

L'UNESCO et la conservation du patrimoine numérique UNESCO and the Conservation of the Numeric Heritage La UNESCO y la preservación del patrimonio digital

Abdelaziz Abid

Volume 49, numéro 3, juillet–septembre 2003

URI : <https://id.erudit.org/iderudit/1030190ar>

DOI : <https://doi.org/10.7202/1030190ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (imprimé)

2291-8949 (numérique)

[Découvrir la revue](#)

Citer cet article

Abid, A. (2003). L'UNESCO et la conservation du patrimoine numérique.
Documentation et bibliothèques, 49(3), 93–101. <https://doi.org/10.7202/1030190ar>

Résumé de l'article

Une large part des vastes quantités d'informations produites dans le monde est d'origine numérique et se présente sous une grande diversité de formats : texte, base de données, son, film, image. Pour les institutions culturelles, qui ont traditionnellement la charge de recueillir et de préserver le patrimoine culturel, il est devenu extrêmement urgent de savoir quels sont les matériaux qu'il convient de conserver pour les générations futures et comment il faut procéder à leur sélection et à leur préservation. Ce gigantesque trésor d'informations numériques, produit aujourd'hui dans presque tous les domaines de l'activité humaine et conçu pour être accessible sur ordinateur, risque d'être perdu si l'on ne développe pas des techniques et des politiques spécifiques en vue de le conserver. L'intérêt que porte l'UNESCO à cette situation n'a rien de surprenant. Cette organisation internationale existe en partie pour encourager et permettre la conservation et la jouissance du patrimoine culturel, scientifique et informatif des peuples du monde entier. Elle ne pouvait négliger l'augmentation du patrimoine numérique et sa vulnérabilité.

L'UNESCO et la conservation du patrimoine numérique

Abdelaziz Abid

Division de la société de l'information

UNESCO

Courrier électronique : a.abid@unesco.org

Site Web : <http://www.unesco.org/webworld/mdm>

Une large part des vastes quantités d'informations produites dans le monde est d'origine numérique et se présente sous une grande diversité de formats : texte, base de données, son, film, image. Pour les institutions culturelles, qui ont traditionnellement la charge de recueillir et de préserver le patrimoine culturel, il est devenu extrêmement urgent de savoir quels sont les matériaux qu'il convient de conserver pour les générations futures et comment il faut procéder à leur sélection et à leur préservation. Ce gigantesque trésor d'informations numériques, produit aujourd'hui dans presque tous les domaines de l'activité humaine et conçu pour être accessible sur ordinateur, risque d'être perdu si l'on ne développe pas des techniques et des politiques spécifiques en vue de le conserver. L'intérêt que porte l'UNESCO à cette situation n'a rien de surprenant. Cette organisation internationale existe en partie pour encourager et permettre la conservation et la jouissance du patrimoine culturel, scientifique et informatif des peuples du monde entier. Elle ne pouvait négliger l'augmentation du patrimoine numérique et sa vulnérabilité.

UNESCO and the Conservation of the Numeric Heritage

A significant part of the vast quantity of information produced worldwide is numeric and comes in several formats: text, data bases, sound, film, and image. For cultural institutions, traditionally responsible for the acquisition and preservation of the cultural heritage, it has become very important to know which materials should be conserved for future generations and how these materials should be selected and conserved. This gigantic treasure trove of numeric data, produced in all domains of human activity and designed to be accessed by computers, could be lost if specific techniques and policies are not developed with its conservation in mind. UNESCO's interest in this issue comes as no surprise. This international organisation exists partly to promote the conservation and the enjoyment of the cultural, scientific and information heritage of humanity. Therefore, it could not neglect the growing numeric heritage and its vulnerability.

La UNESCO y la preservación del patrimonio digital

Una parte considerable de la extensa cantidad de información que se produce en el mundo es digital y se presenta en una gran variedad de formatos: texto, base de datos, sonido, película, imagen. Para las instituciones culturales que tienen tradicionalmente la responsabilidad de preservar el patrimonio cultural, es extremadamente urgente saber cuáles son los materiales que vale la pena conservar para las generaciones futuras y cómo se deben seleccionar y preservar. Ese gigantesco tesoro de información digital que hoy día producen casi todos los sectores de la actividad humana y que ha sido concebido para ser accesible por computadora, corre el riesgo de perderse si no desarrollamos técnicas y políticas específicas destinadas a conservarlo. No es nada sorprendente que la UNESCO se interese en esta cuestión. Parte de la función de esta organización internacional es fomentar y facilitar la preservación y el goce del patrimonio cultural, científico e informático de los pueblos de todo el mundo. La UNESCO no podía desentenderse del aumento del patrimonio digital y de su vulnerabilidad.

L'UNESCO étudie les questions liées à la sélection et à la préservation de l'information numérique afin de guider les efforts de conservation qu'entreprennent les gouvernements à l'âge numérique. La Conférence générale a adopté, lors de sa 31^e session tenue en octobre 2001, la résolution 34 attirant ainsi l'attention sur l'augmentation croissante du patrimoine numérique dans le monde et sur la nécessité de lancer une campagne internationale pour sauvegarder la mémoire numérique des dangers qui la menacent. La Conférence générale avait aussi invité le Directeur général à préparer, pour la session du printemps 2002 du Conseil exécutif, un document à débattre qui comporte une ébauche visant à la fois à préserver des documents d'origine numérique et à encourager l'adoption de politiques natio-

nales accordant une priorité élevée à la conservation du patrimoine numérique.

Au cours de la réunion du Conseil exécutif de l'Organisation en mai 2002, les États membres se sont mis d'accord sur la nécessité d'une action rapide pour sauvegarder le patrimoine numérique. Le débat s'est largement inspiré d'un document de discussion établi pour l'UNESCO par la *European Commission on Preservation and Access* (ECPA), une fondation à but non lucratif basée à Amsterdam, qui a esquissé les problèmes inhérents à la conservation numérique.¹ Le présent article s'inspire largement de ce document et d'une étude plus récente, préparée pour l'UNESCO par l'Institut national de l'audio-visuel (France), à l'occasion du Sommet mondial sur la société de l'information, tenu à Genève en décembre 2003.²

Le champ du patrimoine numérique

Le patrimoine est défini par les documents de l'UNESCO comme étant « *notre héritage du passé, ce avec quoi nous vivons aujourd'hui, et ce que nous transmettons aux générations futures* ». Un patrimoine est quelque chose qui est, ou devrait être, transmis de génération en génération parce qu'il est précieux.

L'idée de patrimoine culturel est familière : ce sont des sites, des objets et des

1. <http://unesdoc.unesco.org/images/0012/001255/125523e.pdf>.

2. La mémoire de la société de l'information. UNESCO, 2003 (Publications de l'UNESCO pour le Sommet mondial sur la société de l'information).

Le texte complet des ouvrages est disponible à : <http://www.unesco.org/wsis>.

Pour résoudre la complexité des problèmes en jeu, la tâche de préservation exige d'impliquer les fabricants de l'information numérique, logiciels compris : ils devraient prendre en compte la conservation lorsqu'ils conçoivent leurs produits.

choses immatérielles qui ont une valeur culturelle, historique, esthétique, archéologique, scientifique, ethnologique ou anthropologique pour des groupes et des individus. La notion de patrimoine naturel est également très familière : elle recouvre des caractéristiques physiques, biologiques et géologiques ; des habitats de plantes ou d'espèces animales et des domaines précieux sur le plan scientifique ou esthétique ou du point de vue de la conservation.

Le patrimoine numérique est, quant à lui, de deux ordres. Une part importante de ce patrimoine est constituée du produit de la reproduction numérique d'œuvres préexistantes, qui peuvent être textuelles, picturales, sonores, audiovisuelles, graphiques, photographiques, cinématographiques, etc., inscrites sur un support matériel stable et défini. Ce « double » numérique ne prétend pas être la copie à l'identique de l'œuvre initiale, mais se contente d'en être sa représentation : c'est une photo, une empreinte, une trace à un moment donné, et dans tous les cas, le fruit d'une politique volontariste de numérisation.

La seconde composante du patrimoine numérique provient des données qui n'existent que sous forme numérique, qu'il s'agisse de sites Internet, de publications électroniques, de productions multimédias, ou encore de banques de données culturelles et scientifiques contenant et organisant des documents textuels ou graphiques, des sons, des images fixes ou des productions audiovisuelles ou multimédias.

Les méthodes de conservation traditionnelles, tel le « dépôt légal » en vigueur dans les bibliothèques nationales pour garantir la conservation d'un exemplaire de tous les matériaux imprimés, ne sont pas applicables, telles quelles, au matériau numérique pour diverses raisons, no-

tamment parce que les « publications » du Web utilisent des données mémorisées sur des serveurs dans différentes parties du monde. Le simple volume des données concernées fait également problème. Internet, estime-t-on, compte un milliard de pages d'une durée de vie moyenne extrêmement brève (évaluée entre 44 jours et deux ans).

Internet, considéré comme le support de publication le plus démocratique qui ait jamais existé, est en extension croissante et mérite, selon certains, d'être préservé intégralement si l'on considère ses pages et ses forums de discussion comme un miroir inestimable de la société.

Garantir que l'objet numérique sauvegardé dans les archives reste accessible sous sa forme originale pose des problèmes techniques. Logiciels et matériels sont constamment remplacés par des générations nouvelles plus puissantes qui ne tardent pas à devenir incompatibles avec leurs prédécesseurs. Résultat : au bout de quelques années seulement, le document numérique — qui comprend souvent du son et des graphiques ou des images animées, ainsi que des liens avec des sites Internet et des bases de données — devient inaccessible.

Le simple volume de données à passer au crible pour sélectionner ce qui mérite d'être préservé est confondant. « *Le montant annuel de la production mondiale de matériaux imprimés, filmiques, optiques et magnétiques nécessiterait environ 1,5 milliard de gigaoctets d'archivage. C'est l'équivalent de 250 mégaoctets par personne pour chaque homme, femme et enfant sur la terre* », selon une étude récente réalisée par l'Ecole de gestion de l'information et des systèmes à l'Université de Californie de Berkeley³.

Un autre problème complexe touche au *copyright*, y compris le *copyright* du logiciel nécessaire pour accéder aux

fichiers numériques. Des sites Web qui offrent un mélange de matériaux de sources diverses peuvent ainsi présenter un ensemble ahurissant de droits. L'accord sur le principe du « droit de reproduire pour sauvegarder » reste encore à mettre au point à l'échelle mondiale.

Alors que de nombreux pays ont pris des initiatives précieuses pour préserver le patrimoine numérique, y compris les sites Web, l'étude de l'ECPA montre les limites de ces efforts et plaide en faveur de normes internationales.

Pour résoudre la complexité des problèmes en jeu, la tâche de préservation exige d'impliquer les fabricants de l'information numérique, logiciels compris : ils devraient prendre en compte la conservation lorsqu'ils conçoivent leurs produits. Le temps n'est plus où la conservation était du ressort exclusif des institutions d'archivage.

Coopération, orientation, direction et partage des tâches sont des éléments clés pour la sauvegarde du patrimoine numérique. Les institutions culturelles ont besoin de la coopération des créateurs de l'information et des fabricants de logiciels. Des ressources et un soutien adéquats au plan politique sont indispensables pour permettre aux générations futures de continuer à avoir accès à la richesse des ressources numériques dans la création desquelles nous avons tant investi au cours des dernières décennies.

L'informatisation de la société

La vague de l'informatisation qui a déferlé sur nos sociétés pendant plus d'un demi-siècle atteint aujourd'hui une forme de maturité. Le phénomène est tellement massif en changeant notre mode de vie que l'on évoque maintenant la Société de l'information. Sans doute y aura-t-il d'autres vagues ? Mais en se retournant vers ce passé récent, on ne peut qu'être saisi par l'étendue des transformations qui auront touché tous les secteurs de l'information, de la création et des savoirs — en fait toute notre culture.

En quatre vagues successives, la même impulsion a créé les conditions, les

3. <http://www.sims.berkeley.edu/how-much-info>.

prémises, et très vite ensuite, l'inflation d'un patrimoine numérique débordant :

- jusqu'aux années 70 : archivage des premières données scientifiques issues des calculateurs scientifiques ;
- dans les années 80 : éclosion de l'édition numérique et de ses supports : au premier rang le CD audio, mais également les CD-ROM multimédia et les jeux vidéo ;
- dans les années 90 : mise en place de la télévision et la radio numérique par satellite ;
- au tournant du millénaire : développement foudroyant d'Internet et de ses applications Web et courrier électronique, notamment.

Brutalement on découvre, portés par cette quatrième vague, les composants constitutifs, lentement tissés, d'un patrimoine numérique abondant, mais dont l'étendue ne se révèle vraiment que dans la clôture avec Internet de ce premier âge de l'informatique.

Changement paradigmatique

Jusqu'au développement d'Internet, on pouvait encore hésiter, il restait alors possible de conserver sous une forme plus traditionnelle ces nouveaux supports déconcertants. Le numérique se répandait très vite dans toutes les sphères de la création ou de la science, mais on pouvait le contourner : le virtuel n'était souvent encore qu'une étape dans un processus qui allait du réel au réel. Même les premiers mondes virtuels, les productions en 3 dimensions, finissaient sur de la pellicule film ou de la bande vidéo pour être rendus publics.

Avec Internet, la question est tranchée : on ne sortira plus de ces espaces virtuels pour les utiliser. On imprime encore souvent sur papier parce que la lecture y reste un peu plus confortable, mais pour combien de temps ? Internet radicalise la question du numérique et la question du patrimoine. Internet nous oblige à repenser toutes nos certitudes sur le sens même du mot conserver, un sens qui nous vient des âges les plus lointains quand l'homme a, pour la première fois, déposé son intelligence sur des objets plus résistants que lui, afin que sa mé-

moire traverse les générations jusqu'à nous.

Toute matière tend à disparaître progressivement, à se dissoudre, se déliter, jaunir, vieillir — pas l'information : elle est ou elle n'est plus. Conserver l'information numérique sera comme conserver la flamme d'un feu : il faudra s'en occuper en permanence, l'entretenir, la nourrir sinon elle s'éteindra et s'effacera. En contrepartie, elle restera éternellement jeune. Ceci ne sera pas sans imposer de lourdes révisions aux institutions chargées de conserver le patrimoine documentaire. Laisser dormir sur des rayonnages dans des conditions adéquates était la meilleure garantie de conservation, et longtemps, la communication des documents a même été considérée comme le pire ennemi de la conservation. Au contraire, la capacité à soutenir la vitesse de rotation de l'information numérique sur des supports toujours nouveaux sera le gage de sa pérennité.

L'actualité de cette nouvelle forme de l'archive est d'autant plus brûlante que l'informatique a jusqu'ici davantage démontré ses capacités d'expansion que celles de durabilité. Les moyens de production et de circulation de l'information ont été démultipliés et le phénomène a pris des proportions inégalées dans l'histoire de l'humanité. Mais, poussées par les cycles de la technologie (si l'on n'y prend garde, le cycle de l'information n'aura jamais été aussi court), nos sociétés vont voir disparaître des pans entiers de mémoire à la charnière des millénaires.

Un patrimoine fragile

Tant que l'information a emprunté des supports physiques pour circuler, elle laissait sa trace. Même sans s'en préoccuper, il reste toujours quelque chose de la trace, quelque chose qui se transforme en archive. Mais l'informatique a ce défaut congénital : si l'on ne sauvegarde pas on efface. Autrement dit la conservation du patrimoine doit dorénavant devenir un acte délibéré et volontaire, organisé au présent.

De plus, l'informatique inverse les propositions qui semblaient les plus assurées : la survie d'un document ne tient pas à la pérennité de son support, mais à sa capacité à en changer le plus souvent possible. La forme numérique du

patrimoine peut s'avérer la meilleure et la pire des choses selon ce qu'on en fera :

- soit une solution fondatrice d'un nouvel âge pour tout ce qui touche à la conservation et permette de penser les problèmes de la pérennité, voire de l'éternité, en des termes totalement nouveaux ;
- soit l'anarchie d'une information débridée dans un éternel présent, fondant des sociétés sans mémoire, s'effaçant au gré de l'obsolescence des machines ou des formats.

La société saisie par le numérique

La numérisation de l'ensemble des données produites par l'intelligence humaine, quelle qu'en soit leur forme d'origine — écrits, sons, images fixes ou animées — affecte tout à la fois les processus de création de contenus, leurs modalités de diffusion, ainsi que les moyens de conservation nécessaires à leur pérennité.

Depuis la mise au point des premiers ordinateurs, de l'invention du traitement de texte, de la conception des premiers logiciels d'images de synthèse et d'outils de publication assistée par ordinateur, de la fabrication des premiers CD numériques à haute capacité, etc., une suite ininterrompue d'innovations technologiques a pénétré, ces dernières décennies (sans même qu'on s'en aperçoive), des pans entiers de la production culturelle, scientifique et artistique.

Actuellement, les institutions scientifiques doivent faire face à une incroyable quantité de données à gérer dans leur diversité — pour certaines, plusieurs centaines de giga-octets par jour — résultant des expérimentations de laboratoire, de manipulations en grandeurs réelles ou d'observations provenant de différents instruments (satellites, radars, télescopes, sondes, capteurs, micro-caméras, etc.), dont certaines constituent de véritables événements historiques qui ne pourront plus jamais être renouvelés dans le temps. Il en est ainsi, par exemple, des phénomènes météorologiques qui doivent être impérativement conservés pour permettre le développement de la prospective climatologique par l'analyse de données accumulées durant plusieurs dizaines d'années.

Et si la sauvegarde de ce capital-information représente pour le monde de la science un enjeu aussi important que sa création et son interprétation, le phénomène apparaît, dans de nombreux pays, encore très dispersé entre de multiples laboratoires, peu interprétable par d'autres et donc peu transmissible. Faute de structure spécialisée qui pourrait prendre en charge ces masses croissantes d'information, la pérennisation du « patrimoine scientifique », lui-même à l'origine de nouvelles découvertes, reste encore un véritable défi pour nos sociétés modernes : sa perte en serait une régression irrémédiable.

Les mutations du régime du texte et de l'écrit

Progressivement, l'ordinateur va sortir des sites informatiques et va pouvoir gérer du texte dans les activités les plus quotidiennes. La fin des années 70 et le début des années 80 auront vu cette double filiation de la bureautique, enfant illégitime de la machine à écrire qui s'électrisait, et des sites centraux qui en devenant micro s'introduisaient dans le bureau de la secrétaire. Outre les activités de bureau, le monde de la presse et de l'édition est alors massivement touché par cette révolution de la mise en page et de l'impression.

Une nouvelle forme de lecture que certains qualifient d'« hyper lecture » provoque une circulation éclatée, un raisonnement ouvert, une logique relationnelle, qui ne prend plus appui sur une pensée singulière mais sur un ensemble de motifs. Les conditions d'acquisition des connaissances et les modalités d'accréditation du savoir s'en trouvent profondément modifiées d'autant plus que, pour la première fois dans l'histoire de l'humanité, tous les savoirs du monde, ou presque, sont virtuellement accessibles.

Depuis quelques années, les bibliothèques numérisent leurs collections et les mettent en ligne ; des moteurs de recherche toujours plus performants fouillent des millions de pages pour satisfaire notre curiosité. Sous nos yeux, une gigantesque bibliothèque universelle est en cours de constitution. Encore faut-il que nos sociétés ne soient pas frappées d'amnésie !

Images et sons

L'extraordinaire développement des capacités de stockage, à temps d'accès différenciés, selon les technologies employées, va permettre à l'informatique d'envahir des champs d'activité qu'autrefois, ses faibles performances, lui interdisaient : on pense ici au monde de l'image et du son.

Pionnier dans le domaine du numérique, le son a été, très tôt, un champ ouvert au développement de l'informatique musicale : d'abord pour la production de sons synthétiques, puis dans l'élaboration et la mise au point d'outils de composition assistée par ordinateur. Centrés initialement sur la synthèse et la manipulation analytique du son à des fins de création, ces travaux ont généré de nouvelles applications touchant, tout à la fois, le système de captation, d'enregistrement et de diffusion. Aujourd'hui, la numérisation généralisée du signal est en cours ; un nouveau système technique se déploie et vient bouleverser profondément l'univers sonore et la musique sous toutes ses formes.

Du côté de la création, la banalisation des logiciels de composition, désormais accessibles à un large public, la généralisation de systèmes de type « *home studio* », renouvelle considérablement les conditions de la production musicale et réintroduit la continuité rompue par les techniques de reproduction électromécanique, entre créateur et consommateur.

En matière de diffusion, le développement d'Internet, conjugué à la généralisation des normes de compression MPEG et MP3, ouvrent de nouvelles possibilités d'accéder en ligne à des fonds musicaux. À tel point, qu'aujourd'hui, la musique est la première industrie culturelle en nombre de fichiers échangés sur Internet (en 2002, plus de 3 millions de fichiers en MP3 ont été quotidiennement échangés).

L'image de synthèse

Comme pour la musique, la numérisation commence toujours du côté de la synthèse (synthèse sonore, MIDI, images 2D et 3D), et en temps différé, avant de se généraliser à l'enregistrement, au traitement et la restitution d'images et de sons naturels. C'est par les techniques de la synthèse, largement modelée sur les

techniques de la simulation militaire et aéronautique, que la numérisation a pénétré l'audiovisuel. Dans les années 80, une rapide extension de ces applications aux besoins de la télévision et du cinéma a permis la généralisation des effets spéciaux — « Tron », le premier film réalisé intégralement en images de synthèse sort en 1982 des studios de Walt Disney — jusqu'à ce que les techniques d'encodage et de compression numériques des images cinématographiques et vidéographiques rendent possible, dans le domaine de la post-production, l'hybridation entre images de synthèse et images naturelles.

À partir de là, et avec la mise au point des normes de compression et de traitement de l'image numérisée, on retiendra les travaux de normalisation du groupe JPEG pour l'image fixe et du groupe MPEG pour l'image animée à partir des années 80 — la numérisation du système audiovisuel va toucher progressivement toutes ses composantes ; de la production au montage, du montage à la régie, de la régie aux réseaux de diffusion et, enfin, jusqu'au récepteur du particulier.

L'ère des réseaux : commerce, services et gouvernance électroniques

En 1969 sont jetées les bases d'Internet, et, en 1989, est conçu le World Wide Web par Tim Berners Lee au CERN. Les ordinateurs communiquaient déjà, mais cette communication était affaire de spécialistes et les réseaux publics étaient encore très marqués par le monde de la télécommunication. La normalisation mondiale de facto des protocoles de communication a ouvert la voie à une interconnexion généralisée de tous les ordinateurs du monde.

Avec cette généralisation commence réellement l'aventure du patrimoine numérique. Auparavant, l'ordinateur n'était qu'un moyen pour obtenir des résultats ou créer des objets dans le monde réel, qui peuvent donc être archivés dans leur état final hors de leur être numérique. Il s'insère comme une étape transitoire dans une boucle qui va du réel au réel.

Internet, qui a connu ces cinq dernières années, notamment dans ses

applications Web et Mail, un développement foudroyant, affichant de 1998 à 2002 un taux de croissance de 217 % (de 2,851 millions de sites en 1998 à 9,040 millions en 2002 — sources OCLC—), entre aujourd'hui dans l'âge adulte :

- c'est le plus gigantesque réservoir de données constitué par l'interconnexion dynamique de multiples banques de données fouillées par de puissants moteurs de recherche ;
- c'est surtout l'outil de la convergence, qui, en réalisant la fusion entre textes, images fixes, sons, images animées, audiovisuel, offre de nouvelles capacités d'expression et de formalisation de la pensée et de la création humaine.

Aujourd'hui, les services publics et l'administration en ligne font partie de notre paysage. Ils ne sont cependant que la face visible du processus d'informatisation des organisations. À l'interne, le déploiement des messageries électroniques et des intranets contribue à modifier profondément les conditions de production et de circulation de l'information. Les centres traditionnels de pouvoir, fondés sur la détention d'une information rare et précieuse, sont ébranlés, tout comme les organisations pyramidales du travail sont remises en cause au profit de formes de travail plus interactives.

La communication électronique entre les personnes provoque une sorte de lissage des relations hiérarchiques, alors que les mécanismes de prise de décision et l'identification des responsabilités, jadis clairement affichés dans un organigramme, ou dans une procédure, se diluent progressivement, ou du moins, apparaissent moins visibles. Cette dilution n'est pas sans conséquence sur les pratiques archivistiques qui, méthodologiquement, s'appuient, par exemple, sur l'origine et la place hiérarchique du producteur d'information : plus le document est issu d'un centre de décision élevé dans la hiérarchie, plus sa valeur de témoignage est grande et, en conséquence, plus il mérite d'être conservé. L'affaiblissement de ce critère de sélection conduira-t-il les archivistes à vouloir tout conserver de peur de perdre l'essentiel ? Encore faudrait-il le pouvoir !

Aussi puissant, utile et efficace qu'il soit, Internet a sa part d'ombre, et non la

[Les bibliothèques] exploitent les potentialités d'Internet pour élargir l'accès à leurs collections en mettant en ligne leurs bases de données bibliographiques, et progressivement, leurs fonds d'ouvrages numérisés.

moins : celle de son absence de mémoire. Dans 50 ou 100 ans, quand des historiens se pencheront sur cette époque des technologies triomphantes de l'information et de la communication, ils risquent de se trouver devant un « grand fossé digital » qui aura englouti des millions d'informations à valeur scientifique, culturelle, historique ou sociologique, voire commerciale et industrielle. Il est donc essentiel de contrer l'amnésie du réseau et de se doter de moyens pour organiser la mémoire de la toile.

De nouvelles formes d'apprentissage et de consommation culturelle

Fort de ses attributs, Internet trouve donc naturellement sa place dans le champ de la culture et de l'éducation, modifiant durablement les circuits d'accès à l'information et à la connaissance, tout en favorisant l'éclosion de nouvelles pratiques culturelles.

Les musées virtuels représentent à cet égard une manifestation des plus novatrices dont le dynamisme a abouti, en 2001, à la création d'un nom de domaine « museum », réservé à la communauté des musées, afin de leur permettre d'améliorer leur visibilité et leur présence sur Internet. Ces musées virtuels contribuent, (par la démultiplication des accès aux œuvres et aux expositions, et par l'organisation de dispositifs muséaux pédagogiques tirant parti de l'interactivité et de l'hypermédia), à renouveler les modes d'appropriation et de compréhension des biens culturels. En réduisant la barrière de la distance géographique et en permettant l'éclosion de nouvelles pratiques culturelles, ils constituent de puissants appuis aux politiques de démocratisation de la culture.

Ce qui est vrai pour les musées, l'est aussi pour les bibliothèques qui, depuis plusieurs années, exploitent les potentialités d'Internet pour élargir l'accès à leurs collections en mettant en ligne leurs bases de données bibliographiques, et progressivement, leurs fonds d'ouvrages numérisés. Bien plus, elles permettent d'accéder à des ouvrages rares et précieux, souvent conservés dans les réserves pour qu'ils ne soient pas exposés aux dégradations physiques. Ce sont des ressources jusqu'alors réservées qui sont accessibles au plus grand nombre. Quand verrons-nous émerger un domaine « bibliotheca » permettant d'améliorer la visibilité des bibliothèques et leur présence sur Internet ?

On entrevoit ici tout le bénéfice que cet accès, élargi aux sources culturelles, peut apporter au monde scientifique et éducatif. En particulier, il ouvre la voie à de nouveaux modes de coopération, notamment dans le domaine de la formation, où l'on a vu se développer, ces dernières années, classes virtuelles et programmes de *e-learning*. Outil de communication, Internet contribue aussi, via des sites dédiés et des forums, à cimenter des communautés d'intérêt autour d'une thématique ou d'un champ du savoir. Ces communautés virtuelles s'échangent informations, analyses et points de vue sur l'objet qui les réunit.

Tout ceci contribue à la formation d'un patrimoine numérique en croissance rapide.

Un patrimoine infini ?

Une part sans cesse croissante de l'information, produite aujourd'hui dans presque tous les domaines de l'activité humaine, est sous forme numérique, soit stabilisée sur un support, soit accessible en ligne via Internet ou des intranets locaux. À l'inverse, la part totale de l'infor-

[Avec Internet] les méthodes traditionnelles de collecte ou d'acquisition ne s'appliquent plus et il n'y a guère d'autres solutions [...] que de mettre en place des dispositifs de captations automatiques.

mation produite dans le monde sur des supports traditionnels comme l'imprimé, la bande magnétique ou le film, diminue chaque année.

Ainsi, d'un point de vue quantitatif, l'ensemble de la production annuelle numérique, ou potentiellement numérisable, peut être estimée à plus de 15 000 téraoctets⁴. Ce volume considère l'ensemble de la production d'écrits faisant l'objet d'une publication (livres, périodiques, littérature grise) de 230 To, l'édition de CD et de DVD de 31 To, les œuvres cinématographiques, environ 16 To, et enfin les productions radiophoniques — 800 To — et télévisuelles — 14 000 To.

Le Web, lui-même, peut être estimé à 150 téraoctets. L'activité privée d'échange de courriers électroniques est très supérieure à celle du Web, à elle seule, elle englobe 10 et 20 000 téraoctets par an. Précisons que ces estimations ne prennent pas en compte les immenses bases de données scientifiques, plusieurs centaines de téraoctets qui constituent ce qui est communément désigné sous l'appellation de « Web profond ».

Pour autant, toute cette production a-t-elle la vocation de faire partie du patrimoine ? Et si tel est le cas, quels chemins doit-elle emprunter, quels traitements doit-elle subir pour intégrer le champ du patrimoine ? Est-elle laissée au hasard des progrès technologiques et de la robustesse des outils qui président à sa création en garantissant sa sauvegarde, ou est-elle le fruit d'une démarche patrimoniale volontariste et maîtrisée ?

Si l'on considère la production issue des programmes de numérisation des institutions culturelles, on est de toute évidence, sur un terrain connu : les œuvres sont définies, identifiées, inventoriées, même si la technique, qui implique des compétences particulières, n'est pas encore complètement familière. De fait, ces opérations de numérisation peuvent être prises en charge par l'entreprise elle-

même, au sein d'un service spécialisé créé à cet effet, ou sous-traité à un prestataire externe, en particulier lorsque les équipements techniques en jeu impliquent de lourds investissements dans un univers technologique parfois instable, comme c'est le cas actuellement pour la recopie numérique des fonds audiovisuels. Cependant, ces opérations de numérisation peuvent déjà être enclenchées pour l'acquisition de nouvelles œuvres enrichissant les collections déjà constituées, ou s'attacher prioritairement à la numérisation des fonds anciens, après en avoir effectué un diagnostic et après avoir mis en place une stratégie de sélection.

Cette stratégie va permettre de définir des priorités basées sur trois types de critères :

- des critères techniques : seront numérisés les fonds les plus fragiles ;
- des critères de contenu : l'attention sera portée sur l'intégralité d'une collection ;
- des critères d'usages : il s'agira de numériser les documents les plus demandés.

Dans tous les cas, l'objectif est de disposer, pour chaque objet, d'un fichier numérique identifié qui sera enregistré, soit sur un support physique amovible, soit dans un serveur de données.

Un flux insaisissable

Mais l'approche est toute autre quand on considère Internet. Là, l'unité du document se perd dans les hyperliens, le flux se substitue à l'objet fini. Dans cet univers, les méthodes traditionnelles de collecte ou d'acquisition ne s'appliquent plus et il n'y a guère d'autres solutions pour les entreprises patrimoniales que de mettre en place des dispositifs de captations automatiques. Ceux-ci reposent sur des logiciels « moissonneurs » qui parcourent la toile et en effectuant des enregistrements réguliers. Leur travail est guidé par un plan d'aspiration qui permet de sélection-

ner les pages à rapatrier en vue d'en assurer la conservation.

Différentes procédures peuvent être mises en œuvre. Ainsi, des prélèvements aléatoires peuvent être effectués, le logiciel d'aspiration procédant alors à une ponction du Web à un instant donné, qui donne une sorte de photo, un *snapshot*, d'un état momentané. C'est ainsi qu'ont procédé les pionniers américains de l'archivage du Web, *Internet Archives* de Brewster Kahle.

D'autres entreprises patrimoniales mettent en œuvre des stratégies de sélection basées sur des critères bien définis. Ceux-ci sont soit thématiques, soit formels ou encore nationaux. Ils permettent de faire des partitions dans l'ensemble du Web de manière à en maîtriser la masse sur le long terme. Ils permettent aussi de piloter le robot « moissonneur » à l'intérieur d'un site, dans sa navigation de lien en lien.

Lorsque l'archivage du Web est pris en charge par un État, dans le cadre par exemple de l'application d'une loi sur le dépôt légal, la sélection s'effectue sur l'adresse du nom de domaine, ce qui revient à faire un sous-ensemble du Web par territoire géographique. Cette pratique est celle de la Bibliothèque royale de Suède qui collecte les sites relevant du domaine suédois.

Les services d'archives nationales s'inscrivent également dans cette logique de continuité des fonds en prenant en charge les sites Web des ministères et des institutions publiques.

Dans cet univers numérique, on le voit, tout l'effort des institutions patrimoniales s'attachera à domestiquer ce flux, à l'endiguer à l'intérieur de typologies thématiques, géographiques ou formelles, et à organiser ce gisement de données multiformes et prolifiques.

Des supports jetables

Ces données numériques, fabriquées ou collectées, ne deviendront patrimoine que lorsqu'elles seront stabilisées, authentifiées, référencées et maintenues acces-

4. Ces données sont issues d'une expertise menée en 2002, pour l'Ina par Dominique Pignon, Laboratoire de Physique théorique de l'Ecole normale supérieure de Paris.

sibles dans le cadre d'un archivage pérenne.

Or, un des points de leur vulnérabilité réside dans l'extraordinaire dispersion technologique qui préside à leur constitution. Cette dispersion touche tout à la fois les supports, qui, à un moment ou un autre du processus, seront réceptacles de ces données, et les contenus codés sous forme de bits, en application de normes et de standards d'encodage qui couvrent une gamme très étendue. D'une part ceux-ci varient selon la forme du document d'origine — image fixe, texte, graphique, page Web, fichier sonore, animation 2D ou 3D, flux audiovisuel — et d'autre part, à l'intérieur de la catégorie, selon les évolutions technico-industrielles qui propulsent sur le marché des applications aux formats propriétaires, aux durées de vie de plus en plus courtes et le plus souvent, incompatibles entre elles.

Les méthodes et pratiques de la conservation patrimoniale ont, de tout temps, considéré essentiellement la conservation des supports : celle du papier et de l'encre, des différentes générations de disques, des bandes magnétiques ou encore des émulsions pour le film, la photographie ou le microfilm. Avec le numérique la famille s'agrandit, en particulier grâce à l'arrivée au début des années 80, des CD audio, mis au point par Philips et Sony, puis des CD informatiques sous l'appellation CDRom et CDWorm, puis les CDI multimédia grand public, les photo CD et les vidéo CD. Et finalement, l'arrivée en 1996 des DVD, pour lesquels la bataille de formats fait actuellement rage. Les institutions patrimoniales ont dû intégrer au sein de leurs collections ces nouveaux supports d'édition ou de stockage.

Plus fragiles que les supports utilisés jusqu'alors, parce qu'ayant une espérance de durée de vie encore plus courte, les nouveaux supports utilisés pour les documents électroniques imposent des actions de conservation, de restauration, et des rafraîchissements périodiques.

Des formats éphémères

Dans l'univers de l'analogique, les opérations de transfert menées pour réaliser des copies de sauvegarde généraient toujours une déperdition de la qualité du contenu. Si le numérique permet d'échap-

Les nouveaux supports utilisés pour les documents électroniques imposent des actions de conservation, de restauration, et des rafraîchissements périodiques.

per à la perte de la qualité du signal, il n'est pas pour autant garant de sa survie. En effet, il expose le contenu à devenir « atone » alors même que son support aurait conservé toute son intégrité, et ce, en raison de l'impossibilité de le décoder compte tenu de la grande instabilité des normes et formats d'encodage dans lequel il est né. Le contenu étant constitué par l'ensemble du codage informatique, une perte de portion du code peut provoquer une perte définitive d'interprétation du contenu. Il est donc nécessaire de conserver, avec le support de stockage, son format physique d'organisation des données.

Cependant, le code seul ne suffit pas pour garantir l'accès à long terme. Le codage met en jeu tout un environnement technique qui intervient, non seulement dans la phase de création, mais aussi, lors de l'utilisation, c'est-à-dire à la lecture, en traitant le code pour le rendre intelligible à l'utilisateur, via, par exemple, une simple interface d'affichage. Toutes ces « couches techniques » qui traitent le flux de bits pour en faire un contenu lisible sont particulièrement vulnérables aux évolutions technologiques qui limitent leur durée de vie, rendant inutilisables certains contenus numériques dès qu'ils sont transposés dans un nouvel environnement. Ainsi, la durée de vie des applications logicielles et des plate-formes techniques est souvent beaucoup plus courte que celle du support lui-même. L'archivage sera toujours à la traîne d'une évolution technologique qui se préoccupe peu de la pérennité et qui ne la croise que par hasard.

Dématérialisation et délocalisation : des mémoires sans lieu ni substance

L'un des caractères les plus remarquables est la dématérialisation des supports. D'ailleurs peut-on encore parler de

supports quand ceux-ci n'ont plus de substance matérielle palpable ?

L'encodage connaît une opacité grandissante : le support écrit se lit sans intermédiaire, mais ce qui est encore perceptible directement pour le film ne l'est plus pour l'enregistrement magnétique vidéo ou audio, et encore moins lorsqu'il est numérisé. La lecture nécessite un intermédiaire technique. Cette intermédiation est un nouveau point de vulnérabilité pour la conservation, car, à côté du support, il faudrait pouvoir conserver la machine. Mais cette conservation du dispositif technique est de moins en moins envisageable pour des appareils qui ont coûté des sommes considérables en développement et qui, à l'inverse des exigences de durabilité, sont conçus pour une rotation accélérée des cycles technico-économiques des biens d'équipements.

Le système d'interconnexion des serveurs qu'est Internet exacerbe la dématérialisation des messages. On parle maintenant de façon généralisée des « contenus ». Par cette expression, il faut comprendre que si nous sommes à l'ère des contenus, c'est que nous ne sommes plus à l'ère des contenant, et que les deux faces du message se séparent définitivement. La perte, en l'occurrence, est celle du corps, comme si l'on redoublait à l'ère électronique, et par de nouveaux moyens, une vieille topique de la pensée.

La mise en réseau généralisée permet un saut supplémentaire dans la délocalisation : gérer l'ensemble des serveurs de la planète comme un seul système de stockage intégré. C'est ce que proposent des systèmes comme Napster qui, pour l'échange de musique au format MP3, met en relation chaque internaute (technique appelée *peer to peer*) publiant ce morceau de musique pour ceux désirant l'acquérir. Des systèmes s'affranchissent même de serveur central, les méta données (données associées du type auteur, compositeur, titre de l'œuvre) auto référençant le format MP3 se recomposent

Autant dire qu'un contenu publié aujourd'hui sur Internet n'a aucune chance d'être lisible dans dix ans, si on ne lui a pas, entre temps, organisé des migrations régulières de format.

sans intervention centrale, l'objectif affiché étant l'anonymat des échanges.

L'instabilité technologique

L'informatique est suffisamment répandue aujourd'hui pour que chacun ait appris à connaître ces problèmes de formats, de versions de logiciels, et d'incompatibilité. Même si le matériel restait stable (ordinateurs), même si les supports de stockage étaient pérennes, même si les langages informatiques de base étaient stabilisés (ASCII, etc.), un texte, par sa mise en forme, pourrait être illisible si l'on ne possède pas la bonne version du logiciel l'ayant créé.

Pour ce que l'on a pu en connaître jusqu'à présent, Internet est en évolution technologique permanente. Le prix à payer de la dématérialisation est cette course technologique d'une abstraction toujours plus grande du code par rapport au support, d'une appréhension aussi bien spatiale que temporelle des messages. L'effort physique pour graver l'information devient infime si on le compare aux stèles antiques ou même aux chaînes industrielles lourdes de la chimie des supports film, aux rotatives de presse ou au pressage des disques vinyles. La rotation technologique, et l'obsolescence rapide des machines qui s'ensuit, est inversement proportionnelle : quelques milliers d'années pour les mégolithes, quelques centaines d'années pour le livre, quelques dizaines d'années pour l'audiovisuel, quelques années au plus pour le réseau des réseaux. Telles sont, à peu de chose près, les durées de vie de l'innovation technologique. Autant dire qu'un contenu publié aujourd'hui sur Internet n'a aucune chance d'être lisible dans dix ans, si on ne lui a pas, entre temps, organisé des migrations régulières de format.

Place des pouvoirs centraux

Qu'il s'agisse des lois sur la presse, des monopoles de la radio-télédiffusion, de la gestion régaliennne du spectre des fréquences, etc., les réflexes centralisateurs des États, justifiés par les impératifs de sécurité nationale ou de protection des personnes (protection de l'enfance, de la vie privée, lutte contre la diffamation, etc.), ont élaboré de solides verrous assurant la maîtrise des territoires de l'information, le pouvoir de censurer, d'interdire la publication, de contraindre l'information, de la répandre, mais aussi de protéger ses auteurs, les individus, le pluralisme des courants de pensée et des opinions, etc.

Ce qui est vrai pour les États l'est également pour chaque ramification de pouvoir. Dans chaque entreprise, les systèmes de courriel bousculent des circuits établis et rôdés. Le courriel s'est modelé, à la fois, sur le circuit traditionnel du courrier administratif et sur le système de la communication privée téléphonique. Le second modèle, beaucoup plus labile, l'emporte dans les pratiques, en déstabilisant l'ensemble des systèmes de certification et d'archivage de l'entreprise.

Les composantes principales des anciens cheminements sont en voie de recombinaison sous la forme de ce que l'on appelle des *workflows*, au moins pour les fonctions productives de l'entreprise. En définitive, nul ne sait où nous conduira la mutation. En attendant ce sont des pans entiers de l'information des entreprises qui sont en train de disparaître.

Le droit d'auteur et le nouveau régime de la copie

Tandis que Beaumarchais jetait en 1777 les bases d'une organisation des auteurs pour défendre leurs droits, Con-

dorcet dans les *Fragments sur la liberté de la presse* dénonçait, en 1776, le droit d'auteur comme contraire à la liberté de l'information. La surenchère technologique sur les réseaux prolonge, avec une violence peut-être sans précédent, ce débat entre Beaumarchais et Condorcet.

L'imprimerie a rendu possible l'acquisition de textes publiés dans un état final. Chacun sait comme il est difficile de corriger un texte manuscrit, sa volubilité le rendant vite illisible, et l'habitus scolaire a marqué le passage du brouillon au propre comme l'une de ses valeurs fondamentales. Le texte dactylographié a pu bénéficier des évolutions successives du « Typex », mais son usage est demeuré restreint, et le courage manquait le plus souvent pour reprendre une page entière. Passé un certain stade, seules les corrections essentielles pouvaient être admises. Le traitement de texte a radicalement inversé cette proposition : un texte n'y est jamais fini tant la retouche y est facile, tentante. Seule l'impression sur papier oblige encore à marquer un stade final. Le texte n'a plus de frontières. Il vit dans un océan et entretient de multiples rapports avec les liens hypertextuels entrants, sortants, indexation dans des outils massifs de repérages. Qu'est-ce que l'œuvre, si elle n'a plus de frontières dans le temps comme dans l'espace, et que faudra-t-il en conserver ?

La disparition des différences entre une copie et son original, la facilité à réaliser des copies et à s'en procurer ont conduit à de profondes révisions dans de nombreuses législations⁵. Tant du côté des éditeurs et producteurs que du côté des auteurs, on se mobilise davantage pour contrer le piratage, en général par des techniques de cryptage et des dispositifs anti-copie, notamment sur les DVD. On retourne donc vers la notion de l'original, mais un original artificiel, par dégradation volontaire de la possibilité de copie. Ces systèmes de protection risquent cependant de supprimer les traditionnelles exceptions pour copie privée, ce qui rendrait, difficile, la mission des institutions de conservation.

5. Le *Digital Millennium Copyright Act* aux USA, la Directive Européenne sur « certains aspects du droit d'auteur dans la société de l'information », par exemple.

C'est pour cela que le droit d'auteur se trouve parfois aujourd'hui au ban des accusés et c'est toute une frange de la société civile qui réclame toute entrave à la libre utilisation des œuvres et développe à cet effet des trésors d'inventivité. Le cœur de cette inventivité, c'est l'anonymisation des parcours individuels sur le réseau et la dilution des lieux de stockage de l'information. Il existe un rapport très étroit entre le stockage de l'information et ses modes d'appropriation.

La campagne de l'UNESCO

L'UNESCO a élaboré une stratégie fondée sur les constatations précédentes pour encourager la conservation numérique. Cette stratégie a pour axes principaux :

- une vaste opération de consultation avec les gouvernements, les acteurs politiques, les producteurs de l'information, les institutions et les experts du patrimoine, l'industrie des logiciels ainsi que les organisations qui fixent les normes ;
- la diffusion de directives techniques ;
- la mise en place de projets pilotes ; et,
- l'adoption d'une charte sur la conservation du patrimoine numérique par la Conférence générale lors de sa 32^e session, tenue en octobre 2003.

Les principes directeurs ou Directives⁶ relèvent du volet technique de la campagne à long terme menée par l'UNESCO pour améliorer l'accès au patrimoine numérique chez tous les peuples du globe, et pour garantir que chaque communauté a entre les mains les moyens de préserver son patrimoine numérique. Des cours régionaux, fondés sur ces Directives sont prévus. Le premier de ces cours a eu lieu en décembre 2003 à la Bibliothèque nationale du Brésil. Il a réuni pendant une semaine près de 60 participants venus de 25 pays d'Amérique latine et Caraïbes.

La portée et l'ambition de ces Directives sont toutefois limitées. Dans un domaine où l'évolution est si rapide, mais dont le champ est déjà vaste et complexe, elles ne peuvent présenter qu'une petite quantité d'informations. Les buts recherchés sont de fournir des indications à des

[L'UNESCO mène une campagne] pour améliorer l'accès au patrimoine numérique chez tous les peuples du globe, et pour garantir que chaque communauté a entre les mains les moyens de préserver son patrimoine numérique.

individus et des organisations qui envisagent d'assumer leurs responsabilités dans la conservation du patrimoine numérique, ceux-ci étant souvent placés dans une situation caractérisée à la fois par une absence de moyens et une pléthore d'informations. Aussi a-t-on délibérément adopté une façon d'aborder les principes qui se présente comme un récapitulatif (assez étendu) des problèmes et des possibilités à prendre en considération dans les programmes.

Il est impossible de répondre à tous les problèmes techniques et pratiques qui surgiront dans la gestion des programmes de conservation numérique. Il sera donc plus utile de voir, dans ces Directives, un guide des questions auxquelles les responsables de programme ont besoin de trouver des réponses. Mais ce guide s'appuie sur la ferme conviction qu'il est temps de poser les questions qui débouchent sur une action positive, plutôt que de continuer à poser celles qui ne font que souligner les difficultés.

Il faut espérer que ces Directives, en complément avec la richesse des informations techniques disponibles dans les sources fournies par la liste bibliographique, aideront les responsables de programme de conservation à clarifier les décisions et les mesures à prendre, les principes à retenir et les considérations pratiques à envisager.

En guise de conclusion

L'ancienne gestion de la mémoire était fondée sur le recensement exhaustif, le catalogage fin et soigneux de l'information. La double filiation, patrimoniale et policière, des systèmes d'archivage est pour une large part responsable de la compulsion exhaustive. L'explosion d'une information facilement accessible réduit à néant le rêve exhaustif. La façon d'appré-

hender ces masses d'information demeure un problème majeur. De nouveaux outils, de nouveaux concepts devront être développés pour travailler ces grandes masses, les appréhender selon des approches en logique floue et non plus les maîtriser au niveau de l'atome documentaire. Ce n'est sans doute pas nouveau, l'inflation du papier ne permettait déjà plus l'approche à l'unité, mais la prise de conscience ne s'en était pas encore radicalement manifestée. Les outils actuels d'indexation du Web ouvrent la voie d'une révolution quantique dans la gestion des sources.

Nos sociétés ont vécu du paradigme de l'archive écrite. Il s'est développé sur des centaines d'années, et tout au long du XX^e siècle, les nouveaux médias sont venus se ranger, sagement et modestement, dans cette lignée prestigieuse. Ce paradigme a déjà basculé, et les dispositifs en place sont incapables de prendre en charge cette avancée brutale des technologies de l'information, son étrangeté, et l'inflation quantitative qu'elle apporte.

Au-delà des institutions spécialisées dans la gestion de la mémoire, c'est tout un nouveau régime de l'information qui devra rapidement se construire, bouleversant totalement les anciens dispositifs de mémorisation et d'archive, car, sans le savoir, nos sociétés fabriquent aujourd'hui des trous dans la mémoire sociale. ■

6. Voir le texte de la Charte sur la conservation du patrimoine numérique : http://portal.unesco.org/fr/ev.php?url_id=17721&url_do=do_topic&url_section=201.html.