

Avant-propos

Archiver du numérique n'est pas stocker les données, mais plutôt les garder de façon intelligente pour pouvoir les exploiter dans le temps tout en conservant leur intégrité et leur authenticité. Avec les transformations rapides dues aux usages des outils informatiques, plusieurs types de documents sont aujourd'hui versés et conservés aux services d'archives : email, bases de données, photographies numériques, audiovisuel numérique, échanges au sein des médias sociaux, etc. Nous assistons ainsi à une évolution des notions et pratiques des sciences humaines et sociales vers ce que nous appelons « les humanités numériques ». Cela a permis de développer de nouveaux outils et applications qui favorisent l'accès aux archives et leur mise en valeur. En parallèle, il y a eu une explosion au niveau des documents et flux d'information émanant des technologies mobiles, des médias sociaux, des transactions en ligne, des capteurs connectés, etc., dont nous devons en assurer la collecte, la conservation et la diffusion. Appelées communément aujourd'hui *Big Data*, ces mégadonnées font appel à des technologies très puissantes, qui ont tendance à tout conserver et pour une longue période. Cependant, ce souci de tout préserver au nom du droit de mémoire est confronté à un autre droit, celui de l'oubli. Ce droit à l'oubli est souvent lié à la liberté des individus qui se voient condamnés à vivre sans vie privée et sans liberté en faisant face aux fautes de leur passé.

Cet ouvrage tente de dresser un panorama général de la complexité d'évolution du concept d'archives numériques à travers trois dimensions : humaine et sociale, technique et juridique. Il présente dans un premier temps les contours et éléments nécessaires de ses définitions, ainsi que les méthodes et stratégies de l'archivage numérique tout en présentant le paysage normatif régissant ce domaine. Par la suite, l'ouvrage détaille les changements subis par les archives avec les humanités numériques sur le plan des contenus, des supports, de la conservation et de l'accès. Ensuite il essaie de montrer dans quelles mesures l'arrivée des Big Data a bouleversé le travail des archivistes. Leur

défi de taille est de traiter plus de données, avec un faible coût et dans une durée plus courte. Dans ce contexte, la *blockchain*, en tant que registre décentralisé et distribué est présentée comme une solution complémentaire à l'archivage numérique. Enfin, une attention particulière sera donnée à la dimension juridique des traitements des archives à travers le droit à l'oubli.

Introduction

Les archives représentent aujourd'hui, dans presque toutes les acceptions des différents pays, l'ensemble des documents, quels que soient leur date, leur nature ou leur support, produits ou reçus par toute personne physique ou morale dans le cadre de ses activités administratives. Ces documents sont conservés pour leur valeur d'information, de preuve et/ou de témoignage. Le terme « archives » renvoie également au lieu où l'on conserve ces documents (bâtiment ou local de conservation). Elles peuvent également constituer, outre des preuves et des garants de droits, des objets de recherche historique et scientifique. Ainsi, la majorité des États ont développé des services d'archives ouverts au public et des textes juridiques qui obligent les administrations à verser à ces services les documents qui n'ont plus de valeur administrative, financière et/ou juridique et qui se dotent d'une valeur historique et/ou scientifique.

Les archives peuvent être publiques, c'est-à-dire créées par les organismes publics ou chargées d'une mission de service public, telles que les établissements publics, les services consulaires, les officiers ministériels, etc. Elles peuvent être privées aussi provenant d'individus, de familles, d'associations ou d'entreprises et sont données aux services d'archives publics à travers des dons, des dépôts, des legs, etc. La distinction entre ces deux catégories n'est pas toujours aisée, et ce, à cause des chevauchements que peuvent avoir des documents par exemple d'un homme politique entre ses fonctions officielles et ses activités privées découlant de sa responsabilité d'un parti ou du parlement.

Par ailleurs, quelle que soit leur catégorie, les archives sont sujettes à deux types de dispositions, d'une part des dispositions concernant les archives courantes et intermédiaires en se basant sur la théorie des trois âges¹ [PER 61] : une première étape

1. Théorie développée par l'archiviste et historien français Yves Pérotin en 1961. Ce modèle connaît une reconnaissance juridique en France avec l'adoption de la loi du 3 janvier 1979 sur les archives.

active dans laquelle elle se dote d'une valeur primaire (administrative principalement avec parfois une valeur financière et/ou légale) et une deuxième étape intermédiaire dans laquelle elle perd légèrement sa valeur primaire et est accessible seulement en cas de besoin. D'autre part des dispositions spécifiques organisant la conservation et la communication des archives définitives par les services d'archives publics (le dernier stade de l'archivage pérenne selon la théorie des trois âges).

À l'heure actuelle, les archives connaissent un passage d'un environnement principalement analogique à un univers de plus en plus numérique, et par conséquent d'un format matériel à un chiffrement en bits et en octets. Ce passage nous met en confrontations au défi d'adapter nos méthodes de travail à ce nouvel environnement dit hybride. Nous parlons ainsi de plus en plus d'archives numériques. Ce terme, dont nous reparlerons en détail plus tard, annonce *a priori* la libération de l'information de ses supports matériels et la dissociation entre le contenu et le médium. Suite au déploiement, dès les années 80, des outils informatiques dans les administrations, les archives nativement numériques se sont considérablement accrues.

Par ailleurs, les services des archives sont appelés à mettre en place des outils et des méthodes permettant de collecter, classer, conserver et communiquer au public ces nouveaux types d'archives. Archiver du numérique n'est pas stocker les données, mais plutôt les garder d'une façon intelligente afin de pouvoir les exploiter dans le temps et de conserver leur intégrité et leur authenticité.

En outre, avec ces transformations rapides dues aux usages des outils informatiques, d'autres types de documents sont aujourd'hui versés et conservés aux services d'archives : email, bases de données, photographies numériques, audiovisuel numérique, échanges au sein des médias sociaux, etc. Avec ces transformations technologiques, nous assistons à une évolution des notions et pratiques des sciences humaines et sociales vers ce que nous appelons « les humanités numériques ». Cela a permis de développer de nouveaux outils et applications qui favorisent l'accès aux archives et leur mise en valeur. Ce champ des humanités numériques peut être repéré à travers le mouvement effréné des technologies mises à la disposition des êtres humains pour produire, gérer et diffuser des savoirs [BOU 17].

De plus, il n'y a pas seulement eu une diversification des documents et des outils technologiques qui ont vu le jour, il y a également eu une explosion au niveau des documents et flux d'information dont nous devons assurer la collecte, la conservation et la diffusion. En effet, les entreprises produisent, chaque jour, un taux très volumineux d'informations qui peut être estimé à des milliers de téraoctets. On parle ainsi de Big Data ou de grosses données sous forme de zetta ou pétaoctets provenant des

différentes technologies mobiles, des médias sociaux, des transactions en ligne, des objets et des capteurs connectés [KAR 14].

Cependant ce souci de tout préserver sur une longue période représenté en général par le droit de mémoire est confronté à un autre droit, celui de l'oubli. Ce droit est souvent lié à la liberté des individus qui se voient condamnés à vivre sans vie privée et sans liberté en faisant face aux fautes de leur passé [ARR 16]. Le droit à l'oubli, généralement présenté comme un droit aux individus à pouvoir tout effacer, faire disparaître et/ou ne pas conserver des informations les concernant, heurte des expressions liées à la mémoire telles que la pérennisation et l'imprescriptibilité.

Nous allons dans cet ouvrage mettre l'accent sur l'éventuelle conciliation entre ces deux principes liés à la mémoire et à l'oubli. Une tâche qui ne semble pas facile, mais possible lorsque nous connaissons déjà tout le travail qui a été fait sur la protection des données personnelles (informatique et libertés) face aux droits d'information, d'accès aux documents administratifs, de transparence, etc. Il faut travailler davantage sur la conformité des textes juridiques liés à la fois au droit de mémoire et au droit à l'oubli. « À partir de là, il n'est plus question de “la mémoire à tout va” [LEG 02], mais d'une mémoire collective respectueuse des libertés des individus et de leur aspiration légitime à se faire oublier de la société » [ARR 16].

Le chapitre 1 présente le concept d'archives numériques, les éléments de ce concept et les outils de base de gestion de celles-ci, à savoir le répertoire des dossiers types, le plan de classification et le calendrier des délais de conservation. Ce chapitre expose également les différentes relations entre les archives numériques et des technologies telles que la gestion électronique des documents et le *Records management*.

Le chapitre 2 traite de façon plus détaillée le concept d'archivage numérique, ses différentes acceptions et les différents standards et normes le régissant tels que la norme NF Z 42-013, devenue la norme internationale ISO 14641, qui met l'accent sur les spécifications et mesures techniques et organisationnelles relatives à l'enregistrement, l'archivage, la consultation et la communication de documents numériques. Il aborde également la norme NF 461 qui représente une certification de la conformité des organismes ainsi que la norme NF Z 42-013 et son équivalent ISO 14641-1 qui concerne le fonctionnement d'un système d'archivage numérique. Par la suite, il présente la norme ISO 14721, appelée aussi « OAIS » (*Open Archival Information System*), la norme ISO 19005, appelée « PDF/A » se rapportant à la pérennisation des données, ainsi qu'une autre série de standards liés aux processus de gestion et de métadonnées des documents d'activité. Une autre partie du chapitre est consacrée à la méthodologie de mise en place d'un système d'archivage numérique. Enfin, une attention particulière est donnée aux actions et processus d'archivage des documents

audiovisuels et des courriers électroniques à l'image de leurs spécificités et leurs places dans l'entreprise d'aujourd'hui.

Le chapitre 3 essaie de positionner la question des archives numériques dans ce que nous appelons aujourd'hui l'ère des humanités numériques. Le défi est de voir dans quelles mesures les humanités numériques ont fait évoluer les archives, leurs outils, leurs fonctionnalités et leurs méthodes de pérennité des documents. Ainsi, une présentation des différentes définitions liées aux humanités numériques ainsi que ses liens avec les archives numériques sera détaillée avec des exemples de plateformes et logiciels tout en montrant comment placer les humanités numériques au cœur de la conservation à long terme.

Le chapitre 4 expose la problématique de la gestion et l'archivage du Big Data. Cette technologie représente l'explosion évolutive des données et des documents numériques dans tous les domaines d'activité ainsi que le rythme effréné de leur production. L'enjeu est de démontrer l'obligation de restructuration des processus et des méthodes d'archivage des données caractérisées par l'important volume, la variété et la vitesse. Une sélection des principaux outils et technologies du Big Data est présentée, et attention particulière est portée à la technologie de la *blockchain* en tant que technologie de traçabilité des données qui, accouplée aux normes de conservation, pourrait représenter l'avenir de l'archivage numérique à l'ère des grosses données.

Pour finir, le chapitre 5 présente une analyse du droit à l'oubli numérique, en opposition au droit de mémoire que représente, à un niveau primaire, l'archivage numérique. Néanmoins, le droit à l'oubli est présenté comme un des principes de l'archivage, tout en faisant la conciliation avec d'autres droits liés à la protection des éléments de la vie privée et des données à caractère personnel. Deux défis sont avancés dans ce contexte, le premier est technique faisant appel aux méthodes et processus techniques d'effectivité. Le deuxième est juridique faisant appel à différents arsenaux juridiques mettant en application ce droit. Des exemples sur les archives publiques et privées en relation avec le droit à l'oubli, ainsi que les efforts fournis par la firme américaine Google dans ce sens, seront également présentés.