

La mémoire de la société de l'information

La mémoire de la société de l'information

Publications de l'UNESCO pour le Sommet mondial sur
la société de l'information

Auteurs :

Jean-Michel Rodes, Geneviève Piejut, Emmanuèle Plas

Responsables éditoriaux :

Isabelle Vinson, Abdelaziz Abid

Les désignations employées dans cette publication et la présentation du matériel adoptée ici ne sauraient être interprétées comme exprimant une prise de position du secrétariat de l'UNESCO sur le statut légal d'un pays, d'un territoire, d'une ville ou d'une région, ou de leurs autorités, non plus que sur le tracé de ses frontières. Les idées et opinions exprimées sont celles des auteurs de ce rapport et ne reflètent pas nécessairement la position officielle de l'Organisation.

Publié en 2003 par l'Organisation des Nations Unies pour l'éducation, la science et la culture 7, place de Fontenoy, F-75352 Paris 07 SP

Composition et impression : Jouve, Paris

© UNESCO/2003

Imprimé en France

Introduction générale

L'UNESCO soutient pleinement, depuis le début, le processus de préparation du Sommet mondial sur la société de l'information (SMSI). A ce titre, elle est parvenue à définir et promouvoir ses positions tout en articulant sa contribution à la Déclaration de principes et au Plan d'action que le Sommet adoptera. Les éléments que propose l'UNESCO à fin d'inclusion dans la Déclaration de principes et dans le Plan d'action sont fondés sur son mandat, qui la mène à promouvoir le concept de *sociétés du savoir*, plutôt que celui, général, de *société de l'information*. Il lui semble en effet que se contenter de renforcer les flux d'information n'est pas suffisant pour saisir l'ensemble des opportunités qu'offre le savoir mis au service du développement. Il est ainsi nécessaire de définir une vision plus complexe, complète et holistique de l'utilisation des technologies de l'information au service du développement.

Les propositions que l'UNESCO a ainsi formulées sont des réponses aux principaux défis posés par l'édification des sociétés de l'information : il est tout d'abord nécessaire de réduire le fossé numérique qui augmente les disparités dans le développement, excluant des bénéfices de l'information et du savoir des groupes sociaux et nations entiers ; ensuite, de garantir la libre circulation de et l'accès équitable aux données, à l'information, aux bonnes pratiques et au savoir dans la société de l'information ; enfin, de bâtir un consensus international sur les normes et principes qu'il est désormais nécessaire de défendre.

Les sociétés du savoir doivent en effet être bâties sur un engagement solennel en faveur des droits de l'homme et des libertés fondamentales, au premier rang desquelles la liberté d'expression. Elles doivent par ailleurs assurer l'entier respect du droit à l'éducation et des autres droits culturels. De la même manière, l'accès au domaine public de l'information et au savoir à des fins éducatives et culturelles doit être aussi large que possible

au sein des sociétés de la connaissance et permettre la consultation d'une information fiable, diversifiée et de haute qualité. Une attention particulière doit enfin être portée à la diversité des cultures et des langues.

En outre, la production et la diffusion de contenus éducatifs, scientifiques et culturels, la conservation du patrimoine numérique, la qualité de l'enseignement et de l'apprentissage sont autant de composantes essentielles des sociétés de la connaissance. Il faut enfin favoriser le développement de réseaux de spécialistes et de groupes d'intérêt virtuels, qui sont autant de vecteurs d'échange et de coopération réels et efficaces dans les sociétés de la connaissance. Les technologies de l'information et de la communication sont en effet autant une discipline à maîtriser qu'un outil pédagogique au service de systèmes éducatifs efficaces et adaptés aux besoins.

Enfin, ces technologies ne sont pas uniquement des outils, elles informent et modèlent nos modes de communication, mais également nos manières de penser et de créer. Comment agir de telle manière que cette révolution mentale et instrumentale ne soit pas le privilège d'un petit nombre de pays économiquement très développés ? Comment assurer que tous puissent avoir accès à ces ressources informationnelles et intellectuelles, en déjouant les obstacles sociaux, culturels ou linguistiques ? Comment promouvoir la mise en ligne de contenus toujours plus diversifiés, susceptibles d'être une source d'enrichissement pour l'ensemble de l'humanité ? Quelles opportunités pédagogiques présentent ces nouveaux moyens de communication ?

Il s'agit là d'autant de questions cruciales dont les réponses devront être trouvées pour que les *sociétés du savoir* soient une réalité et offrent un espace d'interaction et d'échange mondial. Ce sont également des questions auxquelles doivent répondre ensemble les acteurs du développement de ces technologies, Etats, entreprises privées et société civile.

A l'occasion du Sommet mondial sur la Société de l'information, l'UNESCO entend mettre à la disposition de tous les participants une série d'ouvrages de synthèse sur certaines des questions les plus préoccupantes que l'on vient d'évoquer. Il s'agit de prendre la mesure des bouleversements induits par l'apparition des nouvelles technologies de l'information et de la

communication (NTIC), en évoquant les promesses de développement mais aussi les difficultés rencontrées, les solutions possibles, et les projets mis en œuvre par l'UNESCO et ses nombreux partenaires.

Abdul Waheed KHAN

Sous-Directeur général de l'UNESCO
pour la communication et l'information

Table des matières

Introduction générale	5
Introduction	13
1 L'urgence numérique	17
1.1. Genèse du nouveau modèle de mémoire	17
1.2. La société saisie par le numérique	19
L'ère du calcul : émergence d'une informati- que scientifique et technique	20
L'ère du dialogue et des interfaces : Les mutations du régime du texte et de l'écrit	22
L'ère de la mémoire : Images et sons	26
<i>La discrétisation du signal</i>	27
<i>Son et musique</i>	27
<i>L'image de synthèse</i>	28
<i>L'audiovisuel numérique</i>	29
<i>La télédiffusion numérique</i>	30
L'ère des réseaux : commerce, services et administration	31
<i>De nouvelles formes de consommation cul- turelle</i>	32
<i>La communication électronique</i>	33
<i>La diffusion des contenus</i>	34
<i>L'action publique</i>	35
<i>Les fractures du numérique</i>	35
2 Un patrimoine fragile	37
2.1. Le domaine du patrimoine numérique	37
L'œuvre et son double	39
Né numérique	41

Un patrimoine infini ?	42
Le Web insaisissable	44
Des supports jetables.....	46
Des formats éphémères	47
2.2. Les troubles de la mémoire	48
La dématérialisation	49
<i>Contenus et contenant</i>	49
<i>La fragilité des supports</i>	49
<i>L'archive nomade</i>	50
La délocalisation : des mémoires sans lieu	50
La déterritorialisation.....	51
<i>La fin des territoires d'archives</i>	51
<i>La place des Etats</i>	52
<i>Les entreprises déstructurées</i>	52
<i>Les nouvelles résistances à l'appropriation des oeuvres</i>	53
<i>Le nouveau régime de la copie</i>	54
La délinéarisation.....	54
L'ouverture du texte, une mémoire sans début ni fin.....	55
L'instabilité technologique	57
L'information envahie par le bruit	57
La fin de l'exhaustif.....	59
3 Les institutions de mémoire face au numérique l'état des pratiques sur le terrain	61
3.1. Collecte numérique	62
Musée virtuel et musée du virtuel	62
L'e-archive	63
La bibliothèque électronique	65
La toile et le flux : Le Web.....	67
<i>La gestion de la masse :</i>	68
<i>Constituer une hyperarchive :</i>	68
<i>Endiguer la volatilité des contenus</i>	68
<i>Eclater la convergence</i>	69
<i>De la non-finitude des objets à l'unité du document</i>	70

<i>Remédier à l'instabilité technologique</i>	70
<i>Conserver et communiquer l'immatériel</i>	71
Les immémoriaux du numérique	71
3.2. Éléments pour la numérisation	72
Elaboration d'une stratégie	73
Les techniques de numérisation.....	74
Les formats de fichiers.....	76
Les supports de données	78
3.3. Les nouvelles lois de la conservation	79
Assurer la migration continue des supports....	79
Assurer l'évolution des formats : migration, émulation, encapsulation.....	81
3.4. Documentation	83
3.5. Des projets pilotes de conservation pérenne	85
4 L'action des décideurs	91
Une législation adaptée.....	91
L'application de la législation.....	94
La coopération entre les acteurs du numéri- que	94
Le droit d'auteur et le copyright.....	95
L'organisation	97
Retrouver l'unité perdue de l'archive du Web .	98
Le financement	98
La recherche et la formation	99
La coopération internationale	100
5 Les recommandations	103

Introduction

Les cinquante dernières années ont vu le développement très rapide d'une informatisation de la société. Le phénomène est si massif, change tellement les contours de nos civilisations que l'on parle maintenant de Société de l'information.

En quatre vagues successives la même impulsion a créé les conditions, les prémisses et, très vite ensuite, l'inflation d'un patrimoine numérique :

- jusqu'aux années soixante-dix : archivage des premières données informatiques issues des sites centraux et des calculateurs scientifiques,
- dans les années quatre-vingts : développement très rapide des supports d'édition numérique : au premier rang le CD audio, mais également les CDRom multimédia et les jeux vidéo sur console,
- dans les années quatre-vingt dix : mise en place de la télévision et la radio numériques par satellite,
- au tournant du millénaire : interconnexion généralisée des réseaux et développement foudroyant d'Internet et de ses applications Web et mail notamment.

Jusqu'au développement d'Internet on pouvait encore hésiter, il restait 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 rendues publiques.

Avec Internet la question est tranchée : le terme est proche où l'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 mémoire traverse jusqu'à nous les générations.

Pour cet homme là, l'informatique sera la plus mauvaise matière à mémoire : les supports sont de mauvaise qualité, destinés à une diffusion de masse et non à être conservés, une machine chasse l'autre pour alimenter les cycles de la consommation, les formats de fichier changent, à chaque nouvelle version logicielle otage de concurrence exacerbée. En quelques décennies on aura vu plus de normes de signes alphabétiques que depuis l'invention de la pierre taillée.

Le risque est grand, si une volonté politique et des mesures à la hauteur des enjeux ne sont pas prises rapidement par les décideurs, que toute notre société de l'information explose sans laisser plus de traces que la bulle Internet. Nos sociétés de l'information se réduiraient à des sociétés du présent dotées d'une petite mémoire de travail, autocentrées dans leur frénésie de communication et tournant le dos aux générations à venir, rompant le fil de la transmission.

Pourtant la numérisation est en même temps un formidable vecteur de démocratisation de la culture, de diffusion des oeuvres. Il reste encore d'énormes fossés à combler en terme d'équipement, de réseaux et de formation avant que chacun sur Terre ait accès à tous les savoirs du monde, mais le chemin est tracé.

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, elle s'effacera. En contrepartie elle restera éternellement jeune.

Ceci ne sera pas sans poser 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.

Au passage les lieux de la lecture publique, les centres d'archives, les musées seront peut-être devenus des serveurs de contenus, bouleversant les métiers et qualifications, les modes d'organisation, les pratiques de lecture, les modes de financement, les partages entre droit à l'information et respect de la propriété des auteurs et éditeurs et les droits de la personne.

Et puis le réseau nous a offert cette nouvelle vision d'un savoir mondial rassemblé, de partout accessible. Pour retrouver cette unité quand on en sera à lui construire une profondeur historique, quand de partout les initiatives individuelles et collectives auront permis de conserver des fragments de la Toile, il ne faudra pas oublier un puissant effort pour recomposer le puzzle. Un effort soucieux de ne pas écraser les aspirations légitimes de chaque peuple à maîtriser sa mémoire, mais traversant toute la nécessaire diversité des institutions patrimoniales.

Emmanuel HOOG

Président Directeur Général de
l'Institut National de l'Audiovisuel (France)

Chapitre 1 L'urgence numérique

La vague de l'informatisation qui a déferlé sur nos sociétés pendant plus d'un demi siècle atteint aujourd'hui une forme de maturité ; sans doute y aura-t-il d'autres vagues, mais en se retournant vers ce passé d'avant le flux 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.

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

L'actualité de cette nouvelle forme de l'archive est d'autant plus brûlante que l'informatique a jusqu'ici plus montré ses capacités d'expansion que ses qualités d'installation dans la durée. Les moyens de production et de circulation de l'information ont été démultipliés dans des proportions jusqu'ici totalement inouïes dans l'histoire de l'humanité. Cependant, poussé par les cycles de la technologie, le cycle de l'information n'aura jamais été aussi court et, si l'on n'y prend garde, nos sociétés risquent de voir disparaître des pans entiers de mémoire à la charnière des millénaires.

Plus on communiquera et moins on transmettra.

1.1. Genèse du nouveau modèle de mémoire

L'informatique contenait en germe les traits spécifiques qui vont lui permettre :

- de se substituer à l'homme dans les domaines les plus divers,
- d'en constituer la nouvelle prothèse à un niveau inimaginable antérieurement,
- d'être ce langage dans lequel tous les autres peuvent se résoudre,
- de s'interposer comme outil générique de production des savoirs et de l'information,
- d'émerger dans de nombreux domaines tout à la fois comme l'instrument, la matière et la surface de travail de la création,
- de devenir progressivement le médium dominant entre les hommes,

- et en définitive de s'imposer comme le lieu d'une nouvelle archive.
- Pour en arriver à ce stade achevé quatre conditions devaient être remplies :
- machine automate permettant le calcul et l'exécution de programmes à des vitesses très supérieures au temps « réel » de la perception humaine,
- machine communiquant et interagissant avec l'homme sous les formes les plus diverses,
- système technique capable de stocker et indexer d'énormes quantités d'information,
- système d'échange généralisé entre machines construisant un univers autonome de publication et de communication.

Toutes les fonctions sont déjà présentes à l'état embryonnaire dans les premières machines – calcul, programmation, périphérique de sortie, mémoire, interconnexion – de fait, la chronologie de l'innovation technologique va privilégier le développement plus rapide de telle ou telle, déployant au passage les nouveaux usages qu'elle autorise et suscite.

Et puis la dernière décennie du vingtième siècle, avec le développement rapide du réseau mondial, va clore le premier chapitre de l'informatique et poser avec une urgence inattendue la question du nouveau genre de patrimoine que cette histoire révèle.

Il est étrange et impressionnant rétrospectivement de constater à quel point les pionniers étaient conscients du bouleversement mémoriel qu'apporterait l'informatique. On pourrait ainsi citer Vannevar Bush qui, dès juillet 1945, dans « As We May Think » décrit le Memex (MEMory EXtender), vaste hypertexte mémoriel, ou encore J.C.R. Licklider qui, en 1962, lance les premiers travaux sur le réseau ARPANET ancêtre de l'Internet et délimite avec une extraordinaire lucidité, dans ce qui n'est encore pour beaucoup qu'une très grosse machine à calculer, ce nouvel univers que les hommes habiteront.¹

1. "The ARPA theme is that the promise offered by the computer as a communication medium between people, dwarfs into relative insignificance the historical beginnings of the computer as an arithmetic engine."

Au delà encore de cet espace propre de création d'information, l'informatique s'est également mise en mesure d'absorber une grande part de l'archive pré-numérique par l'échantillonnage de plus en plus fin et la binarisation des signaux analogiques qui proviennent du monde réel. Avec, bien sûr, des nuances que nous étudierons, toute la mémoire documentaire de l'humanité peut être assimilée dans l'espace numérique.

On y découvre au passage une constante de l'histoire de l'informatique : d'abord le développement de la capacité de synthèse (texte, image, musique, son, objets 3D, ...) investissant les dernières acquisitions technologiques dans de nouvelles interfaces, de nouvelles formes de représentation et de nouveaux périphériques de sortie, ensuite la mise au point des outils de capture, d'analyse et d'intelligence du réel.

Avec les réseaux le numérique devient un monde en soi, qui n'a plus besoin de s'extérioriser. Un monde virtuel selon l'appréhension la plus courante. Dès lors, toute activité humaine qui pourra se traduire sous la forme de bits doit concevoir son avenir patrimonial sous une forme numérique.

La boucle est bouclée, et le champ est large pour cette nouvelle possibilité de s'exprimer. Car, au passage, chaque vague d'innovation technologique a capté une grande part des modes de l'expression humaine les versant du côté de la numérisation. Il ne leur manquait plus que ce territoire final pour s'autonomiser définitivement. Les mutations ne sont certes pas toutes achevées, mais la voie est désormais claire, l'achèvement de la boucle technologique révèle ce qui était resté caché dans cette déferlante du numérique en à peine un demi siècle de notre histoire, une révolution du patrimoine.

1.2. La société saisie par le numérique

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

(ARPA draft, III-24. 1963, Ce document de l'ARPA (Advanced Research Project Agency) est considéré comme la pierre fondatrice de l'ARPANET, ancêtre de l'Internet).

Elle est en œuvre, avec des degrés variables, dans toutes les sphères d'activités, qu'il s'agisse de la production et de la commercialisation des biens et des services, de la création artistique, intellectuelle et scientifique, ou encore de l'administration de la vie publique.

Elle est liée à la généralisation et l'amélioration des performances de l'informatique amplifiées ces dernières années par le développement des capacités des systèmes de communication, dont on ne mesure pas encore totalement les effets sur nos modes de production et d'accès à la culture et à la connaissance.

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 (Compact Disk) numériques à haute capacité..., une suite ininterrompue d'innovations technologiques a pénétré, ces dernières décennies, sans qu'on n'y prenne garde, sans même qu'on s'en aperçoive, des pans entiers de la production culturelle, scientifique et artistique.

Rétrospectivement on observe qu'il y a eu une association, un couplage très fort et irréversible pour chacune des fonctions (calcul, mémoire, interface, réseau) entre son arrivée à maturité et l'informatisation d'un nouveau continent de la sphère sociale.

L'ère du calcul : émergence d'une informatique scientifique et technique

A la fin de la Seconde guerre mondiale les premiers calculateurs électroniques lointains héritiers des bouliers et de la machine de Pascal entament une mutation rapide vers l'ordinateur. S'ouvre alors une ère de profonde transformation qui atteint aujourd'hui sa maturité.

La fin de cette mutation ouvre grand le champ d'un nouveau régime de mémoire pour l'humanité.

Le premier usage de l'ordinateur prend sa source dans ses capacités calculatoires. Il a d'abord été construit pour des usages militaires – cryptographie ou décodage, calculs balistiques – puis pour des besoins scientifiques – tables astronomiques...

Cette première fonction va amorcer le déploiement généralisé de l'informatique dans les sciences et techniques, les tâches statistiques, et les activités de gestion faisant directement appel au calcul.

On sait depuis l'étonnante progression de la capacité et de la vitesse de calcul des microprocesseurs, qui doublent tous les 18 mois (loi de Moore), que chaque accroissement de cette vitesse va permettre à l'informatique d'aborder de nouveaux champs de l'activité humaine.

Le premier domaine dans lequel s'est massivement déployé le numérique a été celui de la science, porté par ses applications aux technologies militaires et spatiales.

D'abord dans les sciences dites « dures » – physique, mathématiques par exemple – qui ont constitué tout à la fois, un champ d'application des avancées technologiques mises au point par les informaticiens et un puissant moteur aux innovations informatiques. Cette fertilisation croisée fut d'autant plus facile qu'il existait, de fait, une proximité intellectuelle et théorique entre les chercheurs qui bien souvent se côtoyaient sur les mêmes campus, voire dans les mêmes laboratoires.

Au cœur de la bataille donc, la maîtrise de l'information scientifique, celle qui est produite par les laboratoires, collectée par de multiples dispositifs d'enregistrement, traitée par des systèmes toujours plus puissants, mais aussi l'information échangée, distribuée, stockée pour répéter et enrichir sans cesse la chaîne de la production de connaissances.

Il a fallu tout d'abord répondre aux besoins des chercheurs de disposer d'une puissance de calcul de plus en plus grande pour traiter des masses de données de plus en plus volumineuses dans des durées de traitement visant le temps réel. Aujourd'hui, avec des puissances dépassant le gigaflops (milliard de calculs en virgule flottante par seconde) on est bien loin de l'ancêtre de 1946, le super-calculateur ENIAC et ses 30 tonnes pour effectuer 330 multiplications par seconde.

Dans le même temps, de nouveaux systèmes de production d'informations numériques étaient mis au point – par exemple dans le domaine de l'imagerie médicale, ou encore dans celui de la recherche spatiale – tout comme de nouveaux outils de visualisation et de représentation de ces données, tels les

logiciels de simulation ou de production d'images de synthèse en 2D et en 3D. Un nouveau regard, une nouvelle façon de voir le monde s'est construite.

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 sur des dizaines et des dizaines d'années.

Et si la sauvegarde de ce capital-information représente pour le monde de la science un enjeu aussi important que leur création et interprétation, il apparaît, dans de nombreux pays, encore très dispersé entre de multiples laboratoires, peu interprétable par d'autres et donc peu transmissible.

Face à ce problème, les communautés scientifiques ont créé le Conseil international pour la science (ICSU) ; l'ICSU gère environ 49 centres de données mondiaux, couvrant toutes les disciplines. Ces centres de données mondiaux ont été mis en place au début des années cinquante pour favoriser les échanges et la création d'infrastructures permettant un meilleur accès aux données et aux archives.

Toutefois 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.

L'ère du dialogue et des interfaces : Les mutations du régime du texte et de l'écrit

Il y a toujours eu, nécessairement, des périphériques d'entrée et de sortie – cartes, bandes, imprimantes diverses, – autour des unités centrales.

L'écran, dans sa forme primitive, remonte à 1951. Dans les années soixante-dix les écrans et le temps partagé s'imposent, ouvrant à une réelle interaction instantanée entre l'homme et la machine.

Progressivement l'ordinateur va sortir des sites informatiques et devenir capable de gérer du texte dans les activités les plus quotidiennes. La fin des années soixante-dix et le début des années quatre-vingts auront vu cette double filiation de la bureautique, enfant illégitime de la machine à écrire qui s'électronisait et des sites centraux qui en devenant micro s'introduisaient dans le bureau de la secrétaire. L'imprimante rapide, et à qualité laser, permettra l'extinction définitive de la première filiation.

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. Les interfaces graphiques accélèreront ce bouleversement de l'imprimerie lourde et son passage à la PAO (Publication Assistée par Ordinateur).

Peu à peu se sont mis en place les outils d'une mutation sans précédent du texte et de l'imprimé. Contemporains de la révolution des interfaces les supports numériques de distribution commencent à remplacer les supports analogiques (le CD audio supplante rapidement le disque vinyle). Le CDROM va plus loin en proposant pour la première des modes de lecture et des systèmes de navigation radicalement nouveaux. Ce sont les premières pierres d'un patrimoine exclusivement numérique.

1975, BRAVO, le premier traitement de texte WYSIWYG (What You See Is What You Get) est mis au point ; quelques années plus tard, les micro-ordinateurs domestiques s'appêtent à conquérir le grand public. Aujourd'hui, on dispose de e-books et l'on accède, depuis son domicile, aux plus grandes bibliothèques du monde. En à peine un quart de siècle, le numérique se sera affirmé, en proposant une nouvelle forme au livre, comme un nouveau support à la culture de l'écrit. Après l'écriture manuscrite et la publication imprimée, il introduit – sans pour autant les remplacer – une nouvelle forme d'inscription du texte et avec elle, de nouvelles modalités de production et de transmission des savoirs. Il introduit aussi de nouvelles pratiques de lecture.

Cette révolution apportée par l'informatique au régime de l'écrit constitue, dans l'histoire de la pensée, une révolution de même amplitude que

celle introduite, entre le 1^{er} et le 4^{ème} siècle de notre ère, par le codex lorsqu'il s'est substitué aux rouleaux, au volumen.

Plus tard, grâce aux imprimeurs, le codex va devenir livre, et en l'espace de quelques dizaines d'années, l'imprimerie, forme massive de reproduction et de diffusion, intégrera et consolidera tous les acquis antérieurs, la pagination, le sommaire, le chapitrage..., contribuant à transformer profondément, et pour des siècles, les usages du texte et du livre. Avec la généralisation de l'objet livre, une distinction forte et visible s'opère alors entre auteur et lecteur, entre écriture et lecture, entre texte et livre.

Or c'est cette répartition des tâches que le numérique commence à brouiller. Sur support informatique, le texte se soumet à toutes les manipulations, à toutes les révisions : le lecteur peut l'annoter, le couper, le copier, le recomposer, s'en approprier des fragments, le ré-écrire et devenir à son tour co-auteur. Et de ce fait, chacun pouvant devenir acteur d'une écriture plurielle, c'est la notion même de création, définie jusqu'alors comme un acte individuel singulier et original, qui est remise en cause¹.

C'est aussi la profession d'éditeur qui a été profondément bouleversée. En premier lieu, ce sont ses méthodes et outils de travail – de la généralisation des techniques de la PAO à la distribution en ligne – qui ont basculé ces dernières années dans le tout numérique, avec pour conséquence immédiate, une optimisation croissante des capacités de publication. En second lieu, elle affecte l'acte d'édition lui-même. Cet acte qui fige le texte à un instant donné, est d'abord un acte de sélection opéré au milieu d'une production foisonnante. Expression d'un choix éditorial, il se pose comme une promesse de qualité et d'aboutissement. Rien de tel dans l'univers numérique où les textes ne sont plus que des fragments d'un flux mondial, anonyme et mouvant.

Dans l'univers de l'imprimé, les textes sont incarnés dans des formes matérielles qui induisent des attentes particulières, des contrats de lecture différents selon qu'on se trouve par exemple, devant l'article d'un jour-

1. R. Chartier, *Culture écrite et société*, Paris, Albin Michel, 1997.

nal, une lettre, une revue savante, une page d'encyclopédie. La classification et la hiérarchisation des discours opérées grâce à cette matérialité de l'imprimé s'effacent dans l'univers numérique où le texte apparaît à la surface d'un écran sans profondeur, uniforme et banalisé. En perdant leur identité physique, les différents genres textuels deviennent semblables dans leur apparence et équivalents dans leur autorité¹.

C'est pour dépasser ces limites qui, à bien des égards, nous ramènent vers le volumen, que de nombreux développements se sont concentrés, ces dernières années, avec des succès mitigés d'ailleurs, sur le livre électronique, une tentative d'offrir aux textes un support stable, daté, signé, certifié. Adossés à des services en lignes ou à des sites spécialisés, ces développements sont particulièrement prometteurs dans le domaine de la formation et de l'éducation. où se profilent déjà de nouvelles formes d'apprentissage et d'éducation à ces nouveaux mode de lecture².

Cette lecture que certains qualifient d'hyperlecture ne se fonde plus sur la logique linéaire et déductive qui soutenait les cheminements de l'argumentation ou de la démonstration classiques. La multiplication des liens hypertextuels suscite, à l'opposé, une circulation éclatée, un raisonnement ouvert, une logique relationnelle, qui ne prend plus appui sur une pensée singulière mais sur un ensemble de mot-clefs. Les conditions d'acquisition des connaissances et les modalités d'accréditation du savoir s'en trouvent profondément modifiés 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.

1. C. Vanderdorpe, *Du papyrus à l'hypertexte*, Paris, La Découverte, 1999.

2. A. Cordier, *Rapport sur le livre numérique*, issu d'une commission de réflexion sur le livre numérique, initiée en 1999 par le Ministère français de la culture dans le cadre du programme d'action gouvernemental, « préparer l'entrée de la France dans la société de l'information ».

L'ère de la mémoire : Images et sons

L'ordinateur ne se conçoit pas sans mémoire : mémoire vive et mémoire de masse. Si la mémoire vive n'est qu'une mémoire de travail volatile qui s'efface quand l'ordinateur s'éteint, la mémoire de masse permet quant à elle de stocker des données ou des programmes sous une forme pérenne.

Il y a toujours eu des mémoires de masse, de fortune au début – pellicules cinéma 35 mm pour le Z1 en 1938 (premier calculateur à relais, créé par Konrad Zuse), cartes ou bandes perforées d'IBM dès 1928, lointaines descendances du métier Jacquard, bandes magnétiques à partir de 1949 – mais ce n'est véritablement qu'avec le disque dur (1956) que l'ordinateur a accédé à ce qui nous semble de l'immédiateté. En 1975 le disque dur Winchester se banalisera dans les sites informatiques, mais il faudra attendre la décennie 1980 pour que les micro-ordinateurs grand public en soient largement équipés.

Depuis les capacités de stockage se sont développées dans des proportions considérables, tant sous la forme des cartouches magnétiques (S-DLT, LTO, ...) qui passeront sous peu la barrière du téra-octet, sous la forme des disques optiques qui permettent de façon peut-être plus pérenne de stocker plusieurs giga-octets et bénéficient de l'extraordinaire diffusion du CD et du DVD, que sous la forme des disques durs dont les prix au méga-octet ont chuté vertigineusement et rendent aujourd'hui possible ce qui eut été hier inconcevable.¹

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

1. DLT : Digital Linear Tape ; S-DLT : Super Digital Linear Tape ; LTO : Linear Tape Open ; DVD : Digital Versatile Disk.

La discrétisation du signal

L'audiovisuel depuis l'origine et dans chacune de ses composantes était analogique. Une dénomination qui n'a réellement émergé que lors de la naissance du numérique, pour différencier un état antérieur de ce qui apparaissait clairement comme une évolution radicale :

Photos ou films, procédés photochimiques, sont analogiques dans la mesure où leur enregistrement varie selon des grandeurs continues en fonction de la quantité de lumière reçue. La bande magnétique (pour le son ou la vidéo) et le disque vinyle (pour le son) accueillent, par magnétisation, par gravure ou par pressage, l'empreinte d'une modulation électrique elle-même traduction et amplification continue de grandeurs physiques (fréquences lumineuses, fréquences sonores).

En analogique l'opérateur humain n'est pas capable d'intervenir et de maîtriser le processus qui atteint chacune des unités discrètes, ni au niveau de l'empreinte (microcristaux d'halogénure d'argent, particules métalliques des bandes magnétiques, granulose du vinyle et élongation du sillon) ni au niveau du signal (gestion fine des fréquences).

Or, l'aspect essentiel de la numérisation est le caractère manipulateur rendu possible par la constitution d'unités discrètes : échantillonnage des phénomènes physiques continus selon des fréquences (nombre d'échantillons pris par seconde) et des niveaux de définitions (taille en bit de chaque échantillon) variables ; ainsi, plus la fréquence et la définition seront élevées et plus la qualité de l'enregistrement se rapprochera de la qualité originale. Celle-ci est elle-même un absolu indéfinissable dépendant des moyens de captation (optiques, micros) et de l'objectif visé, sachant que tout support, y compris analogique, et tout dispositif technique a ses propres limites, son grain.

Son et musique

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 qui ont contribué à définir les contours d'un nouvel espace musical défriché par Pierre Schaeffer, puis dans l'élaboration et la mise au point d'outils de composition assistée par ordinateur

trouvant leur aboutissement dans la généralisation, dans les années quatre-vingts, de logiciels d'échantillonnage du son et autres « lutherie électronique ».

Centré 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 au public le plus large, la généralisation de système de type « home studio », renouvellent considérablement les conditions de la production musicale et réintroduisent la continuité rompue par les techniques de reproduction électromécaniques, entre créateur et consommateur.

En matière de diffusion, le développement d'Internet, conjugué à la généralisation des normes de compression MPEG et MP3, ouvre 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 (pas moins de 3 millions de fichiers en MP3 échangés par jour en 2002)¹. Déstabilisant les modes de diffusion traditionnelle, qu'il s'agisse de la radiophonie ou de l'industrie du disque, Internet permet non seulement d'accéder, en un ou deux clics, à la plus vaste bibliothèque sonore jamais constituée mais aussi d'accéder à tout un environnement informatif (par exemple les noms des auteurs et interprètes, les titres, les thèmes...), de naviguer dans différents modes de représentation visuelle ou graphique, par exemple par la visualisation synchronisée de la partition, bref, de s'aventurer dans de nouveaux paysages de la musique.

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

1. MPEG : Moving Pictures Experts Group, groupe ISO de normalisation international pour la vidéo, le MP3 est un format MPEG pour l'audio.

temps différé, avant de se généraliser à l'enregistrement, au traitement et la restitution d'images et de sons naturels.

Fluidité des calculs, maîtrise des logiques vectorielles, ergonomie des interfaces et capacité de stockage vont permettre le dessin sur ordinateur. Les applications en seront très larges, notamment en matière de Conception Assistée par Ordinateur ou de création artistique.

Pour des raisons de vitesse de traitement et de capacité de stockage, la numérisation s'attache d'abord au traitement de l'image fixe avec le développement des scanners et des logiciels de traitement de l'image de plus en plus ouverts au grand public. De même, le poids relativement plus faible de l'image fixe ouvrira très tôt la possibilité de circulation en réseau pour les agences d'images.

C'est par les techniques de la synthèse largement issues des techniques de la simulation militaire et aéronautique que la numérisation a pénétré l'audiovisuel. Dans les années quatre-vingts, 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 image de synthèse et images naturelles.

A 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 (Joint Photographic Experts Group) pour l'image fixe et du groupe MPEG pour l'image animée à partir des années quatre-vingts – 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'audiovisuel numérique

Apparaissant clairement comme une évolution radicale d'un état antérieur, le mouvement de numérisation du système audiovisuel a commencé dans la production, par la phase du montage. Dès le début des années

quatre-vingt-dix, le montage virtuel remplace progressivement le montage vidéo. Ce même mouvement s'est produit dans l'industrie cinématographique où la chaîne de post-production image (montage virtuel, effets spéciaux et étalonnage) et la chaîne sonore, sont passées au numérique il y a plus de dix ans.

Quant aux systèmes de prise de vue, les premières caméras numériques ont d'abord été utilisées par les rédactions et les équipes des journaux télévisés qui en ont très vite reconnu les avantages (miniaturisation, souplesse d'utilisation, montage simplifié). Elles vont également séduire les créateurs et cinéastes, et peu à peu, pénétrer l'univers cinématographique, qui au cours d'un siècle d'existence n'avait changé ni de support ni de principe d'enregistrement. Ici, leur utilisation contribue au renouvellement des approches et du regard comme en témoignent les œuvres des cinéastes du groupe Dogma. Avec le numérique, l'expression audiovisuelle et la création cinématographique disposent de nouveaux outils de tournage, de production, de réalisation et de jeu inédits, sans oublier les formidables possibilités en matière d'effets spéciaux.

La télédiffusion numérique

Autre maillon du système audiovisuel à basculer dans le numérique : celui de la diffusion, et en particulier de l'acheminement du programme jusqu'au poste de télévision du téléspectateur. En 1994, Direct TV, la première chaîne de télévision numérique grand public est lancée aux Etats-Unis. En Europe, l'organisation DVB Project (Digital Video Broadcasting) regroupant diffuseurs, industriels, administrations établit, dès 1993 des spécifications techniques et des normes européennes permettant le démarrage de la télévision numérique terrestre, et avec elle, le développement de nouvelles fonctionnalités et de nouveaux services. A titre d'indication, la compression du débit numérique permet de mettre dans un seul canal, selon le taux de compression retenu, 4 à 16 programmes de télévision. Si l'on ajoute la diminution des coûts d'exploitation, la possibilité d'offrir des services interactifs associés, ainsi qu'une image et un son de meilleure qualité, alors, la TV analogique terrestre apparaît pour beaucoup condamnée et certains pays – USA, Grande-Bretagne, Suède, Espagne – ont déjà franchi le cap de la diffusion numérique terrestre.

Après dix années d'évolutions technologique continues, on peut considérer que désormais, toute la chaîne audiovisuelle, depuis la production jusqu'à la diffusion, est numérique et ne nécessite plus, à aucun moment, de repasser par une phase analogique. Il reste encore des supports physiques, car le numérique audiovisuel, notamment pour des raisons de débits très élevés, utilise des supports numériques spécifiques (DAT¹, Digital Béta, ...) mais le terme est proche où la dématérialisation sera complète.

L'ère des réseaux : commerce, services et administration

En 1969 sont jetées les bases d'Internet, et en 1989 est conçu le World Wide Web (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. Avant, l'ordinateur n'est 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.

Le réseau généralisé absorbe rapidement et étend les activités de service (communication privée, distribution, poste, téléphonie, commerce, échanges financiers, en mode BtoB ou BtoC...) et les communications publiques (publication, diffusion, peer to peer, VOD, ...)². Il se superpose ainsi aux anciens maillages qui avaient mis plusieurs siècles à se constituer.

Internet, qui a connu ces cinq dernières années, notamment dans ses applications Web et courrier électronique, un développement foudroyant, affichant de 1998 à 2002 un taux de croissance de 217 % (de 2,851 millions

1. Digital Audio Tape.

2. BtoB : Business to Business ; BtoC : Business to Consumers ; VOD : Video On Demand.

de sites en 1998 à 9, 040 millions de sites en 2002¹ entre aujourd'hui dans l'âge adulte. Après avoir suscité les spéculations les plus folles quant à ses promesses économiques, après avoir donné lieu à une multiplication débridée d'applications et d'expérimentations, ses usages tendent à se stabiliser sur un socle de grandes fonctionnalités désormais solidement établies :

- c'est un outil de communication et de dialogue permettant une mise en relation directe et simultanée entre différents intervenants dotés d'une capacité d'interaction quasi instantanée
- c'est un outil d'information, 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 un nouveau vecteur de diffusion électronique, complétant, voir parfois se substituant aux modes de diffusion traditionnels
- c'est un outil de distribution, d'intermédiation commerciale ou de service permettant une mise en relation directe entre le produit ou service et son acheteur ou usager potentiel
- enfin, c'est 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.

De nouvelles formes de consommation culturelle

Fort de ces 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 culturelles².

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 » (cf. le site de l'ICOM, International Council of Museums) réservé à la communauté des musées, afin de leur permettre

1. Sources OCLC (Online Computer Library Center) : « Web Characterization », <http://www.oclc.org>

2. P. Lévy, *Cyberculture*, Paris, Odile Jacob, 1998.

d'améliorer leur visibilité et leur présence sur l'Internet. Ils contribuent, par la démultiplication des accès aux œuvres et aux expositions, 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 culture¹.

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 domaines jusqu'alors réservés qui sont donnés à voir au plus grand nombre.

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 qui s'échangent informations, analyses et points de vue sur l'objet qui les réunit.

La communication électronique

Aujourd'hui, les services publics et l'administration en ligne sont en passe de faire partie de notre paysage. Ils ne sont cependant que la face visible du processus d'informatisation des organisations. En interne, le déploiement des messageries électroniques et les intranets contribuent à modifier profondément les conditions de production et de circulation de l'informa-

1. A titre d'exemple, une initiative canadienne en faveur du développement des accès et des usages du patrimoine national en ligne : <http://www.museevirtuel.ca>

tion. Les centres traditionnels de pouvoir, fondés sur la détention d'une information rare et précieuse, sont ébranlés tout comme sont remises en causes les organisations du travail pyramidales au profit de formes de travail plus collaboratives.

La communication électronique entre les personnes provoque une sorte de lissage des relations hiérarchiques alors que les mécanismes de prises de décision et l'identification des responsabilités jadis clairement affichée 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 conduirait-il les archivistes à vouloir tout conserver de peur de perdre l'essentiel ? Encore faudrait-il le pouvoir !

La diffusion des contenus

L'interconnexion puis la généralisation progressive des réseaux haut débit bouleverse le monde de la diffusion. Les produits de l'édition sont maintenant soit mis à disposition en ligne, soit distribués par de gigantesques librairies mondiales (Amazon) réduisant considérablement le processus de commande, soit, sans sortir de l'univers numérique, distribués de serveur à ordinateur personnel sous forme d'échange de fichiers ou, enfin, directement diffusés par une édition en ligne (presse) et affichage dans un navigateur HTML (HyperText Markup Language).

Pour l'audiovisuel, le modèle de vidéo à la demande où l'on va puiser dans de vastes catalogues une œuvre audiovisuelle qui s'affiche sur l'écran de l'ordinateur voit enfin le jour. Bien plus qu'un simple système de distribution électronique, l'Internet permet d'accompagner l'œuvre d'un environnement d'information et de documentation complémentaire. Enfin avec la multiplication récente des Web-TV, Internet n'est pas qu'un vecteur de diffusion supplémentaire : il est un nouveau média qui réalise la fusion entre l'audiovisuel, les sons, les textes.

L'action publique

Nombreux sont les gouvernements qui ont impulsé, vers le milieu des années 90, des politiques volontaristes d'équipement et des mesures d'incitation à l'usage de l'informatique dans l'administration et les services publics bousculant parfois des habitudes de fonctionnement bien établies. Effet de modernité, peut-être, mais bien plus, on peut lire dans ces politiques un souci d'améliorer la relation avec les administrés alors même que certains pans des appareils d'état font l'objet de vives critiques de la part des citoyens ; on peut aussi y voir un besoin réel d'augmenter l'efficacité des services, un moyen de réduction des coûts et des délais dans une économie mesurée¹.

Aussi puissant, utile et efficace qu'il soit, Internet a sa part d'ombre, et non des moindres : celle de son absence de mémoire. Dans 50 ou 100 ans, quand des historiens se pencheront sur cette époque des technologies de l'information et de la communication triomphantes, ils risquent de se trouver devant un « grand fossé numérique » 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.

Les fractures du numérique

Le numérique présente des atouts tout à fait considérables en matière de diffusion de l'information et de démocratisation de la culture. Pendant plusieurs dizaines d'années cependant il continuera à créer des clivages au sein de chaque société entre des couches de la population qui auront les moyens financiers d'y accéder, les moyens culturels de le maîtriser, qui seront dans des lieux à forte densité de réseaux haut débit et des couches de la population qui resteront sur le bas côté de la route. De même le numérique s'est répandu très vite dans les sociétés industrielles et dans quelques niches des pays en développement, mais des zones géographiques entières en sont exclues².

1. Rapport du 3^e forum mondial de l'administration en ligne au service de la démocratie et du développement organisé en mars 2001 par l'OCDE, <http://www1.oecd.org/puma>

2. 605 millions d'internautes en mai 2002, à peine 10% de la population mondiale (0,39% en 1995) : Europe 31%, Asie Pacifique 31%, Amérique du nord 30%, Amérique latine 5,5%, Afrique 1%, Moyen Orient 0,8% (Sources nua.com/surveys).

En matière de patrimoine numérique les études du marché du stockage montrent que l'acquisition de supports de masse se concentre dans les pays industrialisés (disques durs, bandes magnétiques, supports optiques). Tous ces supports ne sont pas supports de patrimoine, mais on peut quand même penser qu'il existe de fortes corrélations entre la quantité d'information créée et stockée dans ces pays et la croissance d'un patrimoine né numérique ou numérisé.

On connaît par ailleurs la prééminence de l'anglais sur Internet (72% des sites, sources OCLC 2002), même si la tendance est à la baisse elle est significative de l'avancée des pays de langue anglaise en terme de diffusion de la culture numérique dans les couches de la population, mais également de la tendance à l'uniformisation des outils de communication planétaire et donc au risque de déshérence des patrimoines minoritaires.

Chapitre 2 Un patrimoine fragile

Tant que l'information a emprunté des supports physiques pour circuler, elle faisait 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.

Et puis 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 multiplication des couches d'abstraction logique entre la matière et le contenu sémantique accessible aux humains ont rendu le code plus éphémère que son inscription.

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

- *enfin 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,*
- *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.*

2.1. Le domaine du patrimoine numérique

Dans son acception classique, le patrimoine peut être défini comme l'ensemble des données (monuments, collections muséales, archives, bibliothèques...) ou pratiques qu'une société hérite de son passé et qu'elle entend préserver et transmettre aux générations futures, dans le but de constituer un socle commun de valeurs et de références sur lequel se développe un sentiment d'appartenance et de partage de valeurs sociales communes.

Ces valeurs sociales sont incarnées dans des biens collectifs, matériels ou immatériels, qui peuvent recouvrir un domaine quasi illimité dès lors qu'ils sont investis, par un groupe social déterminé, d'une fonction identitaire ou

communautaire. Des grandes œuvres de l'humanité qui méritent d'être connues et recensées aux pratiques artisanales d'un groupe ethnique particulier, le domaine du patrimoine reste instable, ses frontières en sont floues et ses critères inévitablement flottants¹.

Dans le sens le plus large, les grands principes de sélection de ces biens qui marquent les limites du domaine et qui guident l'action des organisations culturelles et patrimoniales reposent sur la caractéristique fondamentale suivante : « beaucoup de ces ressources ont une valeur et une importance durables et constituent par conséquent un patrimoine qui doit être protégé et conservé pour les générations présentes et futures. Il peut exister dans n'importe quelle langue, n'importe quelle partie du monde et n'importe quel domaine du savoir humain ou mode d'expression. »². Ils concernent potentiellement toutes les manifestations et formes d'expression de la production intellectuelle et artistique de l'humanité, incluant, au fur et à mesure de leur apparition, les nouveaux vecteurs de cette production. Photo, cinéma, audiovisuel, multimédias font désormais partie, d'une manière incontestable, du champ patrimonial. Aujourd'hui, ce sont les données numériques qui sont appelées à le rejoindre.

En effet, à l'œuvre dans toutes les sphères de la pensée et de la création, la numérisation est en marche. Résultant de ce processus, une part importante de fichiers informatiques stockés sur des centaines de serveurs, constitue déjà de nouvelles ressources culturelles, éducatives ou scientifiques qui complètent, voire se substituent aux composantes traditionnelles du patrimoine, celui du livre et de l'écrit conservé dans les bibliothèques, celui des œuvres d'art et des collections d'objets conservées dans les musées, celui de la mémoire des sociétés et des organisations publiques conservée par les institutions d'archives.

Or, ce patrimoine numérique est fragile, d'autant plus fragile que, dans bien des cas, il n'existe que sous forme numérique. Face à cette fragilité, et pour pouvoir transmettre aux générations futures cette mémoire contempo-

1. Le domaine couvert par la publication de l'UNESCO *La revue du patrimoine mondial* apporte de nombreux exemples de la notion du patrimoine, tout comme la Résolution de l'Assemblée Générale des Nations Unies sur le patrimoine culturel, A/RES/56/8 – 2002.

2. Projet de *Charte sur la conservation du patrimoine numérique*. Ce texte sera soumis à la Conférence générale de l'UNESCO en septembre 2003 pour adoption.

raine, des dispositifs d'archivage et des stratégies de conservation doivent être mis en place. Ils doivent apporter une réponse aux besoins de conservation pérenne d'un domaine encore mal perçu en tant qu'objet de mémoire, parce qu'il renvoie à une problématique encore mal maîtrisée et parce qu'il semble se heurter à une instabilité technologique déroutante ; les politiques de conservation sont appelées à se déployer dans le double paradoxe, celui du stockage de l'immatériel et celui de l'archivage du flux.

Toutefois, ce patrimoine numérique est loin d'être homogène et les conditions de sa conservation ne seront pas tout à fait de même nature selon son origine ; il peut s'agir en effet de la conservation des contenus issus de la copie, ou de la représentation numérique, d'œuvres originales matérielles préexistantes – la question essentielle sera celle des supports et des outils de lecture – ou des contenus qui n'ont d'autre forme que leur configuration numérique ; dans ce cas, les expériences menées sur le terrain nous montrent que la prise en compte des impératifs de la conservation s'impose à toutes les étapes de production.

L'œuvre et son double

Une part importante du patrimoine numérique 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.

Ces campagnes de numérisation, c'est-à-dire d'encodage en données binaires d'un contenu qui, par cette opération, s'affranchit de son support initial, sont menées à l'initiative d'institutions patrimoniales, musées, bibliothèques, centres d'archives, ou encore d'institutions culturelles, éducatives, scientifiques. Mobilisant des moyens spécifiques souvent importants, elles ont pour objectif soit de garantir une meilleure sauvegarde des fonds détenus, soit d'en améliorer les conditions de lisibilité et d'accès.

En effet, la décision de promouvoir de telles opérations de numérisation des fonds peut servir plusieurs finalités ; il peut s'agir tout d'abord de

répondre à la nécessité de lutter contre la dégradation physique des supports entraînant la perte du contenu qui y est inscrit. La numérisation et le transfert sur support numérique constituent ainsi une solution permettant :

- la protection de l'œuvre originale, quand son accès implique des manipulations qui peuvent être sources de détérioration du support unique sur lequel elle existe. Par exemple, la bibliothèque de Göttingen, en Allemagne a entrepris de numériser la Bible de Gutenberg qu'elle conserve pour la protéger de toute dégradation tout en la rendant accessible au plus grand nombre.
- la conservation, sur le long terme, d'œuvres portées sur des supports exposés à des dégradations physiques ou chimiques irrémédiables, par exemple l'acidité pour le papier, le syndrome du vinaigre pour les supports-film ou vidéo : il s'agira ici de sauvegarder un contenu avant qu'il ne s'efface à jamais d'un support matériel en voie de décomposition.
- le maintien des conditions de lisibilité, lorsque l'accès à l'œuvre implique des intermédiaires techniques de lecture menacés d'obsolescence. Il en est ainsi par exemple des collections d'enregistrements analogiques audio et vidéo, un patrimoine mondial évalué par l'UNESCO à près de 200 millions d'heures qui risque d'être définitivement inaccessible faute d'outils de lecture¹.

Dans le même temps, ces politiques de numérisation peuvent servir des objectifs de promotion patrimoniale, de valorisation des collections et d'optimisation de leur diffusion. Ainsi, il s'agira de réaliser :

- la production de catalogues illustrés virtuels ou la constitution d'une mémoire numérique des œuvres ; organisées en banque de données ou éditées sur disques numériques, ces ressources électroniques reproduisant les œuvres permettront de garder et de rendre immédiatement accessible le témoignage d'une création artistique ou intellectuelle. De nombreuses initiatives de musées en fournissent des pratiques exemplaires,
- optimisation des accès grâce à la mise en ligne, via des sites dédiés, de ces données ; musées virtuels, bibliothèques électroniques, banques

1. Sources : Sous-comité de technologie du programme Mémoire du monde de l'UNESCO et le Comité technique de l'IASA (International Association of Sound and Audiovisual Archives)

d'images fixes ou animées, sites musicaux... constituent aujourd'hui autant d'espaces virtuels de fréquentation et de consultation des œuvres qui, plongées dans un environnement numérique, bénéficient de nouvelles fonctionnalités d'accès, de lecture et d'analyse.

Né numérique

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.

Ce patrimoine « né numérique » résulte d'un processus de production initiale « tout numérique », le message étant encodé dès sa création – c'est le cas, par exemple, d'une collection de photos numériques de la planète terre ; avec la banalisation de l'usage de l'informatique et la puissance des capacités de traitement, cette production est en augmentation constante. Selon une étude de l'Université de Berkeley, elle est estimée à environ 1,5 milliards de giga-octets soit une moyenne annuelle de 250 mégaoctets par personne, pour chaque homme, femme, enfant vivant sur Terre.

A cela s'ajoute le produit de la numérisation précédemment évoqué, lorsqu'il ne reste de l'œuvre originale qu'un fichier numérique parce que le support premier n'a pas été, volontairement ou par accident, conservé.

On se trouvera donc, d'un point de vue technique, devant deux sous-ensembles, chacun appelant la mise en œuvre de stratégies de conservation particulière :

- celui pour lequel il existe un support physique sur lequel est enregistré le fichier numérique, par exemple un cédérom multi média éducatif ; dans ce cas, on est devant un objet stabilisé, fini et définissable. L'attention que mobilisera une politique de conservation sera portée sur la conservation physique de ces supports, sur les outils matériels et logiques de lecture permettant d'accéder au contenu, ainsi que sur la maintenance du fichier numérique initial permettant d'en garantir la lisibilité ;

- celui dont les contenus sont des fichiers informatiques hébergés dans des mémoires de masse d'ordinateurs ou de serveurs, par exemple une base de données ou les pages d'un site Internet. Ici le contenu s'est libéré définitivement d'un support physique défini et stable. Il se compose, à la demande, sur l'écran d'un ordinateur, à condition de disposer de la bonne configuration matérielle et logicielle. Seule la mise en œuvre de véritables stratégies de captation permettront de fixer ce flux pour pouvoir l'archiver.

Un patrimoine infini ?

Une part sans cesse croissante de l'information produite aujourd'hui dans presque tous les domaines de l'activité humaine l'est sous forme numérique, soit, comme on l'a vu, stabilisée sur un support, soit accessible en ligne via Internet ou des intranets locaux. A l'inverse, la part totale de l'information 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 téra-octets¹. Ce volume considère l'ensemble de la production d'écrits faisant l'objet d'une publication (livres, périodiques, littérature grise) soit 230 To, l'édition de CD et de DVD, soit 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 téra-octets. L'activité privée d'échange de courriers électroniques est très supérieure à celle du Web, elle est évalué entre 10 et 20 000 téra-octets par an.

Précisons que ces estimations ne prennent pas en compte les immenses bases de données scientifiques, plusieurs centaines de téra-octets qui constituent ce qui est communément désigné sous l'appellation de « Web profond ».

1. 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.

Pour aussi important qu'il soit, le problème du volume, d'un point de vue purement technologique, n'est pas sans réponse puisque les progrès de l'électronique sont tels qu'ils permettent une augmentation constante de la capacité des supports de stockage alors que le coût par méga-octet stocké est en réduction : les capacités des disques durs doublent tous les 18 mois et le prix moyen par méga-octet est passé de \$ 11,54 en 1988 à \$ 0,2 en 1999.

A titre indicatif, la bande magnétique, en 1975, n'offrait qu'une capacité de 30 méga-octets alors qu'aujourd'hui on dispose de cartouches magnétiques de plus de 200 giga-octets ; des capacités de l'ordre du téra-octet sont annoncées par les constructeurs à moins de 5 ans, capacité qui permettra de stocker sur un seul support 1 million de documents écrits de 100 pages chacun ou encore 1000 heures de vidéo compressées en qualité moyenne (2 mégabits/seconde), ou encore 30 000 à 60 000 heures de musique enregistrée en MP3.

Pour autant, toute cette production a-t-elle vocation à devenir patrimoine ? Et quand bien même l'aurait-elle, 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 et qui en garantissent 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 à l'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-traitée à un prestataire externe, en particulier lorsque les équipements techniques en jeu impliquent de lourds investissements dans un univers technologique parfois encore instable, comme c'est le cas actuellement pour la recopie numérique des fonds audiovisuels.

Elles peuvent être menées systématiquement lors de 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.

Celle-ci va permettre de définir des priorités basées sur trois types de critères : des critères techniques - par exemple seront numérisés les fonds les plus fragiles - des critères de contenu - par exemple l'attention sera portée sur l'intégralité d'une collection - ou des critères d'usages, par exemple il s'agira de numériser les documents les plus demandés.

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

Le Web insaisissable

Mais l'approche est toute autre quand on considère l'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 effectuent des enregistrements réguliers. Leur travail est guidé par un plan d'aspiration qui permet de sélectionner 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 par exemple celle de la Bibliothèque royale de Suède qui collecte les sites relevant du domaine suédois. Il peut aussi s'agir d'une sélection basée sur un critère linguistique. Ainsi, la Bibliothèque royale de Suède complète sa sélection nationale en archivant les sites exprimés en suédois. Le principe de collection guidera les stratégies de collecte retenues par certaines bibliothèques dont l'intention est, dans une logique de continuité des fonds, de conserver le prolongement électronique de collection existante ; il s'agira par exemple de compléter des collections de périodiques en archivant leur édition en ligne; les services d'archives nationaux s'inscrivent également dans cette logique de continuité des fonds en prenant en charge les sites Web des ministères et des institutions publiques.

D'autres stratégies de collecte peuvent prendre appui sur des critères de contenu et de thématiques qui permettent de constituer une archive spécifique ; par exemple ont été constituées, à l'initiative de chercheurs de l'Université de l'Etat de New York avec l'appui de la Bibliothèque du Congrès des Etats-Unis et la fondation Internet Archives, les archives du Web du 11 septembre 2001. De son côté, l'Institut d'Histoire sociale, en France a entrepris en 1994 de collecter les documents parus sur l'Internet au sein des forums de discussion, qui se rapportent à la politique, aux affaires sociales et à l'écologie. L'Institut d'Histoire sociale a rassemblé 900.000 messages émanant de 974 forums accessibles sur l'Internet.

Enfin, cette sélection peut s'attacher à des critères formels en considérant la forme d'expression en tant que telle, ce qui renvoie à la nature du média présent dans les sites. En France par exemple l'Ina (Institut National de l'Audiovisuel) envisage de se concentrer sur la conservation du Web radio-tv alors que la BNF (Bibliothèque Nationale de France) s'intéressera en priorité au produit de l'édition électronique.

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 accessibles 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. Ceux-ci, en effet, varient selon, d'une part, 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 quatre-vingts des CD audio mis au point par Philips et Sony, puis des CD informatiques en général désignés CDRom et CDWorm, puis les CDI multimédia grand public, les photo CD et les vidéo CD, et enfin, 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.

Mais chacun sait qu'aucun support n'est éternel : selon l'expertise des institutions patrimoniales comme la Library of Congress ou la BNF¹, il apparaît que le disque en plastique utilisé pour les CD audio, les CDI, les

1. C. Lupovici, *Les besoins et les données techniques de préservation*, présenté au 67^e Conseil et Conférence générale de l'IFLA, Boston 2001.

CD photos, les CD ROM pressés et dupliqués à partir d'une matrice pourraient n'avoir qu'une durée de vie de l'ordre de 10 à 25 ans dans les conditions moyennes de conservation et d'utilisation. Le disque enregistrable n'aurait qu'une durée de vie de l'ordre de 3 ans avant d'être gravé (en raison du vieillissement de la couche sensible de même nature que celle d'une pellicule photographique non encore impressionnée et fixée) et une durée de vie sans altération de l'ordre de 5 à 10 ans une fois gravé. Ces indications proviennent d'essais de vieillissement en laboratoire et sont plus pessimistes que les durées annoncées par les constructeurs. Les supports magnétiques sur bande ou sur disque offrent quant à eux l'inconvénient de la migration magnétique naturelle due aux champs magnétiques à l'intérieur de la couche sensible.

Plus fragiles que les supports utilisés jusqu'alors, 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'échapper 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 de l'information, 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 rapidement inutilisables certains contenus numériques qui seraient 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.

D'autre part, ce vieillissement technologique n'est pas identique pour tous les types d'objets ; il varie selon l'utilisation des standards et des applications ouvertes, c'est-à-dire pour lesquels on dispose des profils techniques qui seront relativement faciles à faire évoluer, ou bien selon l'utilisation d'un format ou d'une application propriétaire indissociable d'une plate-forme technique particulière on parle alors d'application fermée. Se trouve en particulier dans ce cas l'essentiel de la production des contenus numériques faisant l'objet d'une édition en nombre ; à noter qu'en outre, les éditeurs intègrent bien souvent des dispositifs anti-piratage qui augmentent d'autant les difficultés de leur conservation.

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.

2.2. Les troubles de la mémoire

On souligne souvent dans l'explosion de la Société de l'information son caractère de mise en relation généralisée des hommes, d'interconnexion des réseaux, l'avancée démocratique donnant la possibilité à tout un chacun d'accéder à l'information, à tout type d'information, à tout moment, voire de la créer.

On s'inquiète à l'inverse de la sécurité des transactions, de la protection de la vie privée, de la porosité du réseau aux trafics et aux messages illicites (pornographie, nazisme, espionnage, cyber-terrorisme...); Internet déboussole la sphère sociale dans toutes ses composantes.

On mesure moins les bouleversements que cette explosion impose aux régimes traditionnels de mémoire dans nos sociétés. Ces bouleversements sont pourtant sans précédent, provoqués par un faisceau serré de facteurs multiples, et remettent en cause plusieurs siècles d'organisation patiente

d'une des composantes fondamentales de la mémoire sociale, cette mémoire qui s'était organisée autour de l'archivage du système de l'écrit sur un support papier.

La dématérialisation

L'un des caractères les plus remarquables et qui apparaît à tous est la dématérialisation des supports. Peut-on d'ailleurs réellement encore parler de supports quand ils n'ont plus de substance matérielle palpable ?

Contenus et contenant

Cette évolution vient de loin. Avec l'imprimerie, l'expression de la pensée s'est trouvée très tôt clivée en arts qui peuvent être mécaniquement reproduits, conçus pour être distribués à de multiples exemplaires, et arts dont la forme est indissociable d'un support unique¹. La valeur résultant de cette distinction, qui ouvre sur la notion d'original et de copie, est sans commune mesure selon les deux genres artistiques.

L'audiovisuel ajoute une nouvelle dimension à cette versatilité de l'expression, puisque la forme du message et pas seulement son contenu devient reproductible mécaniquement.

Mais la grande révolution à cet égard est apportée par le numérique. La distinction entre original et copie y disparaît totalement dans la mesure où la reproduction se fait à l'identique, sans perte.

La fragilité des supports

Les supports, quant à eux, connaissent une évolution parallèle. Le signal y est enregistré de façon de plus en plus dense, et le support est de plus en plus fragile : des tablettes d'argile au papyrus, au parchemin puis au papier, la perte de durabilité se chiffre en centaines voire en milliers d'années, sans parler des techniques de fabrication mal maîtrisées qui conduisent à une dégradation accélérée (acidité, syndrome du vinaigre). Et

1. Les sémiologues parlent d'arts autographiques et d'arts allographiques (Nelson Goodman, Gérard Genette).

aujourd'hui, personne ne sait prédire la durée de vie des disquettes informatiques ou des disques CD. Pour ces derniers, et selon les experts et les constructeurs, la fourchette varie de 5 ans à 270 ans !

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,

L'archive nomade

L'ensemble des conditions techniques conduit à ce que la dématérialisation du message s'accroît, et que sa survie dépende d'une rotation de plus en plus rapide de ses supports hôtes. Certains¹ imaginent même qu'il n'y ait plus de support de stockage, on pourrait presque dire de stationnement, mais qu'un message vive et se pérennise dans l'éther du réseau, rebondissant de machine en machine, au gré des caches des différents serveurs.

La délocalisation : des mémoires sans lieu

Le système d'interconnexion des serveurs connu sous le nom d'Internet exacerbe la dématérialisation des messages. On parle maintenant de façon généralisée des *contenus*, expression sous laquelle il faut bien entendre que si nous sommes à l'ère des contenus c'est que nous ne sommes plus à l'ère des contenants, et que les deux faces du message se séparent définitivement. La perte en l'occurrence c'est celle du corps, comme si l'on redou-

1. Le colloque FIAT (Fédération Internationale des archives de télévision) /IASA (International Association of Sound Archive) /IAAF (Fédération Internationale des Archives du Film), Santiago, 1999, www.fiatifta.org

blait à l'ère électronique, et par de nouveaux moyens, une vieille topique de la pensée.

Les systèmes numériques, sui generis, n'imposent pas une délocalisation. On pouvait imaginer une utilisation du numérique à l'identique de l'analogique, où le message était résident sur un support, avec une domiciliation fixe et linéaire. Dans le domaine de l'audiovisuel, les choses ont commencé de cette manière, avec des supports numériques à bande reproduisant le fonctionnement de l'analogique et la possibilité de localiser précisément le signal sur un support singulier. Ce fonctionnement s'éloigne pour laisser place à des systèmes de gestion découplant la vue logique de l'information de son implantation physique.

La dispersion maîtrisée des unités élémentaires d'information en des secteurs physiques éloignés est un gage de résistance du message aux agressions que subissent les supports fragiles. Ce fonctionnement impose une forte redondance de l'information, notamment des systèmes d'adressage et de vérification des données, mais l'augmentation des débits réseaux, des vitesses des processeurs et des stockages de masse permettent d'accepter cette inflation des couches internes de gestion de l'information.

Des systèmes robotisés permettent de gérer d'énormes masses de données (on parle de péta-octets soit 1 million de milliards d'octets) vues comme un seul ensemble logique. Seul le système¹ sait encore où est stocké tel ou tel octet et se charge de sa régénération sur un autre support s'il l'estime nécessaire.

La déterritorialisation

La fin des territoires d'archives

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

1. En informatique on parle de HSM, Hierarchical System Management, quand un système logiciel gère de manière transparente pour l'administrateur et les utilisateurs de multiples types de support à vitesse d'accès décroissante (disque dur, disque optique, bandes) dans un même ensemble logique.

Napster qui, pour l'échange de musique au format MP3, met en relation chaque internaute (technique appelée *peer to peer*) publiant un morceau de musique et ceux désirant l'acquérir. Des systèmes s'affranchissent même de serveur central, les métadonnées (c'est ainsi que l'on appelle les données associées du type auteur, compositeur, titre de l'œuvre...) auto référénçant le format MP3 se recomposent sans intervention centrale, l'objectif affiché étant l'anonymisation parfaite des échanges.

La place des Etats

Ce qui est en cause avec l'Internet, c'est donc la possibilité d'une maîtrise des voies de l'information. De tous temps, les pouvoirs se sont développés à partir d'une telle maîtrise. Le contrôle des voies de circulation était le symbole du pouvoir sur un territoire, contrôle des voies de communication donc à des fins de sécurité publique, de préservation de l'intégrité du territoire national ou d'un marché intérieur, mais aussi, et pour les mêmes raisons, contrôle des moyens de communication. Affirmée dans la Déclaration Universelle des Droits de l'Homme en 1948, la liberté de communication a toujours dû composer avec des systèmes d'endiguement. Qu'il s'agisse des lois sur la presse, des monopoles de la radio-télédiffusion, de la gestion régaliennne du spectre des fréquences..., 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...), 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 personnes privées, le pluralisme des courants de pensée et des opinions...

Maintes fois soulignée, l'irruption d'Internet, qui se déploie à l'écart de tout pouvoir étatique classique, bouleverse profondément l'exercice de la puissance publique, conduite à substituer à son pouvoir de réglementation un rôle d'acteur plus ou moins actif de la co-régulation dans un cadre toujours plus internationalisé.

Les entreprises déstructurées

Ce qui est vrai des états l'est également pour chaque ramification de pouvoir. Dans chaque entreprise les systèmes de courrier électronique bous-

culent des circuits établis et rôlés. L'organisation hiérarchique pyramidale est déstructurée, les strates intermédiaires d'encadrement, tirant leur légitimité de micro-stratégies de rétention partielle et temporaire de l'information, sont déstabilisées par la nouvelle circulation horizontale et informelle de l'information. Le courrier électronique a emprunté à la fois au circuit traditionnel du courrier administratif et au système de la communication privée téléphonique. Le second modèle, beaucoup plus labile, l'emporte dans les pratiques, mettant à mal du même mouvement l'ensemble des systèmes de certification et d'archivage de l'entreprise.

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

Les nouvelles résistances à l'appropriation des œuvres

Tandis que Beaumarchais jetait en 1777 les bases d'une organisation des auteurs pour défendre leurs droits, Condorcet 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. Le droit d'auteur a longtemps fait figure de victime au cours de l'histoire, écrasé entre des régimes autoritaires et centralisés réduisant l'appropriation individuelle de la pensée au bénéfice du collectif, et des systèmes libéraux laissant se développer l'aliénabilité privée des œuvres de l'esprit comme tout autre marchandise.

Mais aujourd'hui le droit d'auteur est au banc des accusés et c'est toute une frange de la société civile qui récusé 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. C'est là une preuve supplémentaire du rapport étroit entre le stockage de l'information et ses modes d'appropriation.

Le nouveau régime de la copie

La disparition des différences entre copie et original, la facilité à réaliser des copies et à s'en procurer ont conduit à de profondes révisions dans de nombreuses législations¹. Tant du côté des éditeurs et producteurs que du côté des auteurs, il existe une mobilisation croissante pour faire barrage au piratage, en général par des techniques de cryptage et des dispositifs anti-copie, notamment sur les DVD. On va donc vers un retour de la notion d'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, voire de rendre difficile pour les institutions de conservation l'accomplissement de leur mission.

La délinéarisation

On n'en a jamais fini de la délinéarisation du message. Il semble qu'une loi génère des vecteurs de communication dans un premier temps linéaire, avec un déroulement temporel, sous la forme d'un flux, puis dans un second temps les spatialise, c'est à dire en permet une représentation inscrite dans l'espace. La parole se spatialise dans des stratégies de mémoire, dans des figures de rhétorique, se déploie sous la forme écrite. L'écriture se spatialise en passant des rouleaux au codex, puis en inventant des marqueurs spatiaux (pagination, table des matières, ...). Le texte informatique est passé du défilement à l'écran des origines aux interfaces graphiques.

Plus avant encore le texte s'hypertextualise pour offrir d'autres parcours que son déroulement linéaire premier, il se déplie selon des voies propres à chaque lecteur. Ce qui n'était que poésie d'avant-garde² devient un mode habituel d'écriture et de lecture.

La numérisation a permis la spatialisation du flux audiovisuel. Il est maintenant possible de l'arrêter, de le capturer, de l'annoter, de le structurer, de le relire à un rythme autre que la vitesse contrainte par la machine.

1. Par exemple le *Digital Millenium Copyright Act* mis en place en 1998 par l'administration Clinton aux Etats-Unis d'Amérique, ou encore la *Directive européenne relative à l'harmonisation de certains aspects du droit d'auteur dans la société de l'information*, adoptée en 2001.

2. R. Queneau, *Cent Mille milliards de poèmes*, Paris, Gallimard, 1961.

L'organisation généralisée des serveurs en réseau permet, au delà, de transformer progressivement en un gigantesque stock tous ces flux. Les messages radio, télévisés, mais aussi l'ensemble du système de l'écrit vont vivre le temps arrêté. La possibilité du retour en arrière, d'un accès instantané à toute l'épaisseur historique du temps qui autrefois passait. Les nouvelles du jour chassaient celles d'hier : les journaux mettent maintenant leurs archives en ligne. Et toutes ces informations sont indexées au même titre, les plus récentes comme les plus anciennes, par des moteurs de recherche qui *crawlent* en permanence sur la Toile.

Une information ne disparaîtra que s'il n'y a plus de place sur le serveur. Ce qui n'est pas sans poser problème, certains invoquant un droit à l'oubli face à cette éternité électronique au vu et au su de tous. Internet démultiplie ce phénomène, un texte peut être lié, repris et recopié de multiples fois, et résonner comme en écho d'un bout à l'autre de la Toile, puis s'éteindre selon les lois d'une nouvelle physique qu'il reste à inventer.

La protection des données personnelles devient une préoccupation forte dans ce contexte tant il peut facilement y avoir des appropriations privées de l'information et constitutions occultes de banques de données nominatives pour des usages pas toujours licites.

L'ouverture du texte, une mémoire sans début ni fin.

Depuis l'imprimerie, nous vivons un régime du texte rendu public dans un état final. Des versions ultérieures peuvent réouvrir cet état final et autoriser des modifications, des repentirs, mais le poids du système de l'édition rend cette pratique malaisée.

L'informatique a réouvert le texte dans un inachèvement natif.

Chacun sait comme il est difficile de corriger un texte manuscrit, les surcharges 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 l'usage ne pouvait en être que 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 radi-

calement 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.

Avec la publication électronique en réseau plus rien ne vient entraver l'évolution permanente d'un même texte dont on ne connaîtrait avec certitude que son état à un instant T. Le régime fondamental voulant qu'un texte soit publié dans un état final et datable n'a plus de point d'appui dans le système technique, s'il doit survivre ce sera pour d'autres raisons.

Et puis le texte n'a plus de frontières. Il vit dans un océan de textes avec lequel il entretient de multiples rapports – 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-il en conserver ?

C'est tout un nouveau régime de la citation qui peut en naître. Autrefois l'œuvre incorporait l'extrait cité dans son propre corps. Quel besoin de procéder encore ainsi ? Les références de bas de page peuvent déjà être remplacées par des URLs (Uniform Resource Locator), le texte lui-même pourrait n'être qu'un patchwork virtuel des diverses citations. Mais que deviendra ce texte si l'une des adresses référencées n'est plus active ou si le texte en a changé ? on pourrait imaginer, dans un conte à la Borgès, que le changement d'un texte modifie instantanément par des effets en cascade des milliers d'autres textes, voire toute la Toile.

Cette facilité à citer est source de la violence soulignée plus haut entre libre utilisation et droit d'auteur. Pour beaucoup le réseau est devenu une gigantesque carrière où l'œuvre se construit de ses multiples emprunts à d'autres œuvres prises comme matière première. Ce phénomène est particulièrement sensible aujourd'hui en musique, il se développera rapidement en audiovisuel dès que les débits réseau le permettront.

Tous ces phénomènes ont toujours existé, les œuvres se citaient, elles pouvaient avoir plusieurs versions, on pouvait les lire en diagonale, à l'envers, les confronter à d'autres, les auteurs se nourrissaient les uns des autres,...mais tout ceci se passait à une autre échelle de temps, incommensurable. Les textes étaient assignables et datables.

L'instabilité technologique

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, d'une abstraction toujours plus grande du code par rapport au support, d'une appréhension aussi bien spatiale que temporelle des messages, c'est cette course technologique. 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 mégalithes, 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 quelque chose près les durées de vie de l'innovation technologique.

Autant dire qu'un contenu publié aujourd'hui sous 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.

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,...) un texte pourrait être illisible, et plus encore sa mise en forme, si l'on ne possédait pas la bonne version du logiciel l'ayant créé.

L'information envahie par le bruit

Les masses d'information produites actuellement sont sans commune mesure avec celles précédant les systèmes électroniques. Il y a un changement radical d'échelle, contemporain d'une refondation documentaire.

Les systèmes d'archivage actuels sont hérités du dix neuvième siècle, et une bonne part des techniques de la recherche, notamment historique, y sont proportionnées. Déjà à la fin de l'ère du papier on peut s'interroger sur le rapport que peut entretenir le chercheur avec ces masses documentaires :

plus de 50 000 livres versés par an à la Bibliothèque Nationale de France, plusieurs dizaines de milliers de collections de périodiques, près de 110 kilomètres de rayonnages d'archives administratives versées par an aux Archives de France... face à ces masses considérables les instruments de référencement, catalogues, inventaires et fichiers, sont restés très modestes jusque dans les années quatre-vingts.

Internet amène avec lui une dimension nouvelle, il s'autoréférence. De partout naissent des annuaires, des moteurs de recherche indexant le Web, des index d'index, ... après le grand silence des catalogues arrive l'ère bruyante, proliférante, protubérante des index.

Les moteurs *full text*, les systèmes d'indexation, étaient jusqu'à récemment réservés au petit monde de la documentation. D'un point de vue informatique il s'agissait d'un marché de niche face au monde des SGBD de gestion. Et puis le développement du Web a changé la donne, il s'agit aujourd'hui de l'un des secteurs les plus porteurs et où la recherche est la plus vivace. Elle en a bien besoin, car face à l'immensité de la Toile, au moins en nombre de documents, si l'on ne veut pas être envahis par les résultats à la moindre question il faut que les moteurs développent des stratégies de pertinence, fondées par exemple sur l'environnement hypertextuel d'un texte comme pondération sémantique¹. Tous les textes ne sont pas égaux aujourd'hui sur Internet : comme dans une galaxie il y en a qui sont près du cœur, là où tous les liens hypertextuels partent et arrivent, et d'autres qui sont dans des régions froides, voire dans des galaxies inconnues du centre. Cette galaxie ressemblerait paraît-il à un nœuds papillon².

Qu'une science topologique modélise la répartition mouvante de l'information au sein du cyber-espace montre bien que l'on a changé d'échelle dans notre façon d'appréhender les savoirs humains. Les avancées des sciences humaines au cours des années soixante pour s'éloigner d'une interprétation du texte et construire une science des savoirs devaient concéder à leur objet des heures de dépouillement minutieux en archives ou en bibliothèques.

1. Système PageRank utilisé par Google par exemple.

2. Cartographie des liens sur le Web mise en évidence en 2000 par une étude de chercheurs d'IBM, Compaq et Altavista, intitulée *Graph structure in the Web*, <http://www.almaden.ibm.com/cs/k53/www9.final/temp.gif>

ques, et n'appréhender en somme que des corpus étroits. Aujourd'hui Internet offre à une telle investigation la présence immédiate et totale des savoirs.

La fin de l'exhaustif

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. Mais la tension de la recherche elle-même portait vers l'exhaustivité. Tout travail de recherche sérieux se devait d'avoir fait le tour pointilleux d'une question.

L'explosion d'une information facilement accessible réduit à néant le rêve exhaustif. Le problème essentiel devient avec évidence la possibilité d'appréhender ces masses d'information. 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 maîtrisée 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 le paradigme de l'archive écrite. Il s'est développé sur des centaines d'années, et tout au long du vingtième 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 nos sociétés fabriquent aujourd'hui sans le savoir des trous dans la mémoire sociale.

Chapitre 3 Les institutions de mémoire face au numérique : l'état des pratiques sur le terrain

On rapporte que le Museion d'Alexandrie, le temple des muses, fut l'ancêtre de nos musées, que Pisistrate créa la première bibliothèque publique, à Athènes au sixième siècle avant Jésus Christ. Vraisemblablement contemporaine des premiers états fondés sur l'écriture, l'archive c'est à la fois ce qui est ancien et ce qui réside dans les lieux du pouvoir¹, la trace des actes du pouvoir que gardaient les archontes.

Cette division tripartite des lieux de conservation de la mémoire collective a traversé les siècles, elle a pris un caractère universel en s'étendant à l'ensemble des civilisations et elle traverse les grands partages entre action publique et initiative privée. On touche là sans doute à des invariants qui délimitent le champ des modalités possibles de conservation collective de la mémoire².

Même si nous avons insisté sur le caractère invasif du numérique dans l'ensemble de la sphère sociale et pour une grande part des modes de l'expression humaine, chacune de ces grandes fonctions de mémoire n'est pas concernée au même titre par la conservation d'un patrimoine numérique. Les différences fondamentales tiennent à la nature de document au cœur de chaque institution et au degré de numérisation de chacune des phases de circulation du document.

Il existe toute une gamme de nuances depuis le document dont on peut estimer que sa copie numérique en est un substitut complet au point que l'original ne s'en distingue plus, celui qui subit une dégradation plus ou moins lourde, et le document d'une nature radicalement étrangère au numérique et dont la copie n'en sera jamais qu'une image réductrice.

1. J. Derrida, *Mal d'archive*, Paris, Galilée, 1995.

2. Il faudrait y ajouter le patrimoine monumental, mais on peut le considérer ici comme une autre modalité du musée, et la transmission orale, autre modalité de la bibliothèque. « Un vieillard qui meurt c'est une bibliothèque qui brûle », disait Amadou Hampaté Bâ.

3.1. Collecte numérique

Musée virtuel et musée du virtuel

Le musée conserve des objets souvent uniques, rares (on parlait de cabinet de curiosités ou de raretés jusqu'au XVIII^e siècle) et toujours précieux. Tout le marché de l'art est construit sur une valorisation des objets physiques. Or, on connaît mal le modèle économique qui pourra sous-tendre la valorisation d'un objet immatériel. Déjà les objets reproductibles mécaniquement, comme la photographie, devaient recréer assez artificiellement une notion d'original daté, marqué pour être intégrés dans le marché de l'art.

De manière générale, l'objet de musée est à plus de deux dimensions et la notion d'original y est prépondérante. Le numérique n'offre pas, pas encore, de modalité permettant d'imaginer qu'un objet numérique virtuel pourrait se substituer totalement à un objet physique réel tridimensionnel. C'est a fortiori encore plus vrai pour le patrimoine monumental.

Le numérique ne touchera donc le musée qu'à la marge, essentiellement dans ses fonctions de diffusion, parfois de préservation des contenus après la mort des objets, peu de conservation. Le numérique est un atout extraordinaire de diffusion et de démocratisation en permettant le développement très large d'une fonction de représentation de l'œuvre, en ligne ou sur des supports multimédias développant un appareillage critique, ou même localement quand elle est devenue trop fragile pour être présentée, mais cette représentation n'est pas l'œuvre.

Reste que le numérique est par ailleurs de plus en plus présent à la source dans la création artistique : manuscrits d'auteurs, dossiers de conception d'architectes, de designers, travaux de photographes, esquisses de décors, de costumes, fichiers de dessins animés en 3D ou 2D, maquettes d'urbanisme, graphisme, créations artistiques utilisant toute la gamme possible de traitement de l'image...

Généralement il s'agit encore des phases préliminaires ou intermédiaires de l'œuvre qui selon les cas deviendra monument, tirage papier, film ou vidéo numérique, objet industriel, ... ou site Web, toutes formes de publication dont on peut ramener la conservation à des schémas connus.

Mais que faire de tous ces matériaux d'origine, trace d'une génétique de l'œuvre dont notre époque est si inquiète ? Leur situation est entre deux mondes : celui de l'objet unique tant qu'ils restent dans le circuit fermé de la création, celui du produit infiniment multipliable de par leur nature numérique. Leur origine privée aura de plus conduit à la plus grande dispersion dans l'emploi des formats informatiques rendant leur pérennité très problématique.

L'e-archive

Résultant de l'activité des organismes publics ou privés, les archives constituent un patrimoine social et historique de premier ordre qui permet de transmettre aux générations futures aussi bien la trace de l'exercice de la chose publique que le témoignage de l'activité des personnes et des groupes. A la fois instrument de preuve essentiel au fonctionnement d'une société moderne et support privilégié de la mémoire collective, elles sont le plus souvent confiées à une institution ou à un service spécialisé dont la mission est de rassembler et conserver un ensemble de pièces, de documents, études, rapports, données statistiques... tous supports d'écriture d'actes ou de faits susceptibles de renseigner les hommes d'aujourd'hui ou de demain sur leurs activités ou celles des générations précédentes.

Dûment sélectionnés, classés, catalogués, stockés, ces fonds, traités selon des méthodes éprouvées depuis plusieurs siècles – l'archive n'est-elle pas née avec l'écriture en Mésopotamie ? – vont occuper sagement les mètres linéaires de rayonnages des magasins de conservation, dans l'attente qu'un lecteur vienne les consulter.

Or, cette chaîne de traitement parfaitement maîtrisée jusqu'alors, chancelle. Les prémisses de ces bouleversements étaient observables dès les années soixante, lorsque de grandes organisations ont commencé à utiliser des moyens informatiques pour traiter l'information, en particulier pour traiter des grandes masses de données, par exemple les données démographiques. Ce mouvement s'est accéléré ces dix dernières années : non seulement l'informatique a pénétré tous les rouages de l'activité des organisations mais en plus, celles-ci s'en sont emparé à des fins de communication interne et d'interface avec l'extérieur.

En effet, si l'on considère l'accroissement quantitatif de l'information produite, induite tant par l'augmentation des possibilités de traitement – par exemple multiplication des indicateurs dans le domaine économique – que par la facilité d'écriture et de reproduction générant à la fois forte redondance et brouillage de l'original, et si l'on considère dans le même temps l'extrême volatilité de cette information – courrier, notes, rapports ne se fixent sur un support matériel que si émetteur ou destinataire en prennent l'initiative – on perçoit aisément les troubles que le numérique fait subir au régime de l'archive¹.

On est encore loin d'une société du zéro papier, le terme a même semblé s'en éloigner en proportion inverse de l'informatisation de la société. Pour les archives le résultat le plus concret de l'informatique aujourd'hui, paradoxalement, aura été l'accroissement considérable de la consommation de papier et donc de sa conservation, du fait de l'exceptionnel développement de la production de textes permis par la chaîne bureautique.

Mais, comme on l'a vu (cf. chapitre 1), cette phase de l'informatique est en train de se refermer et depuis quelques années le mail, avec ses fichiers attachés, tend à envahir les systèmes d'administration et de gestion. De plus, la plupart des législations légalisent des systèmes d'authentification du document électronique supprimant une des dernières fonctions nécessaires réservées jusqu'ici à la version papier.

Alors que le mouvement de numérisation est irréversible, très peu d'institutions aujourd'hui se posent la question de l'organisation de leur mémoire pour les générations futures. Dans le passé, cette responsabilité était déléguée aux conservateurs qui, au plus près des organisations, collectaient des documents lorsque leur valeur d'usage (Durée d'Utilisation Administrative) était épuisée, les sélectionnaient selon des critères de tri bien établis, les organisaient dans des fonds parfaitement identifiés et en respectant leur composition organique. Par cette intervention externe à son contexte de production, le document acquérait le statut d'archive². Or, avec le numérique, il apparaît de plus en plus que le niveau de conservation doit être intégré à la

1. Plusieurs dizaines de millions de mails auraient été versés au NARA (Archives Nationales des Etats-Unis) à la fin du mandat du président Clinton.

2. C. Dhérent, *Les archives électroniques*, Paris, Direction des Archives de France, 2002.

chaîne de production dès la création du document, pas seulement comme aujourd'hui en terme de classification de contenu mais également en terme de support et de format, faute de quoi, si aucune disposition n'est prise, il ne restera rien, il n'y aura pas d'archive.

Il appartient donc aux institutions de prendre les mesures nécessaires à la sauvegarde des informations numériques qu'elles produisent, en coopération étroite avec les institutions patrimoniales qui auront la tâche d'en assurer la pérennité.

La bibliothèque électronique

La bibliothèque va être profondément touchée par la révolution du livre, par toutes les phases de cette révolution depuis l'acte de création jusqu'à la diffusion numérique sur les réseaux. On parle ici bien sûr de la bibliothèque dans sa fonction de conservation et de communication d'objets édités en nombre, quel qu'en soit le media, pas de la fonction muséologique souvent également tenue par des bibliothèques (incunables, sceaux et monnaies, manuscrits originaux, ...)

Les grandes bibliothèques ont toutes entamé des programmes de numérisation des collections anciennes, mais, au delà, le livre (ou le journal) né numérique et lu sur le Web modifie en profondeur les fonctionnements de la bibliothèque (collecte, conservation, communication).

De nombreuses législations sur le dépôt légal ont prévu le dépôt des bases de données et des multimédia sur support. La plupart ont, jusqu'à une date récente, écarté le dépôt des documents en ligne¹.

Parmi les produits numériques de l'édition conservés dans les bibliothèques certains ne posent pas de problèmes particuliers de conservation : les CD audio notamment sont facilement duplicables sans perte, grâce à une sortie numérique.

Les DVD vidéo en revanche vont progressivement présenter des difficultés accrues en raison des dispositifs anti-copie qui les équipent et qui commencent à être légalisés par les lois sur le droit d'auteur ou le copyright.

1. Le Danemark, la Finlande, la Norvège et la Suède l'ont intégré dans leur législation. La France s'apprête à le faire.

La complexité des supports multimédia est plus grande encore. Ils n'existent pas en l'absence d'une station d'accueil (micro ordinateur, console de jeux) et des couches logicielles propres à cette station. Le risque est donc très fort que le support ne soit plus jouable dès que la station, voire même la version du système d'exploitation de la station change. C'est déjà le cas pour bon nombre des premiers jeux vidéo dont les consoles ont disparu ou ne sont plus en état de marche (Amiga, Atari, etc).

Une solution à ce problème est de créer une émulation des anciens systèmes d'exploitation sur une machine contemporaine. Il y a encore quelques développeurs indépendants nostalgiques qui développent bénévolement des émulateurs pour faire tourner ces anciens jeux ou logiciels. On peut douter que cette solution soit viable dans la durée, et de plus si les bibliothèques doivent prendre en charge une telle fonction de façon permanente le coût risque d'en être exorbitant.

De façon plus générale la conservation des langages informatiques en état de marche, avec leurs éditeurs de code source, leurs bibliothèques, leurs compilateurs, ... est une question qui n'est pas aujourd'hui traitée sauf par leur éditeur – or beaucoup disparaissent rapidement – ou quelques collectionneurs amateurs.

Si la collecte des produits multimédias distribués peut être ramenée aux catégories connues liées aux supports édités en nombre, en revanche le développement de la librairie électronique en ligne transforme les modalités de la collecte et celles de la conservation.

D'une part cette collecte en adressant des objets nés numériques évite la tâche de numérisation et allège le poids du stockage puisque le format des documents est en mode natif et non en mode image. Pour les documents textuels structurés ce mode de collecte ouvre à un catalogage beaucoup plus fin et automatisable. D'autre part les éditeurs ont tendance là aussi à développer des dispositifs interdisant les copies (morceaux de musique téléchargeables en mode streaming...), la charge de la collecte est inversée et la bibliothèque doit devenir pro-active là où avant elle attendait un dépôt. Enfin la bibliothèque est contrainte de gérer un stock numérique en ligne et donc l'ensemble de la logistique informatique qui lui est nécessaire.

Si cette dernière charge est facilement absorbée par les grandes bibliothèques où la collecte des produits en ligne rejoint la numérisation des docu-

ments anciens dans de grandes bases de données, elle change en revanche radicalement la nature de l'investissement nécessaire pour la réalisation des bibliothèques de petite taille et remet en question la notion de prêt et les modalités de la lecture.

Dans un premier temps la consultation des collections numériques peut s'organiser sous forme de réseaux d'ordinateurs internes aux bibliothèques mais l'évolution devrait conduire naturellement à la bibliothèque virtuelle. Le frein principal, pour les documents qui ne sont pas tombés dans le domaine public, sera la gestion des droits de l'éditeur et de l'auteur. A terme la notion même de bibliothèque de lecture se dilue dans ce qui pourrait devenir de nouvelles modalités de financement de la lecture publique dépendant d'une part directement des sites des éditeurs et d'autre part des sites de quelques grandes bibliothèques centrales. Ainsi, c'est toute une économie de l'intervention publique qui est à reconstruire dans le souci de la démocratisation de la lecture.

En matière de conservation, l'audiovisuel a constitué depuis toujours un véritable casse-tête pour les institutions patrimoniales contraintes de conserver, dans des conditions optimum, non seulement une multitude de supports menacés de dégradation physique à court terme – films et bandes magnétiques en de multiples formats et standards – mais aussi tout un parc d'outils de lecture capable encore de les lire. Sans parler de la perte irrémédiable de la qualité du signal au fur et à mesure des cycles de recopies. En réponse à cet ensemble de fragilité à la conservation, la voie ouverte par le numérique réside précisément dans l'immatérialisation : détaché d'un support matériel, le message se prête à tous les transferts sur des supports hôtes et homogènes, dont les formats auront été définis par l'institution de conservation elle-même. Si l'on ajoute que les vitesses de transfert des fichiers informatiques sont sans commune mesure avec le temps réel de la copie analogique, pour la première fois peut-être dans son histoire, l'audiovisuel peut espérer maîtriser sa mémoire. Mais le prix du passage sera élevé.

La toile et le flux : Le Web

Immatériel, instable, volatile, multi et hypermédia, non fini et infini, le Web pose de redoutables questions aux sciences du patrimoine dont les

méthodes et les pratiques se sont patiemment constituées sur la collecte et la conservation des supports matériels stables.

Pourtant aujourd'hui, les initiatives de conservation se multiplient, et même si tout n'est pas encore résolu, des ébauches de réponses indiquent sérieusement le chemin à suivre pour, enfin, organiser la mémoire de la Toile.

La gestion de la masse

Un des premiers obstacles souvent mis en avant est celui de sa masse, d'un volume supposé si important qu'il serait impossible de le maîtriser. De fait, et s'il est vrai que le Web a affiché de forts taux de croissance, il est en passe de se stabiliser. Pour faire face à cette masse, la réponse des institutions patrimoniales consiste, en fonction de leur secteur de compétence ou centre d'intérêt, à opérer des partitions du Web, à constituer des sous-ensembles soit par territoires, par thématiques, par langues. Les grandes bibliothèques nationales ou les services d'archives, par exemple, s'attachent à la prise en charge de la conservation de la portion du Web correspondant aux limites de leur patrimoine national.

Constituer une hyperarchive

Les pratiques de partition du Web devront avoir pour corollaire l'interopérabilité des systèmes – c'est-à-dire la capacité d'accéder, d'une manière normalisée, à tous les fonds constitués ; cette démarche est déterminante pour redonner au Web toute sa dimension. Elle s'impose d'autant plus que l'univers du Web est caractérisé par l'hypertextualité. Celle-ci repousse les frontières des documents pour ouvrir sur une infinité de parcours où le sens se construit par cheminement successif de lien en lien, de balise en balise soit au sein d'un site, soit entre sites. A l'hypertextualité du Web doit répondre la constitution généralisée d'une hyperarchive.

Endiguer la volatilité des contenus

Autre difficulté à affronter, celle de la volatilité des données : média de flux marqué par l'éphémère, le Web affiche une extraordinaire instabilité des contenus : 70 % des pages ont une durée de vie inférieure à 4 mois et

seuls 10 % des sites en .com restent stables plus d'un mois. Des réponses à cette instabilité des contenus peuvent être trouvées du côté des procédures et des outils de collecte. Ainsi, lors des campagnes d'aspiration, les robots de collecte peuvent détecter les modifications et rafraîchissements des pages des sites visités. Ces informations alimentent un plan d'aspiration qui pilote la collecte en fonction des taux de modification des pages : plus le site est instable, plus les fréquences de collecte seront élevées, et à l'inverse, plus le site est stable, plus les fréquences de visites seront basses.

Eclater la convergence

La convergence des médias constitue une autre caractéristique d'Internet qui n'est pas sans poser de sérieuses difficultés. Dès lors que tous les types d'informations, qu'elles soient textuelles, iconographiques, sonores, audiovisuelles, peuvent exister sous forme de données numériques manipulables, celles-ci trouvent naturellement les chemins de leur diffusion dans le Web qui accueille toutes les formes d'expression des médias antérieurs, parfois en direct pour la Web-télévision ou la Web-radio. La convergence des médias ne se résume pas seulement à la co-existence des médias antérieurs, mais donne lieu le plus souvent à leur fusion dans un régime de relations d'interdépendance où le sens, la plénitude de l'information, s'élaborent de façon interactive par rebond d'une forme d'expression sur l'autre. Cette convergence, cimentée par les hyperliens, fonde l'Internet comme média à part entière et non comme simple vecteur des médias précédents.

Aujourd'hui, l'état des pratiques et réflexions apportent une amorce de solution : il s'agit en particulier d'instaurer un traitement différencié selon le type de média. Ainsi, lors de la collecte, des fenêtres de captations autonomes seront dédiées à l'enregistrement en continu des données sonores ou audiovisuelles diffusées en « streaming », selon des principes somme toute assez identiques à ceux mis en œuvre pour la captation patrimoniale des programmes de radio-TV.

Mais pour permettre à ces contenus de retrouver trace de leur dynamisme lors de la consultation tout en évitant une énorme redondance de données, chaque objet sera traité d'un double point de vue distinguant contenu et structure. Ainsi lors de la collecte, une distinction sera faite entre, d'une

part, le contenu informationnel qui donnera lieu à une description documentaire, et d'autre part son organisation, c'est-à-dire sa structure d'écriture, de mise en page, de mise en relation avec son environnement, qui donnera lieu à une description de structure. Cette voie fait encore l'objet d'expérimentation sur de gros volumes ; elle devrait faciliter la restitution de la réalité dynamique du Web.

De la non-finitude des objets à l'unité du document

Dans un univers non fini, la question de l'unité du document est posée avec une acuité particulière.

Si l'on se place d'un point de vue technique, certaines institutions patrimoniales considèrent que l'objet de l'archive est défini par la page ou encore un fichier repérable grâce à des données d'URL et daté selon un ordre chronologique d'édition. D'autres institutions considéreront le site, voire un regroupement de sites, un portail par exemple, comme unité documentaire qui fera l'objet d'une description et d'une datation. Il faut toutefois considérer que, si l'on se place du point de vue de la restitution, de nouveaux niveaux de granularité constitués par l'utilisateur lui-même, vont venir se superposer aux objets ainsi définis, par exemple en fonction de regroupement thématiques autour de mots clefs ou de relations de proximité. Il faut donc ménager des systèmes de restitution capables de rendre tout son sens à l'archive.

Remédier à l'instabilité technologique

Sur un autre plan, l'instabilité technologique est souvent avancée comme un lourd handicap à la constitution d'une mémoire du Web. De fait, la vitalité du réseau, où un nouveau logiciel en chasse un autre, où une nouvelle application étend le champ des possibles, apporte son lot de perpétuelles évolutions des standards d'encodage et des formats d'écriture qui fragilisent encore un peu plus la conservation de l'archive numérique. Là aussi les institutions patrimoniales prenant la mesure de la difficulté ont avancé des solutions – migration, émulation et encapsulation – permettant d'envisager une conservation à long terme des données. Mais il faudra sans doute avancer résolument sur la voie de la définition d'un format de conservation normalisé pour toutes les institutions d'archivage.

Conserver et communiquer l'immatériel

Enfin constituer une archive du Web implique de fixer les contenus sur des supports de mémoire de masse et par là même de relever le dernier défi du Web, celui de son immatérialité. Toutefois, le choix des supports et de l'architecture de stockage, bien que cette question nous ramène en terrain connu – on connaît les avantages et inconvénients respectifs des supports de mémoire de masse – n'est pas sans se heurter à de nouvelles contraintes, celles en particulier des modalités d'accès et de restitution. Il faudra en effet concevoir une organisation de l'archive permettant d'accéder tant à sa profondeur historique qu'à sa dimension spatiale, de naviguer d'une manière dynamique dans des objets du passé interconnectés entre eux. Intégrer dans l'organisation de l'archive cette dimension spatio-temporelle n'est pas aisée et mérite encore un effort de recherche et développement qui devra s'appuyer largement sur les usages de ce nouveau patrimoine.

Les immémoriaux du numérique

La société numérique rassemble, par touches successives, tout un savoir sur les personnes. Au début il s'agissait surtout de l'informatisation des données produites par l'administration – données d'état civil, revenus et impôts, casier judiciaire, données de sécurité sociale, cursus scolaires, ... – à quoi il fallait ajouter toutes les données issues des relations que l'individu entretient comme client avec différents fournisseurs – consommation, facturation, comptes bancaires, assurances.

La première vague se limitait donc au recueil de données de gestion. De nombreuses législations, pour préserver la liberté individuelle et la vie privée, ont élevé des barrières à l'interconnexion des différents fichiers face à la crainte d' "Etat Big Brother" qui hantaient les esprits.

Aujourd'hui les données numériques vont beaucoup plus loin : cela peut concerner l'ensemble du courrier électronique, des pages personnelles, l'enregistrement des parcours sur les sites, des destinations et consommations téléphoniques et il pourrait même s'agir à l'avenir des communications elles-mêmes, des données de télésurveillance, ... La société numérique multiplie les traces de chaque parcours individuel.

Mais les données numériques peuvent être encore plus intimes et concerner la personne dans son corps propre. Il faut penser ici à l'auscultation de plus en plus numérique du corps par la médecine. Les développements du génie génétique et de l'imagerie médicale vont multiplier dans des proportions considérables les traces numériques laissées par l'individu.

Aujourd'hui la majeure partie de ces données ne sont pas conservées ou sont dispersées. Ni consolidées ni capitalisées, elles ne sont pas encore le dossier numérique de chacun. Mais techniquement rien n'interdit de le penser. Quelles seront les limites à la numérisation de la personne, dans toutes ses composantes, à la constitution d'un double numérique de soi-même tout au long de la vie ? Depuis l'empreinte génétique jusqu'aux lettres d'amour, des photos de vacances jusqu'aux Web-vidéos, toute une vie pourra s'écrire dans la même matière numérique infiniment et indéfiniment manipulable et restituable.

Il y a en germe tous les éléments d'un mémorial absolu de l'Humanité, depuis le patrimoine génétique commun jusqu'à chacune de ses occurrences concrètes. Pour éviter que le mort ne saisisse le vif, pour éviter d'être enseveli sous un trop plein de mémoire, il faudra bien trouver les nouvelles règles d'un travail de deuil, les nouveaux rites funéraires de l'âge de l'information.

3.2. Éléments pour la numérisation

Musées, bibliothèques, centres d'archives, centres scientifiques ou institutions culturelles, un certain nombre d'institutions se sont déjà engagées, à des degrés variables, sur la voie de la conservation à long terme.

Certaines ont publié des guides et recommandations¹ – nous en exposons ci-après les grandes lignes – ; d'autres se sont associées dans des programmes de recherche et d'expérimentation ; et toutes s'accordent à reconnaître la nécessité de prendre en compte les finalités d'une conservation pérenne dès la naissance du document numérique – on parle alors de pré-

1. Voir par exemple, *Le guide de l'information numérique*, http://europa.eu.int/information_society et *Principes directeurs de la conservation du patrimoine numérique*, UNESCO, 2003.

servation préventive – et de maintenir cette préoccupation à chacune des phases de la vie du document.

Elaboration d'une stratégie

Lorsque l'on crée, met à jour ou fournit une information numérique, une stratégie globale sur le long terme, prenant en compte les besoins d'une conservation pérenne, doit être définie, en raison principalement :

- des capacités d'accès – matériel et logique – au contenu doivent être maintenues, celui-ci pouvant être réutilisé longtemps après la fin de son utilisation courante (par exemple pour des raisons légales ou historiques)
- des dispositifs de conservation doivent garantir la lisibilité des données quels que soient les traitements matériels qui leur sont affectés
- des utilisations ou des mises à jour peuvent être réalisées par d'autres personnes que celles qui ont créé l'information
- des multi-utilisations, en liaison avec d'autres sources, doivent pouvoir être menées

Cette stratégie doit être développée en impliquant tous les acteurs concernés (conservateurs ou archivistes, documentalistes, ingénieurs informatiques, techniciens...). Une des solutions consiste à mettre en place dès le départ un groupe pluridisciplinaire qui définit et maintient cette stratégie. Celle-ci repose sur la prise en compte des éléments suivants :

- la sélection et l'identification des données numériques à conserver
- l'établissement des règles d'organisation et de classement, en particulier la définition de l'unité documentaire et des ensembles qui lui correspondent
- la définition des normes et des spécifications à utiliser pour assurer l'indépendance des données par rapport aux supports et garantir la pérennité
- l'identification et l'attribution de responsabilités pour chacune des étapes
- la prise en compte des besoins des utilisateurs
- l'expertise juridique sur la propriété des fonds et leurs conditions d'utilisation et de diffusion

- la mise en place d'un plan d'évaluation pour contrôler et corriger le programme
- la mise en place d'une cellule de veille technologique et de suivi de l'évolution des systèmes
- la définition d'une politique de formation et d'information interne
- l'échange d'information et le suivi des bonnes pratiques d'institutions voisines

A ce stade, il est fondamental de définir les procédures et les règles de traitement matériel et documentaire de l'objet numérique pour pouvoir, par la suite, en assurer le suivi à chacune des étapes de sa vie.

Chaque document devra être clairement identifié et décrit en considérant l'ensemble des données fondamentales qui le constituent.

Les techniques de numérisation

Pour transformer un document-papier en un format numérique, il existe en général deux solutions principales.

Tout d'abord, le mode-image qui consiste à scanner le document pour en obtenir une image d'une plus ou moins grande résolution ; les scanners actuels permettent des résolutions de 300 à 600 dpi (nombre de points par pouce).

Par ailleurs, le mode alphanumérique, qui consiste à coder dans un format électronique avec l'aide d'un outil de reconnaissance optique de caractères, un document déjà scanné.

Il existe aussi des outils de vectorisation qui seront plutôt utilisés pour le traitement des graphiques.

En mode-image, la taille du fichier à conserver est plus grande, 50 kilo-octets par page contre quelques kilo-octets dans le cas du codage alphanumérique. Une possibilité consiste à combiner les deux solutions : conserver les images en mode-image et traiter le texte dans un système de reconnaissance optique des caractères (OCR).

Dans le cas d'une reconnaissance optique de caractère, un contrôle doit toujours être fait avec l'aide de logiciels de correction, par un opérateur.

Le traitement d'une image fixe se fait également par scannérisation, avec l'adoption de normes de résolution variables selon la qualité et l'usage ; elle peut ensuite être compressée, afin de réduire la taille du fichier pour en faciliter le stockage ou la diffusion par réseaux.

Pour l'image fixe, l'image animée ou le son va se poser le problème de la compression du signal numérique directement issu de la numérisation (270 mbits/s pour la vidéo). La compression est considérée comme non destructive si elle est totalement réversible, c'est-à-dire si l'on peut reconstituer à l'identique le fichier sans compression depuis le fichier compressé.

Quel que soit le media, mais c'est donc encore plus vrai pour l'audiovisuel, il faut avant toute opération de numérisation se poser la question de la destination pour définir le degré de résolution de la numérisation et le taux de compression dans le but de :

- réaliser une copie de sauvegarde d'un support analogique ancien qui va disparaître à court ou moyen terme. L'idéal dans ce cas est de se positionner en haut de gamme de la résolution et sans compression, en tout cas sans compression destructive (c'est le cas en vidéo pour le bêta numérique, et les débits compressés à partir de 50 méga bits/s, à 1,5 Kbits/s pour l'audio, ...). A partir de cette position de principe des compromis devront souvent être effectués pour des raisons financières, de quantité, de qualité de l'original qui pourrait ne pas mériter pas un tel investissement, ... de fait en Mpeg 2 à partir de 8 Mbits/s on approche une qualité de diffusion.

Cette étape peut être l'occasion de procéder à des corrections ou restauration du support d'origine.

- réaliser une copie d'exploitation, c'est à dire destinée à être réutilisée (édition papier, vidéo, audio, télévision, radio...). La qualité retenue dépendra des fourchettes hautes et basses de qualité en usage dans l'univers d'exploitation.
- réaliser une copie de consultation. Le choix peut être large selon que l'on sera dans un environnement local où les débits peuvent être élevés (jusqu'à plusieurs mégabits en vidéo Mpeg 1 ou Mpeg 2) ou sur le Web

où les débits ne dépasseront pas quelques dizaines de kbits/s sur le réseau téléphonique, et quelques centaines de kbits/s sur les réseaux « haut débit » (Mp3, Mpeg 4, Real audio ou vidéo, Windows media, Quicktime).

Lors d'une opération de numérisation on réalisera fréquemment les différents types de copies en parallèle. Chaque copie sera stockée sur un support, choisi en fonction de l'usage auxquels on la destine. Les supports sont très fréquemment différents (disque dur, bande magnétique, disque optique) selon les types de copie (sauvegarde, exploitation, consultation).

À l'issue des chaînes de traitement, il est nécessaire de procéder à des contrôles et vérification de la qualité de l'enregistrement et des données associées.

En cas d'acquisition de matériels et de logiciels nécessaires à la numérisation, il est recommandé d'inclure une clause d'achat imposant aux fournisseurs de mettre à disposition les moyens permettant de récupérer les données engendrées avec leurs applications pour faciliter leur conservation et leur utilisation à long terme. Il est évident que l'achat de produits fondés sur des normes stables et ouvertes offre de meilleures garanties de suivi.

En cas de sous-traitance, il est nécessaire que le cahier des charges garantisse que les personnes qui maintiendront ces données disposeront des matériels, logiciels et de la documentation nécessaires à une utilisation sur le long terme.

Les formats de fichiers

Face au nombre et la variété des normes et standards, il est recommandé de sélectionner, dès le début, l'ensemble des standards et de s'y tenir dans la mesure du possible. Il faut par ailleurs toujours choisir des formats standardisés internationalement, ouverts et multi plates-formes. Ceux-ci sont regroupés en fonction du type d'objet qu'ils traitent :

- L'image en mode point, les formats les plus répandus sont les suivants : le format TIFF (Tag Image File Format), qui correspond aux données traitées par scanner ; le format GIF (Graphics Interchange Format.), que l'on trouve dans les applications Internet ; le format JPEG, un format de compression qui tend à devenir dominant sur Internet.

- Les formats graphiques qui permettent de conserver la structure du graphique ; on retiendra par exemple, le format CGM (Computer Graphics Metafile) qui, normalisé pour les graphiques vectoriels, apporte une bonne garantie de pérennité.
- Les formats vidéo, en particulier les standards MPEG-1, destiné aux CD-ROM, MPEG-2 pour la télévision numérique et le DVD ; Le format MPEG-4 plus adapté aux applications interactives liées au multimédia et au Web, MPEG-7 plutôt comme standard de description orienté recherche ; les formats AVI (Advanced Visual Interfaces), QuickTime, RealVideo sont des formats propriétaires utilisés essentiellement pour la diffusion sur Internet.
- Les formats son, en particulier le format MP3, très répandu sur le Web, le format Wav (de Wave audio), les formats Mpeg pour leur partie son.
- Les formats textes destinés à traiter la structure du texte, les caractères et la présentation. La norme UNICODE permet par exemple, de traiter les alphabets et les idéogrammes et remplace progressivement le vieux système ASCII (American Standard Code for Information Interchange). Le format RTF (Rich Text Format) permet de conserver une version en format riche d'un texte tout en étant en principe convertible dans tous les autres formats. Un grand nombre de formats propriétaires existent autrement dans ce domaine où les nouvelles versions chassent rapidement les générations antérieures. Aucun format n'apporte réellement de garantie de pérennité ; à noter que le format PDF (Portable Document Format) permet la consultation sur diverses plates-formes. Tout en étant un format propriétaire (Adobe) c'est un format ouvert et il tend à se généraliser comme un standard de fait pour la consultation des documents textuels volumineux, ce qui l'a fait adopter par de nombreuses bibliothèques ou sites Web. Il existe également des normes de description et de traitement logique d'un document textuel qui propose un langage de balisage ayant fait l'objet d'une norme internationale tel que le langage SGML (standard Generalized Markup Language) et son extension XML (Extended Markup Language) ainsi que le langage HTML, ce dernier étant très répandu sur Internet.

Les supports de données

Il existe deux familles de supports de conservation utilisés pour les données numériques :

- les supports à bande magnétique, les plus anciens et les plus fragiles ; il s'agit principalement des disquettes, cartouches magnétiques, cassettes DAT (Digital audio Tape) et DLT (digital linear tape). Les supports à bande magnétique ont pour principal atout leur capacité de stockage (300 giga octets en 2003) et le coût le plus faible au kilo octet, et pour principal défaut leur temps d'accès. En effet, ils ne permettent pas une lecture directe de l'information mais doivent être déroulés et le fichier recopié sur disque dur pour pouvoir être utilisé. Ces supports sont donc réservés soit à des fonctions de sauvegarde, soit utilisés en association avec des supports à accès plus rapide dans des ensembles robotisés permettant des temps d'accès différenciés, soit dédiés aux medias audiovisuels très consommateurs de capacité de stockage. Sur les supports de très haute capacité l'offre aujourd'hui est concentrée autour du S-DLT de Quantum, du LTO d'IBM et Hewlett Packard et du S-AIT de Sony les trois standards proposant des solutions proches supérieures à la centaine de giga octets.
- les supports optiques, en général des disques qui permettent un accès direct à l'information mais qui possèdent de plus faibles capacités de stockage. Il a existé de nombreuses solutions appelées généralement DON (Disque Optique Numérique) à base de disques optiques. Ces solutions non standardisées et n'occupant chacune qu'un marché de niche n'ont pas eu de continuité. En revanche les déclinaisons du CD audio lancé par Philips et Sony en 1980 ont eu un succès considérable :
 - Il s'agit des formats CD : CD-Rom (pressés), CD Worm, qui sont des disques inscriptibles une seule fois « write once read many » ; les disques optiques réinscriptibles qui combinent les deux technologies de stockage pour obtenir une rapidité d'accès, la densité et la possibilité de réinscrire les données,
 - des formats DVD dont la version vidéo bénéficie d'une très large et très rapide diffusion. Le DVD Rom, qui offre une capacité huit fois supérieure au CD Rom et les différents formats inscriptibles (DVD-R, DVD Ram, DVD-RW, ...) pour lesquels la bataille de format fait rage aujourd'hui entre les constructeurs.

Au delà de ces supports de masse externes la baisse continue du coût des disques durs commence à en faire un support de conservation crédible.

Le choix d'un support est lié essentiellement à la nature des données, en particulier à son besoin en place de stockage. Son choix dépend aussi des usages et des fonctionnalités de consultation et d'accès envisagés. Dans tous les cas, il est recommandé de choisir un support, ou un nombre restreint de supports afin d'homogénéiser le plus possible les fonds.

Il est indispensable de respecter certaines règles de température et d'humidité pour assurer la conservation physique de ces supports. Les organismes de normalisation (l'organisation internationale de normalisation, l'institut américain de normalisation, le conseil international des archives) recommandent des fréquences de réinscription de l'ordre de 10 ans.

3.3. Les nouvelles lois de la conservation

Assurer la migration continue des supports

Répetons le encore, à l'âge de l'analogique la pérennité d'une oeuvre dépend de la résistance physique au vieillissement des éléments qui la composent : support, encre, magnétisation, ... mais quelle que soit cette résistance tout support est destiné à mourir. Cela prendra beaucoup plus de temps pour une matière très dure, homogène et dont la technique d'inscription est fruste (stèle en granit, pyramide, tablettes d'argile) que pour une cassette vidéo par exemple. Mais s'il n'est pas régénéré périodiquement le support mourra parce que les cycles de régénération analogiques effacent progressivement l'inscription.

A l'âge du numérique il serait très hasardeux de parier sur la durabilité d'un quelconque support, marqué comme la grande majorité des objets de consommation modernes par des cycles de vie courts et par une obsolescence accélérée des technologies. En revanche, le signal numérique binaire est très résistant à la copie. La conservation d'un fichier informatique va donc dépendre d'une organisation systématique de cycles de migration d'un support à un autre, et pour une garantie plus grande des copies de sauvegarde seront faites et stockées dans des lieux différents.

Une information numérique peut donc atteindre une forme d'éternité, mais à condition d'en avoir un souci constant . Ceci change radicalement le processus de la conservation : auparavant les phases essentielles étaient celles de collecte, de stockage puis de communication. La conservation devient maintenant un poste capital puisque toute archive, en survie artificielle, nécessite un environnement contrôlé (hygrométrie, température, lumière) et surtout une administration de sa conservation. Eléments positifs cependant qui facilitent ces migrations cycliques : les coûts de stockage baissent très rapidement en fonction des progrès de la technologie, et les vitesses de transfert d'un support vers un autre sont en constante accélération.

On n'est en ce domaine qu'à l'aube des possibilités technologiques. Déjà des systèmes intégrés et robotisés de conservation d'informations numériques incluent une option de surveillance de la dégradation des supports physiques (analyse des taux d'erreurs de la couche magnétique) et de migration automatique vers un nouveau support quand certains seuils sont dépassés.

L'autre grande action mise en oeuvre par l'informatique pour garantir l'information est la redondance. On l'a déjà abordée en parlant de copies de sauvegarde, mais l'informatique a généralisé ces systèmes avec la mise en place de codes correcteurs d'erreur, de redondance partielle ou totale sur les disques durs et stockée dans des secteurs physiques éloignés.

Dans le monde numérique la pérennité sera donc entretenue par la vitesse : vitesse de rotation de l'information, vitesse d'accès à l'information, vitesse d'analyse et de correction des erreurs, vitesse à recomposer l'intégrité de l'information.

A l'âge de l'analogique, conserver signifiait poser, arrêter le temps, suspendre le cycle de vie des objets; renversement à l'âge du numérique où conserver devra signifier mettre en mouvement.

Il n'est pas sûr que le coût global de conservation de l'information soit plus élevé au final que dans le monde analogique, la concurrence technologique tire fortement à la baisse sur de nombreux postes, mais la structure des coûts a radicalement changé imposant à la plupart des institutions de profondes mutations des métiers. La conservation de l'analogique pouvait être négligente, pouvant se contenter d'un stockage sur étagère, l'informatique ne le permet pas.

Cette nouvelle problématique de gestion des archives a rapidement fait naître une offre de solutions logicielles baptisée par les constructeurs et éditeurs de logiciels le *Digital Asset Management*.

Assurer l'évolution des formats : migration, émulation, encapsulation

Nous avons déjà vu que les difficultés les plus grosses posées par la société numérique reposaient sans doute sur la multiplicité des formats logiques, souvent propriétaires – systèmes d'exploitation, fichiers de données, programmes exécutables – et leur évolution très rapide.

Aussi vives soient-elles, ces difficultés ne sont pas restées sans réponse, et aujourd'hui des solutions techniques permettent d'entrevoir les conditions d'un archivage sur le long terme des formats numériques. Les solutions que certaines institutions patrimoniales commencent à mettre en œuvre sont de trois ordres : la migration, l'émulation, l'encapsulation. Elles peuvent être effectuées séparément ou donner lieu à des pratiques combinées.

La solution la plus communément utilisée est la migration, une opération au cours de laquelle les bits d'un fichier ou d'un programme sont modifiés afin de pouvoir être lus par un nouveau système d'exploitation ou une nouvelle version d'une application. Il s'agit d'un retraitement des données anciennes visant à les rendre compatibles avec un nouvel environnement.

Cette solution peut être mise en œuvre dans le cas d'un transfert de support physique ayant un format logique de données différent, ou au niveau du système d'exploitation ou des formats de fichiers, ou bien encore au niveau du format des données ou de l'application. Toute migration se fait donc en modifiant une partie du codage des contenus initiaux pour qu'ils puissent être exploités par des systèmes récents.

Compte tenu de l'instabilité technologique déjà soulignée, cela signifie que l'on planifie une migration systématique et régulière de tous les documents à chaque changement de version. A titre d'exemple, la BNF a entrepris la migration de 15 millions de pages d'imprimés numérisés en mode image pour l'application de consultation sur Internet Gallica. Cette migration d'environ 1 To de données sans changement du format des données mais avec modification du format lié au changement de type de support

prend environ 10 jours en tenant compte de tous les problèmes techniques qui se posent en cours de traitement¹.

Une autre solution, l'émulation, consiste à créer un programme simulant l'environnement informatique nécessaire à l'exploitation des données sur une plate-forme courante alors que leur environnement technique est obsolète. Il s'agit de simuler, à l'aide d'un logiciel appelé « émulateur » le comportement d'un système qui a disparu. La migration transforme le stockage des données elles-mêmes, l'émulation s'interpose lors de leur restitution. Cette technique a l'avantage de garder intact, sans aucune modification, les bits originaux d'un fichier ou d'un programme ; mais elle ne permet pas toujours de retrouver les conditions optimum de leur exploitation, par exemple des mises en pages seront bouleversées ou la synchronisation images/son sera faussée. Si, sur le plan théorique elle apparaît comme une solution moins destructive, elle fait encore l'objet de travaux de recherche importants, notamment pour pouvoir créer les émulateurs du futur. Elle pose par ailleurs le problème de la conservation en cascade de diverses couches d'émulateurs.

Une autre solution prometteuse, sur laquelle travaillent des consortium de bibliothèques et centres d'archives à travers le monde, est l'encapsulation, désignée aussi sous l'appellation « voie de référence ». Cette technique est issue du modèle OAIS, modèle mis au point pour la conservation des données spatiales par le Comité Consultatif pour le Système de Données Spatiales et qui vient de faire l'objet d'une norme ISO².

Le modèle OAIS (Open Archival Information System) élaboré dans le cadre d'un groupe de normalisation dédié aux échanges de données satellitaire, s'est peu à peu imposé comme un modèle à caractère général offrant un cadre conceptuel permettant d'identifier les composants fonctionnels élémentaires des archives numériques et établissant les concepts et la terminologie pour décrire les architectures et les modèles de données.

L'encapsulation se présente donc comme un moyen de regrouper autour du contenu lui-même, les informations contenant les instructions nécessaires

1. C. Lupovici, *Les principes techniques et organisationnels de la préservation des documents*

2. C. Huc, *La pérennité des documents électroniques*, <http://www.archivesdefrance.culture.gouv.fr>, 2001

au décodage de ses bits dans le futur par n'importe quel système. Elle propose une succession de couches d'instruction, couche externe sur laquelle sera porté un texte lisible par tous décrivant le contenu de l'élément encapsulé ainsi que la manière de l'utiliser, une couche interne sur laquelle figurent les caractéristiques du logiciel, du système d'exploitation et du matériel à reconstituer en vue de la lecture de l'objet lui-même. L'encapsulation, qui permet de rendre autonome le contenu et l'information qui lui est attachée, semble être une méthode relativement viable pour la conservation à long terme en particulier pour des fichiers texte. Elle est encore incertaine pour les autres documents qui souffrent de la dispersion technologique et du trop grand nombre de nouveaux logiciels, de systèmes de compression ou de formats mis sur le marché chaque année.

Au final la solution résidera sans doute dans la définition d'un format de conservation édicté par l'institution d'archivage, et qu'elle devra suivre dans la durée. L'homogénéité de format permettra ensuite l'automatisation de la conversion à la génération suivante.

3.4. Documentation

Nous avons déjà évoqué la révolution documentaire amenée par le numérique. L'explosion du nombre des unités documentaires générée par la société de l'information a pour corollaire une capacité, sans précédent, à y accéder par leur contenu.

- D'un point de vue documentaire l'analogique était une matière morte. Elle ne nous apprenait rien sur elle même, ne renvoyait aucune information. Toute l'information devait être générée par différentes professions qui produisaient une information catalographique, bibliographique, archivistique, documentaire permettant de structurer des fonds passifs et de créer des clés d'accès pour les lecteurs. Etant donné les quantités à traiter, quantités par ailleurs déjà croissantes à l'âge de l'analogique, il s'agissait souvent d'un travail de Sisyphe, toujours inachevé, laissant de grandes zones d'ombre dans les catalogues les mieux gérés, et donnant au final une information somme toute modeste au lecteur.

Avec le numérique non seulement le document peut être instantanément présent sans qu'un magasinier doive le chercher sur une étagère, mais son

contenu est par ailleurs directement compréhensible par un ordinateur et le document primaire lui-même, sous l'angle de la recherche documentaire, n'est autre que l'addition de clés d'accès. Cette évolution considérable a fait la fortune des moteurs de recherche.

Cette aptitude du numérique est pour le moment réservée à l'information textuelle. Mais déjà il existe des programmes de recherche très avancés permettant de transcrire en information textuelle la parole, y compris dans des environnements très bruités comme en radio. Il existe également de nombreux programmes de recherche sur l'analyse de l'image, fixe ou animée, dont on peut espérer qu'un jour ils approcheront la sémantique humaine et nos capacités de reconnaissance, de désignation, de mise en résonance et de connotation dans de grandes bases de connaissances.

- Le problème n'est donc plus le silence des documents mais le bruit, le trop plein d'information. Et chacun peut constater que la plupart des requêtes adressées aux moteurs de recherche devient vite inexploitable. Ce phénomène a conduit les fondateurs du Web à préconiser un Web sémantique¹, dont les voies passeraient par une plus grande structuration des données, données propres au document ou données l'accompagnant (on parle alors de métadonnées). Cette évolution nécessaire passera par une plus grande implication des producteurs de données dans la qualité des données produites.

Sans entrer dans le détail on doit mentionner l'intense activité de normalisation mondiale qui règne dans ce domaine de structuration des documents. Le document textuel a été très tôt structuré selon le format SGML ; le format HTML, format standard sur le Web n'a pas repris une structuration aussi forte, vue comme trop contraignante. Mais l'évolution du Web ramène à ces problématiques avec la diffusion actuelle de la norme XML. Dans le domaine de l'image animée les normes Mpeg 7 et Mpeg 21 sont très orientées sur la normalisation des différents types de métadonnées (informations documentaires, juridiques, techniques, d'usages, ...), aptes à suivre l'œuvre et à s'enrichir tout au long de sa carrière.

1. T. Berners-Lee, *The Semantic Web*, présenté à La 9e conférence internationale du World Wide Web, Amsterdam 2000.

3.5. Des projets pilotes de conservation pérenne

Dès le début des années 90, alors que le numérique achevait de pénétrer toutes les formes d'expression et de création humaine, des programmes de recherche et d'expérimentation généralement portés par des regroupements d'institutions de recherche scientifiques, d'archives ou de bibliothèques, ont commencé à explorer les manières de prendre en compte cette nouvelle donne dans les politiques d'archivage et de conservation.

Pionnier en ce domaine, Internet Archives, fondation américaine qui a vu le jour en 1996 en tant qu'entreprise privée à but non lucratif, rassemble toutes les pages Web librement disponibles dans le monde entier, un fonds qui comprend plus de 10 milliards de pages, cinq fois la quantité de documents détenus par la Library of Congress. La fondation a mis en place « une machine à remonter le temps » en octobre 2001 pour donner libre accès aux données sur le Web¹.

Etat des lieux de pratiques, état des lieux des besoins, état de l'art des outils, évaluation de solutions techniques et procédurales, élaboration de modèle, le terrain à explorer était, et reste, vaste ; et si d'ores et déjà, des enseignements se dégagent, il reste encore beaucoup d'interrogations et d'incertitudes auxquelles sont confrontées les pratiques locales et les programmes de recherche et d'expérimentation.

Les initiatives sont nombreuses et mériteraient chacune qu'on s'y arrête pour en tirer tout l'enseignement et valoriser au mieux les acquis. Nous retiendrons simplement quelques programmes désignés comme pilote en raison de leur positionnement institutionnel et scientifique, en particulier parce qu'ils mobilisent des expertises internationales et que la diffusion la plus large de leur travaux permet une mise en commun de leurs résultats.

• **Pandora (Preserving and Accessing Networked Documentary Resources of Australia)**² constitue un des premiers programmes d'envergure soutenus par les pouvoirs publics australiens ; il a pour objectif la mise en œuvre effective d'un système d'archivage des publications électroniques et des sites australiens du Web. A l'issue d'une première étape de travaux

1. <http://www.archive.org>

2. <http://pandora.nla.gov.au/index.html>

menés dès 1996 par la Bibliothèque nationale d'Australie, un ensemble de recommandations ont donné lieu à publication ; elles définissent notamment des principes et méthodes de sélection ainsi que les modes d'organisation et de gestion de l'archive. La publication y est définie en termes généraux : toute chose parue sur Internet est considérée comme une publication, seule la documentation interne étant expressément exclue. Les sites sélectionnés aux fins de conservation doivent concerner l'Australie ou traiter d'un sujet hautement pertinent pour le pays et conçu par un Australien. La sélection se fait sur la base du contenu et « un rang de priorité est accordé à des publications qui font autorité et ont une valeur à long terme pour la recherche ».

Une seconde phase de travaux « PADI » (Preserving Acces to Digital Information) a élargi le champ des partenaires en associant , outre un large panel d'experts internationaux, des bibliothèques des Etats d'Australie ainsi que les Archives nationales du film et du son, le musée national d'Australie, des associations professionnelles des TIC, des centres de recherches et des universités australiennes, etc... Son objectif est de concevoir et mettre en œuvre un modèle national d'archivage partagé des données électroniques et du du Web australien.

Pour compléter le dispositif, le Public Record Office et les Archives nationales australiennes, ont élargi la gestion des archives électroniques aux sites Web des instances gouvernementales (sites publics ainsi que sites Intranet) et élaboré des principes directeurs énonçant les meilleures pratiques. Les Archives nationales appellent l'attention sur le fait que les documents contenus dans des sites Web ne sont pas toujours considérés comme des archives. Ces sites doivent donc faire l'objet d'une gestion rigoureuse. Dans le monde de l'Internet aussi, l'identification et la gestion du matériel documentaire impliquent des responsabilités et sont assujetties à certaines procédures.

• **Le programme PRESERV¹** : ce programme est animé par le Research Libraries Group (RLG), la Digital Library Federation (DLF) et On-line Computer Library Center (OCLC). Il fait suite à différentes études, enquêtes et réalisations expérimentales menées dans l'univers des bibliothèques confrontées d'une part, à la montée en charge des publications électroniques, et d'autre part, à la gestion des fonds numériques dont elles ont la res-

1. <http://www.oclc.org/research/preserv>

ponsabilité. Il explore les conditions d'archivages tant des publications électroniques, des archives numériques que des documents issus de politiques de numérisation, d'un point de vue méthodologique et pratique ; il s'attache en particulier à définir les attributs de l'archive numérique au regard des finalités de recherche, à clarifier la notion de collection de données hétérogènes, la question de la certification, à structurer les métadonnées et leur utilisation dans le cadre d'une politique de conservation à long terme.

• **NEDLIB (Networked European Deposit Library)**¹ est un programme qui a été soutenu par la Commission Européenne de 1998 à 2001. Il regroupe huit grandes bibliothèques européennes : la Bibliothèque Royale des Pays-Bas, la Bibliothèque nationale de Norvège, la Bibliothèque nationale d'Allemagne, la Bibliothèque nationale du Portugal, la Bibliothèque nationale de Suisse, la Bibliothèque de Florence, la Bibliothèque universitaire d'Helsinki, la Bibliothèque nationale de France. Sont également impliquées les sociétés d'édition Elsevier Science, Kluwer Academic et Springer Verlag. Le programme vise la constitution d'un modèle de spécifications fonctionnelles et techniques relatif aux documents électroniques publiés sur support ou diffusés sur le Web. Il a considéré l'ensemble des fonctions, depuis la collecte via la mise au point d'un prototype, la documentation et le catalogage, jusqu'à la conservation incluant la mise à plat des techniques de migration et d'émulation.

• **Le programme CEDARS**² est développé par un consortium de bibliothèques de recherche universitaires britanniques (Oxford, Leeds et Cambridge) et américaines (Université du Michigan). Démarré en avril 1998, il se propose d'explorer les questions relatives à l'acquisition des publications électroniques, à leur conservation, à leur description et enfin à leur accès ; son objectif est de mettre à la disposition de la communauté des bibliothèques un guide décrivant, tant d'un point de vue méthodologique qu'opérationnel, les bonnes pratiques permettant d'intégrer dans les collections pré-existantes des corpus de textes numériques. En octobre 1999, le programme a mis en place un nouveau volet visant à approfondir les problématiques des techniques de l'émulation. Ce nouveau programme appelé CAMILEON (Creative Archiving at Michigan and Leeds Emulating the Old on the New) se déroule

1. <http://www.kb.nl.coop/nedlib>

2. <http://www.leeds.ac.uk/cedars>

jusqu'en septembre 2003. Il explore les conditions de recours à l'émulation comme stratégie de conservation capable de remédier à l'obsolescence des systèmes de production de données numériques et de maintenir les possibilités d'accès à ces contenus dans leur configuration de création initiale. Il a pour objectif le développement et l'expérimentation d'outils d'émulation ainsi que l'élaboration d'un ensemble de recommandations concernant tant la migration que l'émulation.

• **Le programme INTERPARES (International Research on Permanent Authentic Records in Electronic Systems)**¹ constitue actuellement un des plus importants programmes de recherche et développement consacré à la conservation à long terme des données numériques. Regroupant des équipes de chercheurs, des institutions d'archives, des producteurs de contenus et des industriels issus de plus d'une vingtaine de pays, il est piloté par la British Columbia University du Canada. Il se déroule en deux phases : Interparès 1 démarré en 1999 et achevé en 2001 par la publication des modèles théoriques et méthodologiques du domaine, suivi par Interparès 2 dont l'échéance est prévue fin 2006. Son objet porte sur l'identification et l'analyse des paramètres physiques et intellectuels affectant les conditions de maintien des données numériques, et sur les nouvelles méthodes et procédures d'archivage et de conservation qui en découlent. Il s'agit au final de proposer des standards, des principes d'organisation et des recommandations pour la mise en place par les gouvernements, par les institutions culturelles et patrimoniales et par les acteurs industriels, des actions concourant à la conservation des données numériques.

A côté de ces programmes de coopération internationale, de nombreuses institutions scientifiques, culturelles et patrimoniales nationales se sont investies dans l'exploration et la mise en œuvre de dispositif d'archivage des données numériques dont elles ont la responsabilité. Ainsi, par exemple la Bibliothèque nationale autrichienne s'est associée à l'Université technique de Vienne pour concevoir un système de collecte et de conservation des données numériques ; la Bibliothèque nationale de Chine pilote un programme ayant pour objectif la mise en place d'une bibliothèque numérique ; aux Pays-Bas, les universités d'Amsterdam, de Tilburg et de Twente mènent, en collaboration avec la société IBM un programme de recherche et déve-

1. <http://www.interpares.org>

loppement appuyé largement sur le modèle OAIS ; en Afrique du Sud, ce sont les principales institutions culturelles et scientifiques du pays qui explorent les conditions de faisabilité de constitution d'une archive numérique de la nation sud-africaine. A Taiwan, neuf institutions nationales se retrouvent dans un projet de mise en place des archives nationales numériques ; aux Etats-Unis, la Bibliothèque du Congrès est chargée de piloter et d'animer le National Digital Program doté de 100 millions de \$. Au Canada, le Réseau d'information sur le patrimoine, soutenu par le gouvernement fédéral, met à la disposition des institutions culturelles et éducatives des guides, conseils et ressources pour la numérisation et la conservation des fonds. En Europe, des projets et des réalisations concrètes sont en cours de développement notamment en Norvège, en Finlande, en Suède, en Grande-Bretagne, en France. Ces initiatives sont largement appuyées par la Communauté européenne dont le Conseil a publié en mai 2002 une résolution relative à la « préservation des contenus numériques »¹.

On le voit, peu à peu les sociétés et les organisations prennent conscience des risques d'amnésie qui les menacent, sans toutefois disposer encore du remède universel qui permettrait d'y faire face. C'est pourquoi, plus que jamais il est nécessaire de poursuivre les efforts de recherche et développement, plus que jamais il est nécessaire d'organiser des coopérations, de partager les connaissances, de mettre en commun les acquis : les défis de ce nouveau patrimoine nous l'imposent : il se déploie à l'échelle mondiale, sa conservation aussi.

Si des progrès incontestables ont été réalisés sur la voie de la conservation sur le long terme des documents numériques, il reste encore de grandes zones d'ombre relatives aux conditions permettant de garantir la pérennité des contenus et de leurs accès.

Ainsi, la conservation du patrimoine numérique demeure actuellement terra incognita pour la plupart des institutions. Lorsque des responsabilités leur seront officiellement confiées dans ce domaine, il faudra qu'elles adaptent leurs organigrammes et redéfinissent les tâches de leurs agents.

1. <http://europa.eu.int>

En effet, l'immatérialité de la production numérique oblige à concevoir de nouvelles formes de conservation patrimoniale, à redéfinir de nouvelles procédures et dispositifs de collecte, de stockage, d'indexation et de consultation adaptés à ces cyber-contenus générés dans la plus grande instabilité technologique.

Il importe donc que, dans le même temps où le média se constitue, s'élaborent les outils de sa mémoire.

Chapitre 4 L'action des décideurs

La conservation du patrimoine numérique dépend d'un certain nombre de décideurs dont les premiers sont les gouvernements des états. En effet, les forces du marché ne sauraient à elles seules garantir la conservation et la promotion du patrimoine numérique mondial. De ce point de vue, la prééminence des pouvoirs publics, en partenariat avec le secteur privé et la société civile est essentielle.

La volonté politique des gouvernements de préserver le patrimoine numérique doit les conduire logiquement à mettre en œuvre dans chaque pays un cadre législatif approprié définissant les rôles et les responsabilités des différentes institutions chargées du patrimoine au plan national. Elle doit les conduire aussi à dégager les moyens financiers indispensables à cette conservation et à la consultation de ce patrimoine.

Une législation adaptée

Les gouvernements et les décideurs doivent comprendre que la conservation du patrimoine numérique est un problème urgent qui ne peut être résolu du jour au lendemain. Le risque de perdre des matériels essentiels dans lesquels des ressources précieuses ont été investies est extrêmement réel (selon la Bibliothèque du Congrès, la durée de vie moyenne d'une page sur Internet est de deux mois, et presque la moitié de la totalité du « contenu » affiché sur Internet disparaît en un an).

Traditionnellement, la conservation du patrimoine culturel obéit à des cadres et des procédures juridiques qui reposent - sur des critères bien définis. Les bibliothèques nationales rassemblent et conservent les publications par le biais, en général, du dépôt légal de la production nationale. D'autre part, une vaste législation sur les archives définit les missions des institutions d'archives et les objets qu'elles doivent collecter et préserver. Complétant la tâche des Archives nationales, les archives spécialisées et les musées ont la responsabilité de collecter et de conserver à des fins patrimoniales les matériels sonores, photographiques, cinématographiques ou audiovisuels.

La législation peut varier considérablement selon les pays qui attribuent ces responsabilités à telle ou telle institution. Elle varie aussi grandement en

ce qui concerne les catégories de documents conservés. Il existe, cependant, un large accord sur deux principes de base : le patrimoine culturel doit être préservé et les citoyens doivent pouvoir le consulter. C'est pourquoi les institutions d'archivage ont besoin d'une législation appropriée en matière de patrimoine numérique leur permettant de définir les tâches, de sélectionner et collecter les documents à conserver.

Les juristes remarquent que l'espace numérique, en particulier celui des réseaux, n'est pas naturellement celui du droit. « Celui-ci, d'application territoriale, s'appuie sur des comportements, des catégories homogènes et stables, tous les éléments qui font défaut dans le cas de l'Internet. Cet antagonisme avec le droit aurait même, selon certains, favorisé l'essor initial du réseau, libre de toutes contraintes hormis celles fixées par la communauté des chercheurs qui sont à l'origine de sa création. » souligne, en juillet 1998, une étude du Conseil d'Etat (France), intitulée *Internet et les réseaux numériques*.

Les législations existantes doivent donc être adaptées ou étendues au nouvel et complexe environnement numérique. De plus, proposer une nouvelle législation est une tâche longue et difficile et il n'existe pas de recette miracle.

La conservation patrimoniale est parfois régie par une loi spécifique de dépôt légal, d'autres fois par une loi sur le droit d'auteur ou le copyright et d'autres fois encore par la loi qui régit l'institution dépositaire d'archive nationale.

La France, la Finlande et la Suède relèvent de ce premier cas de figure. En France, une nouvelle loi sur le dépôt légal a été adoptée en 1992 : elle s'applique aux documents imprimés et non imprimés notamment les programmes de la radio-télévision mais aussi les productions multimédias et les bases de données éditées sur support ; un projet d'extension de cette loi aux sites Web français est actuellement en préparation. En Finlande, la loi sur le dépôt légal de 1980 couvre l'imprimé, le son et l'image et une loi distincte, de 1984, régit le dépôt légal des images animées, des films et des bandes vidéo. En Suède, la loi de 1978 sur le dépôt légal porte aussi création des Archives nationales du son et de l'image animée. Un texte plus récent, de 1993, étend l'obligation de dépôt légal aux documents électroniques « portatifs » et autres types de documents non imprimés.

En Australie, au Royaume-Uni et aux Etats-Unis, les dispositions régissant le dépôt légal font partie de la législation sur le droit d'auteur et le copyright. En Australie, les textes relatifs au dépôt légal des « documents de bibliothèque » font partie de la loi de 1968 sur le droit d'auteur, qui confère à la Bibliothèque nationale le statut de dépositaire national desdits documents. Mais la loi entend par « document de bibliothèque » les seules publications imprimées sur papier. La Bibliothèque nationale et les Archives nationales du film ont présenté une proposition commune tendant à élargir la portée de la loi, tout en recommandant que les dispositions relatives au dépôt légal soient maintenues dans la loi sur le droit d'auteur et du copyright, et non incorporées à la loi relative à la Bibliothèque nationale ou érigées en texte législatif distinct. En Australie, le dépôt légal est toujours allé de pair avec le droit d'auteur et le copyright, aussi a-t-on jugé plus simple de maintenir ce lien et de légiférer sur le premier à l'occasion d'un réexamen de la législation relative au second. Les textes qui régissent le dépôt légal au Royaume-Uni font partie de la loi sur le droit d'auteur de 1911. La British Library a recommandé que dans les cas de publications non imprimées, la nouvelle législation soit tout à fait distincte du droit d'auteur et du copyright, afin de bien marquer les différences entre ces deux domaines. Aux Etats-Unis, les obligations en matière de dépôt légal figurent dans la loi sur le droit d'auteur et du copyright de 1976 qui confère au Bureau du droit d'auteur et du copyright des Etats-Unis, relevant de la Bibliothèque du Congrès, le pouvoir d'édicter des règles exigeant le dépôt de la « meilleure édition » des œuvres sur tout support. Au Canada et en Allemagne, ce sont les lois sur la Bibliothèque nationale qui rendent le dépôt légal obligatoire. En Espagne, la loi de 1971 sur l'Instituto Bibliografico Hispanico couvre le dépôt légal.

D'une manière générale, les bibliothèques nationales abordent le problème de l'environnement numérique sous l'angle de la législation sur le dépôt légal. Le dépôt de produits numériques hors-ligne, par exemple les CD-Rom, est déjà une obligation légale dans plusieurs pays. Les bulletins électroniques en ligne sont considérés comme le prolongement d'une longue tradition de publications sur le support imprimé, que les bibliothèques ont toujours collectées et préservées. La législation sur le dépôt légal est donc un élément clé de la politique nationale de conservation. Comme dans le cas des médias traditionnels, le citoyen doit pouvoir accéder aux matériels

numériques déposés sans porter préjudice à l'exploitation légitime du patrimoine dans l'environnement numérique. Il faut donc proposer une législation qui soit la plus large possible et veiller à ce qu'elle s'applique aussi bien aux publications qui ne sont produites que sous forme électronique qu'à celles qui sont produites « en parallèle ». S'il y a le moindre doute sur la pertinence d'un objet, mieux vaut également ne pas établir de distinction entre les formes connectées (en ligne) et autonomes de publications électroniques, et les inclure toutes les deux lorsqu'il y a risque d'évolution rapide vers les publications électroniques en ligne. Il appartiendra ensuite à l'institution depositaire de déterminer quels éléments doivent être retenus pour la collection nationale.

Cette législation doit être élaborée non seulement pour les matériels et les publications mais également, par exemple, pour les données de recherche, en faisant, peut-être, du dépôt une condition des subventions à la recherche.

L'application de la législation

Ces principes posés par la loi, cette dernière doit être appliquée. Les pays qui ont commencé à le faire ont adopté pour cela une démarche progressive, mais il est difficile de donner des consignes précises étant donné l'importance que revêtent à cet égard les pratiques déjà suivies pour le dépôt légal des publications imprimées, le cadre proposé pour la nouvelle législation applicable au patrimoine électronique, les types de produits numériques à inscrire dans le champ d'application de la législation, et le nombre de documents électroniques pouvant être assujettis au dépôt légal chaque année.

La coopération entre les acteurs du numérique

L'instauration d'un véritable partenariat à l'intérieur d'un même pays est indispensable à la conservation du patrimoine numérique : les créateurs de documents numériques et les professionnels des technologies de l'information et de la communication, doivent être associés au processus de préservation parce que leur coopération peut éliminer une partie du fardeau qui pèse sur les institutions chargées du patrimoine. Il faudra encourager les

créateurs à utiliser des standards ouverts et à documenter convenablement les fichiers. Il faudrait également convaincre les professionnels des TIC de l'intérêt que présentent les logiciels à code source non protégé, et de la nécessité de publier une documentation détaillée et complète sur leurs produits pour faire en sorte que ceux-ci puissent être utilisés par la suite dans un contexte de conservation.

En effet, les exigences des systèmes de conservation du patrimoine numérique amènent, en amont, les créateurs à prendre conscience du fait que les choix qu'ils font au moment de la création de documents déterminent les possibilités d'archivage ultérieur et peuvent aider à leur maintenance future : l'emploi de normes et de formats ouverts, la description et la documentation convenables et l'utilisation de noms permanents pour les ressources en ligne facilitent la conservation à long terme et contribuent à en réduire les coûts.

Pour cela, un effort de communication de la part des décideurs politiques en direction de leur opinion publique nationale est indispensable. Il est important de lui présenter une argumentation claire des enjeux de cette mission. Il est tout à fait possible de susciter un réel enthousiasme en mettant en avant la nécessaire prise en compte du patrimoine numérique national et la possibilité à l'avenir d'en conserver l'essentiel. Ce travail d'explication et de mobilisation doit être mené de pair avec les études de faisabilité afin que les institutions soient pleinement motivées sur cette activité prioritaire et les partenaires privés prêts à jouer leur rôle.

C'est dans le cadre de cette coopération entre les acteurs du numérique que les problèmes de droit d'auteur et de copyright devraient trouver une issue positive grâce à des accords entre institutions patrimoniales et éditeurs visant à concilier les intérêts des deux parties, en autorisant la copie, aux seules fins de conservation tout en limitant l'accès.

Le droit d'auteur et le copyright

Le droit d'auteur et le copyright constituent, en effet, un enjeu important. La législation actuelle impose des limitations tellement strictes en matière de copie que les bibliothèques ne peuvent même pas, dans certains cas extrêmes, préserver les revues électroniques auxquelles elles s'abonnent

sans empiéter sur les droits des propriétaires et des créateurs. Les éditeurs reconnaissent certes que le droit d'auteur et le copyright représentent peut-être un obstacle à la conservation à long terme, mais ils sont aussi très méfiants à l'égard de tout arrangement qui perturberait leurs intérêts commerciaux en facilitant l'accès sur les réseaux au matériel en dépôt.

L'Organisation Mondiale de la Propriété Intellectuelle (OMPI) veille à l'adaptation des conventions internationales existantes pour tenir compte des problèmes soulevés par le développement des réseaux numériques. En particulier, c'est sous son égide qu'ont été conclus deux traités en décembre 1996 sur le droit d'auteur d'une part, et sur les « interprétations et exécutions et les phonogrammes » d'autre part. Un nouveau traité est par ailleurs en cours de négociation actuellement sur les droits des artistes interprètes. Ces nouvelles conventions visent notamment à élargir le champ de protection des œuvres afin d'y inclure les nouveaux supports et modes de transmission, notamment numériques. Toutefois, elles n'ont pas, à ce jour, résolu les difficiles questions liées au champ des exceptions au droit d'auteur sur les réseaux ou à la loi applicable.

Par ailleurs, un accord très important a été conclu, dans le cadre de l'Organisation Mondiale du Commerce (OMC) sur les aspects des droits de propriété intellectuelle qui touchent au commerce.

Face au désir des auteurs de protéger toujours davantage leurs œuvres contre la contrefaçon, notamment par des moyens techniques, de nombreux intellectuels redoutent que la logique économique ne restreigne la circulation des idées et l'accès de tous à la culture. Cette préoccupation est partagée par certaines organisations internationales comme l'UNESCO. Elle est également reprise par des gouvernements qui ont une tradition très forte d'exception au droit exclusif de l'auteur permettant la libre circulation des œuvres à des fins d'éducation et de recherche, en particulier dans les pays anglo-saxons (exception dite de « fair use ») et dans les pays du nord de l'Europe, comme l'observe l'étude du Conseil d'Etat.

Cependant, les documents exprimant officiellement et publiquement les intentions des différents gouvernements quant à l'adaptation de leur régime national de la propriété littéraire et artistique aux spécificités des réseaux numériques sont encore rares.

La gestion des droits est en train de devenir une opération extrêmement complexe : la loi et les accords entre institutions de conservation et éditeurs ne sauraient en couvrir tous les aspects. Par exemple, lorsqu'un produit numérique fait appel à un logiciel qui est la propriété exclusive de tiers, le créateur du contenu n'est généralement pas détenteur de droits. Il faut noter que les fournisseurs de logiciels ne se sont jusqu'ici guère associés aux efforts de conservation, et les logiciels ne sont généralement pas couverts par la législation sur le dépôt légal.

Le créateur et le conservateur ont chacun une responsabilité en matière de préservation. Dans la mesure où les créateurs ne sont pas toujours conscients de tous les risques, les institutions chargées du patrimoine doivent solliciter leur coopération pour permettre, dès la création du document, les conditions de sa conservation pérenne.

L'organisation

L'édification d'une solide infrastructure capable d'assurer le fonctionnement d'un système réparti d'archives numériques suppose l'existence d'organisations fiables capables de maintenir des matériels en vie sur le long terme. Les bibliothèques et archives nationales, ainsi qu'un certain nombre d'instituts de recherche spécialisés et d'archives de données assument actuellement ce rôle. Mais il y a aussi toute une série d'autres institutions qui pourraient avoir à intervenir dans la conservation de certains types de documents (photographies, sons, œuvres d'art, documents télédiffusés, etc.) sous forme numérique ou dans la conservation de documents à l'intention d'une communauté bien déterminée (institutions ayant une compétence locale ou régionale, instituts de recherche dans une discipline déterminée, etc.).

Si les archives numériques doivent être prises en charge par des organisations fiables, c'est d'abord dans l'intérêt et pour le bénéfice culturel et démocratique de la collectivité nationale et au delà, de l'humanité. En prenant l'initiative de tester des modèles de conservation, les institutions nationales peuvent aussi aider d'autres organisations à comprendre les exigences d'un système de conservation opérationnel et à mettre en place des systèmes dans leurs propres domaines.

Retrouver l'unité perdue de l'archive du Web

La Toile est aujourd'hui mondiale, elle est une. Or, la plupart des projets de conservation découpe dans cet espace global des parcours thématiques, ou reconstituent des territoires à la mesure des pays et de leurs institutions nationales de mémoire. Cette perspective semble normale et légitime mais elle fait peser le risque d'une perte de substance en perdant la totalité : à savoir tous les liens qui pointent par delà les frontières et seront rompus.

On pourrait rêver d'une archive centrale d'Internet à l'échelon mondial. Une telle tentation sous tend le projet d'Internet archive¹. Mais il paraît peu concevable que les nations abandonnent toute souveraineté sur leur patrimoine. A la réflexion ce modèle est par ailleurs très éloigné d'Internet dont le concept de base est l'organisation fédérée des réseaux sans concentration pyramidale et centralisée.

L'avenir de l'archive se situe plus dans l'organisation concertée de normes internationales d'archivage permettant de construire l'hyper-archive par fédération progressive d'un patrimoine en réseau et l'interconnexion mondiale des institutions de mémoire.

A l'évidence cette hyper-archive, Web historique apportant une troisième dimension au Web, ne devra pas se confondre avec le Web vivant pour préserver les droits des éditeurs de sites, mais le suivre comme sa trace.

Cet objectif ne sera possible qu'à travers une coopération internationale et une forte motivation de chacun des acteurs.

Le financement

Pour pouvoir assurer une conservation adéquate du patrimoine numérique et l'accès général à celui-ci, des efforts soutenus sont donc nécessaires de la part des gouvernements, des auteurs, des éditeurs, des industries concernées et des institutions chargées de la conservation du patrimoine documentaire. Aussi, les institutions chargées de la conservation à long terme du matériel numérique doivent-elles avoir une pérennité financière assurée.

1. <http://www.archive.org>

Les institutions patrimoniales auxquelles est conféré le droit de recueillir les documents numériques doivent disposer des fonds nécessaires pour acquérir les installations, le matériel et les locaux nécessaires au stockage de ces données et recruter du personnel suffisamment qualifié pour les prendre en charge et les rendre accessibles. Or, dans la majorité des cas, sinon dans tous, l'institution elle-même n'est pas financièrement en mesure d'entretenir les collections de documents numériques recueillis et d'assurer les services de consultation sans une dotation financière supplémentaire affectée à cette fin.

À noter, en effet, que la complexité du stockage des documents numériques n'annule pas les coûts de stockage, mais les transforme et que les institutions patrimoniales devront gérer de front deux types de documents : les supports traditionnels et l'immatériel.

La technologie nécessaire à la conservation des documents numériques exige des investissements en recherche-développement qui sont certes importants dans l'absolu mais demeurent négligeables si on les compare aux ressources investies dans la création des matériels eux-mêmes ou au coût que la perte de ces matériels représenterait pour la société si des systèmes convenables de conservation ne sont pas mis au point.

La recherche et la formation

Un financement public doit être prévu aussi pour soutenir résolument les recherches sur les technologies et modèles prometteurs afin de mettre au point le plus rapidement possible des systèmes pleinement opérationnels de conservation du patrimoine numérique.

Enfin, un financement doit être trouvé pour la création de vastes programmes de formation. Ces derniers sont nécessaires pour diffuser largement les compétences et expériences réunies à ce jour parmi les cadres et le personnel des institutions chargées du patrimoine. La conservation du patrimoine numérique exige, en effet, de nouvelles formes d'organisation, de nouvelles méthodes de travail et de nouveaux modes de réflexion. Il faudra que ces programmes traitent non seulement les aspects techniques mais offrent également au personnel une formation qui lui permette de s'adapter à un environnement en évolution et aux réorientations qu'il impose.

Comme l'ont été et le seront de plus en plus la conservation et l'exploitation des patrimoines culturels traditionnels, on peut envisager que la conservation du patrimoine numérique sera créatrice de développement.

La coopération internationale

Considérant le fossé numérique existant, il est nécessaire de renforcer la coopération et la solidarité internationales pour permettre à tous les pays, en particulier aux pays en développement et aux pays en transition, d'assurer la conservation de leur patrimoine numérique et l'accès continu à celui-ci, par la mise en commun des expériences, la diffusion des résultats et des meilleures pratiques et la conclusion d'accords de jumelage.

Ainsi, la coordination des institutions doit être menée en parallèle avec une collaboration active au sein des regroupements internationaux des acteurs déjà investis dans la conservation des données numériques. Cette collaboration donnera une vue claire des travaux déjà réalisés et qui peuvent permettre à chaque état de bénéficier d'expériences assez significatives.

Par exemple, l'Union européenne a mis en place, en avril 2001, lors de la conférence de Lund, en Suède le Groupe des représentants nationaux sur la numérisation du patrimoine scientifique et culturel (GRN). Le GRN doit permettre de coordonner les politiques de numérisation des différents Etats-membres. Les représentants nationaux sont des responsables administratifs et politiques nommés par leur gouvernement. Le GRN a établi « les principes de Lund », déclaration d'intention des Etats pour faire évoluer leur politique de numérisation du patrimoine dans un sens commun. Parmi ces principes, on trouve la constitution d'une structure commune de coordination de la numérisation dans les différents secteurs culturels (musées, bibliothèques, archives, archéologie...), la conduite d'une politique de recherche pour la valorisation du patrimoine, l'évaluation des pratiques de numérisation et l'adoption de standards communs. Les réunions du GRN permettent de discuter de la mise en pratique des principes de Lund et de définir les priorités.

Le patrimoine culturel de toutes les régions et de tous les pays devrait être préservé et rendu accessible afin de parvenir avec le temps à une représentation équilibrée et équitable de tous les peuples, nations et cultures. Il convient aussi d'encourager les groupes minoritaires au sein des nations à

voir dans leur patrimoine numérique un élément du patrimoine numérique global du monde.

C'est pourquoi, comme dans le domaine strictement économique, il est indispensable de réduire les inégalités entre pays développés et pays en développement en permettant à ces derniers de pouvoir exercer une persévérance dans leur action patrimoniale. Ainsi, le président du Sénégal, Abdoulaye Wade propose-t-il le concept de « solidarité numérique » comme une stratégie ayant pour but la résorption du fossé numérique ». En effet, la conservation du patrimoine numérique mondial n'est pas uniquement une question de patrimoine culturel. À plus longue échéance, elle aura une incidence sur la nature des sociétés du savoir qui sont en train de voir le jour.

Cela étant, le domaine est si nouveau et l'expérience, pour le moment, si limitée qu'il faudra des efforts extraordinaires pour édifier l'infrastructure nécessaire. Des ressources suffisantes et un appui au niveau décisionnel sont indispensables pour faire en sorte que les générations futures puissent encore accéder aux abondantes ressources numériques notamment créées au cours des dix dernières années.

Chapitre 5 Les recommandations

La conservation du patrimoine numérique de l'humanité est une nouvelle tâche qui incombe aux Etats et aux acteurs de la société de l'information.

Il s'agit d'une activité continue qui exige l'engagement et la participation non seulement des institutions qui s'occupent du patrimoine mais également des pouvoirs publics, des autres décideurs, des producteurs et utilisateurs de l'information, des fabricants de logiciels et des organisations et associations professionnelles internationales.

Les solutions supposent de :

1. S'appuyer sur l'expérience acquise en matière de conservation d'autres formes du patrimoine mondial, matériel ou immatériel, tels que les manuscrits et les documents imprimés ou audiovisuels, pour créer un cadre institutionnel, juridique et technique susceptible d'optimiser la conservation et l'accessibilité du patrimoine numérique

2. Fixer des objectifs clairs – tant qualitatifs que quantitatifs – en matière de conservation, permettant de déterminer quels documents ou familles de documents doivent être conservés et si cette conservation doit être exhaustive et systématique ou seulement périodique, par échantillonnage

3. Garantir l'accès en permanence aux documents du patrimoine numérique, en particulier ceux qui appartiennent au domaine public, pour donner au fil du temps une image équilibrée et équitable de tous les peuples, nations, cultures et langues

4. Favoriser les programmes d'enseignement et de formation, les accords en matière de partage des ressources, ainsi que la diffusion des résultats de la recherche et des meilleures pratiques afin de démocratiser l'accès aux techniques de conservation numérique

5. Soutenir résolument les recherches sur les technologies et modèles prometteurs afin de mettre au point le plus rapidement possible des systèmes pleinement opérationnels de conservation du patrimoine numérique

6. Renforcer la coopération et la solidarité internationales en vue de permettre à tous les pays, en particulier aux pays en développement et aux pays en transition, d'assurer la création, la diffusion et la conservation de leur patrimoine numérique ainsi que l'accès permanent à celui-ci