

---

## Introduction

Produire des documents dans le cadre d'une activité administrative ou industrielle, créer des supports de communication pour diffuser des idées, constituer des collections de sources d'information pour les mettre à disposition du public, sont autant d'initiatives courantes et anciennes, séculaires voire millénaires, mais surtout en pleine expansion grâce aux énormes possibilités des techniques numériques et des réseaux.

Pour le gestionnaire de l'information qui, quel que soit le support, doit collecter les documents et les mettre à disposition des utilisateurs, les technologies numériques sont paradoxales : d'un côté, elles décuplent les possibilités de production, de diffusion et de stockage, à en donner le vertige ; de l'autre, elles fragilisent les documents/données qui, comparativement à l'ère analogique, perdent en fiabilité et en durabilité ce qu'ils gagnent en puissance d'expression.

C'est ainsi qu'est apparu un nouveau concept pour l'archivage, celui de la pérennisation car le problème numéro Un n'est plus la place de stockage ou la constitution de répertoires et d'index ; c'est : comment « faire passer les années » à tous ces objets numériques qui tracent et documentent l'activité des hommes et auxquels on aura besoin de se référer, demain ou après-demain, pour pouvoir affirmer son droit, savoir ce qui s'est passé ou éviter de refaire ce qui a déjà été fait, parfois à grand frais.

A vrai dire, la préoccupation de la pérennisation des documents après leur archivage n'est pas tout à fait nouvelle dans la mesure où la conservation matérielle a toujours été un souci des conservateurs du patrimoine archivistique et des responsables d'archives d'entreprise. Mais, avec la relative stabilité des supports traditionnels, elle occupait une place plus discrète en termes d'enjeux et de coûts. Les archives endommagées par l'usure du temps, les champignons ou les catastrophes, même si les mesures de conservation et les coûts de la restauration du papier ne sont pas négligeables, semblaient un avatar inévitable qui touchait 5 à 10 % des fonds et qu'il fallait régulièrement contrer. Pour les films, la menace s'est fait plus présente au cours des dernières décennies, sans doute à cause de l'impact du document audiovisuel, mais on pouvait encore voir le document abîmé, apercevoir son contenu, le toucher. Avec les bandes magnétiques, l'inquiétude s'est accrue du fait de l'intermédiaire que constitue le codage ; c'est comme si tous ces documents se trouvaient sur une île sans qu'on ait de bateau pour l'atteindre...

Avec le numérique, le mot pérennisation, dans sa double acception de préservation matérielle et d'accès intellectuel, devient un des termes fondamentaux du monde des archives, alors qu'il ne faisait pas partie du vocabulaire technique des archivistes il y a encore dix ans. Cette évolution sémantique n'est bien évidemment pas neutre. Il y a dans « pérennisation » une connotation plus active, plus dynamique que dans le terme classique de conservation, plus statique. Et les archivistes ne sont plus les seuls à s'occuper d'archives.

Ce numéro spécial de *Document numérique*, après un premier numéro sur l'archivage en 2001, veut souligner par son titre « Archivage et pérennisation » l'importance des deux démarches successives mais solidaires de « mettre en archive » les documents qui sont appelés à témoigner demain (archivage) puis d'accompagner ces objets d'information dans le temps (pérennisation). Archivage et pérennisation ne sont pas des démarches ponctuelles que l'on effectue une fois par an ou une fois tous les cinq ans, parce qu'il le faut bien. Dans l'environnement électronique, l'attention aux archives est permanente, sans solution de continuité, et vise non seulement les documents/données mais aussi les chaînes matérielles et logicielles (données + format + support + application de gestion + appareil de lecture). Or, la capacité des objets numériques à perdurer dépend moins de leurs caractéristiques intrinsèques que de l'aptitude des technologies futures à préserver la chaîne d'accès à l'information. L'« indexation » de la pérennisation sur les progrès technologiques dont, par définition, on ignore aujourd'hui les développements est l'élément novateur de l'archivage électronique.

Ainsi, le fait d'archiver et de pérenniser des documents ou données numériques, quel qu'en soit le motif, est indissociable de la notion de maintenance : maintenir les documents/données existants dans le futur, maintenir l'accès, maintenir le sens. L'archivage n'est plus une opération que l'on effectue « après » ou « à côté » de la vie active des données mais, de plus en plus, une mesure qui s'opère « avec » la production et l'exploitation du document, un accompagnement normal qui touche aussi bien le contenu que la forme et le support.

Le défaut de durabilité des matériels et logiciels renforce le besoin de normalisation et les exigences de conservation plaident pour le rapprochement et la collaboration des différents décideurs et acteurs de l'environnement technologique dont les intérêts, finalement, convergent : institutions publiques, éditeurs, industriels, chercheurs, praticiens de l'archivage. Ces rapprochements et les échanges autour des expérimentations en cours ou abouties mettent en évidence l'importance des réponses organisationnelles et méthodologiques en parallèle des solutions techniques. L'analyse de ce qu'on veut archiver, de la finalité de cet archivage et des implications juridiques, conduit à rechercher des logiques d'identification, de sélection, de classement et de description adaptées aux nouveaux types d'objets d'information produits par le numérique, et à construire une interactivité entre l'offre technologique et la demande des utilisateurs d'un système sécurisé, fiable et efficace : préciser la granularité de l'information archivée et pérennisée, définir des critères de gestion et d'accès.

Le comité de lecture réuni pour ce numéro spécial regroupe des professionnels de l'archivage, français et étrangers, experts et opérationnels, du secteur public et du secteur privé : Françoise Banat-Berger (direction des Archives de France), Guylaine Beaudry (Université de Montréal), Didier Courtaud (Commissariat à l'énergie atomique, Paris), Marc Fresko (Cornwell Management Consultants, Londres), Claude Huc (Centre national d'études spatiales, Toulouse), Julien Masanès (Bibliothèque nationale de France), Joël Poivre (Archives nationales de France), Jean-Daniel Zeller (Hôpitaux universitaires de Genève). L'appel à communication, diffusé à la fin de l'année 2003, a suscité de nombreuses propositions intéressantes. La composition du numéro, eu égard aux contraintes de publication, a conduit à n'en retenir qu'une dizaine et à solliciter des articles relativement courts.

La sélection de textes proposés au lecteur évoque les aspects fondamentaux de la question, sans toujours proposer des solutions définitives ou même sans nécessairement trancher car de nombreuses études sont encore en cours, les expérimentations n'ont pas livré tous leurs résultats, la technologie n'a pas dit son dernier mot et la méthodologie de l'archivage numérique est loin d'être aboutie. Néanmoins, on trouvera dans ce numéro un panel assez significatif des problématiques de l'archivage électronique en 2004 et des orientations qu'il est légitime de suivre compte tenu des connaissances et expériences actuelles.

Les articles vont de la synthèse méthodologique à la réflexion prospective en passant par la présentation d'expérimentations, de stratégies et de projets d'archivage d'électronique, au niveau national ou local. Ils sont proposés partie par des chercheurs, partie par des responsables opérationnels ou directeurs de projets, qui sont aussi bien des ingénieurs que des archivistes. La moitié des auteurs sont français, les autres sont francophones (Suisse, Belgique) ou anglo-saxons (Australie, Canada). Cette variété des auteurs et des thèmes illustre bien le caractère pluridisciplinaire et international du défi de l'archivage électronique et de la pérennisation des archives numériques.

Le premier article du numéro est proposé par Isabelle Boydens, historienne et docteur en philosophie, chargée de cours à l'Université libre de Bruxelles et consultante dans une société informatique prestataire de services pour l'administration belge. L'auteur expose la problématique de la conservation des données numériques confrontée à divers facteurs matériels et logiciels de détérioration de l'information, et les stratégies élaborées pour y faire face. Le recours aux métadonnées pour consolider l'archivage et aider à la pérennisation pose à son tour le problème de la conservation des métadonnées, qui sont elles aussi des données peu pérennes par nature. Isabelle Boydens insiste sur les réponses méthodologiques et sur la prise en compte du cycle de vie, suggère l'application à l'archivage du concept de « temporalités étagées » de Fernand Braudel et l'intégration de la perspective d'archivage dans la conception des bases de données afin d'anticiper le « continuum » de l'information.

Howard S. Quenault, du Public Record Office de l'Etat de Victoria, en Australie, présente le projet VERS (*Victorian Electronic Records Strategy*) mis en œuvre pour

l'archivage des documents électroniques du gouvernement, avec le concours du monde industriel et de milieux universitaires dans l'analyse de l'existant et les recherches de solutions technologiques compatibles avec les besoins de conservation à long terme. L'article décrit de façon très pragmatique les différentes étapes du projet depuis son lancement en 1995 jusqu'à la mise en place d'un centre d'archives numériques (*Digital Archive*). Le déploiement du système est en cours dans les différents services administratifs de l'Etat : conformité des systèmes d'information pour un archivage à long terme, définition de procédures, choix des outils et formats.

Le projet ARELDA concerne l'archivage des données et documents électroniques de l'administration fédérale suisse et s'inscrit dans l'initiative plus vaste de la cyber-administration de l'Etat. Il est présenté par Thomas Zürcher Thrier et Stephan Heuscher, de l'équipe projet. Le projet, qui a également démarré au milieu des années 1990, est prévu jusqu'en 2008 et est évalué à 8 millions d'euros. Les auteurs insistent sur les quatre exigences de l'archivage numérique : persistance, intégrité, authenticité et accessibilité. Le projet se décompose en plusieurs sous-projets thématiques dont trois sont décrits plus précisément : le projet AMDA pour l'acquisition des métadonnées des enregistrements sonores des débats du Parlement (avec rétro-numérisation pour une collection de plus de cent ans) ; une étude dédiée aux bases de données avec le développement, en collaboration avec une société privée, de logiciels spécifiques regroupés sous l'appellation SIARD (*Software Invariant Archiving of Relational Databases*) ; l'élaboration d'une intéressante typologie des systèmes d'information dans le but de mieux maîtriser les besoins d'archivage : systèmes de gestion électronique de documents (GED), systèmes d'archivage électronique, bases de données statistiques, bases de données de travail.

L'article suivant, proposé par Louis Faivre d'Arcier, conservateur aux Archives de Paris, expose un projet à l'échelon local : l'archivage des données de l'application de gestion des autorisations d'occupation des sols à Paris. La collection papier des dossiers de permis de construire, depuis 1880, représente 100 000 dossiers et 2 kilomètres de rayonnage. Les dossiers ne sont pas (pas encore) dématérialisés mais l'application de gestion GERCO, en place depuis 1988, change la donne de l'archivage. Deux opérations complémentaires sont programmées : d'une part, un versement des données de l'application utiles à la gestion des dossiers papiers sur le long terme, au format XML et conformes à la DTD EAD, données qui serviront d'instrument de recherche aux versements papier ; d'autre part, un archivage des données de la base témoignant « directement » des évolutions de la construction à Paris entre 1988 et 2000. Ces opérations sont l'occasion d'une réflexion archivistique sur des questions qui ne se posaient pas ou se posaient différemment dans l'environnement papier, notamment la notion de clôture de dossiers dans une base de données, le tri des données ou le délai de versement, c'est-à-dire la périodicité d'archivage des données encore utiles à l'administration et déjà constitutives du patrimoine archivistique.

Le projet AMALFI visant à la dématérialisation du livre foncier d'Alsace-Moselle est à la fois un projet national et local : en effet, il ne concerne qu'une petite partie du territoire français, dotée pour des raisons historiques d'un système spécifique d'enregistrement de la publicité foncière, mais il est conduit par le ministère de la Justice français. Il est présenté conjointement par Françoise Banat-Berger, de la Direction des Archives de France, et par Jean-François Blanchette qui suit le projet au nom du groupe de recherche InterPARES (voir ci-dessous). Il ne s'agit pas à proprement parler d'archivage puisque l'informatisation porte sur les documents utilisés au quotidien par l'administration judiciaire mais cette utilisation étant requise sur des dizaines d'années, la question de la conservation à long terme et de la pérennisation est particulièrement sensible dans ce projet, d'autant plus que les documents en cause revêtent une valeur probante forte. Les auteurs abordent donc tout spécialement les difficultés liées à la signature électronique des actes et à la conservation de l'écrit électronique authentique ; ils évoquent par ailleurs les aspects de conservation/suppression des données nominatives en conformité avec la législation (CNIL) et de transfert de données vers un centre d'archives patrimoniales.

Luciana Duranti, directeur du groupe de recherche international InterPARES (*International Research on Permanent Authentic Records in Electronic Systems*), présente dans l'article suivant le projet InterPARES 2, consacré à « la conservation à long terme des documents dynamiques et interactifs dans le domaine des arts, des sciences et de l'administration électronique ». Cette seconde phase du projet de recherche a été lancée en 2002 et doit s'achever en 2006. La première partie de l'article résume la mission d'InterPARES 1, centrée sur la conservation de l'authenticité des documents (*records*) et donne une synthèse très pertinente de la problématique générale de la pérennisation. La seconde partie présente la méthode et énumère les nombreux et très diversifiés projets couverts par le champ d'étude d'InterPARES (voir le site du groupe de recherche ([www.interpares.org](http://www.interpares.org))). Constatant qu'il n'existait aucun article en français sur les travaux d'InterPARES, pourtant fondamentaux, et que, de ce fait, ils n'étaient pas aussi connus du public français qu'ils pourraient l'être, j'ai pris la décision de traduire le texte de Luciana Duranti, avec l'approbation de l'auteur et la collaboration d'une de ses étudiantes francophones, Genevieve Shepherd (la version originale anglaise ne figure pas dans le numéro).

Claude Huc, responsable de l'archivage des données numériques du Centre national d'études spatiales (CNES), propose un modèle pour l'organisation d'un centre d'archives numériques. Basée sur la pratique du modèle de référence OAIS (*Open Archival Information System*), cette organisation se structure en trois services : le service « Collecte et préparation des données », le service « Stockage » et le service « Gestion des données et communication » (où l'on retrouve les trois grandes fonctions de l'archivistique traditionnelle), pilotés par une structure de coordination. L'article passe en revue, pour chacun des trois services, les fonctions, le contexte, les moyens techniques et les compétences requises par la pérennisation des objets numériques. La mise en place d'une telle structure au CNES il y a dix ans

permet un retour d'expérience concret et une conclusion nourrie qui aborde la question de la certification des centres d'archives numériques.

Jean-Daniel Zeller, archiviste principal des Hôpitaux universitaires de Genève, propose une réflexion sur la typologie des documents numériques en vue de leur archivage. L'auteur s'intéresse aux typologies existantes, aux qualités intrinsèques des typologies, à la typologie spécifique des données, au codage, à la structuration, à la validation, aux métadonnées, avant d'aborder quelques exemples de documents composites (système d'information géographique, courriel, site web). Les éléments de réponses recueillis, insatisfaisants dans l'immédiat pour une théorie globale, débouchent sur un besoin d'approfondissement de l'articulation des approches technique et documentaire de la catégorisation des objets archivables, et de l'approfondissement des notions de document et de données.

Abordant la question des professions chargées de l'archivage, Thibaut Girard, expert de EdF-GdF Archives (Electricité de France-Gaz de France), met en regard les tiers de confiance, créés par la nouvelle législation sur la signature électronique, et la communauté des archivistes que l'évolution technologique confronte, par une autre voie, aux mêmes défis. Après une analyse comparée des métiers, l'auteur insiste sur les aspects économiques et juridiques et sur les conséquences du coût de l'archivage numérique. Au-delà des questions méthodologiques et techniques, cette contribution originale et stimulante donne à réfléchir sur les enjeux de l'archivage dans la société et les responsabilités de ceux à qui il est confié.

Le numéro se termine sur un article collectif présentant le groupe PIN (groupe de réflexion français sur la pérennisation de l'information numérique). Constitué en l'an 2000 et regroupant des responsables de l'archivage du secteur public et du secteur privé, des ingénieurs et des archivistes, le groupe PIN se veut un lieu d'échange et de recherche autour des défis de l'archivage numérique, le ferment d'une culture commune aux différents métiers impliqués dans la pérennisation des données, que ces données soient scientifiques, techniques, administratives ou patrimoniales.

Notre comité de rédaction avait envisagé d'assortir le choix d'articles d'un glossaire bilingue, anglais-français, des termes de l'archivage numérique, afin de clarifier à la fois les chevauchements entre les différents concepts manipulés (par exemple, électronique et numérique) et les équivalences entre les termes anglais et français parfois employés indifféremment par le même locuteur (par exemple, support et medium). Ce travail, lourd, n'a pu être formalisé dans les temps impartis mais verra peut-être le jour dans une publication ultérieure.

En dépit de ses lacunes mais grâce à la complémentarité de ses témoignages, *Archivage et pérennisation* est une invitation à l'échange, au partenariat et à la recherche collective et pluridisciplinaire.

Marie-Anne Chabin