative/pratique, probante et historique) sont présentes tout au long du cycle de vie des documents et déterminent conjointement la valeur d'information de ceux-ci. La deuxième est qu'en avançant dans le temps, la valeur historique prend progressivement la place des deux autres valeurs. La troisième est que les valeurs des documents individuels ou des agrégats de documents sont instables. Pour ce qui est de la masse totale des documents d'archives, les tendances valorisante et dévalorisante opèrent indépendamment des cas concrets individuels.

6 Défis d'archives dans le monde analogue et numérique

Dans la première partie de ce livre, nous avons esquissé un condensé de l'évolution de l'archivistique, une succession de défis auxquels devaient répondre les professionnels pour pouvoir sauvegarder les archives et les ouvrir à la recherche. Entre 1840 et 1950, le principe du respect des fonds s'est imposé comme le pivot de la réflexion et de la pratique archivistiques dans tous les pays qui s'étaient dotés d'un système d'archives publiques. La mise en commun du savoir, des expériences et de l'énergie professionnels à l'échelle mondiale a conduit, au cours de la seconde moitié du XX^e siècle, à l'élaboration d'une science archivistique ambitieuse, capable de détecter les priorités émergentes et de les affronter. Grâce à ce capital intellectuel et à la coopération internationale permanente des institutions et des associations professionnelles, l'entrée dans l'ère numérique s'est faite, malgré l'amplitude du changement, sans rupture de continuité.

6.1 L'informatisation, la communication et les documents électroniques

Les trois périodes que l'on peut distinguer dans l'évolution de l'informatique correspondent bien aux trois moments de la dialectique hégélienne : 1) la première (1943-1980) fut l'époque des grands ordinateurs centraux qui gèrent de grandes bases de données (la thèse) ; 2) la deuxième (1981-1995) fut l'époque des PC, quand l'informatique « se démocratisa » et que les données se décentralisèrent sur des millions d'ordinateurs personnels (l'antithèse) ; 3) la troisième (1995-) se caractérise par la révolution d'Internet qui relia les

données décentralisées (la synthèse). En ce qui concerne cette dernière période, on peut la diviser en deux sous-périodes. Dans la deuxième, qui a débuté avec la sortie de l'iPhone (2007), la proportion des outils mobiles (smartphones, laptops) commença à croître rapidement au détriment des outils fixes.

Les périodes et sous-périodes successives produisirent divers types de communication et de documents électroniques. La communication humaine passe actuellement par un changement radical. Il ne s'agit pas seulement de la croissance de la masse des informations disponibles et de la rapidité d'accès en augmentation continue, s'y ajoute une certaine démocratisation : de plus en plus de gens passent sur Internet et y laissent de traces (des millions de pages web privées, des commentaires, des blogs). En fait, on trouve sur le web une empreinte de plus en plus complète de la parole et de l'activité humaines. De par leur fonction et leur style, la messagerie instantanée et les commentaires ressemblent davantage à une conversation (téléphonique) qu'à l'information écrite dont les règles s'étaient développées pendant des siècles. Alors que dans le passé l'archivage de la parole humaine était tenu pour impossible, on s'efforce maintenant de préserver les informations orales numérisées. Mais cela nous confronte à un double dilemme. D'une part, nous sommes incapables de conserver tout, aussi devons-nous éliminer de plus en plus de documents. Or, comme la valeur du document d'archives n'est jamais nulle – il en a été question dans le chapitre 5.2 - nous craignons toujours de détruire des informations de valeur. D'autre part, le fait que la limite entre les secteurs public et privé a tendance à s'effacer pose des questions sérieuses éthiques et juridiques, du fait que l'archivage des « conversations privées » par l'autorité publique viole les droits de l'homme (on verra des exemples ci-dessous.)

L'e-mail et la communication sur les réseaux sociaux ont repris partiellement la fonction de la communication orale quotidienne et, pour cette raison, les règles de l'usage et les caractéristiques formelles de l'écriture ont disparu ou se sont relâchées, ce qui entraîne des conséquences archivistiques. La pratique est devenue très décontractée (« cool ») avec sa « nouvelle orthographe »; les éléments formels, comme le terme d'adresse, manquent souvent ; les annexes sont très variées et colorées (effets, documents sonores et visuels). Il est de plus en plus difficile de tracer la limite entre la vie privée et la vie publique. Au début, les pages web servaient surtout de moyens de communication entre des institutions et des organisations ; aujourd'hui, les personnes privées aussi les utilisent intensément. Les réseaux sociaux ont suivi un sens inverse : ils étaient utilisés d'abord par des personnes privées, des cercles d'amis et des regroupements de la société civile, puis les institutions publiques et les

personnages de la vie publique les ont découverts aussi. Les hommes politiques, par exemple, utilisent souvent Facebook comme moyen de communication pour lancer une action ou faire une déclaration. Dans les administrations, les e-mails privés et officiels sont souvent mêlés dans les comptes de messagerie des employés, ce qui cause un vrai casse-tête pour les records managers et les archivistes quand ils veulent archiver la correspondance électronique.

En ce qui concerne les types de documents, dans l'ère des grands ordinateurs on ne créa pratiquement que des grandes bases de données de structure simple. Dans l'ère des PC, les documents créés par des logiciels de traitement de texte et de tableurs dominèrent le monde numérique. À l'ère de l'Internet, c'est le message électronique, surtout le courriel, qui prédomine, mais on rencontre également sur la toile une grande variété de documents. Il s'agit, en particulier, des produits des nouveaux outils de communication et des réseaux sociaux, qui ont développé des styles et des formes bien différents des documents traditionnels (lettres, télégrammes, etc.). Enfin, il faut mentionner que dans le monde du web la plupart des documents électroniques sont de caractère transitoire ou dynamique et changent rapidement et constamment. C'est autant le cas des documents privés (par exemple la page web de quelqu'un) que des documents publics⁵¹. Les raisons en sont d'une part que, comme il est très facile de créer et de publier des documents électroniques, toute institution publique peut fournir des informations à jour aux demandeurs et gérer les transactions (enregistrements, achats, etc.) sur Internet. D'autre part, les organisations civiles, voire les institutions publiques, font appel de plus en plus à des personnes privées pour créer des documents ou des données. Wikipedia en donne un bon exemple, tout comme les tentatives de fusionner des ensembles de données généalogiques ou encore le projet Domesday de la BBC qui sera évoqué en détail dans le chapitre 16. Les services interactifs et personnalisés du web s'appellent web 2.0.

6.2 La numérisation de la mémoire humaine et l'explosion des données

Dans les cinquante dernières années, l'informatique a progressivement conquis les objets de la mémoire humaine. Au début furent numérisés les grands enregistrements et fichiers, puis vint le tour des documents textuels, tableurs et dessins. On passa ensuite à la numérisation des images fixes, des images en mouvement et du son et, avec les imprimantes en 3D, nous

⁵¹ Anne J. GILLILAN-SWETLAND, "Testing our Truths: Delineating the Parameters of the Authentic Archival Electronic Record" in *American Archivist*, vol. 65 Fall/Winter 2002, pp. 207-208.

sommes parvenus à numériser des objets physiques. Aujourd'hui, nous pouvons dire que l'informatique est capable de créer ou de reproduire la majorité des objets de la mémoire humaine. Il en résulte une croissance effrénée des données électroniques en accélération continue :

- Puisque presque toute communication humaine (texte, image, son) peut être numérisée, de plus en plus d'activités se déplacent dans le monde numérique (correspondance, conversation) et, en même temps, le nombre des producteurs de données croît rapidement. Celui ou celle qui crée un compte de messagerie, un profil Facebook, un album de photos, etc., devient par là même un producteur de données.
- Il est très facile de produire des données numériques, notamment par traitement de texte, enregistrement du son ou photographie. Une partie croissante des données est automatiquement produite par des outils informatiques, comme les données fournies par des instruments de mesures météorologiques ou les métadonnées de transmission des e-mails (informations sur l'expéditeur et le destinataire, etc.).

Bien que le bureau sans papier ne soit pas encore une réalité, ce temps s'approche. À l'ère des PC, la majorité des documents électroniques furent imprimés, ce qui n'est plus le cas aujourd'hui ; nous vivons actuellement dans une période de transition. Pour faire des économies et pour des raisons pratiques, les gouvernements encouragent souvent les administrations à numériser les documents papier reçus et à les intégrer ainsi dans le système informatique. Les citoyens constatent que de plus en plus d'informations et de prestations – y compris des prestations d'État – ne sont accessibles que sur Internet, en forme numérique.

6.3 Les difficultés de la conservation pérenne des documents électroniques

L'histoire de l'usage de l'écriture montre qu'en avançant dans le temps, les supports utilisés (table de pierre, tablette d'argile, papyrus, parchemin, papier [livre], microfilm, disque et bande magnétiques et disque optique) contenaient des informations de plus en plus denses et devenaient aussi de plus en plus vulnérables à cause de cette densité. Avec le microfilm, la densité atteignit un point où l'information n'était lisible qu'à l'aide d'un outil. Les deux forces motrices du passage de l'âge analogique à l'âge numérique, qui avait commencé dans la seconde moitié du XX^e siècle, furent la volonté de pousser à la densification et le souhait d'améliorer la gestion des informations (classement, recherche documentaire). Or, lors de la lecture d'un document numérisé, les outils, à savoir le matériel et le logiciel, ne servent pas à

un simple agrandissement mais à rendre compréhensible l'information qui est codée. Par conséquent, la préservation physique du support de l'information ne suffit plus, il faut conserver également un « dictionnaire de décodage » qui ne sera pas forcément le matériel et le logiciel originels. La conservation authentique des documents dans le long terme, une routine dans le monde analogique, est impossible dans le monde numérique pour des raisons techniques et archivistiques. Les causes techniques sont bien connues et nous y avons déjà fait allusion : 1) pour produire, lire, traiter et conserver les documents électroniques, on a besoin d'un système technique comprenant le matériel, le logiciel et un environnement électronique ; il s'ensuit logiquement que ce système détermine les documents ; 2) puisque, avec le temps, ce système technique change constamment, il est impossible de conserver les documents dans leur état originel.

Tous ceux qui utilisent l'ordinateur depuis longtemps pouvaient constater que les anciens disques et fichiers ne pouvaient être lus par les nouveaux matériels et logiciels, ou ne fonctionnaient pas bien parce que les normes avaient changé. À mesure que grandit l'intervalle, la chance de pouvoir les utiliser se réduit. Il y a vingt ans, l'usage des disquettes souples était général ; les adolescents d'aujourd'hui, tout au plus, en ont entendu parler. Aujourd'hui, de plus en plus d'ordinateurs sont fabriqués sans lecteur de CD/DVD ; quant aux smartphones, qui sont également des ordinateurs, l'espace pour loger un tel support y fait défaut. Seuls des informaticiens spécialistes sont capables de ranimer les fichiers créés par un traitement de texte précoce (de DOS ou des premiers Windows). Du fait de la tombée en désuétude du matériel et du logiciel, les documents électroniques sont en danger cinq à dix ans après leur création.

Le manque comme le trop-plein de normes, aboutissant d'ailleurs au même résultat, posent problème dès à présent. Celui qui ne possède pas le logiciel approprié sera incapable de lire les fichiers qui dépendent de ce logiciel. La standardisation est une exigence dans tous les environnements, dont le monde analogique. Mais là, on considère naturel, ou on ne s'en rend pas compte, que l'œil humain est « standardisé » et c'est ce qui nous met à même de lire sans difficulté les documents analogiques. (Les comprendre, c'est une autre histoire.) Avec les documents électroniques, nous avons pris conscience des risques de l'absence de normes. Les conséquences en sont beaucoup plus graves que dans un environnement technique différent (par exemple l'automobile).

Pour assurer la conservation des documents électroniques authentiques dans le long terme, il y a des conditions archivistiques à respecter. Bien qu'elles soient tout aussi

importantes que les conditions techniques, elles sont moins connues. Leur non-respect peut causer autant de problèmes que le défaut de conditions techniques, parce que l'informatique a reformulé les notions et les règles de la profession. Les principaux changements intervenus dans l'archivistique peuvent être attribués à quatre facteurs : 1) le document électronique n'a pas d'aspect physique ; 2) la structure du document électronique diffère de celle de l'archive papier et dépend de la technologie ; 3) il est facile de créer, modifier, multiplier, combiner et diffuser un document électronique ; 4) pour ces raisons, la quantité des documents électroniques a explosé.

L'objectif de la conservation pérenne des documents électroniques consiste à préserver leurs valeurs réelles et potentielles pour que les archives puissent remplir leurs fonctions. Il ne s'agit pas d'une exigence sectorielle bibliothéconomique ou archivistique, mais d'un intérêt élémentaire de la civilisation humaine, et nous espérons pouvoir le prouver dans notre livre. Dans les chapitres suivants nous exposerons ces questions en détail.

7 Le document d'archives et ses éléments constitutifs

Le document constitue une entité plus complexe que l'information ou la donnée. Pour devenir document, elles doivent répondre à un certain nombre d'exigences. Il y a plusieurs sortes de documents. Pour ce qui concerne le présent ouvrage, seuls importent le concept et la structure du document d'archives, dont la définition généralement acceptée dans la littérature archivistique du dernier demi-siècle figure plus haut (l'ensemble des documents, produits ou reçus par une personne physique ou morale dans l'exercice de leur activité – voir la définition détaillée dans le chapitre 2.3). Peut-être faudrait-il ajouter à cette définition que le document d'archives doit être enregistré. Le document d'archives est constitué de plusieurs éléments en corrélation, à savoir le support, la forme, le contenu, l'identifiant, la structure et le contexte⁵².

7.1 Le support

Dans le monde analogique, nous considérons comme support l'objet physique qui porte et rend visible l'information. Dans le cas des documents papier, les supports sont le papier lui-

⁵² Il y a d'autres articulations, par exemple le projet InterPARES distingue, d'après la diplomatie, quatre éléments principaux: la forme documentaire, les annotations, le support et le contexte. Voir *Requirements for Assessing and Maintaining the Authenticity of Electronic Records, Appendix 2*. Accessible: http://www.interpares.org/display_file.cfm?doc=ip1_template_for_analysis.pdf

même et l'encre ou l'encre d'imprimerie et nous pouvons lire l'information « portée » directement à l'œil nu. Dans le cas des microfilms, le support est également un objet physique, la pellicule ou la microfiche. Si en dernier ressort une loupe peut suffire, pour une lecture confortable on a besoin d'un outil, d'un lecteur de microfilm.

Dans le monde numérique tout cela change radicalement, du fait surtout que le document électronique n'existe pas physiquement. Ce fait est devenu clair pour beaucoup de monde grâce à Internet : le document peut être conservé n'importe où dans le monde sur n'importe quel support (disque dur, CD, clé USB, etc.), en n'importe quelle articulation (dispersé dans plusieurs fichiers). De plus, les bits, qui représentent l'information numérique, ont beau être présents sur un disque dur ou optique ou, éventuellement, dans la mémoire d'un ordinateur, cela ne suffit pas à faire apparaître l'information « portée » ; il faut avoir aussi un système d'outils approprié (matériel, logiciel, environnement électronique)⁵³. Dans le monde numérique, le support du document est étroitement lié au système d'outils, en partie physique, en partie virtuel. Pour compliquer encore les choses, comme nous l'avons vu dans le chapitre 6.3 et grâce au développement de la technologie, le support physique ne cesse de changer et les éléments du document peuvent en être dissociés ; on peut faire une copie identique sur un autre support, le convertir selon diverses normes. Contrairement au monde analogique, pour faire apparaître et conserver le document il n'est pas nécessaire d'avoir recours à un support identique et, comme on le verra, cela est impossible dans le long terme. En bref : l'importance du support comme élément du document a considérablement diminué dans le monde numérique.

7.2 La forme

La forme du document signifie son apparence ; dans le cas d'un document écrit, ses couleurs, ses dimensions, les caractères, les figures, l'en-tête, etc. Dans le monde analogique, ces caractéristiques sont complétées de marques physiques : par exemple, le document est-il en forme de rouleau ou en forme de livre ? Dans le cas des documents analogiques, la forme est inséparable du support ; cela se constate et se comprend automatiquement⁵⁴. Dans le monde numérique, le choix du matériel et du logiciel (par exemple le traitement de texte)

⁵³ La littérature spécialisée – par exemple la Terminologie archivistique multilingue de l'ICA mentionnée ci-dessus ou la norme OAIS – entend par support l'objet physique où les bits sont enregistrés.

⁵⁴ Les documents audiovisuels analogiques sont des exceptions, mais ils jouaient un rôle périphérique dans la vie des archives.

exerce un effet considérable sur la forme : quelles polices seront utilisées, y aura-t-il ou non un en-tête, etc. Quand on lit un document, le matériel et le logiciel font apparaître la forme d'après des spécifications techniques (métadonnées) et, en cas de compatibilité incomplète, si par exemple le traitement de texte du lecteur n'a pas les polices appropriées, la forme changera.

La forme peut être intégrée dans le fichier où se trouve le contenu, mais ces deux éléments peuvent aussi être dissociés et conservés dans des fichiers séparés. Même plusieurs formes (stylesheets) peuvent être liées à un seul contenu ; cela arrive chez les documents XML (voir chapitre 12.1). Étant donné que la forme est séparée du contenu, qu'elle dépend du matériel et du logiciel qui changent par utilisateur et dans le temps, l'importance en diminue dans le monde numérique. Aussi, pour conserver les documents dans le long terme, les archivistes sont-ils contraints d'accepter des compromis et ne pas tenir à la préservation intégrale de la forme. C'est ce qui arrive lorsqu'en convertissant les documents on se contente d'en préserver les caractéristiques essentielles, ce qui signifie souvent le sacrifice de certains éléments d'intérêt secondaire de la forme. L'importance de la forme peut aussi changer par type de document. Alors qu'elle est importante dans le cas d'un document graphique, le type des lettres est souvent indifférent pour faire apparaître les résultats d'une recherche dans une base de données.

7.3 Le contenu

Le contenu est l'information immanente relative au sujet du document. Sa fonction et son importance sont identiques dans le monde analogique et numérique, parce qu'il est dématérialisé dans les deux mondes. Il concorde donc avec la substance du monde numérique. Il convient d'ajouter que la forme, aussi bien que le contexte présenté ci-après, sont en rapport avec le contenu, mais comme ils ne sont pas immanents et ne concernent pas seulement le sujet du document, nous les traitons séparément.

7.4 L'identifiant

L'identifiant a un double rôle : 1) identifier le document individuel, si possible de manière à ne pas devoir le modifier dans l'avenir ; 2) rendre possible de désigner la place précise qu'occupe le document dans le système documentaire en indiquant à quels agrégats documentaires il appartient et quels sont ces agrégats. C'est pour cela que les documents

reçoivent une cote individuelle dans les dépôts d'archives et dans les départements de manuscrits. L'identifiant individuel doit s'insérer dans un système souple, capable d'admettre dans l'avenir des documents de types divers et des identifiants extérieurs. L'identifiant unique et fixe est une exigence fondamentale pour préserver l'identité (l'unicité) et l'authenticité du document.

L'importance des identifiants a considérablement augmenté dans le monde numérique. D'une part, l'intégration des bases de données et des collections numériques de plusieurs institutions sur un portail commun étant devenue une pratique courante, les identifiants non appropriés et non uniques pourraient causer des confusions. D'où les initiatives internationales pour établir les règles des systèmes d'identifiants uniques⁵⁵.

D'autre part, comme nous le verrons dans les chapitres suivants, la structure des documents électroniques est beaucoup plus compliquée que celle des documents analogiques, parce qu'ils contiennent de nombreux identifiants nouveaux, invisibles pour l'utilisateur, qui étaient inconnus dans le monde analogique. Les données d'une grande base de données peuvent se classer dans une centaine de tableaux et le logiciel qui les gère a besoin des identifiants appropriés pour « maintenir l'ordre ». Ces identifiants « invisibles » seront indispensables lors des conversions futures. Les documents (fichiers) électroniques ont aussi des identifiants techniques, qui décrivent leur environnement électronique, par exemple leur format, et ceux-ci sont également nécessaires pour les conversions futures. (D'un autre point de vue, ces identifiants passent pour des contextes.)

Dans la plupart des cas, nous traitons les identifiants en métadonnées et ils peuvent être intégrés, avec le contenu et le format, dans un même fichier, voire dans le nom du fichier, mais ils peuvent être aussi groupés dans un tableau séparé.

7.5 La structure

Étant donné que les documents sont composites dans leur contenu et leur forme, de même que sur les plans logique et informatique, leurs structures peuvent être classées de plusieurs manières, par exemple du point de vue de la diplomatique, de l'archivistique ou de l'informatique. Nous privilégions la structure archivistique, parce que c'est elle qui porte la

⁵⁵ ISADg (voir le chapitre 12.4), parue en 1994, insistait déjà sur l'importance de l'identifiant unique. Dans le monde de l'informatique, il y a plusieurs normes sur les identifiants uniques pour que l'identification d'un objet numérique soit possible même si sa place a été changée. Voir par exemple *l'Universal Resource Names* (URN) ou le *Digital Object Identifiers* (DOI) ou le *Persistent Uniform Resource Locators* (PURL).

cohérence qui est indispensable à la conservation pérenne des documents. Un document a une structure interne et une structure externe et les agrégats de documents sont constitués grâce à cette dernière.

7.5.1 La structure interne

Les informations et les données se structurent en document et cette structure interne fut minutieusement explorée par la diplomatique. Les documents modernes et contemporains ont hérité de nombreux éléments de la structure des actes anciens ; ce thème relève de l'étude des documents courants (records management) et aussi, du moins en partie, de l'archivistique. Il suffit de jeter un coup d'œil sur une lettre traditionnelle pour identifier le destinataire, l'objet, la date et le signataire, c'est-à-dire voir et comprendre la structure interne. Dans le cas de l'envoi d'un e-mail, quand nous remplissons les champs du destinataire ou de l'objet, le logiciel génère automatiquement des métadonnées par lesquelles l'ordinateur du destinataire reconstruira la même structure. Dans le monde papier, nous pouvons visualiser la structure d'un document ou d'un agrégat de documents en mettant en vedette des titres et des sous-titres ; dans le cas d'une page web nous devons instruire expressément l'ordinateur au moyen de balises par exemple.

Les documents électroniques comportent aussi des données techniques qui, dans le monde analogique, n'existaient pas ou n'étaient pas considérées comme importantes. S'agissant du courrier électronique, font partie de ces (méta)données les moments de l'envoi par l'expéditeur et de la réception par le destinataire, ou encore l'itinéraire parcouru par le message, etc. Dans le monde numérique, la structure interne des documents est particulièrement menacée en cas de conversion, quand peuvent disparaître les métadonnées structurelles, en l'absence desquelles la valeur des documents se dégrade fortement.

7.5.2 La structure externe

Les documents d'archives sont liés à d'autres documents : ils se groupent par contenu, par procédure, par fonction ou par forme. La norme internationale de description ISADg (voir chapitre 12.4) propose une structure externe hiérarchisée en six niveaux, appelés de bas en haut la pièce, le dossier qui comprend les documents relatifs à une transaction, la sous-série, la série, le sous-fonds et le fonds. Cette structure correspond en général au plan de classement

originel de l'organisme producteur dont il reflète le fonctionnement avec plus ou moins de précision.

La structure externe ne s'arrête pas au niveau des fonds. Un service d'archives reçoit les versements de nombreux producteurs ; il conserve donc de nombreux fonds qu'il doit intégrer dans une seule grande structure, pour pouvoir établir son système d'enregistrement et de recherche. Nous appelons cette structure au-dessus du niveau des fonds la « macrostructure ». Tout document peut aussi être directement relié à des documents appartenant à d'autres agrégats. Le meilleur exemple est offert par une page web sur laquelle figurent les liens avec d'autres pages web.

Dans la structure externe, les liens peuvent avoir un sens vertical ou horizontal. L'adjectif « vertical » se réfère à une relation hiérarchique, parce qu'en image les rapports entre les niveaux (groupe des fonds, fonds, sous-fonds, etc.) sont représentés par des lignes verticales. L'adjectif « horizontal » se réfère à des relations entre entités de même rang indiquant par exemple quelles séries constituent un sous-fonds ou à quel message répond l'e-mail et à quels documents analogiques il est relié. Les liens entre pages web mentionnés ci-dessus sont considérés comme horizontaux.

Pour garder la valeur informative des documents, il est indispensable de conserver la structure externe. Si nous dispersons les documents d'une série et ne signalons pas leur connexité, même étant intacts, leurs trois valeurs diminuent considérablement car ils ne peuvent pas être attachés à la raison de leur production, à la transaction, à l'organisme ou à la personne qui les a produits. C'est pourquoi il est important de présenter, de documenter et de conserver la structure externe. Dans le monde analogique, les divers enregistrements, descriptions et instruments de recherches servaient à cela.

Dans le monde numérique la situation est plus compliquée parce que les documents électroniques ont davantage de liens externes : les pages web en sont pleines et les e-mails ont des annexes. Le simple fait que, dans le monde numérique, l'espace en trois dimensions se réduit à un espace en deux dimensions présente des inconvénients majeurs qu'on ne peut compenser que par des métadonnées structurelles. Dans les archives papier, la structure externe se perçoit immédiatement, puisque cartons et dossiers se séparent les uns des autres et qu'en faisant des recherches nous pouvons ranger les pièces dans l'ordre logique que nous souhaitons. (Il va de soi que la recherche terminée, l'ordre de conservation sera rétabli.) Une fois numérisés, ces documents, dont chacune des pages forme un fichier, s'alignent dans un ordre rigoureux et séquentiel sur l'écran, qui est beaucoup plus difficile à manipuler. Pour

faciliter la manipulation, on connecte aux images scannées une base de données - c'est-à-dire un ensemble de métadonnées –, qui fait apparaître les liens et les séparations et à l'aide de laquelle on peut aisément manipuler, feuilleter et chercher les documents.

7.6 Le contexte

La littérature archivistique appelle contexte les informations et les données relatives au producteur des documents : son histoire ou sa vie, ses activités, sa compétence, ses fonctions, sa structure organisationnelle (dans le cas d'un organisme) ses relations avec d'autres organismes ou personnes, etc. Fait également partie du contexte tout ce qui entoure le producteur et les documents sur les plans juridique, politique et social⁵⁶. (Si par exemple, la série se rapporte à un événement sociétal important, il faut décrire brièvement l'arrière-plan de cet événement.) Souvent ces éléments contextuels, bien qu'implicitement présents dans les documents, doivent être décelés par l'archiviste qui les complétera parfois à l'aide d'autres sources. Le contexte archivistique est généralement consigné comme métadonnée lors de la description des documents. (Nous en parlerons dans les chapitres 8 et 12.4-12.5.)

Le document électronique est aussi un objet numérique, dont les attributs techniques importants, telles les données sur le format du fichier, sont indispensables autant pour la conservation que pour une utilisation ultérieure. Il importe de décrire avec un soin particulier le contexte technique informatique quand les archives exécutent une opération d'émulation (voir le chapitre 13.2). Dans un tel cas, il convient de documenter de manière particulièrement détaillée le matériel et le logiciel. Dans le monde numérique, l'importance du contexte, en tant qu'élément constitutif du document, a gagné en importance en raison des exigences techniques.

Mentionnons enfin que certains éléments contextuels des documents, qu'ils soient traditionnels ou électroniques, n'appartiennent pas à notre sujet. Il s'agit notamment des corrélations thématiques que nous appelons contexte du chercheur. Ces contextes souvent cachés seront développés par le chercheur plutôt que par l'archiviste. L'art de la recherche

⁵⁶ Il y a un bon résumé sur le contexte archivistique dans l'ISADg. Une partie du contexte est exposé en détail par l'ISAAR (CPF), Norme Internationale sur les notices d'autorité utilisées pour les Archives relatives aux collectivités, aux personnes ou aux familles (Conseil international des archives), 2004. <http://www.ica.org/10241/normes/liste-des-normes.html>. L'approche du groupe InterPARES est différente et il distingue cinq sortes de contexte: juridico-administratif, de provenance, procédural, documentaire, technologique. Voir Luciana DURANTI, Kenneth THIBODEAU, *Interpares 2: Experimental, Interactive and Dynamic Records*, Appendix 2, p. 18. Accessible: http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_appendix_02.pdf

consiste précisément à déceler de tels contextes. Si les contextes archivistiques et techniques sont bien structurés, précis et détaillés, l'utilisateur des documents a toutes les chances de découvrir les contextes de chercheur.

7.7 La séparation des éléments constitutifs des documents électroniques

Dans le monde analogique, on s'accordait à estimer que le document traditionnel était constitué conjointement par les supports (parchemin, papier, encre), la forme et le contenu, formant un ensemble indivisible. La conservation du document signifiait donc la conservation conjointe des trois éléments. Cette conception ne fut entamée que dans une faible mesure par l'expansion des copies analogiques électrostatiques et l'usage des microformes. Le microfilm enregistre 99 % des informations d'un document textuel hors les filigranes et les couleurs. Bien qu'il nécessite peu d'espace de stockage et soit facile à réaliser, le microfilmage de substitution avec élimination des originaux n'a pas su s'imposer, contrairement au microfilmage de sécurité qui préserve les originaux. Le recours restreint au microfilmage de substitution s'explique, outre le coût (voir ci-dessus chapitre 2.4), par la diminution de la valeur probante par rapport aux originaux et la considération que tout changement de support dégrade l'authenticité.

Dans le monde numérique tout cela s'est transformé radicalement. Comme nous l'avons vu, le document s'est dématérialisé, ses éléments constitutifs se sont transmués et ont été complétés par des attributs techniques, leur séparation est devenue possible et leur importance a diminué ou augmenté. Les éléments qui mettent de l'ordre dans ce système compliqué, à savoir l'identifiant, la structure et le contexte, ont vu leur valeur augmenter parce qu'ils sont indispensables pour assurer l'identité, l'intégrité et l'authenticité des documents. Tous ces changements ont rendu les documents encore plus compliqués et vulnérables. La vulnérabilité des fichiers figure parmi les raisons qui rendent si difficile la pérennisation des documents électroniques.

7.8 Le caractère singulier des documents d'archives

Les documents d'archives se distinguent nettement de ceux conservés dans les bibliothèques ou les musées par leur structure hiérarchique complexe et par la singularité du contexte archivistique. Cet arrière-plan complique aussi bien le traitement des documents

d'archives que le travail de recherche. La complexité n'est pas le fruit de l'arbitraire. Elle résulte de la nature du fonds d'archives, le concept fondamental de l'archivistique.

- Le fonds contient des données et des informations, elles-mêmes complexes, relatives à son producteur et à l'environnement de celui-ci en reflétant les relations entre les personnes et transactions concernées.
- Les documents d'un fonds se produisent dans un processus, c'est-à-dire dans le temps ; or, en avançant dans le temps, le producteur et son environnement passent par des changements, que reflètent les documents.

Tant la complexité que le processus ne peuvent être représentés que par relations (structure) et contexte. L'archiviste analyse et décrit les documents en descendant du fonds jusqu'à la pièce, sans perdre de vue leur cohésion, autrement dit leur unité intellectuelle. Les méthodes de la recherche se sont adaptées à ce système. En général, le chercheur atteint l'information qui l'intéresse en suivant des liens verticaux : il s'approche du document recherché pas à pas, passant de niveau à niveau, en restreignant progressivement le champ à explorer. Pour répondre à la demande de renseignement du lecteur, l'archiviste procède en identifiant d'abord l'organisme producteur du fonds, puis le département compétent au sein de cet organisme, à partir des fonctions susceptibles d'avoir été à l'origine de la production de documents portant sur le thème de la recherche. Ce procédé d'accès hiérarchique indirect est à l'opposé de l'accès direct qu'offrent les catalogues des bibliothèques aux noms d'auteurs, aux titres et aux matières. Ceci dit, l'accès direct à l'information, qui se faisait traditionnellement au moyen des index, est singulièrement facilité dans l'ère informatique par des bases de données et des inventaires numérisés. Mais comme les termes désignant les sujets recherchés dépendent d'un choix subjectif, la démarche hiérarchique gardera son importance dans l'avenir.

8 Les métadonnées

D'après une définition courante, les métadonnées sont des données sur des données (documents). Le monde analogique connaissait aussi les métadonnées, appelées catalogue dans les bibliothèques ou instrument de recherche dans les archives. Dans le monde numérique, leur rôle s'est considérablement accru parce que, dans la plupart des cas, les éléments constitutifs du document apparaissent comme des métadonnées : on se réfère au support par métadonnées techniques, l'identifiant est une métadonnée et naturellement on

décrit par métadonnées la structure et le contexte. Elles sont accolées aux documents électroniques natifs comme aux images scannées. Elles ont une importance particulière dans le cas de ces dernières et des fichiers sonores, car faute de pouvoir repérer des mots et des expressions comme dans les documents textuels, toute recherche et tout enregistrement doivent passer par les métadonnées. Elles rendent des services similaires dans le cas des grandes bases de données scientifiques car c'est par elles que les producteurs décrivent comment les données ont été collectées, comment les instruments ont été calibrés et quels paramètres ont été employés pendant le processus. Les recherches scientifiques futures reposent sur elles⁵⁷.

Caractéristique importante des métadonnées : bien qu'elles puissent être rendues lisibles à l'œil humain, leur fonction principale consiste à pouvoir être traitées, interrogées et manipulées par ordinateur⁵⁸.

Les métadonnées peuvent être classifiées de plusieurs manières. Les trois groupes établis par le *Making of America II Testbed Project* dans les années 1990 sont généralement admis dans la littérature spécialisée⁵⁹. Elles peuvent être :

- descriptives : ce sont les données nécessaires pour l'identification des documents et pour mettre à jour les informations relatives à leur contenu ; elles sont donc du même type que les instruments de recherche et les bases de données ;
- administratives : ce sont les données nécessaires pour l'utilisation du document électronique, telles les spécifications techniques, les données sur les conditions de la production ou les restrictions de communicabilité ;
- structurelles : ce sont les données relatives aux connexions techniques et archivistiques entre les composants des objets numériques.

Les métadonnées peuvent également être classifiées selon le cycle de vie des documents électroniques⁶⁰. D'après cette classification les métadonnées peuvent être

⁵⁷ Michael DAY, *Instalment on "Metadata"* (DCC Digital Curation Manual 2005) p. 10.

⁵⁸ Michael DAY, *op. cit.* p. 12.

⁵⁹ Voir par exemple Bernard J. HURLEY, John PRICE-WILKIN, Merrilee PROFFITT, Howard BESSER, *The Making of America II Testbed Project: A Digital Library Model* (1999). Accessible: <http://www.clir.org/pubs/reports/pub87/pub87.pdf>

⁶⁰ Andrew MCDONALDS (éd.), *Les archives électroniques, Manuel à l'usage des archivistes* (ICA Studies 2005, p. 56)

- d'archives courantes : en général, elles sont générées automatiquement par le système de gestion électronique des documents (GED) en même temps que les documents ; elles peuvent contenir le nom de l'auteur, la date de la création, le titre, le niveau de la restriction de communicabilité, les mots-clés du document, etc. ;
- d'archives : en général on les donne après la clôture des dossiers soit chez le producteur, soit après le versement aux archives ; elles peuvent contenir la date du dernier contrôle des documents, le nom de l'organisme producteur, etc. ;
- techniques : données nécessaires pour la reconnaissance et la gestion techniques des documents électroniques ; elles peuvent être générées chez les producteurs ou aux archives ; il semble souhaitable qu'elles se génèrent automatiquement chez les producteurs, mais dans les archives uniquement lors d'une conversion (migration).

Les métadonnées peuvent être intégrées dans le fichier du document comme c'est le cas avec le format PDF/A, « enveloppées » avec lui ou placées dans un tableau séparé.

Techniquement il est recommandé de les garder ensemble pour réduire le risque d'erreur. Les métadonnées placées dans un tableau séparé forment une base de données, qu'il est souhaitable de construire en respectant les règles archivistiques. Telles sont par exemple les bases de données qui décrivent le contenu et le contexte des documents à l'intention des chercheurs. Les métadonnées techniques étant indispensables à l'entretien et à la migration (voir le chapitre 13.1), il faut surveiller quels sont et où se trouvent les fichiers ayant des formats menacés d'obsolescence.

Nous avons vu dans le chapitre 7 que les identifiants, les données qui décrivent la forme et la structure ainsi que le contexte constituent des métadonnées. Elles font donc partie des documents électroniques, qu'elles organisent en système et qui sans elles formeraient des accumulations inutilisables d'objets numériques. Plus grande est l'accumulation et plus important est le rôle que jouent des métadonnées détaillées de bonne qualité. Les métadonnées ont leur logique et leur système qui nécessitent un projet préalable.

La planification, l'interopérabilité et l'échange de données sont grandement facilités par des métadonnées standardisées. L'interopérabilité, c'est-à-dire la faculté d'opérer avec différents matériels et logiciels, réduit la dépendance des documents des produits informatiques. Quant à l'échangeabilité des données, c'est elle qui permet de les exporter, de les importer, de les fusionner et de les intégrer. Des normes internationalement acceptées ont

rapidement suivi les normes nationales de métadonnées. En voici quelques-unes brièvement présentées. (Les normes seront évoquées en détail dans le chapitre 12.)

Le *Dublin Core*, le « noyau de Dublin », a reçu son nom d'une petite ville de l'Ohio où se réunit en 1995 le groupe de travail qui l'a élaboré. Le terme « noyau » indique que seules sont contenues les métadonnées descriptives les plus importantes⁶¹. Le groupe a d'abord élaboré un *Metadata Element Set* comprenant 15 éléments ; actuellement c'est le niveau simple. Ensuite, le groupe a augmenté le nombre des éléments et les a affinés ; aujourd'hui c'est le niveau qualifié. Les éléments du niveau simple sont les suivants : contributeur, couverture spatiale ou temporelle, créateur, date, description, format, identifiant, langue, éditeur, relation, droits, source, sujet, titre, type de ressource.

Le *Dublin Core* n'est pas trop prisé dans la communauté des archivistes parce qu'il ne tient pas compte des spécificités de l'Archive exposées dans les chapitres précédents. Il ne peut donc servir pour décrire les documents d'archives qu'au prix d'un fort compromis. Ses notions ne correspondent pas toujours à celles qu'utilisent les archives et, défaut plus important encore, il ne prend pas en compte leur structure hiérarchisée.

La communauté internationale des archivistes élaborera relativement tôt ses normes de métadonnées descriptives. Les normes mentionnées ci-dessous seront présentées plus en détail dans les chapitres 12.1, 12.4-12.5. ISADg fut la première norme publiée par le Conseil international des Archives (ICA) en 1994. Elle est certainement le document le plus réussi de l'histoire de l'organisation. Deux ans plus tard, l'ICA publia une norme complémentaire de l'ISADg, intitulée ISAAR(CPF) (Norme internationale sur les notices d'autorité utilisées pour les archives relatives aux collectivités, aux personnes ou aux familles). L'EAD (Description archivistique encodée ; voir le chapitre 12.5), la troisième norme appelée à une large utilisation, fut mise en chantier en 1993 à l'Université de Berkeley en Californie. Sa première version parut en 1998, la deuxième en 2002. L'EAD fut élaborée expressément pour l'application informatique dans le langage descriptif XML (voir le chapitre 12.1). De ce point de vue, l'ISADg et l'ISAAR sont neutres ; elles n'emploient pas de notions informatiques. Aussi peuvent-elles être utilisées pour préparer des instruments de recherche traditionnels. Si au début, l'influence de la bibliothéconomie américaine fut sensible, les auteurs s'efforcèrent d'harmoniser la deuxième version d'EAD avec l'ISADg. L'EAC (Contexte archivistique encodé), la version d'XML de l'ISAAR(CPF) ne parut qu'en 2011.

⁶¹ Dublin Core Metadata Initiative. Accessible: <http://dublincore.org/>

Pour terminer, nous allons mentionner deux normes internationales qui concernent les métadonnées de conservation. Le dictionnaire de données PREMIS (*PREservation Metadata : Implementation Strategies*), élaboré par un groupe de travail, devint de fait la norme internationale en la matière⁶² : « Le dictionnaire est organisé autour d'un modèle de données se composant de cinq entités liées à la procédure de conservation : entité intellectuelle, objet, événement, agent et droit. »⁶³ Le PREMIS utilise plusieurs autres normes comme des schémas XML et s'efforce de se conformer au modèle d'information l'OAIS⁶⁴.

Les métadonnées de conservation enregistrées doivent être « enveloppées » ou « emballées » et liées à d'autres métadonnées descriptives, administratives et structurelles. La norme METS (*Metadata Encoding and Transmission Standard*)⁶⁵ définit un format d'emballage, correspondant à un schéma XML ou un paquet d'information de l'OAIS. Un document METS comprend sept sections : en-tête, métadonnées descriptives, métadonnées administratives, liste des fichiers, carte de structure, liens structurels et comportement (un outil pour lier le contenu numérique aux applications).

9 L'intégrité et les caractéristiques essentielles des documents

Intégrité signifie complétude et cohérence. S'il est relativement facile de vérifier la complétude et la cohérence d'un document papier, il en va autrement des documents électroniques plus compliqués dont les éléments constitutifs se séparent. Il est primordial d'assurer, malgré la difficulté, leur intégrité pour préserver leur authenticité et leur utilisabilité. L'intégrité doit être maintenue non seulement au niveau des documents individuels mais aussi à celui des agrégats de documents.

La littérature spécialisée n'est pas unanime au sujet de l'intégrité du document électronique. Plusieurs auteurs⁶⁶ rattachent l'intégrité à la fixité. La fixité est la « Qualité d'un

⁶² PREMIS Data Dictionary for Preservation Metadata, version 2.2, July 2012. Accessible: <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>

⁶³ *Ibid.* p. 2.

⁶⁴ Voir le chapitre 12.3.

⁶⁵ Accessible: <http://www.loc.gov/standards/mets/>

⁶⁶ Par ex. Andrew MCDONALDS (éd.), *Les archives électroniques, Manuel à l'usage des archivistes* (ICA Studies 2005) p. 11 ou Michael DAY, *Instalment on "Metadata"* (DCC Digital Curation Manual 2005) p. 7.

document d'archives qui garantit une forme fixe et un contenu stable. »⁶⁷ Ce rattachement n'est pas heureux. D'une part, pour bien garder la fixité, pour que l'aspect du document ne change en rien, il faudrait conserver le logiciel originel et, d'autre part, la condition est inapplicable aux documents dynamiques qui changent constamment. L'intégrité ne signifie pas que les fichiers et les métadonnées demeurent intacts car dans la pratique pareille stabilité ne peut être atteinte.

Il est beaucoup plus judicieux de considérer l'intégrité comme une complétude conservant tous les éléments essentiels du document. Certes, le support d'un document électronique, disque dur ou CD, peut changer, la forme peut changer partiellement, mais il faut conserver inaltérés le contenu, l'identifiant, la structure et le contexte, parce qu'ils portent la cohérence du document. Les structures interne et externe des documents sont particulièrement menacées en cas de conversion, quand des métadonnées peuvent disparaître, tels les destinataires des e-mails dont l'absence diminue fort la valeur et l'authenticité des documents. Plus les structures sont compliquées et plus la conservation et la reproduction de l'intégrité des documents deviennent difficiles. Les bases de données possèdent la structure la plus compliquée, bien que l'utilisateur ne réalise pas que, par exemple, la liste comportant des données provenant de plusieurs tableaux qui paraît sur l'écran est organisée par l'ordinateur.

Le projet InterPARES, qui a traité d'une manière particulièrement approfondie la question de l'intégrité, donne la définition suivante : « L'intégrité d'un document d'archives se réfère à sa complétude et à sa condition intacte : un document est intègre quand il est complet et inaltéré dans ses aspects essentiels. » Pour nous c'est l'expression « aspect essentiel » qui est important : « Ceci implique que son intégrité physique, tel que le nombre exact des chaînes binaires, peut être modifiée, à condition que l'organisation du contenu, les annotations nécessaires, ainsi que les éléments de la forme documentaire restent les mêmes. »⁶⁸

Selon les Archives nationales d'Australie, il faut conserver l'essence du document (record's essence). Dans le cas d'un document rédigé par traitement de texte, l'essence comprend le contenu, la composition et les graphiques, etc., mais n'y appartiennent pas les barres d'outils, les boutons et les couleurs. Les archivistes australiens reconnaissent que « la

⁶⁷ *Lignes directrices à l'intention des créateurs*, Projet InterPARES 2, p. 4. Accessible: http://www.interpares.org/ip2/display_file.cfm?doc=ip2_creator_guidelines_booklet_french.pdf

⁶⁸ Conditions requises pour évaluer et maintenir l'authenticité des documents d'archives électroniques. InterPARES, p. 3. Il est à mentionner que nos opinions sur la forme documentaire divergent.

définition de l'essence des documents ne relève pas de la science et elle admet des subjectivités ». Ils déterminent l'essence par type de document (e-mails, documents par traitement de texte, etc.), pour que la vérification puisse se faire par un mécanisme formel⁶⁹.

La conception australienne se base sur l'archivistique ; d'autres auteurs mettent l'accent plutôt sur les critères techniques ou la diplomatique des actes contemporains. L'essence varie par type de document aussi bien que par type de fichier ; les spécifications d'un fichier texte diffèrent de celles d'un fichier vidéo. Plusieurs projets, tels l'InsPECT et le PLANETS⁷⁰, ont été lancés pour définir les propriétés significatives de divers objets numériques, indispensables pour en assurer l'intégrité. Cela peut servir de base pour développer des logiciels qui vérifient automatiquement l'existence de ces propriétés après une conversion. L'InsPECT attribue aux textes structurés trente-sept propriétés contenant des informations sur le contenu (titre, producteur, mots-clés, etc.), la forme (saut de ligne, souligné, gras, etc.) et d'autres données (boutons, éléments de liste, etc.)⁷¹.

La littérature spécialisée s'intéresse surtout à l'intégrité des documents individuels, c'est-à-dire à la structure interne, et s'occupe peu de l'intégrité des agrégats de documents, c'est-à-dire de la structure externe. C'est regrettable parce que le document d'archives particulier ne se suffit pas ; son sens et sa valeur ne ressortent qu'en combinaison avec d'autres documents ; il faut donc conserver l'intégrité extérieure aussi.

L'intégrité extérieure doit être assurée d'une manière différente de l'intégrité interne. La majorité des documents produits, les analogiques aussi bien que les électroniques, seront éliminés et cela met en jeu leur intégrité. Mais si le tri et l'élimination se font conformément au plan de classement et aux méthodes professionnelles, en coupant le moins possible de liens et en donnant des précisions sur les critères de l'élimination, l'authenticité et l'intégrité externe seront assurées. Il faut d'ailleurs se résigner à ce que la sauvegarde de l'intégrité externe de certains types de documents soit impossible. C'est le cas des pages web, à partir desquelles, à l'aide des liens énumérés, on peut sauter à d'autres pages. D'une part, les sites changent souvent et, d'autre part, ces pages ont elles-mêmes des liens avec de nouvelles

⁶⁹ Helen HESLOP, Simon DAVIS, Andrew WILSON, *An Approach to the Preservation of Digital Records*, National Archives of Australia, December 2002, pp. 13-15.

⁷⁰ <http://www.planets-project.eu/>

⁷¹ Stephen GRACE, Gareth KNIGHT, Lynne MONTAGUE, *InSPECT Final Report, 2009*, p. 20. Accessible: <http://www.significantproperties.org.uk/inspect-finalreport.pdf>

pages, en sorte que si on les suivait, le nombre des documents à archiver augmenterait en progression géométrique.

Les bases de données peuvent susciter des soucis particuliers, non seulement à cause de leur structure compliquée, mais aussi parce que 1) elles sont dynamiques, leur teneur change rapidement ; 2) celles qui fonctionnent longtemps peuvent changer en structure et aussi en fonction ; 3) les bases de données sont caractérisées par la présence d'enregistrements temporaires, appelés parfois enregistrements transactionnels, produits par exemple par les utilisateurs au cours de la recherche ou de filtrage. Comme il n'existe pas de solution universelle pour ces problèmes, il faut accepter des compromis.

Une des solutions consiste à sauvegarder des instantanés, autrement dit à archiver le contenu d'une base de données à un moment donné. Cette solution présente l'inconvénient de créer une empreinte statique de données dynamiques, qu'il faut refaire de temps en temps. Dans le long terme et avec de nombreuses bases de données, elle exigera un espace de stockage énorme. Une sauvegarde incrémentale, qui n'archive que les enregistrements modifiés, demande moins d'espace et rendra possible de reconstruire avec précision les changements. Toutefois, elle n'aidera guère si la fonction et/ou la structure de la base de données a/ont été considérablement modifiée(s). Il ne vaut pas la peine de conserver les enregistrements temporaires ; par contre la conservation des fonctions et des outils qui rendent possible leur création est souhaitable⁷².

Nous avons vu que la sauvegarde de l'intégrité suppose la conservation des liens entre les documents groupés. Les organismes producteurs collectent souvent séparément les e-mails bien que dans la plupart des cas ils portent sur des transactions et forment un dossier avec d'autres documents (lettres, mémos, etc.) de la même affaire. Les organismes produisent pour une même transaction des documents papier et des documents électroniques, appelés hybrides, dont la connexion est maintenue par le système d'archivage. Les archives accueillent de nombreux documents hybrides et elles-mêmes microfilment et numérisent leurs documents analogiques. Aux archives, l'intégrité extérieure des documents doit être

⁷² Sur les questions ci-dessus voir par ex. Luciana DURANTI, Kenneth THIBODEAU, "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES" in *Archival Science* 6, 1: p. 13-68. (2006) ou Malcom TODD, *File Formats for Preservation* (Technology Watch Report), Digital Preservation Coalition (2009), p. 27-29.

documentée par des enregistrements et des instruments de recherche que préparent les archivistes lorsqu'ils décrivent les documents⁷³.

10 L'authenticité des documents

L'authenticité pose des problèmes, même dans le cas des copies analogiques et malgré le fait que le microfilm, par exemple, soit infalsifiable en ce sens qu'il est impossible de manipuler le texte à l'intérieur d'une vue. L'authenticité a des degrés divers ; un contrat de vente obéit à des règles strictes de contenu et de forme. En Hongrie, la signature ne suffit pas ; un tel contrat n'est considéré authentique que si chaque page est paraphée par le vendeur, l'acheteur et tous les témoins. En revanche, aucune règle ne s'applique aux mémorandums, qui se trouvent en nombre dans toutes les archives courantes. Il est cependant recommandé de les signer car autrement leur authenticité peut être contestée. Au demeurant, un pourcentage non négligeable des documents n'est pas pourvu de signature, sans que ce fait ne rende leur authenticité douteuse. Il existe des techniques bien connues comme l'estampillage et la cotation qui renforcent l'authenticité, mais d'autres facteurs peuvent aussi agir dans le même sens, comme un règlement précis de la gestion des documents, le plan de classement, le registre de distribution, etc.

L'authenticité des documents électroniques est mise en doute en premier lieu parce qu'ils sont facilement altérables sans que la manipulation ne laisse de traces : nous pouvons, par exemple, changer une phrase dans un fichier texte, modifier la couleur d'une image ou effacer le bruit de fond d'un enregistrement sonore. De plus, comme indiqué plus haut, les éléments constitutifs du document, déjà très vulnérables, se dissocient facilement, le support change souvent et la forme peut se modifier lors d'une conversion.

L'authenticité et la force probante ne sont pas des concepts identiques ; la première a une signification plus large, puisqu'elle implique que le public, notamment les chercheurs, reconnaissent l'authenticité du document. D'après une définition courante, un document est authentique s'il est bien ce qu'il prétend être. L'authenticité est souvent liée à d'autres notions comme la fiabilité (la capacité d'un document de servir de preuve fiable), l'identité (les attributs qui le caractérisent et le distinguent des autres documents), l'intégrité ou la fixité.

⁷³ Voir aussi les chapitres 12.4 et 15.

Le sujet de l'authenticité a été examiné dans une série de projets par le groupe InterPARES, d'après lequel l'authenticité doit être assurée à deux niveaux⁷⁴ :

- 1) Au niveau des documents : par leur identité et leur intégrité.
- 2) Au niveau du créateur et du conservateur des documents
 - par les privilèges d'accès pour la création, la modification, l'annotation et l'élimination ;
 - par les procédures de protection contre la perte et l'altération des documents (le support technologique étant important pour cela) ;
 - par la forme documentaire associée à chaque procédure ;
 - par l'élimination ou le transfert de la documentation pertinente aux archives.

Le manuel de l'ICA traite ensemble les exigences d'authenticité, d'identité, d'intégrité, de fiabilité et de durabilité les documents. Ces exigences peuvent être respectées si

- La création des documents, l'accès à l'information et l'altération ou l'effacement des données se font sur la base d'un système de compétences approprié et où les procédures sont contrôlées et documentées dans le journal des activités du système.
- Les documents ont un identifiant unique.
- Tous les documents sont munis des métadonnées appropriées qui assurent un lien entre ceux-ci et les opérations ainsi qu'avec d'autres documents liés.
- Les métadonnées décrivent tout ce qui est arrivé au document depuis sa création (par exemple le changement du format).
- Les documents sont systématiquement contrôlés pour que la conversion nécessaire soit faite à temps.
- Toutes les données peuvent être exportées de façon automatique dans des formats ouverts sans perte essentielle du contenu, de la structure et du contexte des documents.
- Une documentation appropriée décrit les contextes et les structures des documents.
- Les documents et les métadonnées sont bien protégés contre la détérioration et la disparition⁷⁵.

⁷⁴ *Conditions requises pour évaluer et maintenir l'authenticité des documents d'archives électroniques.* InterPARES, pp. 6-9.

L'énumération ci-dessus couvre une série de procédures techniques et archivistiques qui sont nécessaires pour garantir l'authenticité. En font partie la somme de contrôle (checksum), la journalisation des événements qui se fait indépendamment de l'utilisateur qui ne peut l'altérer, l'enregistrement des métadonnées et la signature électronique. De cette dernière, il vaut la peine de parler en détail.

Il est bien connu que c'est la signature électronique ou numérique à deux clés qui s'est propagée dans le monde entier. Chaque personne autorisée a deux clés : une clé d'écriture (chiffrement) avec laquelle le détenteur signe le document et une clé de lecture (déchiffrement) avec laquelle on peut contrôler si la signature a été apposée par la personne autorisée. Les deux clés sont des logiciels, la clé privée a son propre matériel, un token ou une carte à puce aussi.

La clé d'écriture analyse le document à l'aide d'un code strictement confidentiel variant par titulaire puis, à l'issue de l'analyse, génère un code qu'elle attache au document (elle signe). Comme chaque document est différent – à l'exception des fichiers copiés – la « signature » différera par document.

La clé privée chiffre le document écrit, la clé publique le déchiffre. Au cours du déchiffrement, la clé publique contrôle non seulement si le document a été effectivement créé par la personne qui possède la clé privée, mais aussi si le document a été altéré après la signature. La signature est authentique si la réponse est oui à la première question et non à la deuxième.

La signature électronique n'est ni indéchiffrable, ni éternelle. D'après certaines estimations, grâce au développement rapide de l'informatique, les algorithmes de chiffrement seront déchiffrables dans cinq à dix ans. Il faut donc renouveler les clés et rechiffrer périodiquement les documents signés. En ce qui concerne les archives, elles sont incapables de contrôler la validité des documents signés mais, tout bien réfléchi, ce n'est pas leur tâche. Elles assument seulement la responsabilité de préserver l'authenticité des documents à partir de leur versement, en garantissant des conditions d'accueil, de gestion et de stockage appropriées. Cela vaut même pour les documents électroniquement signés qui seront convertis avec les autres documents, donc altérés. Il convient tout de même de mentionner que la conservation des documents chiffrés implique un risque d'accessibilité.

⁷⁵ Andrew MCDONALDS (éd.), *Les archives électroniques, Manuel à l'usage des archivistes* (ICA Studies 2005), pp. 30-31. L'énumération est basée sur la norme ISO 15489-1.

Des politiques différentes peuvent être suivies en la matière. Le Public Record Office de Victoria (Australie) « enveloppe » après conversion les documents avec les métadonnées, puis signe électroniquement les « enveloppes » et répète ces opérations après chaque altération ultérieure⁷⁶. InterPARES recommande de retirer les signatures électroniques et, comme les Archives nationales des États-Unis et d’Australie, de conserver les métadonnées sur celles-ci, documentant ainsi que la signature électronique a effectivement eu lieu⁷⁷.

L’authenticité, la fiabilité et l’intégrité d’un document ou d’un agrégat de documents étant assurées, leurs valeurs administrative/pratique, probante et historique peuvent s’affirmer en fonction de leur cycle de vie (voir ci-dessus chapitre 2.3). Si les valeurs se détériorent plus ou moins gravement, les documents ne pourront fournir de l’assistance aux activités du producteur, servir de preuve dans un contentieux, et n’offriront pas la fiabilité dont a besoin un géologue ou un historien.

11 Les formats de sauvegarde des fichiers

La conservation pérenne dépend avant tout des formats de fichier. Le format de fichier est l’organisation spécifique et préétablie des données encodées dans le fichier. La plupart des logiciels sont développés pour lire et traiter un ou plusieurs formats et ne peuvent lire que partiellement les fichiers d’autres formats.

Un fichier peut couvrir plusieurs formats : ainsi les fichiers XML et SGML (voir chapitre 12.1), qui sont en principe des fichiers texte, peuvent encapsuler des fichiers d’autres formats aussi.

Les formats sont des produits de l’industrie informatique. Il en existe beaucoup et (bien qu’à un moindre degré que les logiciels et les matériels) ils sont changeants. Dans la plupart des cas, on a besoin de formats différents pour les images fixes, les vidéos, les documents sonores, les textes, etc. De nouveaux formats apparaissent et d’anciens disparaissent, les existants sont développés, et un format peut avoir plusieurs versions. Tout cela joue contre la conservation pérenne. Il y a des différences de qualité entre les formats, l’un reproduit mieux le son ou l’image que l’autre, ils peuvent intégrer ou ne pas intégrer les métadonnées, demander plus ou moins d’espace et être complexes ou simples.

⁷⁶ Michael DAY, *Instalment on "Metadata"* (DCC Digital Curation Manual 2005) p. 19.

⁷⁷ *Lignes directrices à l’intention des créateurs*, Projet InterPARES 2, p. 7 et <http://www.archivesdefrance.culture.gouv.fr/static/1051>, p. 3.

Dans les années 1990, il était généralement admis que seuls certains formats servaient à la conservation pérenne. Les avis étaient naturellement partagés quant aux formats mais les institutions patrimoniales devaient s'en tenir aux formats « canonisés ». Deux décennies se sont écoulées depuis et cette approche a changé pour deux raisons. D'une part, certains formats canonisés se sont révélés insuffisants à plusieurs égards ; d'autre part, il est devenu manifeste que les formats devaient être choisis conformément à la fonction et à la stratégie de l'institution et pouvaient donc différer d'une institution à l'autre. Dans les années 1990, la Bibliothèque nationale des Pays-Bas préférait le format PDF parce qu'il permettait de bien représenter les documents entrants. Au fil du temps, la bibliothèque a reçu des documents de plus en plus variés ; elle a donc révisé sa politique et maintenant elle accepte de nombreux formats⁷⁸. Il y a peu, la British Library a converti 80 téraoctets de journaux numérisés sous format TIFF en format JPEG2000, réduisant ainsi considérablement le coût du stockage⁷⁹.

Dans le long terme, les archives ne pourront certainement pas accueillir trop de formats car, comme les matériels et logiciels, ils peuvent tomber en obsolescence. Mais la réduction du nombre de formats, une sorte de normalisation au niveau de l'institution, débouche sur un dilemme. Si on accepte trop de formats, on se heurtera tôt ou tard à des difficultés d'usage. Si on en réduit trop le nombre, de graves problèmes de qualité peuvent se poser, dès à présent, lors des conversions.

La limitation du nombre de formats peut être réalisée de trois manières :

1. Le système de gestion des documents électroniques du producteur ne génère que les formats demandés – des règles juridiques peuvent imposer cette solution.
2. Le producteur travaille avec d'autres formats, mais les convertit et verse aux archives les formats demandés – cette solution suppose aussi une contrainte juridique.
3. Le producteur verse tous les formats qu'il produit aux archives où est faite la conversion.

Du point de vue des archives, la première solution est la plus avantageuse parce qu'elles reçoivent les documents tels qu'ils étaient produits. Les solutions N° 2 et 3 présentent l'inconvénient de la conversion causant inévitablement des pertes d'information. La solution N° 3 est la plus défavorable car elle met à la charge des archives les travaux de conversion souvent compliqués et d'un coût élevé. Lors des migrations ultérieures, les archives

⁷⁸ Judith ROG, Caroline VAN WIJK, *Evaluating File Formats for Long-term Preservation*, p. 1. Accessible: http://www.kb.nl/sites/default/files/docs/KB_file_format_evaluation_method_27022008.pdf

⁷⁹ *Preservation and Long Term Access via NETWORKED SERVICES*, Planets Project, p. 14. Accessible: <http://www.planets-project.eu/>

affronteront la tâche plus simple de convertir (voir le chapitre 13.1) un ensemble relativement homogène de formats.

Toute institution qui entend conserver des documents électroniques dans le long terme doit élaborer sa politique de formats. Certaines institutions ne formulent leurs exigences qu'en des termes généraux, d'autres précisent quels sont les formats qu'elles peuvent accepter. Dans la littérature spécialisée, nous pouvons lire beaucoup d'exigences et de points de vue ; sont généralement admis les critères suivants⁸⁰ :

- *La diffusion du format.* Les formats largement adoptés présentent l'avantage de pouvoir être utilisés par beaucoup de logiciels et les convertisseurs disponibles sont en général meilleurs.
- *L'indépendance du format* de tout logiciel et de tout matériel. L'indépendance est un avantage évident.
- *La publication des spécifications du format.* Si les spécifications sont bien documentées et publiées, c'est un avantage parce qu'on peut développer de meilleures applications.
- *La transparence du format.* La compression, l'encapsulation, le chiffrement (signature électronique comprise) peuvent empêcher l'identification et le contrôle automatique des attributs du format.
- *Les métadonnées du format.* La richesse en métadonnées et la facilité de leur exploitation sont des avantages évidents.

Outre les critères ci-dessus, un grand nombre d'experts insistent sur ce que le format

- doit être simple – à condition qu'il soit capable de bien présenter le contenu et de résister aux erreurs ; il doit posséder la faculté de détecter et de signaler automatiquement les éventuels problèmes ou erreurs de conversion ;
- doit assurer la compatibilité ascendante s'il appartient à un cycle de développement ;
- si possible, ne soit pas propriété intellectuelle protégée car cela empêche l'usage libre et peut coûter cher.

Les critères (positifs et négatifs) peuvent se combiner dans le cas de certains formats. Le format PDF, propriétaire et publié, est très répandu ; le fichier propriétaire DOC de Microsoft

⁸⁰ Malcom TODD, *File Formats for Preservation* (Technology Watch Report), Digital Preservation Coalition (2009), pp. 13-15.

est aussi répandu, mais n'est pas publié. Les formats peuvent être complexes ou simples, structurés ou non structurés.

Les institutions qui conservent des dizaines de millions de documents doivent tenir compte du besoin d'espace de stockage. Certains fichiers, surtout les images et les images en mouvement, sont fort volumineux. Pour réduire l'espace de stockage et faciliter la télétransmission, on a développé de nombreux algorithmes de compression qui peuvent être classés en deux groupes : 1) les compressions sans perte, comme par exemple la LZW (Lemple-Zif-Welch) ou la RLE (Run Length Encoding), qui réduisent le volume des images TIFF ; 2) les compressions avec perte, dont la plus connue est la JPEG (Joint Photographics Experts Group). Les compressions sans perte réduisent moins le volume mais, comme leur appellation l'indique, la qualité du fichier ne change pas. Quant aux compressions avec perte, il faut payer l'efficacité de la réduction avec la dégradation de la qualité ; par exemple, on perd les détails fins d'une image. Les compressions répétées peuvent causer des pertes considérables.

Dans les années 1990, les institutions patrimoniales furent très réservées à l'égard de la compression, mais depuis lors l'explosion du nombre de documents électroniques imposa un compromis. Les Archives nationales du Royaume-Uni acceptent la compression avec perte même dans le cas des masters, c'est-à-dire des copies qui ne sont pas utilisées tous les jours, à condition que la dégradation de qualité ne soit pas perceptible (Guidance 2008). Par ailleurs, la perte passe pour tolérable si les originaux papier des documents électroniques comprimés sont disponibles⁸¹.

Faut-il préférer le format simple au format complexe ? Les avis sont partagés. Les formats simples génèrent moins d'erreurs, ils abîment donc moins les documents et protègent mieux leur accessibilité dans le long terme lors des conversions. En revanche, les documents complexes ne peuvent être bien représentés que par formats complexes, autrement leur qualité se dégrade. Il faut donc peser le pour et le contre et éventuellement conclure un compromis.

Toutefois, plusieurs archives et bibliothèques publient la liste des formats qu'elles ont l'intention d'employer dans le long terme. Certaines institutions, comme les Archives fédérales suisses, limitent strictement le nombre des formats acceptés ; d'autres, à l'instar des

⁸¹ Voir *Image Compression. Digital Preservation Guidance* note 5, p. 10. Accessible: <http://www.nationalarchives.gov.uk/documents/information-management/image-compression.pdf>, et *Digitization at The National Archives* (2013), p. 9. Accessible: <http://www.nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf>

Archives nationales du Royaume-Uni et des États-Unis, ont adopté une position beaucoup plus permissive. Sans prétendre à l'exhaustivité, il vaut la peine d'énumérer, à titre d'information seulement, les formats populaires que, de nos jours, plusieurs institutions prestigieuses recommandent pour la conservation pérenne :

- Textes non structurés : PDF/A, TXT
- Textes structurés : XML, ODT, PDF/A
- Tableaux : XML, ODS, CSV
- Bases de données : XML, CSV
- Images fixes : TIFF, JPEG2000, PNG
- Audio : WAV, WMA
- Vidéo : MXF, MOV, AVI
- E-mail : XML, MBOX, EML

Il n'est pas aisé d'identifier et de vérifier les formats. L'extension du fichier (par exemple TIFF ou PDF) qui en principe indique le format ne signifie pas pour autant que le fichier est ce que nous croyons. Il existe dans le monde de nombreuses variantes de formats dont beaucoup ne correspondent pas aux spécifications et aux structures publiées. Il peut en résulter des problèmes graves, parce que même une petite différence peut rendre inaccessible le document électronique ou la base de données lors de la lecture ou de la conversion du fichier par notre logiciel. L'institution conservatrice patrimoniale ne peut appliquer sa stratégie de migration (voir le chapitre 13.1) que si elle identifie et vérifie les fichiers entrant dans son système avant la conversion. Les registres de formats constituent des outils importants de cette procédure. Il y en a plusieurs dans le monde. Nous allons présenter brièvement le registre PRONOM, développé par les Archives nationales du Royaume-Uni ; il est accessible de nos jours par tous sur Internet⁸².

Le PRONOM contient la description précise et les spécifications techniques de 800 formats de fichier, ainsi que des informations sur les logiciels qui supportent les formats en question. Le logiciel DROID, par lequel nous pouvons identifier et vérifier les spécifications techniques de nos fichiers, fait partie du PRONOM. Comme nous l'avons mentionné ci-dessus, les formats sont des produits de l'industrie informatique, ils ont leur cycle de vie, ils

⁸² Voir <http://www.nationalarchives.gov.uk/aboutapps/pronom/#documentation>

sont développés et redéveloppés, on en fait des variantes, on arrête leur développement, ils deviennent obsolètes. Pour ces raisons, il importe que les registres de formats soient tenus à jour et complétés continuellement, comme c'est le cas de PRONOM dont les Archives nationales du Royaume-Uni assument l'entretien.

12 Les normes

Dans les chapitres précédents, nous avons vu les défis auxquels la communauté professionnelle des archivistes devait faire face au cours des cent vingt dernières années, ainsi que les réponses professionnelles, théoriques et pratiques, qu'elle a données. Jusqu'à l'avènement de l'ère des documents électroniques, au début des années 1990, les grandes étapes du développement de la profession ont été marquées par des ouvrages de théorie archivistique influents comme le manuel hollandais ou les études de Theodore R. Schellenberg. Mais dans le monde numérique, l'archivistique, tout à fait logiquement, emprunte beaucoup à la science de l'information et ce rapprochement se traduit par le fait que, dans cette nouvelle ère, l'évolution de l'archivistique est jalonnée d'adoptions de normes.

Les normes ont une importance particulière dans l'informatique, parce que la capacité d'abstraire de l'ordinateur, tout en progressant, reste très limitée et que tout écart de la régularité programmée aboutit à une erreur. C'est pourquoi les logiciels ne reconnaissent pas le format « étranger » et les programmes « étrangers » ne peuvent collaborer l'un avec l'autre. La solution passe par la normalisation, mais nombreux sont les obstacles à surmonter.

Le monde informatique présente la dichotomie intéressante de la présence parallèle de la prolifération et de l'homogénéisation et de leur équilibre dans le long terme. Dès les débuts, mais surtout à partir de l'expansion de l'Internet, ce monde est caractérisé par la liberté, fondée sur la facilité de produire, changer, altérer, multiplier, combiner et diffuser les informations. Mais une liberté excessive par rapport au logiciel, au matériel, au réseau, au format, etc., aboutit à l'anarchie et empêche la compréhension et la diffusion. Il s'ensuit qu'à un point donné l'homogénéisation s'enclenche automatiquement au travers, notamment, de la législation et de l'adoption de normes.

Les normes peuvent être canonisées, c'est-à-dire acceptées et publiées par une autorité nationale ou internationale compétente, comme c'est le cas des formats JPEG2000, PDF/A, TIFF ou ODF, qui sont des normes de jure. Il y a des normes de facto, c'est-à-dire pas canonisées par les autorités mais acceptées et répandues dans la communauté concernée,

comme les formats du Microsoft Office. Une norme peut être appelée norme, modèle ou exigence mais, pour qu'elle ait des effets réels, elle doit être une norme de facto, la canonisation n'est pas exigée.

Les normes sont aussi les produits du monde informatique, donc elles sont aussi caractérisées par la dichotomie mentionnée ci-dessus. Avec la prolifération qui se trouve à l'opposé de leur objectif originel, leur nombre augmente sans cesse et elles diffèrent souvent l'une de l'autre, alors que l'homogénéisation apporte, dans ce cas, fusion, intégration et superposition.

Voilà quelques exemples : au début des années 1990, il y avait d'innombrables normes nationales portant sur la description archivistique, puis l'ISADg publiée par le Conseil international des Archives allait exercer une très grande influence sur les normes nationales, et plusieurs d'entre elles allaient, par la suite, en absorber des éléments. Deux ans après l'ISADg parut l'ISAAR qui exploite une partie de la zone de contexte de l'ISADg, et rend ainsi possible que les archives qui ne suivent pas le principe de provenance puissent aussi utiliser l'ISADg⁸³. En même temps, aux États-Unis on élaborait l'EAD qui, contrairement à l'ISADg et l'ISAAR, ne peut être appliqué que dans un environnement électronique, du fait qu'il utilise le langage descriptif XML⁸⁴. Tandis que la première version de l'EAD était liée au format MARC, élaboré par la Bibliothèque du Congrès à l'intention des bibliothèques, la deuxième version (2002) s'en est éloignée et s'est approchée de l'ISADg en fait jusqu'à l'intégrer. L'ISAAR connut le même sort en 2010, quand on a élaboré l'EAC, en tant que pendant de l'EAD. On constate des chevauchements, des intégrations et des parallélismes pareils entre les normes OAIS, METS et PREMIS⁸⁵.

Les normes ont aussi leur durée de vie, longue ou courte ; elles changent aussi. Cette antilogie apparente tient à l'essence même du monde numérique, en changement et en évolution constants. Les normes ayant une longue durée de vie évoluent de version en version et, bien que chaque nouvelle version soit basée sur la version précédente, ces révisions mettent en cause la fonction primordiale des normes, la durabilité.

Le succès d'une norme se mesure par sa longévité d'une part, par sa diffusion et le nombre de normes qui s'y adaptent, d'autre part. La norme la plus réussie est sans doute le

⁸³ Voir l'ISADg et l'ISAAR dans le chapitre 12.4.

⁸⁴ Voir le compte-rendu sur l'XML dans le chapitre 12.1 et sur l'EAD dans le chapitre 12.5.

⁸⁵ Le METS et le PREMIS, ces normes de métadonnées, ont été exposés dans le chapitre 8; voir l'OAIS dans le chapitre 12.3.

XML qui se répand irrésistiblement depuis les années 1990. Dans les milieux des archives, le Dublin Core (mentionné dans le chapitre 8) fut relégué à l'arrière-plan d'abord par l'ISADg et l'ISAAR, puis par l'EAD et l'EAC. Le MOREQ⁸⁶, ce modèle européen de gestion des documents, et l'OAIS, une norme pour la conservation pérenne des documents et des données, remportent des succès incontestables. L'OAIS s'est imposé comme référence dans la communauté patrimoniale ; tout le monde s'efforce de rendre compatibles sa norme ou ses recommandations avec l'OAIS. Dans la suite de ce chapitre, nous allons faire le tour de ces normes réussies.

Il est incontestable que les institutions patrimoniales ont besoin de normes pour mener leur politique de collecte, pour échanger des données et surtout pour la conservation pérenne de leurs fonds et collections. Il faut les choisir avec soin, en ayant toujours en vue deux critères primordiaux : 1) l'utilité, c'est-à-dire si la norme est conforme à l'objectif et si elle est aisément applicable ; 2) la sécurité, c'est-à-dire si la norme est utilisable dans le long terme, si elle est largement répandue et s'il y a des organisations pour l'entretenir et la développer (bien que les normes qui changent souvent soient désavantageuses). Il faut prendre acte que les normes ne sont pas impérissables, qu'elles peuvent changer et que, dans le long terme, il sera inévitable de passer à des normes nouvelles, donc à convertir les documents, ce qui entraînera forcément des pertes de données.

12.1 Norme de langage : XML

Le XML⁸⁷ (Extensible Markup Language, c'est-à-dire : langage de balisage extensible) est membre d'une famille de langages. Le premier avait été le SGML (Standard Generalized Markup Language, c'est-à-dire langage standard général de balisage) ; il devenait norme en 1986 (ISO 8879), par la suite on en développait le HTML (Hypertext Markup Language) et finalement le XML en 1998. Les fichiers écrits en langage appartenant à cette famille peuvent être lus par n'importe quel logiciel de lecture. C'est pourquoi on les considère « indépendants de logiciel ».

Le langage XML fut développé pour la gestion des textes structurés dont les éléments (titres, sous-titres, cotes, dates extrêmes, sujets) peuvent être représentés par balises encadrées par des chevrons (< >). Les informations entre chevrons font partie des métadonnées. Pour

⁸⁶ Voir le chapitre 12.2.

⁸⁷ Le World Wide Web Consortium (W3C) s'occupe de l'XML. Voir <http://www.w3.org/standards/xml/>

représenter les documents très complexes, les balises sont complétées par des attributs (avec des valeurs) et des commentaires.

La structure du document et les balises (éléments) sont définies par le créateur - conformément à des règles sévères – dans un DTD (définition de type de document) ou schéma ; c'est pourquoi le XML s'appelle un langage auto-descriptif. Le XML permet de faire paraître la forme aussi dans une feuille de style (stylesheet).

Le XML a déclenché une révolution dans le monde informatique par son utilité, sa simplicité (texte plat) et sa durabilité. Les diverses professions, l'une après l'autre, ont créé leur DTD ou schéma. Ceux de la communauté d'archivistes, l'EAD et l'EAC sont détaillés ci-dessous.

12.2 Norme pour les archives courantes : MOREQ

Dans les années 1990, la Commission européenne avait chargé le DLM Forum⁸⁸ d'élaborer un recueil d'exigences générales pour l'organisation de la gestion des documents. Le résultat de ce travail, le MOREQ (Model Requirements for the Management of Electronic Records), a paru en 2001. Depuis, deux nouvelles versions ont été publiées. MOREQ est devenu le standard à succès de la gestion des documents électroniques qu'on applique hors d'Europe aussi. Les gouvernements qui veulent standardiser la gestion des documents dans le secteur public s'y réfèrent, tout comme les programmeurs qui développent des systèmes de gestion électroniques des documents, dans les phases active et semi-active.

La version la plus récente, c'est le MOREQ2010⁸⁹ qui, comme ses prédécesseurs, s'inspire de la norme de gestion des documents ISO 15489, mais est beaucoup plus détaillé. Il articule les exigences fonctionnelles en neuf « services », par exemple le service de l'utilisateur et de groupe, le service de classement, le service de tri, les services de recherche et de rapport. En dehors de cela, il y a une série d'exigences non fonctionnelles relatives à la vie privée, à la sécurité, à la fiabilité, etc.

Le MOREQ2010 recommande le recours au XML qui facilite l'exportation et l'importation des données.

⁸⁸ Le DLM Forum (Document Lifecycle Management) a été fondé par la Commission européenne en 1994. Son site: <http://www.dlmforum.eu/>

⁸⁹ Accessible: http://www.moreq.info/index.php?option=com_content&view=article&id=17&Itemid=22&lang=e

12.3 Norme pour la conservation des données : OAIS

Le Modèle de référence pour un système ouvert d'archivage d'information (OAIS - Open Archival Information System)⁹⁰ fut élaboré par le Conseil de Management du Comité Consultatif pour les Systèmes de Données Spatiales (CCSDS). Il devint une norme ISO (14721) en 2002. Il formule les règles à suivre par les archives pour la conservation pérenne et la communication des informations.

Le modèle d'OAIS traite les informations en paquets, un paquet comprenant le contenu et ses métadonnées qui forment une unité logique. À chaque procédure d'archives (versement, préservation, communication, etc.) correspond un paquet. Les paquets se répartissent en trois types :

- Paquet d'informations à verser (SIP) : ce que le producteur envoie aux archives lors du versement.
- Paquet d'informations archivé (AIP) : ce que les archives retiennent des SIP pour conservation pérenne. (C'est le plus important.)
- Paquet d'informations diffusé (DIP) : ce que l'utilisateur reçoit quand il demande des informations.

Outre les procédures, l'OAIS propose un modèle de fonctionnement. Il regroupe dans une série « d'entités » les tâches qui incombent à un service d'archives dans l'exercice de chacune de ses fonctions : « Entrées », « Stockage », « Gestion de données », « Administration », « Planification de la pérennisation », « Accès », etc.

12.4 Norme pour la description des archives 1 : ISADg et ISAAR(CPF)

Le Conseil international des archives (ICA) publia la Norme générale et internationale de description archivistique (ISADg) en 1994⁹¹. Son élaboration est devenue nécessaire dès la première phase de l'informatisation des archives 1) pour que les données sur les fonds écrits puissent s'échanger et 2) pour créer des instruments de recherche en ligne. Venue en temps utile, elle est utilisée dans le monde entier pour la révision des méthodes de description et la construction des bases de données.

⁹⁰ Accessible: http://pin.association-aristote.fr/lib/exe/fetch.php/public/documents/norme_oais_version_francaise.pdf

⁹¹ La deuxième édition de l'ISADg est parue en 2000. Accessible: <http://www.ica.org/10225/normes/isadg-norme-gnrale-et-internationale-de-description-archivistique-deuxieme-edition.html>

L'ISADg respecte le principe de provenance et rend possible une description équilibrée des quatre éléments constitutifs des documents (identifiant, contenu, structure et contexte). La norme est conçue pour la description des agrégats de documents hiérarchisés (fonds, sous-fonds, série, etc.), et pour fixer la place de chacun dans le macrosystème (voir chapitre 7.5.2)⁹².

L'ISADg groupe vingt-six éléments de données sur sept zones d'information :

- identification
- contexte
- contenu et structure
- accès et utilisation
- sources complémentaires
- notes
- contrôle de la description.

Deux ans après la parution de l'ISADg – en 1996 – l'ICA publia la Norme internationale sur les notices d'autorité utilisées lors de la description des archives des collectivités, des personnes et des familles (ISAAR[CPF])⁹³. Cette norme complète la zone de contexte de l'ISADg (les informations sur le producteur) et recommande l'application simultanée des deux normes⁹⁴. L'ISAAR contient trente et un éléments groupés sur cinq zones d'information.

12.5 Norme pour la description des archives 2 : EAD et EAC

Dans les années 1980, pour créer des instruments de recherche électroniques, les archives américaines avaient recours au format MARC (Machine Readable Cataloging, développé par la Bibliothèque du Congrès dans les années 1960)⁹⁵. Il s'est avéré que ce format conçu pour le

⁹² Supposant que la description suit l'ordre des documents. Quelquefois ce n'est pas le cas, par exemple quand on unifie un fonds dispersé par description.

⁹³ La deuxième édition de l'ISAAR est parue en 2004. Accessible: <http://www.ica.org/10230/normes/isaar-cpf-norme-internationale-sur-les-notices-dautorite-utilisees-pour-les-archives-relatives-aux-collectivites-aux-personnes-ou-aux-familles-2me-edition.html>

⁹⁴ La question se pose de savoir pourquoi on n'a pas complété la zone d'information appropriée de l'ISADg au lieu de créer une nouvelle norme. À notre avis, il n'y a qu'une explication logique: il fallait rendre possible la description d'un « fonds » ou « sous-fonds » qui a plusieurs producteurs. Mais cela s'oppose au principe de provenance.

⁹⁵ Voir www.loc.gov/marc/. Plus précisément, elles utilisaient la version développée pour les archives. Les archives américaines étaient toujours proches des bibliothèques pour des raisons historiques. Les institutions d'archives américaines se sont organisées relativement tard, les Archives nationales seulement en 1934, sous la pression des chercheurs et des historiens. Antérieurement, les documents d'archives restaient chez les producteurs, à moins d'être transférés à une bibliothèque. La raison peut être attribuée à la vision sociétale et à la

catalogage de collections ne convenait pas à la description des fonds d'archives⁹⁶. Donc la communauté des archivistes américains se mit à développer une nouvelle norme.

Cette norme qui reçut le nom EAD (Description archivistique encodée) fut fortement influencée par le format MARC. Elle fut élaborée (1998) dans le langage SGML disponible sur le marché. Cette première version n'ayant pas donné entière satisfaction, les travaux ont continué et ont abouti à la mise au point de la deuxième version (2002) écrite en langage XML⁹⁷. Cette version incorpore beaucoup d'éléments de l'ISADg qu'elle adapte à l'informatique. Mais son approche de la description archivistique diffère de celle d'ISADg : elle considère le produit de la description comme un document électronique. La grille de l'EAD est beaucoup plus détaillée (cent quarante-six éléments) que celle de l'ISADg et l'ISAAR (vingt-six et trente et un). La richesse de la méthode confère à l'EAD une souplesse qui rend possible la description de toutes les catégories documentaires et permet de convertir les inventaires et répertoires analogiques en instruments de recherches numériques.

L'EAD est devenu une norme de facto acceptée et utilisée par la communauté professionnelle.

Relativement tard, en 2010, on développa l'EAC-CPF (Contexte archivistique encodé – Collectivités, personnes, familles)⁹⁸, une « traduction informatique » (schéma XML) de l'ISAAR. Il s'utilise pour compléter l'EAD.

philosophie économique américaine. D'après cette vision, l'homme est originellement un être libre, les hommes et les communautés sont responsables d'eux-mêmes, donc il leur faut assurer la plus grande liberté de s'organiser et d'agir, donc réduire autant que possible la compétence de l'État. La grande dépression ébranlait cette conception et on peut dater du New Deal la construction d'un État fédéral plus centralisé et le renforcement des interventions. Cela amenait automatiquement une bureaucratie plus étoffée et une production de documents à grande échelle qui exigeait l'établissement d'un réseau de services d'archives. Malgré cela, aux États-Unis, l'influence des bibliothèques sur les archives reste remarquable. Les bibliothèques publiques et privées conservent une très grande quantité de fonds d'archives et il arrive souvent que les personnalités publiques importantes lèguent leurs archives à des bibliothèques.

⁹⁶ Helen R. TIBBO, Lokman I. MEHO, "Finding Finding Aids on the World Wide Web" in *American Archivist*, vol. 64. Spring 2001, p. 63.

⁹⁷ Voir les deux premières versions du EAD: <http://www.loc.gov/ead/>

⁹⁸ La version (anglaise) complète de l'EAC est accessible: <http://eac.staatsbibliothek-berlin.de/>. Une traduction française est accessible: <http://www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/eac/>

13 Stratégies pour la conservation pérenne des documents électroniques

Dans les chapitres précédents, nous avons exposé les difficultés que rencontre la conservation pérenne des documents électroniques : obsolescence du matériel, du logiciel et du format, absence ou non-application des normes, structure complexe des documents, vulnérabilité de la structure, séparation des éléments constituants, facilité de manipulation, problèmes du droit d'auteur et caractère dynamique des documents. Par la suite, nous mentionnerons d'autres difficultés comme les virus, le chiffrement ou la protection des documents, par exemple par mot de passe.

Au vu de ces faits et à la lumière des expériences des dernières décennies, nous pouvons affirmer avec certitude que les documents électroniques ne pourront être conservés dans l'état originel pendant des centaines ou seulement des dizaines d'années. Si on s'y essaie, après un certain temps les documents deviendront illisibles. Le plus que nous pouvons faire, c'est de mettre et de garder les documents dans un état qui permettra de les reproduire dans un avenir prévisible⁹⁹. Toutes les institutions qui entendent s'occuper de la conservation dans le long terme doivent élaborer leur stratégie, c'est-à-dire adopter des mesures pour que leur capacité de reproduction puisse être maintenue.

La stratégie de conservation pérenne ne connaîtra le succès que si les archives reçoivent des producteurs des versements convenablement classés. Les documents sont bien classés s'ils sont rangés dans des catégories préétablies, conformément au plan de classement. Qu'ils soient analogues ou électroniques, on ne peut évaluer les documents non classés, leur tri est impossible, ils forment un amas confus et inaccessible. Même s'ils restent lisibles, c'est comme s'ils étaient détruits.

Condition supplémentaire, il faut que les documents reçoivent un traitement archivistique convenable, qu'ils soient intégrés dans les systèmes d'archives, c'est-à-dire groupés en sous-fonds, fonds et agrégats de fonds. Dans le cas des documents analogiques, ce système est représenté par l'état des fonds, dans le cas des documents électroniques par une base de données.

Il faut coordonner la stratégie de conservation pérenne avec les autres stratégies de l'institution, en particulier avec la politique de collecte (quels documents sont collectés et

⁹⁹ Les paroles de Kenneth Thibodeau citées par Terry EASTWOOD, "The Appraisal of Electronic Records: What is New?" in *Comma* 2002/1-2, p. 77.

quelles sont leurs caractéristiques) et avec la politique de communication (qui seront les lecteurs, dans quel but et dans quelle mesure les documents seront accessibles sur l'Internet, etc.). Le but poursuivi consiste à garder les documents accessibles et exploitables par les chercheurs en ligne, ce qui demande du matériel, du logiciel, des bases de données et des formats appropriés. La révolution de la communication des archives et des bibliothèques est déjà une réalité, elle a étendu la salle de recherche pratiquement au globe entier. De plus, l'intégration des archives et des bibliothèques du monde est en cours au moyen des portails, Europeana ou Apenet¹⁰⁰, qui donnent accès aux données et aux documents numériques de nombreuses institutions patrimoniales.

En évaluant les risques, les menaces qui pèsent sur les documents, on peut estimer les coûts et l'espace de stockage à prévoir ainsi que les besoins en matière d'expertise. Or, comme les stratégies ne sont pas éternelles, il vaut la peine d'examiner leur efficacité tous les cinq à dix ans, période pendant laquelle le monde informatique peut changer considérablement. En y apportant les corrections nécessaires et si les conditions le permettent, on peut aboutir à un changement de stratégie.

Au cours des dernières décennies, les archives ont essayé plusieurs stratégies de conservation. Nous exposerons dans les chapitres suivants celles qui ont réussi, c'est-à-dire se sont avérées durables. Il y en avait des moins réussies, qui méritent tout de même notre attention. Celle dite de conservation technologique proposait de conserver non seulement les fichiers originaux mais le logiciel et le matériel aussi. Cette stratégie est pratiquement abandonnée parce que même un service d'archives ayant un champ de collecte modeste ne saurait faire face au coût qu'exige le maintien en état de fonctionnement d'un parc de matériels et de logiciels croissant sans cesse. À cela s'ajoute que les machines devenant obsolètes, les pièces ne sont plus fabriquées au bout de quelques décennies. Or, le matériel, donc les machines et les documents qu'elles gèrent, ne vit que tant que les pièces fonctionnent. L'entretien de plusieurs centaines de logiciels accumulés poserait également des problèmes insolubles. Dans certains cas cependant, la conservation technologique est justifiée. Ainsi il y a des raisons valables pour conserver, pendant une ou deux décennies, des éléments de matériel récemment dépassés, qui peuvent servir dans l'avenir. Grâce à des ordinateurs munis de lecteur de disquette, par exemple, on pourra lire d'anciens fichiers et les copier sur des supports conformes aux normes actuelles.

¹⁰⁰ Voir <http://www.europeana.eu/portal/> et <http://www.archivesportaleurope.net/>

L'approche post-custodiale était un sujet populaire aux conférences et colloques professionnels dans les années 1990, au début de la révolution de l'Internet. D'après cette vision, les services d'archives n'accueilleront pas dans l'avenir les documents électroniques, qui resteront chez les producteurs. Dans l'ère de la Toile mondiale, peu importe où ils sont stockés physiquement. Les organismes producteurs s'occuperont de l'entretien et, si nécessaire, de la conversion des documents, alors que la tâche des archives se limitera à en assurer, via l'Internet, le contrôle intellectuel et la description en mettant l'accent sur l'élaboration des contextes¹⁰¹. Cette stratégie ne s'est pas répandue parce que les producteurs, appartenant en majorité au secteur public, n'étaient nullement intéressés à prendre en charge les dépenses pour conserver et maintenir en état d'accessibilité les vieux documents pendant des décennies voire des siècles. Au demeurant, cette tâche ne leur incombait pas.

Finalement, il faut mentionner une solution qui rappelle en partie l'approche post-custodiale. Il s'agit du *cloud computing*, un mécanisme tout nouveau. On manque encore d'expérience mais toutes les institutions, les organismes producteurs comme les archives, s'y intéressent. À vrai dire, il s'agit plutôt d'une stratégie partielle puisqu'elle concerne seulement le stockage et la mise en service technique des documents électroniques, tandis que la conservation est une politique globale couvrant un domaine beaucoup plus vaste.

Le stockage dans le nuage signifie que les documents électroniques ne sont pas versés au service d'archives compétent, mais envoyés loin de là, sur l'ordinateur d'une compagnie spécialisée qui sera responsable de l'entretien des documents et de la mise en marche des logiciels nécessaires, alors que le service d'archives aura accès à ses propres documents par l'Internet. Un cloud service typique est par exemple le Gmail opéré par Google. Le cloud offre plusieurs avantages. Il réduit les dépenses, les archives sont libérées des soucis d'infrastructure (matériel, logiciel) et ne sont pas tenues d'embaucher des informaticiens, mais ces avantages sont accompagnés d'un risque énorme : les documents ne sont plus sous le contrôle des archives. Tout le travail professionnel comprenant l'élaboration des stratégies, le développement du système d'enregistrement et de recherche, la vérification, la validation, le tri et la description reste aux archives, qui délèguent une partie des tâches informatiques à la firme s'occupant du stockage. Bien que dans les pays développés les prestations offertes par ces firmes soient réputées sûres, les institutions patrimoniales qui songent à s'y adresser

¹⁰¹ Un compte rendu de la théorie: Terry COOK, "Archives in the Post-Custodial World; Interaction of Archival Theory and Practice since the Publication of the Dutch Manual in 1898" in *Archivum XLIII*, pp. 205-207. Une telle stratégie fut appliquée par les Archives nationales d'Australie dans les années 1990, voir http://www.naa.gov.au/Images/an-approach-green-paper_tcm16-47161.pdf, p. 6.

doivent peser le pour et le contre, comparer bénéfices et risques. Si on opte pour le cloud, il faut s'entourer de certaines garanties, tels la production de plusieurs copies ou l'export sans heurts des données du cloud vers les archives.

13.1 Migration

La majorité des services d'archives qui pratiquent l'archivage électronique a choisi la stratégie de migration. La migration signifie que les fichiers sont convertis de formats obsolètes en formats conformes aux normes nouvelles. Lors du choix des nouveaux formats, on tiendra compte de la qualité et de la durabilité d'abord et aussi des critères exposés dans le chapitre 11. Puisque les formats changent, il faut opérer la migration de temps en temps.

Au cours de la migration, la perte de données ne saurait être évitée, mais la quantité et la qualité des données perdues ne sont pas indifférentes. S'il faut veiller à réduire au minimum les pertes de contenu, de métadonnées et de données qui représentent des fonctions importantes, en ce qui concerne les fautes d'affichage de la forme du document, nous pouvons être plus tolérants. Dans le chapitre 9 sur l'intégrité, nous avons mentionné la conception australienne sur l'essence du document qui, à leur avis, doit être conservée à tout prix.

Les migrations consécutives, qui convertissent encore et encore les documents déjà migrés, accumulent les pertes ; il est donc souhaitable de les éviter. Chaque répétition menaçant la sécurité des documents, il faut y procéder le plus rarement possible, en intercalant de longs intervalles. Pour détecter les pertes, il ne suffit pas de pratiquer un sondage ; le logiciel qui fait la conversion doit signaler automatiquement les documents ayant été touchés dans leur intégrité.

La méthode préférée pour éviter les migrations consécutives consiste à faire toujours la conversion sur les fichiers originels ou les masters. Plusieurs services d'archives possèdent parallèlement : 1) les fichiers originels ; 2) les fichiers de préservation ; 3) les fichiers de communication. La conservation des fichiers originels sert à la sécurité parce que si la conversion rate, on peut répéter l'opération. C'est la politique des Archives nationales du Royaume-Uni et le projet CEDAR recommande la même méthode¹⁰². Les Archives nationales d'Australie ont développé une plate-forme de logiciels pour la conservation numérique, par

¹⁰² Voir <http://www.nationalarchives.gov.uk/aboutapps/pronom/#documentation>, ainsi que David HOLDSWORTH, *Instalment on "Preservation Strategies on Digital Libraries"* (DCC Digital Curation Manual 2007) p. 12-13.

laquelle passent les versements des documents électroniques. Le plus important élément de cette plate-forme est le logiciel XENA (*XML Electronic Normalising for Archives*), qui convertit les fichiers entrants en format de conservation. C'est une sorte de normalisation à grande échelle. La *Victorian Electronic Records Strategy* (VERS) développée par le Public Record Office de Victoria applique des principes pareils : elle convertit les fichiers versés en formats « de long terme », de base XML comme c'est le cas aux Archives de Canberra¹⁰³.

Il y a intérêt à faire la migration le plus tôt possible, juste après le versement, car à ce moment-là l'information sur les fichiers et l'expertise sont encore disponibles sur le marché et auprès de l'organisme producteur. Si la conversion ne réussit pas, on peut la corriger. Ne devront pas être accueillis et convertis les fichiers chiffrés, protégés par mot de passe ou infectés par virus parce qu'ils resteront inaccessibles.

Il est recommandé de vérifier de temps en temps les fichiers déjà migrés et, comme dans le long terme l'environnement informatique change, de nouvelles migrations seront inévitables. Elles devront intervenir par périodes déterminées, par exemple tous les dix ans, ou avant le changement du système d'exploitation, et non s'improviser, pour pallier le danger de ne pouvoir convertir les formats obsolètes. Si nous choisissons des formats durables, la nécessité de migration arrivera plus rarement.

Les formats uniques résultant de la migration permettent aux archives de développer un système général de recherche pour l'ensemble de leurs collections numériques, ce qui facilite considérablement le travail des archivistes et des chercheurs.

13.2 Émulation

L'émulation signifie que nous imitons l'environnement informatique originel, le matériel et le logiciel, par des logiciels spéciaux qui s'appellent émulateurs. Elle rend possible l'utilisation des documents électroniques par leur logiciel originel malgré le nouvel environnement informatique.

L'émulation nécessite une expertise par informaticiens qualifiés, parce qu'il faut faire fonctionner non seulement le système d'exploitation et le logiciel d'application mais aussi tous les fichiers nécessaires pour le fonctionnement de ce dernier. En plus, il faut décrire les spécifications d'émulateur, c'est-à-dire tous les attributs (par exemple vitesse, attributs

¹⁰³ Voir http://www.naa.gov.au/Images/Digital-Preservation-Software-Platform_tcm16-60788.pdf, illetve <http://prov.vic.gov.au/government/vers>

d'affichage, outils et périphériques) du matériel originel qui rendront possible la reprogrammation future de l'émulateur actuel¹⁰⁴. Selon ses partisans, le plus souvent des informaticiens, en ayant recours à l'émulation on peut éliminer les inconvénients de la migration, par exemple le changement de la forme des documents, la détérioration de l'authenticité et de l'utilisabilité, et les pertes accumulées qui résultent des conversions consécutives dans le long terme.

Tout cela est vrai, et pourtant les archives rechignent à appliquer l'émulation. Les archivistes estiment que les inconvénients mentionnés ci-dessus ne sont pas bien graves. C'est pourquoi ils ont défini l'essence du document. Ils pensent que la forme inchangée ou la chaîne binaire intacte n'assurent pas forcément l'authenticité et que la conversion consécutive peut être évitée. Sauvegarder l'utilisabilité ne signifie pas un avantage en soi, parce que dans les systèmes de gestion des documents des producteurs, il y a des fonctions (par exemple enregistrer, modifier et effacer des données, interrogations relatives à certaines procédures) qui sont inutiles, voire indésirables pour les archives. D'après David Bearman, sont à conserver les dossiers et pas les fonctions du système informatique ; il ne faut pas confondre dossier et information¹⁰⁵.

L'émulation a d'autres inconvénients aussi :

- Apprendre à utiliser les logiciels d'application originels dont le nombre augmente sans cesse peut causer de grandes difficultés aussi bien aux archivistes qu'aux chercheurs.
- En quelques décennies, un service d'archives peut accumuler plusieurs centaines de logiciels.
- À cause du changement continu de l'environnement du matériel et du logiciel, il faut produire de plus en plus d'émulateurs qui à leur tour doivent être renouvelés encore et encore à la suite des changements intervenus.
- Les logiciels propriétaires impliquent souvent une redevance de licence qui peut coûter très cher en cas d'accumulation.

¹⁰⁴ Jeff ROTHENBERG donne un bon résumé sur l'émulation (bien qu'il manque d'objectivité) dans *Avoiding Technological Quicksand: Finding a Viable Foundation for Digital Preservation* (Council on Library and Information Resources, Washington 1999).

¹⁰⁵ Cité par Stewart GRANGER, "Emulation as a Digital Preservation Strategy" in *D-Lib Magazine*, October 2000, vol. 6. No. 10. Accessible: <http://www.dlib.org/dlib/october00/granger/10granger.html>

- Les nombreux agrégats de données distincts, traités par autant de logiciels, empêchent ou rendent difficile de développer un système de recherche unifié et compliquent l'accès aux données par l'Internet.

Malgré ces arguments, il ne faut pas exclure l'émulation de la vie des archives, parce que certains documents électroniques peuvent avoir des fonctions importantes qui ne peuvent être conservées par migration mais seulement par émulation. Plusieurs projets d'émulation ont réussi, par exemple le *Global Access to Emulation Services* (GRATE) développé par l'Université de Fribourg-en-Brisgau (Allemagne) dans le cadre du projet européen PLANETS, l'émulateur Dioscuri¹⁰⁶, développé sous la direction des spécialistes hollandais pour PC de x86 ou encore les résultats et les outils du projet KEEP (*Keeping Emulation Environments Portable*)¹⁰⁷ financé par l'Union européenne.

13.3 Conservation en forme analogique et hybride

Partant du fait qu'il n'y avait pas de différence essentielle entre les documents traditionnels et électroniques, jusqu'aux années 1990, les archives imprimèrent souvent les documents électroniques importants et conservèrent les copies papier sans s'occuper des versions numériques. Mais quand les bases de données et les réseaux devinrent de plus en plus répandus, elles comprirent vite que cette méthode était tout à fait impropre à conserver et à reproduire la complétude et les structures compliquées des documents électroniques. Un e-mail imprimé perd non seulement sa fonctionnalité mais aussi les fichiers implantés et les annexes. Dans le cas d'une page web nous ne pouvons rien faire avec les liens internes, et vouloir imprimer une base de données est tout simplement déraisonnable.

Après le tournant du millénaire, la tendance s'inversa. Aujourd'hui, il est plus typique que les organismes et les compagnies numérisent les documents papier qu'ils reçoivent et qu'ils éliminent ensuite, créant ainsi un système numérique unifié et homogène. Dans plusieurs pays développés, cette pratique reçut l'appui de la législation, quand on mit au point les conditions légales de l'authenticité des documents électroniques et établit l'infrastructure nécessaire pour l'émission des signatures électroniques. Certains gouvernements incitent carrément les administrations à réduire la quantité des documents papier. C'est ainsi que le

¹⁰⁶ Accessible: <http://www.planets-project.eu/>, http://www.planets-project.eu/docs/reports/Planets_PA5-D7_GRATE.pdf, <http://dioscuri.sourceforge.net/>

¹⁰⁷ Accessible: <http://www.keep-project.eu/ezpub2/index.php>

gouvernement australien lança son *Government's Digital Transition Policy* en 2011, alléguant comme motif l'économie¹⁰⁸.

Il y a tout de même des documents dont la conservation en papier est souhaitable, voire stipulée par la loi (voir ci-dessus, chapitre 3.4). Au vu de nos expériences des dernières décennies, nous pouvons être sûrs que les livres et les documents subsisteront simultanément sous formes papier et électronique dans un avenir prévisible. D'une part parce que le papier reste toujours un support et un affichage plus naturels que l'écran. Dans le domaine des documents sonores et visuels, où cette différence n'existe pas, la technique analogique a complètement disparu. D'autre part, la majorité des producteurs ne peut pas se permettre ou n'éprouve pas le besoin d'acheter et d'utiliser un système cher et compliqué de gestion électronique des documents, grâce auquel il pourrait réduire radicalement le volume des documents papier produits.

Les microfilms peuvent aussi avoir un rôle important dans la conservation pérenne des documents électroniques. Le microfilm a trois grands avantages : 1) il condense l'information ; 2) il est très stable ; si on le conserve aux conditions appropriées il peut durer plusieurs centaines d'années ; 3) il est relativement indépendant de la technologie, c'est le cas même des COM (voir ci-dessous).

La technologie *Computer Output Microfilm* (COM) photographiant les codes numériques sur microfilm avec une caméra spéciale fut populaire dans les années 1980. Les codes numériques ainsi conservés de manière analogique peuvent être, si c'est nécessaire, reconvertis (numérisés). Bien que la technologie COM ait reculé dans les dernières décennies, il serait prématuré de l'enterrer parce que les avantages mentionnés demeurent. Les Archives nationales de Suède introduisaient, il y a quelques années, le COM pour la conservation de leurs documents électroniques et, d'après leur calcul, le coût du stockage des microfilms ne représente qu'une fraction de celui des supports numériques¹⁰⁹.

Le microfilm peut aussi jouer un rôle important lors de la numérisation des documents papier. D'après la stratégie des Archives fédérales d'Allemagne, les copies numériques remplissent à la fois une fonction de communication et une fonction de préservation,

¹⁰⁸ Voir <http://www.naa.gov.au/records-management/digital-transition-policy/scanning-incoming-paper.aspx>

¹⁰⁹ Heather BROWN et al., Instalment on "The role of microfilm in digital preservation" (DCC Digital Curation Manual 2011) p. 15.

puisqu'en les communiquant aux chercheurs on ménage les documents papier. Le microfilm sert aussi de copie de sécurité¹¹⁰.

13.4 Documents spéciaux

Il y a des documents électroniques qui n'ont pas d'équivalents dans le monde analogique, ou bien s'il en existe, ils en diffèrent profondément. Nous en évoquerons trois qui, en raison de leur importance et de leur caractère spécial, méritent une attention particulière lors de l'élaboration de la stratégie de conservation.

13.4.1 *E-mails*

Aujourd'hui, les e-mails forment probablement la masse la plus volumineuse des documents électroniques produits dans le monde. D'après les estimations du Radicati Group, en 2011, le nombre de comptes de messagerie électronique était de 3,1 milliards et les hommes d'affaires envoyaient en moyenne 33 messages par jour¹¹¹. Les e-mails ont pris la place non seulement des lettres traditionnelles mais aussi celle des notes administratives ; ils sont présents à tous les niveaux de décision, dans toutes les procédures administratives. Par rapport au rôle qu'ils jouent, les organismes ne leur accordent que peu d'attention malgré l'importance des informations qu'ils portent comme peuvent en témoigner les scandales et les procès récents, dont les affaires Enron, Arthur Andersen, Wikileaks, Apple contre Samsung dans lesquels les e-mails ébruités ou mis à jour jouaient un rôle décisif¹¹². Malheureusement, pour plusieurs raisons, les archives reçoivent souvent des organismes des e-mails incomplets ou mal classés.

En principe, un e-mail produit ou reçu dans un bureau est un document officiel. Il devrait avoir le même statut que les autres documents créés dans le cadre de la même procédure ; en d'autres termes, il appartient organiquement aux archives courantes. La pratique montre autre chose. Nous avons mentionné que la communication électronique, e-mails compris, a repris, à

¹¹⁰ *Digitalisierung im Bundesarchiv 2011-2016*. Accessible: http://www.bundesarchiv.de/imperia/md/content/bundesarchiv_de/fachinformation/informationstechnologie/digitalisiertesarchivgut/digitalisierungimbarch2011_2016fv.pdf

¹¹¹ *The Radicati Group Inc. Email Statistics Report, 2011-2016*, p. 2-3. Accessible: <http://www.radicati.com/wp/wp-content/uploads/2011/05/Email-Statistics-Report-2011-2015-Executive-Summary.pdf>

¹¹² Cités par Christopher J. PROM, *Preserving email* (DPC Technology Watch Report 11-01, 2011), p. 5.

de nombreux égards, la fonction de la communication orale, tout en modifiant les règles de l'usage et les caractéristiques formelles de la correspondance écrite. L'omission de l'objet, par exemple, abîme la structure interne du document et rend difficile son classement. Les employés reçoivent en général leur compte de messagerie sous leur nom et beaucoup d'entre eux utilisent ce compte pour leur correspondance privée. Il en résulte que les e-mails officiels et privés se confondent ce qui rend très difficile l'archivage, d'autant plus qu'il faut tenir compte des règles légales qui protègent la vie privée¹¹³. Les e-mails peuvent être stockés sur le serveur central de l'organisme ou sur l'ordinateur de l'employé. Dans ce dernier cas, il arrive que l'employé vide son compte quand il quitte l'organisme. Tout cela aboutit à faire des e-mails du bureau des documents privés plutôt qu'officiels.

Il y a tout de même des solutions qui renforcent le caractère officiel et l'authenticité des e-mails produits dans les bureaux : si l'e-mail est créé et conservé par le serveur central, si l'indication de certaines métadonnées est obligatoire, car autrement les systèmes ne permettent pas d'envoyer l'e-mail, si le compte de messagerie se réfère à la fonction de l'employé et pas à son nom (par exemple chef-departement@ministere...), le successeur en héritera.

Les e-mails sont souvent gérés dans un système séparé et il en résulte que les relations entre les e-mails et les autres documents sont coupées. Quand une affaire est vivante, les employés savent quels documents sont liés, mais comme le temps passe il est de plus en plus difficile de reconstruire un dossier.

Naturellement il y a des solutions pour ce problème aussi. La meilleure solution consiste à intégrer le système e-mail dans le système général de gestion électronique des documents (GED), où les messages sont mis à la place appropriée. Si le producteur n'a pas de GED, il faut que tous les e-mails reçoivent leur identifiant d'archives qui les lient aux autres documents de la même affaire, et il faut créer une structure de répertoire synchronisée avec le plan de classement des archives.

Les annexes, qui sont souvent plus importantes que les messages eux-mêmes, peuvent causer de gros soucis à la gestion et surtout à l'archivage des e-mails, parce que les formats en sont fort variés. Si elles sont placées dans la GED ou dans le système d'archives, il convient

¹¹³ Cela dépend des pays, en France la protection est très forte. Voir Frédéric BRÉGIER et al., *L'archivage des messageries électroniques, Preuve de concept VITAM*, Version 1.1, 2013, pp. 15-16.

de les conserver séparément des messages, en y joignant les métadonnées indiquant les liens, ou bien de les convertir au format standard.

Après le versement aux archives, les e-mails connaîtront le même sort que les autres documents électroniques. Ils seront vérifiés, convertis au format standard (la littérature spécialisée recommande le plus souvent le format XML¹¹⁴, en revanche l'émulation n'est pas préconisée), décrits et intégrés dans le système unifié d'enregistrement et de recherche des archives.

13.4.2 Bases de données

Le guide pour la gestion archivistique des documents électroniques du Conseil international des archives donne la définition suivante de la base de données : « un regroupement organisé (structuré) de données interreliées. Elle comprend deux parties essentielles : les éléments ou données particulières (le contenu) et les structures selon lesquelles ces données sont organisées. [...] les bases de données modernes ne se limitent plus à la structure relationnelle car on y intègre notamment des fichiers-textes, des images, des sons, etc. »¹¹⁵

Probablement les bases de données sont les documents électroniques dont l'archivage est le plus difficile pour plusieurs raisons, déjà mentionnées dans le chapitre 9 :

- leur structure est très compliquée ;
- selon la façon dont sont filtrées et interrogées les données, on crée des enregistrements temporaires (transactionnels) ;
- la plupart des bases de données sont dynamiques, c'est-à-dire que leur contenu change très vite ;
- Les bases de données qui fonctionnent longtemps peuvent changer de structure et de fonction.

¹¹⁴ Voir par ex. Frédéric BRÉGIER et al. *op. cit.* pp. 36-37. et 48-52, Christopher J. PROM *op. cit.* pp. 23-24 ainsi que Maureen PENNOCK, Instalment on "Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages" (DCC Digital Curation Manual 2006) pp. 48-52. (Dans ces œuvres, les auteurs font un tour d'horizon des recherches internationales portant sur le sujet en question.)

¹¹⁵ *Guide pour la gestion archivistique des documents électroniques*, Conseil international des archives, Études 8 (1997), p. 43.

La solution passe nécessairement par un compromis : il faut sauvegarder pour l'archivage le plus possible de la structure et de l'organisation des données (qui s'appelle le schéma de base de données) ; il faut, de temps en temps, demander des versements incrémentaux (qui contiennent seulement les enregistrements changés) et il faut consigner aussi les changements de structure ou de fonction.

Dans le cas des bases de données, on ne doit avoir recours à l'émulation qu'exceptionnellement, quand c'est la seule possibilité d'en conserver les fonctions essentielles. Les arguments contre l'émulation que nous avons exposés dans le chapitre 13.2 – accumulation des logiciels d'application, redevances de licence, bases de données isolées – sont valables ici aussi. Au demeurant, il arrive souvent que le contrat entre l'organisme producteur et le développeur du logiciel interdise la livraison du logiciel à une tierce partie, donc aux archives aussi¹¹⁶.

Or, comme les archives qui poursuivent une stratégie de migration ne peuvent créer un logiciel pour chaque base de données, elles utilisent des systèmes de recherche et de gestion simplifiés, communs pour plusieurs bases de données. Il est de toute façon souhaitable que les archives créent un système de recherche unifié pour toute la collection.

Peut-être à cause de la complexité du problème, peu de projets envisagent d'offrir une solution universelle pour l'archivage des bases de données. Le SIARD (*Software Independent Archiving of Relational Databases*)¹¹⁷ est un logiciel de préservation pour l'archivage des contenus de bases de données relationnelles élaboré par les Archives fédérales de Suisse. Par SIARD, on convertit non seulement le contenu mais les métadonnées et les relations entre les données aussi, et par cela il est possible de construire une nouvelle base de données. Les archivistes suisses ont développé un format spécifique nommé aussi SIARD qui se base sur des normes ouvertes : XML, SQL, ZIP et UNICODE. « Contrairement à une exportation à plat d'une base de données, SIARD permet de coder la structure de la base (liste des tables, description pour chaque table de la liste de ses champs, description pour chaque champ de son type). Le format permet aussi de décrire les relations entre les tables ainsi que des informations de gestion. »¹¹⁸

Les Archives fédérales de Suisse rendent accessibles le logiciel sur leur site Internet.

¹¹⁶ Il est à mentionner que dans certains cas la base de données est aussi propriétaire, ce qui peut être un obstacle à la conversion.

¹¹⁷ Accessible: <http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=fr>

¹¹⁸ <http://www.archivesdefrance.culture.gouv.fr/static/4163>, p. 2.

Le système CHRONOS¹¹⁹, développé en Bavière, offre une autre approche et ses objectifs principaux sont les suivants :

- séparer les données du logiciel originel ;
- conserver la structure et la sémantique originelles des données en forme lisible à l'œil humain ;
- intégrer les bases de données dans une seule base de données d'archives ;
- accroître d'une façon incrémentale les données ;
- assurer l'accès aux données en ligne par le web ;
- créer des fonctions d'exportation et d'importation des fichiers textuels ;

Il convient de souligner que CHRONOS crée des archives électroniques en format de pur texte (ASCII/UTF pour les données primaires, XML pour les métadonnées). C'est un facteur important de sécurité parce qu'il rend les données indépendantes du matériel et du logiciel, et grâce à cela on peut les lire par un navigateur simple aussi.

13.4.3 Pages web

Les bibliothèques aussi bien que les archives collectent les pages web. Bien qu'il y ait un certain parallélisme entre les actions qu'elles mènent, les bibliothèques sont beaucoup plus avancées et les points de vue de collecte diffèrent souvent. Tandis que les bibliothèques considèrent les pages web comme des publications de plein droit, les archives les traitent généralement comme faisant partie du fonds du producteur, qu'il soit une personne physique ou morale¹²⁰. Aussi, la politique de collecte des archives est-elle plus ciblée. Tandis que la British Library construit les archives web du Royaume Uni avec l'objectif de documenter la société totale, le but du *Web Continuity Project* des Archives nationales du Royaume-Uni est « l'archivage complet du domaine web du gouvernement ».

Bien que les pages web d'une institution fassent partie, à coup sûr, du fonds de celle-ci, il y a tout de même des différences entre les documents produits par l'activité administrative ordinaire et les informations mises sur le web. Les informations de l'institution accessibles sur

¹¹⁹ Stefan BRANDL, Dr. Peter KELLER-MARXER, *Long-term Archiving of Relational Databases with Chronos* (First International Workshop on Database Preservation, Edinburgh 2007). Accessible: http://homepages.inf.ed.ac.uk/hmueller/presdb07/papers/Germany_CHRONOS_PresDB07.pdf

¹²⁰ Il y a tout de même des archives – surtout au Canada et au Pays-Bas – qui pratiquent la conception des archives totales qui signifie qu'elles collectent non seulement les fonds des organismes et des individus, mais tous les documents (manuscrits, livres, photos, etc.) relatifs à leur compétence (pays, région, thème, etc.).

le web comportent toujours une partie publique que l'institution entend montrer d'elle-même, où les informations bien structurées sont tenues à jour. De plus en plus d'activités des institutions sont mises sur le web, comme les publications, les expositions, les recherches par catalogues et des documents numériques. En liaison avec la partie publique, beaucoup d'institutions organisent sur le web des forums professionnels ou sociaux qui s'adressent à des publics précis, en rassemblant les informations et leur servant de points d'orientation. Souvent clos, le forum de communication est réservé aux membres du groupe-cible, ce qui a des conséquences lourdes pour l'archivage.

La structure compliquée et l'instabilité des pages web rendent difficile leur archivage, mais leur complexité et volatilité diffèrent de celles des bases de données. Les pages web contiennent beaucoup de documents incorporés, toutes sortes de formats et énormément de liens se référant à de nouvelles pages et, comme nous l'avons mentionné dans le chapitre 9, la conservation des liens actifs est sans espoir parce que si on les suivait, le nombre des documents à archiver augmenterait en progression géométrique.

Le dynamisme des pages web a quelque chose de déconcertant. On a maintes fois essayé d'exprimer en chiffres cette volatilité. D'après un sondage de 1999, 44 % des pages de l'année précédente ne s'affichaient plus ou elles étaient considérablement modifiées¹²¹ ; d'autres recherches estimaient la durée de vie moyenne des pages entre quarante-quatre et cent jours. Évolution incompatible avec les règles de la science, dans les publications scientifiques en ligne, un tiers des références et des citations étaient incontrôlables au bout de dix ans¹²². À quelle fréquence faut-il faire des instantanés d'une page volatile ? Le dilemme est permanent. La décision sera prise en fonction de l'importance de la page, des événements particuliers qui s'y présentent et du rythme de changement du contenu. Nous avons évoqué plusieurs fois le problème de la croissance rapide de la quantité des documents électroniques à archiver. La responsabilité revient, en partie non négligeable, à ces instantanés ; de plus, ils contiennent beaucoup de redondances : d'après certaines enquêtes 25 % des informations sont répétées¹²³.

¹²¹ Peter LYMAN, "Archiving the World Wide Web" in *Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving*; Council on Library and Information Resources and Library of Congress; Washington D.C. 2002, p. 38

¹²² Maureen PENNOCK, "Web-Archiving" (DPC Technology Watch Report 13-01, 2013), pp. 3-4.

¹²³ Daniel GOMES, André L. SANTOS, Mário J. SILVA, *Managing Duplicates in a Web Archive*, p. 4. Accessible: <http://visibilidade.net/daniel/docs/presentations/gomes06duplicatesPPT.pdf>

Nombreux sont les contenus protégés par un mot de passe et les pages web renvoient souvent à des bases de données, d'où peuvent être tirées les informations. Les bibliothèques et les archives qui pratiquent la collecte générale des pages web ne peuvent guère atteindre ces contenus « cachés ». La collecte peut suivre deux méthodes techniques : opération côté client et opération côté serveur. La collecte côté client signifie que le contenu est copié « du dehors ». C'est la démarche la plus fréquente ; elle se fait automatiquement par des moteurs de recherche. Les contenus « cachés » ne sont accessibles que par la collecte côté serveur, quand la copie est faite directement du serveur, ce qui suppose la coopération du producteur¹²⁴.

En collectant des pages web, on se heurte facilement au droit d'auteur. Obtenir l'autorisation de chaque possesseur de droit est difficile, voire impossible si la collecte porte sur un grand nombre de pages. Les institutions qui opèrent à grande échelle, comme l'*Internet Archive* qui fait une collecte mondiale¹²⁵, poursuivent souvent une politique « opt out », c'est-à-dire qu'elles effacent le contenu si le possesseur du droit le demande. Les Bibliothèques et les Archives nationales sont, en général, habilitées par la loi à collecter les pages qu'elles entendent conserver.

Comme l'apparence et la forme revêtent une plus grande importance dans le cas des pages web que dans celui des documents électroniques ordinaires, il faut s'efforcer d'archiver la forme aussi. (Les formats généralement utilisés sont HTML et XML.) Le fait que l'apparence dépende du navigateur peut causer du souci, mais puisqu'on ne peut envoyer un navigateur optimal pour chaque utilisateur, il faut trouver une solution de compromis.

14 La préservation des supports physiques des documents électroniques

Bien que nous ayons souligné ci-dessus l'importance de la conservation archivistique et technique des documents électroniques, cela ne signifie pas pour autant qu'il faille négliger la conservation physique des supports qui portent les données.

Nous pouvons distinguer deux sortes de supports utilisés pour stockage de longue durée, le support magnétique et le support optique. Dans le premier groupe, une bande ou un disque (dur) magnétique contient les données, dans le deuxième c'est un CD ou un DVD,

¹²⁴ Maureen PENNOCK, *op. cit.* p. 7.

¹²⁵ Une donnée de 2013: il archive plus de 150 milliards de pages web. Maureen PENNOCK, *op. cit.* p. 5.

l'enregistrement et la lecture se faisant par rayon laser. Lors du choix du support, il faut tenir compte des critères suivants : stabilité, durée de vie, diffusion (normalisation), capacité, maniabilité et aussi altérabilité des données. Ce dernier critère concerne la vigilance sur l'authenticité. En général, on peut dire que les bandes et les disques magnétiques, enclins à la démagnétisation, sont moins stables que les supports optiques, leur durée de vie est plus courte, ils sont difficilement maniables (l'accès aux données est lent, disques durs exceptés) et les données sont modifiables, mais ils sont fort répandus et leur capacité est grande. Les disques optiques, quoique très vulnérables aux effets mécaniques, sont plus stables, leur durée de vie est plus longue, ils sont utilisés à grande échelle, leur capacité est relativement petite (mais elle croît rapidement), on peut les manipuler facilement et les données ne sont pas altérables (sauf les disques réinscriptibles).

Les supports aussi changent et se développent très rapidement, les normes tombent en désuétude en dix à quinze ans, ce qui signifie que les supports que nous utilisons aujourd'hui ne seront probablement plus lisibles dans vingt ans. Pour cette raison, les affirmations des fabricants de certains supports, CD et DVD notamment, selon lesquelles leurs produits dureront cent ans, n'ont qu'une importance théorique. En effet, l'obsolescence du matériel et du logiciel sera certainement plus rapide. Il ne sert donc à rien d'avoir un support durable si le matériel et le logiciel compatibles font défaut. Cela ne signifie pas pour autant que nous pouvons utiliser n'importe quel support, parce qu'un support de mauvaise qualité peut se dégrader en quelques années¹²⁶.

Il y a un rapport étroit entre la capacité et la maniabilité du support : une quantité d'information de 10 téraoctets peut être écrite sur 13000–16000 CD (650-850 Mo) ou 1200-2200 DVD (4,7-8,5 Go) ou 140 bandes DAT/DLT (72 Go). Comme il faut surveiller la collection de façon systématique, c'est-à-dire tester chaque support de temps en temps, ce n'est pas la même chose de contrôler 16 000 CD ou une centaine de bandes.

On oublie souvent que les opérations de conservation exigent matériel et logiciel. Si on travaille avec beaucoup de CD, on a besoin d'un outil approprié, par exemple un juke-box. On peut acheter des logiciels spéciaux pour la surveillance et la conservation (*repository management software*). Ces logiciels identifient automatiquement des données, contrôlent leur intégrité, procèdent en temps voulu au rafraîchissement et, conformément aux instructions, à

¹²⁶ Un sondage en 2005 dans les collections de CD des archives départementales en France a montré que plus de 20 % des disques étaient en mauvais état de conservation. Accessible: <http://www.archivesdefrance.culture.gouv.fr/static/1053>

la migration des fichiers¹²⁷. Il est impératif de fixer à l'avance, dans un règlement, la périodicité des contrôles ainsi que le niveau quantitatif et qualitatif des erreurs qui déclencheront l'intervention. Il y a des normes et des recommandations internationales sur le mesurage des erreurs, il importe d'en tenir compte¹²⁸.

Il est fortement recommandé de conserver les documents électroniques en plusieurs exemplaires. La solution idéale consiste à garder l'original dans des conditions optimales, à conserver une copie de sécurité dans un lieu éloigné et à disposer d'une copie d'utilisation. Les trois exemplaires peuvent être sur des supports différents, par exemple l'original sur DVD, la copie de sécurité sur bande, et la copie d'utilisation sur disque dur.

Il faut réduire au minimum l'utilisation des bandes magnétiques. Afin d'atténuer les tensions accumulées dans les bandes, il faut les rembobiner de temps en temps (tous les ans ou tous les deux ans). Il faut veiller à l'entretien et surtout au nettoyage des appareils, parce qu'un lecteur sale peut irrémédiablement endommager le support.

Afin d'éviter les pertes d'information, les supports magnétiques ne doivent pas être stockés à proximité de câbles, de blocs d'alimentation ou d'autres appareils qui génèrent des champs magnétiques. Le volume des pertes éventuelles dépend de la force du champ, de la distance et de la durée de l'exposition, en sorte que même un effet faible peut être nocif s'il dure longtemps. Les disques et les bandes (cartouches) doivent être stockés en position verticale.

Il faut éviter la température et l'humidité relative élevées ainsi que leurs fluctuations, parce qu'elles peuvent fragiliser la matière du support, décoller la couche magnétique ou agglutiner la bande. Les normes climatiques diffèrent, selon qu'il s'agisse de l'original, de la copie de sécurité ou de la copie d'utilisation. Pour cette dernière, les conditions normales d'un bureau (20-24 °C, 55 % RH) sont acceptables. En revanche, pour le stockage pérenne, les exigences sont beaucoup plus sévères : pour les supports magnétiques la température recommandée est de 18±1 °C et l'humidité relative doit être de 45±5 %¹²⁹.

¹²⁷ Kevin BRADELY, "Risks Associated with the Use of Recordable CD and DVD as Reliable Storage Media in Archival Collections – Strategies and Alternatives", UNESCO, Paris 2006, p. 21.

¹²⁸ Kevin BRADELY, *op. cit.* p. 15.

¹²⁹ Il faut dire que les normes nationales recommandent des paramètres bien différents surtout dans le cas de l'humidité relative.

Si la différence entre les températures du lieu de stockage et du lieu d'utilisation dépasse 8°C, il faut interposer un intervalle d'acclimatement, c'est-à-dire un stockage à une température intermédiaire pendant plusieurs heures.

Les gaz corrosifs de l'air urbain, même en basse concentration, peuvent endommager les bandes magnétiques. Pour cette raison, l'application de filtres efficaces dans les salles de stockage est recommandée. Les supports magnétiques sont aussi à protéger de la lumière, surtout du soleil. Il faut veiller à la propreté parce qu'une surface magnétique poussiéreuse ou grasse peut induire des processus chimiques nocifs.

Les supports optiques craignent particulièrement la flexion et la rayure. Il est recommandé de porter des gants doux quand on manipule les disques afin de les protéger des empreintes qui peuvent les rendre illisibles. Au nettoyage, il ne faut pas utiliser des solvants, la poussière ou la pollution légère doivent être éliminées par pinceau ou par textile. Il est très important que les lecteurs et les graveurs soient propres et entretenus. Il ne faut pas laisser les disques dans les lecteurs.

Les exigences climatiques pour le stockage des disques optiques sont pareilles à celles des bandes magnétiques. La conservation à basse température ralentit le processus de vieillissement des supports. La stabilité des conditions climatiques est importante aussi. La fluctuation de l'humidité relative ainsi que de la chaleur accélèrent les processus nocifs. L'humidité relative au-dessus de 60 % facilite le développement des moisissures. Un léger mouvement d'air dans le magasin peut empêcher l'établissement des microclimats nocifs.

Le soleil direct est dangereux pour les disques aussi. Il ne faut pas les stocker en tas parce qu'ils peuvent se déformer. Leurs boîtes plastiques d'origine les protègent de la déformation.

15 La transformation de l'archivistique à l'ère des documents électroniques

Le monde numérique a transformé l'archivistique, parce que les documents eux-mêmes se sont transformés : leur structure est devenue plus compliquée et leurs éléments, dont l'importance a changé, se sont séparés. Les métadonnées ont reçu un rôle éminent, les caractéristiques techniques des documents sont passées au premier plan des préoccupations du conservateur et la normalisation est devenue une exigence fondamentale. Nous avons parlé plus haut des difficultés d'assurer l'intégrité et l'authenticité des documents (chapitres 9 et

10) ; cette fois-ci nous voudrions mettre en lumière deux facteurs qui ont aussi influencé considérablement la transformation des méthodes du travail archivistique : d'une part la perte de la troisième dimension, d'autre part l'explosion documentaire.

Les documents électroniques, s'ils sont affichés, se présentent en deux dimensions, ce qui signifie que l'orientation dans une collection est très difficile. Si nous trouvons des milliers de fichiers en désordre dans un répertoire de notre ordinateur, nous pouvons les considérer comme perdus. Par contre, le classement de la même quantité (quelques cartons) de documents papier en désordre sera probablement faisable. Nous pouvons mettre de l'ordre dans les documents électroniques, nous pouvons nous y orienter et y faire des recherches, à condition que chaque pièce soit munie des métadonnées appropriées, ce qui n'est pas une exigence dans le monde analogique.

Nous avons mentionné plus haut que pendant des siècles, les archivistes s'étaient appliqués à décrire les documents pièce par pièce. La description, dont le rôle est repris dans le monde numérique par les métadonnées, sert en même temps à accéder aux données recherchées et à assurer l'authenticité des documents. La cote (l'identifiant) d'un document fixe aussi sa place, sa position dans le système d'archives. Faisaient partie des descriptions d'archives typiques, médiévales et modernes, les catalogues, répertoires et inventaires. Les règles de l'analyse scientifique des documents individuels ont été formulées par la diplomatie.

La progression de l'usage de l'écriture amena la croissance rapide de la quantité des documents à conserver, ce qui nécessita le développement de nouveaux systèmes d'enregistrement et de recherche. La bureaucratie ecclésiastique et étatique, de plus en plus étoffée, abandonna peu à peu la description pièce par pièce et passa à l'organisation de séries regroupant les documents d'après les sujets ou des critères formels. Puis, dans la plupart des pays européens, les administrations adoptèrent la méthode visant à constituer des dossiers contenant l'ensemble de pièces créées et reçues concernant la même transaction.

Dans le monde analogique, le dossier est l'élément de base sur lequel se fondent toutes les entités imposées par le principe de provenance.

Il faut dire qu'au milieu du XX^e siècle, il y avait un certain retour en arrière dans l'approche des archivistes, quand ils nouaient des relations étroites avec les records managers des organismes producteurs afin de s'assurer que les versements à venir fussent de bonne qualité. Le records manager s'occupe essentiellement des documents individuels et son

horizon ne s'étend pas au-delà du fonds. Quant à l'archiviste, une fois le versement accueilli, il a la tâche de l'intégrer dans le système d'archives qui englobe un grand nombre de fonds. Pour cette raison, les structures externes des documents, les relations entre les documents et les agrégats de documents, sont beaucoup plus importantes pour l'archiviste que la structure interne des documents.

Avec les documents électroniques, la situation a changé : la modification de la structure, la séparation du contenu, de la forme et du support, l'apparition des spécifications techniques (informatiques), la vulnérabilité de tous les éléments, ainsi que les erreurs cachées ont, de nouveau, mis au premier plan l'analyse de la structure interne des documents individuels¹³⁰. Puisqu'à cause de « l'explosion documentaire »¹³¹ le traitement manuel des documents électroniques est devenu hors d'atteinte, on avait besoin de méthodes (normes) et d'outils (logiciels) rendant possible l'automatisation des procédures : munir les documents de métadonnées, contrôler l'intégrité des documents lors du versement et de la conversion, détecter les erreurs cachées¹³². Comme les méthodes et techniques disponibles de l'archivistique s'avéraient inadaptées à ces tâches, les archivistes devaient recourir aux méthodes de la diplomatique et, plus encore, à l'aide de la science de l'information. Le plus grand apport de la science de l'information était de pouvoir organiser les données de façon à rendre possible l'automatisation. (En ce qui concerne l'approche, document individuel ou agrégat, la science de l'information est neutre parce qu'elle traite des deux). L'influence de la diplomatique est évidente, notamment, sur le projet InterPARES qui s'en réclame¹³³, sur certains acquis de la science de l'information telles les propriétés significatives de l'InSPECT (voir le chapitre 9) ou sur l'OAIS complet.

La mise en place d'un système d'archives, se composant de nombreux fonds, demande un travail de synthèse qui ne relève ni de la diplomatique ni de la science de l'information. Quand l'archiviste, ayant étudié le fonctionnement d'un organisme à travers les documents produits, le situe dans la conjoncture politique et économique du moment, puis intègre le

¹³⁰ Ce développement de l'archivistique peut aussi être décrit par la dialectique hégélienne (thèse, antithèse et synthèse).

¹³¹ Sur le nombre des documents électroniques produits au monde, voir par ex. http://www.planets-project.eu/docs/comms/PLANETS_BROCHURE.pdf, p. 3. ou les données statistiques sur les e-mails mentionnées au chapitre 13.4.1.

¹³² Dans le monde analogique cela n'était pas nécessaire, seule la conservation de la structure externe des documents était importante, la sauvegarde de la forme ou la structure interne étaient évidentes si le document restait.

¹³³ Voir par ex. http://www.interpares.org/display_file.cfm?doc=ip1_atf_report.pdf

nouveau fonds dans la macrostructure, il fait ce travail de synthèse. Il fait aussi un travail de synthèse quand il résume les résultats de cette étude dans l'histoire de l'organisme (contexte) et fournit ainsi aux chercheurs futurs des informations importantes, il est vrai subjectives. Dans un avenir prévisible, l'ordinateur ne pourra effectuer un tel travail de synthèse, car son manque de souplesse le prive de la capacité d'abstraire comme un homme créatif. Il est toutefois capable d'opérer des abstractions limitées, par exemple filtrer un océan de données pour en ressortir les corrélations sélectionnées et programmées au préalable. Dans le monde numérique, l'archiviste doit se servir des méthodes et des outils de la science de l'information sans perdre de vue sa tâche de synthèse d'évaluation et de description aux niveaux élevés des agrégats de documents (fonds, sous-fonds). S'il ne parvient pas à maintenir l'équilibre, la complétude de son travail d'archiviste s'ébrèche et dans ce cas nous pouvons parler d'un conflit entre le paradigme d'archives et le paradigme de records management.

Selon les tenants du paradigme d'archives, les archives constituent l'ensemble organisé des documents gouvernementaux et privés devenus inactifs, indépendamment du fait qu'ils soient versés aux archives ou non. Dans cette conception, les fonctions probante et mémorielle sont en équilibre et sont également importantes. Par contre, selon les tenants du paradigme de records management, les documents sont en premier lieu les preuves des procédures suivies et des transactions opérées par le producteur et les organismes ou personnes intervenant dans l'affaire. Leur conservation est motivée, avant tout, par des raisons administratives et légales¹³⁴. Les partisans de cette conception érigent la fonctionnalité, la valeur probante, le contexte et les caractéristiques formelles et structurelles des documents en valeurs absolues et, en même temps, dégradent le rôle du contenu, la fonction mémorielle. Quand ils analysent ou évaluent les documents, ils s'efforcent de s'appuyer, autant que possible, sur des faits objectifs (fonctionnalités, règles juridiques, caractéristiques formelles), par lesquels ils croient arriver à « la vérité absolue » des documents (que sont-ils ? lesquels sont précieux et lesquels sont sans valeur ?)¹³⁵. Ajoutons que cette conception qui se focalise sur les documents individuels néglige plus ou moins la structure externe, les agrégats des

¹³⁴ Mark A. GREENE, "The Power of Meaning: The Archival Mission in the Postmodern Age" in *American Archivist*, vol. 65. 2002, p. 45. Selon Greene un représentant éminent du paradigme d'archives courantes est Luciana Duranti, le spiritus rector du projet InterPARES.

¹³⁵ Mark A. GREENE, *op. cit.* p. 53.

documents. En repensant à l'attachement étroit du projet InterPARES à la diplomatie, nous arrivons à la conclusion que le conflit de paradigmes esquissé existe bel et bien¹³⁶.

16 L'enjeu de la conservation pérenne des documents électroniques

Après la conquête de l'Angleterre, Guillaume le Conquérant ordonna d'évaluer et de recenser les biens de son nouveau royaume ; c'est ainsi que le Domesday Book, cette source incomparable de l'Angleterre du XI^e siècle naquit en 1086. Aujourd'hui, il est l'un des plus précieux trésors des Archives nationales du Royaume-Uni. En 1986, le BBC a lancé un nouveau projet appelé Domesday¹³⁷ se fixant comme objectif de documenter - à l'aide de l'informatique - la vie ordinaire du Royaume-Uni à la fin du siècle passé. On partageait le pays en secteurs et, à l'aide des écoles ainsi que des communautés locales, on réunit sur ces secteurs cent cinquante mille pages de texte et vingt-trois mille photos d'amateur. Toute la documentation était publiée sur des disques optiques de 12 pouces, une technologie très moderne en ce temps-là. Les organisateurs envisageaient de répéter le projet tous les vingt-cinq ans, pour illustrer les changements. Mais en 1999, il s'est avéré que l'environnement informatique (matériel et logiciel) du projet de 1986 n'existant plus, pratiquement toute la documentation était devenue inaccessible. (Alors que les documents de Guillaume restaient bien lisibles après plus de neuf cent treize ans). En 1999, on a lancé le projet CAMiLEON¹³⁸ et, grâce aux spécialistes britanniques et américains, on a réussi, après plusieurs années de travail, à ranimer (par émulation) la documentation. Le projet Domesday est devenu le symbole de l'obsolescence numérique.

Au tournant de 2002 et 2003, les informaticiens de l'Office exécutif du président des États-Unis ont changé le matériel et le logiciel du bureau. Le résultat fut la disparition de vingt-deux millions d'e-mails¹³⁹.

Les deux exemples ci-dessus illustrent très bien les conséquences choquantes de la destruction massive des documents électroniques. Dans le chapitre 5.2, nous avons exposé

¹³⁶ Certaines études d'InterPARES mentionnent d'une façon marginale les fonds et l'importance de la description archivistique aux hauts niveaux des documents. Voir par ex. Luciana DURANTI, Randy PRESTON (ed.), *InterPARES2: Experimental, Interactive and Dynamic records* (Associazione Nazionale Archivistica Italiana, 2008), p. 682.

¹³⁷ Voir <http://www.bbc.co.uk/history/domesday/story>

¹³⁸ Voir <http://www.ariadne.ac.uk/issue29/camileon>

¹³⁹ D. GEWIRTZ, *Where Have All the Emails Gone?*, Palm Bay, Fla.: Zatz. (2007). Cité par Christopher J. PROM, *Preserving email* (DPC Technology Watch Report 11-01, 2011), p. 37.

l'importance des valeurs administrative/pratique, probante et historique des documents. Ces notions professionnelles arides couvrent cependant des valeurs fondamentales ainsi que des risques et des conséquences graves.

Si des documents d'archives importants disparaissent, la garantie des droits des citoyens décline et la possibilité de connaître le passé diminue. Les conséquences en sont : 1) le contrôle des pouvoirs publics s'affaiblit, ce qui facilite les abus ; 2) l'étrécissement du champ des recherches historiques affectera nécessairement la conscience politique et la psychologie collective des générations à venir. C'est pourquoi on tient l'accès aux documents d'archives et, par là même les archives, pour des garants de la démocratie. On peut infailliblement reconnaître les systèmes politiques autoritaires : si le pouvoir essaie d'empêcher les citoyens par des textes législatifs ou réglementaires d'avoir accès aux documents – les arguments sont toujours abondants – la démocratie est en danger.

La préservation des documents et l'accès aux sources historiques ont un aspect beaucoup plus général qui affecte l'humanité dans son ensemble. Pour que la civilisation humaine persiste et progresse, le savoir acquis doit se conserver et se répandre aussi largement que possible sur les plans géographique et social, en sorte que les générations successives puissent accumuler des connaissances nouvelles. Dans l'histoire de la civilisation, trois impulsions de vigueur exceptionnelle ont fait avancer à pas de géant la sauvegarde et la diffusion du savoir : l'écriture, l'imprimerie et l'informatique. Par l'écriture, la plus grande invention dans l'histoire des civilisations, le savoir put être enregistré, jour après jour. En ce qui concerne l'imprimerie, elle joue un double rôle : elle conserve et en même temps elle diffuse le savoir. L'informatique créa un nouvel environnement pour le développement. En premier lieu, en incorporant des mécanismes de la réflexion humaine dans les équipements et en combinant ceux-ci avec des machines effectuant un travail physique, elle accéléra le développement de façon spectaculaire. (Pensons par exemple aux robots commandés par logiciel dans l'industrie d'automobile). De notre point de vue, il est plus important que, grâce à l'informatique, la fixation et la diffusion du savoir soient devenues incroyablement efficaces. Étant au début du processus, nous n'entrevoions qu'une fraction des potentialités.

Nous avons déjà évoqué le thème de la numérisation de la mémoire et de la communication humaine dans le chapitre 6.2. Les archives et les bibliothèques abritent, traditionnellement, les plus grandes concentrations du savoir humain. À ces trésors viennent de s'ajouter les bases de données scientifiques. Pendant sa courte histoire de quelques décennies, l'informatique a donné une impulsion géante à l'astronomie, à la physique

nucléaire, à la biologie, à la médecine, à la géologie, etc., qui ne pourraient plus exister sans bases de données, sans logiciels pour modeler des processus et sans réseaux professionnels¹⁴⁰.

Toutefois, la continuité du développement n'est nullement assurée, parce que, comme nous l'avons vu dans les chapitres précédents, la conservation est le point faible de l'informatique, et plus grande est la quantité de l'information numérique (le savoir) et plus grande peut être la perte. Si le savoir accumulé disparaît, les nouvelles générations perdront la base sur laquelle elles pourraient bâtir et, puisqu'elles ne posséderont plus le savoir de leurs prédécesseurs, cet appauvrissement conduira inévitablement au déclin. Bien sûr, ce danger ne surviendra pas du jour au lendemain, mais tout document, base de données ou image détruit ou devenu inaccessible mutile le capital de savoir sur lequel l'humanité peut fonder le progrès¹⁴¹.

Ayant reconnu ce danger, le monde scientifique est intervenu activement. L'OAIS, norme d'archives électroniques généralement utilisée, a été développé à l'origine pour l'archivage des systèmes de données spatiales. Parmi les contributeurs actuels, il n'y a aucun service d'archives. Le Dublin Core est le fruit du travail des bibliothécaires, le PREMIS est né de la coopération de bibliothécaires, d'archivistes et d'experts universitaires ; la Bibliothèque du Congrès apparaît comme la principale institution organisatrice de l'élaboration de nombreuses normes ; dans les groupes de travail du MOREQ participent des archives, des universités et des firmes informatiques, et nous pourrions continuer l'énumération. Les gouvernements, les organisations professionnelles informatiques et l'Union européenne ont investi des centaines de millions d'euros et beaucoup d'énergie dans la recherche pour assurer la préservation des documents électroniques. Nous sommes les témoins d'une coopération sans précédent dans l'histoire des archives. Bien que la percée ne soit pas encore intervenue, nous pouvons être beaucoup plus optimistes que par le passé. Et surtout, il faut persévérer.

¹⁴⁰ D'après certaines estimations, la quantité des bases de données scientifiques double chaque année grâce aux capteurs de plus en plus bon marché. Voir <https://www.sciencenews.org/article/preserving-digital-data-future-escience>

¹⁴¹ Naturellement, le développement a des conditions politiques, économiques, sociales, culturelles même ; il dépend de l'environnement naturel aussi, mais ceci se situe en dehors de nos recherches.

BIBLIOGRAPHIE SÉLECTIVE

(Toutes les références aux sites Internet datent d'avril 2014)

Première partie : Archives et histoire

Actes de la vingt-deuxième Conférence internationale de la Table ronde des archives. Bratislava 1983. L'archiviste et l'inflation des documents contemporains. Conseil international des Archives, Paris, 1984, 204 p. (pp. 104: Version française; pp. 105: English version).

Gérald ARLETTAZ, *Libéralisme et société dans le canton de Vaud, 1814-1845.* Lausanne 1980, 742 p. (Bibliothèque historique vaudoise, N° 67).

David BEARMAN, „An Indefensible Bastion: Archives as a Repository in the Electronic Age” in David BEARMAN (ed.), *Archives and Museum Informatics, Technical Report 13*, 1991.

A. de BOÜARD, *Manuel de diplomatique française et pontificale. I. Diplomatie générale.* Aguste Picard, Paris, 1929, 397 p.

John BURROW, *A History of Histories. Epics, Chronicles, Romances and Inquiries from Herodotus ans Thucydides to the Twentieth Century.* Penguin Books, Lonon, 2009, xix-553 p.

E.H. CARR, *What is History? The George Macaulay Trevelyan Lectures Delivered in the University of Cambridge, January-March 1961.* - London, Penguin Books, 1973, 169 p. (1ère édition: Macmillan, 1961).

Jean-Claude CARRIÈRE, Umberto ECO, *N'espérez pas vous débarrasser des livres.* Grasset, Paris 2009, 283 p.

Marie-Anne CHABIN, Peut-on parler de diplomatique numérique.

<http://www.marieannechabin.fr/diplomatique-numerique/>

A. CHAMPOLLION-FIGEAC, *Notice sur les Archives départementales en France*, Grenoble 1887, p. 8-9.

Sophie CŒURÉ, *La mémoire spoliée. Les archives des Français, butin de guerre nazi puis soviétique (de 1940 à nos jours).* Payot, Paris, 2007, 270 p.

Terry COOK, *Archival Science and Postmodernism: New Formulations for Old Concepts* <http://www.mybestdocs.com/cook-t-postmod-p1-00.htm>

Terry COOK, "What is Past is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift" in *Archivaria*, 43, p. 17-63.

Marie CORNU - Jérôme FROMAGEAU (dir.) *Archives et patrimoine 1-2.* L'Harmattan, Paris, 2004, 188 et 188 p. Actes du colloque, Ajaccio, décembre 2001.

Christian DELACROIX, François DOSSE, Patrick GARCIA, *Les courants historiques en France. XIX^e-XX^e siècle.* Édition revue et augmentée. Gallimard, Paris, 2007, 724 p. (Folio Histoire).

Bruno DELMAS, *La société sans mémoire. Propos dissidents sur la politique des archives en France.* Bourin, Paris, 2006, 203 p.

Paul DELSALLE, "L'archivistique sous l'Ancien Régime, le Trésor, l'Arsenal, et l'Histoire" in *Histoire, économie et société*, 1993, 12-4, pp. 447-472.

Direction des Archives de France, *Manuel d'archivistique. Théorie et pratique des Archives publiques en France*. S.E.V.P.E.N., Paris, 1970, 805 p.

Michel DUCHEIN, *Les Bâtiments et équipements d'archives*, Paris, Conseil international des Archives, 1966, 314 p.

Michel DUCHEIN, *Les Bâtiments d'archives, construction et équipements*, Paris, Direction des archives nationales, la Documentation française, 1985, 256 p.

Christian DUVERGER, *Cortés et son double. Enquête sur une mystification*. Seuil, Paris, 2013, 317 p.

Lucie FAVIER, *La mémoire de l'État. Histoire des Archives nationales*. Fayard, Paris, 2004, V-466 p.

Fondation Singer-Polignac, *Les archives au fil du temps*. Actes du colloque, Paris, le 26 février 2002, Perrin, Paris, 2002, 242 p.

Carlo GINZBURG, *Le fromage et les vers. L'univers d'un meunier du XVI^e siècle*. Flammarion, Paris, 1980, 220 p.

Patricia Kennedy GRIMSTED, *Trophies of War and Empire. The Archival Heritage of Ukraine, World War II and the International Politics of Restitution*. Harvard University, Cambridge, Mass., 2001, xlvii-749 p.

Petra HEINE - Jürgen BACIA, "Plädoyer für die Bewahrung des Geschichte von unten" in *Der Archivar*, 50, 1997, 2, col. 299-326.

Charles KECSKEMÉTI, „Disputed archival claims. Activities of UNESCO and ICA since 1976” in *Sovereignty Disputed Claims Professional Cultures. Essays on Achival Policies*, Brussels, 2000, 365 p.

Simon-Daniel KIPMAN, *L'oubli et ses vertus*. Albin Michel, Paris, 2013, 225 p.

György KÖVÉR, *A tisztaeszlári dráma* [Le drame de Tiszaeszlár]. Osiris, Budapest, 2011, 749 p.

Pierre Camille LE MOINE, *Diplomatique pratique, ou traité de l'arrangement des archives et trésors des chartes*, chez Joseph Antoine, Metz, 1765, in 4° 396 p. Sous-titre: ouvrage nécessaire aux commissaires à terriers, aux dépositaires des titres des anciennes seigneuries, des Évêchés, des Chapitres, de Monastères, des Communautés, des corps de ville, et à tous ceux qui veulent s'adonner à l'étude des Monuments de l'Antiquité.

Elio LODOLINI, *Archivistica. Principi e problemi*, Franco Angeli, Milano, 1987, 315 p.

Elio LODOLINI, *Storia dell'arcivistica italiana. Dal mondo antique alla metà del secolo XX*. - Milano, Franco Angeli, 2001, 316 p.

Richard STAPLETON, "Jenkinson and Schellenberg: a Comparison" in *Archivaria*, 17, pp. 75-85.

Alexandre SUMPF, Vincent LANIOL (dir.). *Saisies, spoliations et restitutions. Archives et bibliothèques au XX^e siècle*. Presses Universitaires de Rennes, Rennes, 2012, 383 p.

Alexis de TOCQUEVILLE, *L'Ancien Régime et la Révolution dans Œuvres III*, Gallimard, 2004, p. 41-313 (Bibliothèque de la Pléiade).

Enzo TRAVERSO, *Le passé modes d'emploi histoire, mémoire, politique*. La fabrique, Paris, 2005.

Peter WALNE (ed.), *Dictionary of Archival Terminology, English and French with Equivalents in Dutch, German, Italian, Russian and Spanish*. K.G. Saur, München –New York-London-Paris, 1984, 226 p. (ICA Handbooks Series. Volume 3).

Deuxième partie: Les Archives dans le monde numérique

Kevin BRADELY, *Risks Associated with the Use of Recordable CD and DVD as Reliable Storage Media in Archival Collections – Strategies and Alternatives*, UNESCO, Paris 2006

Frédéric BRÉGIER et al., *L'archivage des messageries électroniques, Preuve de concept VITAM*, Version 1.1, 2013

Heather BROWN et al., *Instalment on "The role of microfilm in digital preservation"* (DCC Digital Curation Manual 2011). Accessible:

http://www.dcc.ac.uk/sites/default/files/documents/Microfilm_2011_Final.pdf

Adrian CUNNINGHAM: "Dynamic Descriptions, Australian Strategies for the Intellectual Control of Records and Recordkeeping Systems" in P. J. Horsman – F. C. J. Ketelaar – T. H. P. M. Thomassen [ed.] *Naar een nieuw paradigma in de archivistiek*, Jaarboek 1999 Stichting Archiefpublicaties, 's-Gravenhage 1999

Michael DAY, *Instalment on "Metadata"* (DCC Digital Curation Manual 2005). Accessible: <http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/metadata/metadata.pdf>

Luciana DURANTI, Randy PRESTON (ed.), *InterPARES2: Experimental, Interactive and Dynamic records* (Associazione Nazionale Archivistica Italiana, 2008) Accessible: http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_complete.pdf

Luciana DURANTI, Kenneth THIBODEAU, "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES" in *Archival Science* 6, 1 (2006)

(EAD) <http://www.loc.gov/ead/>

Terry EASTWOOD, "The Appraisal of Electronic Records: What is New?" in *Comma* 2002/1-2

Mark A. GREENE, "The Power of Meaning: The Archival Mission in the Postmodern Age" in *American Archivist*, vol. 65. 2002

Guide pour la gestion archivistique des documents électroniques, Conseil international des archives, Études 8 (1997)

Helen HESLOP, Simon DAVIS, Andrew WILSON, *An Approach to the Preservation of Digital Records*, National Archives of Australia, December 2002. Accessible: http://www.naa.gov.au/Images/an-approach-green-paper_tcm16-47161.pdf

David HOLDSWORTH, *Instalment on "Preservation Strategies on Digital Libraries"* (DCC Digital Curation Manual 2007). Accessible:

<http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/preservation-strategies/preservation-strategies.pdf>

InterPARES. *Conditions requises pour évaluer et maintenir l'authenticité des documents d'archives électroniques*. Accessible: http://www.interpares.org/display_file.cfm?doc=ip1_authenticity_requirements%28french%29.pdf

ISAAR (CPF) *Norme Internationale sur les notices d'autorité utilisées pour les Archives relatives aux collectivités, aux personnes ou aux familles* (Conseil international des archives), 2004. Accessible: <http://www.ica.org/10241/normes/liste-des-normes.html>

ISAD(G): *Norme générale et internationale de description archivistique* (Conseil international des archives) Ottawa, 2000. Accessible: <http://www.ica.org/10241/normes/liste-des-normes.html>

Andrew MCDONALDS (éd.), *Les archives électroniques, Manuel à l'usage des archivistes* (ICA Studies 2005). Accessible: <http://www.ica.org/10801/studies-and-case-studies/ica-study-n16-electronic-records-a-workbook-for-archivists.html>

MOREQ. Accessible: http://www.moreq.info/index.php?option=com_content&view=article&id=17&Itemid=22&language=e

OAIS. Accessible: http://pin.association-aristote.fr/lib/exe/fetch.php/public/documents/norme_oais_version_francaise.pdf

Maureen PENNOCK, *Instalment on "Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages"* (DCC Digital Curation Manual 2006). Accessible: <http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf>

Maureen PENNOCK, *Web-Archiving* (DPC Technology Watch Report 13-01, 2013) Digital Preservation Coalition. Accessible: <http://www.dpconline.org/advice/technology-watch-reports>

Christopher J. PROM, *Preserving email* (DPC Technology Watch Report 11-01, 2011). Accessible: <http://www.dpconline.org/advice/technology-watch-reports>

Jeff ROTHENBERG, *Avoiding Technological Quicksand: Finding a Viable Foundation for Digital Preservation* (Council on Library and Information Resources, Washington 1999). Accessible: <http://www.clir.org/pubs/reports/rothenberg/contents.html>

Malcolm TODD, *File Formats for Preservation* (Technology Watch Report, Digital Preservation Coalition 2009). Accessible: <http://www.dpconline.org/advice/technology-watch-reports>

XML. Accessible: <http://www.w3.org/standards/xml/>