

Diplôme national de master

Domaine - sciences humaines et sociales

Mention – sciences de l'information et des bibliothèques

Spécialité – archives numériques

## **Types, Stratégies et Enjeux des nouveaux supports de l'information**

**Perrine MERCIER-GUYON**

Sous la direction de Pascal Robert  
Professeur des universités - ENSSIB



## ***Remerciements***

*Je tiens à remercier Monsieur Pascal Robert pour ses conseils et pour m'avoir orientée dans mes travaux de recherche.*

*Merci à Bénédicte pour son soutien, à Tommy pour sa patience et son aide tout au long de ce mémoire, et à Catherine pour sa précieuse correction.*

### **Résumé :**

*L'évolutivité des technologies numériques actuelles se caractérise par un investissement intense dans la recherche et le développement de nouvelles techniques de sauvegarde permettant de stocker d'importantes quantités d'informations. Devenues support de notre mémoire, les données numériques réécrivent l'identité de l'archiviste et bouleversent les stratégies de conservation pérenne.*

### **Descripteurs :**

*Archives -Technologies – Nouveaux supports - Numérique – Numérisation – Stratégies - Enjeux*

### **Abstract:**

*The evolution of today's digital technologies is characterized by an intense research investment and the development of new backup technical to save a lot of data. It's become a support for our memory; digital data rewrite the identity of the archivist and change perennial conservation strategies.*

### **Keywords:**

*Archives -Technologies - New Media - Digital - Scanning - Strategies - Issues*

### **Droits d'auteurs**



Cette création est mise à disposition selon le Contrat :  
« **Paternité-Pas d'Utilisation Commerciale-Pas de Modification 2.0 France** »  
disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/3.0/deed.fr> ou  
par courrier postal à Creative Commons, 171 Second Street, Suite 300, San  
Francisco, California 94105, USA.



# Sommaire

<b>SIGLES ET ABREVIATIONS .....</b>	<b>7</b>
<b>INTRODUCTION.....</b>	<b>11</b>
<b>LES TECHNOLOGIES DE LA MEMOIRE.....</b>	<b>13</b>
<b>Les supports matériels .....</b>	<b>13</b>
<i>Qu'est ce qu'une mémoire? .....</i>	<i>13</i>
<i>Les progrès de la mémoire informatique .....</i>	<i>18</i>
<i>Graver l'information .....</i>	<i>24</i>
<b>Des limites physiques .....</b>	<b>31</b>
<i>L'obsolescence programmée.....</i>	<i>31</i>
<i>La durée de vie des supports en archive.....</i>	<i>32</i>
<i>L'altération du matériel .....</i>	<i>33</i>
<b>L'information devient immatériel .....</b>	<b>36</b>
<i>Les formats de l'information et leurs évolutions.....</i>	<i>36</i>
<i>Les solutions pour préserver l'intégrité d'un document.....</i>	<i>38</i>
<i>Les défis de la duplication.....</i>	<i>38</i>
<b>LES STRATEGIES ARCHIVISTIQUES.....</b>	<b>40</b>
<b>Redéfinir le document.....</b>	<b>40</b>
<i>Du papier au numérique .....</i>	<i>40</i>
<i>La sélection de l'archive, un processus intellectuel .....</i>	<i>41</i>
<i>La nature du produit numérique .....</i>	<i>43</i>
<b>Des approches de conservation en évolution .....</b>	<b>45</b>
<i>Fonctionnement actuel, quels projets?.....</i>	<i>45</i>
<i>Des outils de conception.....</i>	<i>50</i>
<i>Exemple du fonctionnement de la plate-forme PAC.....</i>	<i>53</i>
<b>Cadre, limites et recommandations pour le numérique : .....</b>	<b>58</b>

<i>Le contexte législatif et réglementaire</i> .....	58
<i>La préconisation de l'Open Access</i> .....	59
<i>Les risques du libre échange</i> .....	62
<b>L'AVENIR DE L'INFORMATION</b> .....	<b>67</b>
<b>Les enjeux émergents des nouveaux supports</b> .....	<b>67</b>
<i>L'évolution de l'information</i> .....	67
<i>La «logistication» du document numérique</i> .....	68
<i>Nouveaux enjeux, nouvelles missions</i> .....	70
<b>Quels effets sur la mémoire sociale ?</b> .....	<b>75</b>
<i>L'information comme bien commun</i> .....	75
<i>Les risques éthiques</i> .....	78
<i>Les conséquences sur la mémoire sociale</i> .....	83
<b>Donner du sens : l'utilisation de la sémiotique</b> .....	<b>85</b>
<i>Enjeux de la capitalisation de l'information</i> .....	85
<i>Le rôle de l'archiviste et l'analyse discursive</i> .....	86
<i>Le développement d'une archive sémantique</i> .....	88
<b>CONCLUSION</b> .....	<b>95</b>
<b>SOURCES</b> .....	<b>97</b>
<b>BIBLIOGRAPHIE</b> .....	<b>99</b>
<b>TABLE DES ANNEXES</b> .....	<b>105</b>
<b>GLOSSAIRE</b> .....	<b>107</b>
<b>TABLE DES ILLUSTRATIONS</b> .....	<b>109</b>
<b>TABLE DES MATIERES</b> .....	<b>111</b>



## *Sigles et abréviations*

ADN : Acide Désoxyribonucléique

AFNOR : Association française de normalisation

ANSSI : Agence nationale de la sécurité des systèmes d'information

ASCII : American Standard Code for Information Interchange

BNF : Bibliothèque nationale de France

CNIL : Commission nationale de l'informatique et des libertés

CRM : Customer Relationship Management

DON : Disques optiques numériques

ENSSIB : École nationale supérieure des sciences de l'information et des bibliothèques

INA : Institut national de l'audiovisuel

JPEG: Joint Photographic Experts Group

MARC: Machine-Readable Cataloging

PDF: Portable Document Format

SGML: Standard Generalized Markup Language

SIB : Science de l'information et des Bibliothèques

UNESCO : Organisation des Nations unies pour l'éducation, la science et la culture

XML : Extensible Markup Language



# INTRODUCTION

---

Il est aisé de remarquer l'obsolescence des technologies de notre quotidien, les disquettes sont désormais reléguées au placard et d'autres supports subissent progressivement le même sort. Excepté un regain d'intérêt vintage comme c'est le cas avec les vinyles, on peut facilement prédire que les supports d'aujourd'hui seront tôt ou tard dépassés. C'est non, seulement l'émergence d'un problème de pérennité de support, mais également une course à la qualité et à la performance que se lancent les industriels.

C'est grâce à cette dynamique, qui finance les recherches technologiques à la pointe, que les capacités de stockage augmentent chaque année davantage. C'est un marché devenu concurrentiel dans lesquels les politiques d'archivage doivent mettre en place une veille dynamique de ces évolutions afin d'utiliser le meilleur support. Les distributeurs font souvent l'éloge de la durée de vie de leurs matériels mais dans les faits la plupart se détériore en quelques années. Les nouveaux supports se caractérisent ainsi non seulement dans leurs aspects matériels mais aussi dans le défi qu'ils représentent pour le domaine de l'archivage numérique. Les archives sont, selon la définition de Foucault « *le système général de la formation et de la transformation des énoncés* »<sup>1</sup>, en d'autres termes un témoignage matériel d'une idée, discours qu'il qualifie d'énoncés. Selon le code du patrimoine, article L. 211-1 il s'agit de l'ensemble des « *documents, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité* ».

On emploie plus naturellement le terme des archives au pluriel puisqu'on distingue les archives, les sources, les sources imprimées et les traces du passé. Elles sont

---

<sup>1</sup> Michel FOUCAULT, L'archéologie du savoir, Gallimard, Paris, 2002 cité dans l'article de Éric KETELAAR, « (Dé) Construire l'archive Matériaux pour l'histoire de notre temps » ; « Matériaux pour l'histoire de notre temps », N° 82, p65-70, 2006, p138, (<http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2-page-65.htm>, consulté le 15 mai 2014)

traditionnellement constituées en fonds qui vont permettre de comparer l'évolution dans le temps des documents et des informations. Les archives deviennent donc garant de la préservation d'une mémoire sociale. La question se pose alors, comment une mémoire sociale pourrait-elle résister aux dégâts du temps si nous ne nous interrogeons pas sur la pérennité des supports aujourd'hui?

Pour bien comprendre les enjeux de l'archivage de demain, il est tout d'abord nécessaire de faire un tour d'horizon des technologies disponibles et de celles qui sont encore à l'état de projet. Sur de nombreux aspects le numérique domine désormais l'analogique et face à ce changement les institutions responsables de la préservation de l'information doivent mettre en place de nouvelles stratégies de conservation.

Ces projets qui émergent et font de la pérennité un enjeu central pour notre société seront le sujet de discussion de notre seconde partie. Les archives doivent en effet assurer un rôle de conservation patrimoniale et dans un environnement numérique où la copie est identique à l'original, une politique de sélection doit être réfléchie. L'archiviste qui travaillait autrefois en amont du cycle de vie du document doit travailler désormais au plus proche de sa création pour lui permettre de suivre et d'identifier les étapes du cycle de vie du document et ainsi de pouvoir lui associer un contexte qu'un support matériel nous apportait auparavant. Aujourd'hui on a la possibilité d'archiver en temps réel, et donc de penser aux modalités du futur archivage c'est une véritable transformation du rapport entre le passé et le présent.

Les processus de sélection de ce que les générations futures connaîtront du passé place l'archiviste à un poste central dans la construction de l'histoire. La quantité d'informations permet d'avoir une vue globale des tendances et non plus fragmentaire donnant naissance à des enjeux informationnels parfois politiques ou commerciaux. Ces flux informationnels deviennent alors des flux sociaux, l'information et la communication apportent à la fois une transparence dérangeante et à la fois un énorme bénéfice pour le partage du savoir. Les questions éthiques et politiques seront ainsi traitées dans une troisième partie pour nous permettre de faire émerger les mécanismes complexes de cette machine archivistique

# LES TECHNOLOGIES DE LA MEMOIRE

---

## LES SUPPORTS MATERIELS

Des tablettes d'argiles aux parchemins, l'archéologie et l'histoire témoignent de la nécessité ancienne de fixer l'information sur un support constitutif d'une mémoire sociale. L'arrivée du numérique en bouleverse la nature ancestrale dès l'arrivée au XIX<sup>ème</sup> siècle des métiers à tisser à cartes Jacquard. Ces cartes contrôlaient via un système de trous l'altération des mouvements des crochets des métiers à tisser et rendaient donc automatique la fabrication des textiles. Un peu plus tard en 1834, un mathématicien anglais du nom de Charles Babbage va avoir l'idée d'utiliser ce même système pour sa machine à calculer. Les cartes étaient alors prévues pour automatiser les opérations à effectuer<sup>2</sup>.

Aujourd'hui les ordinateurs et le développement de nouveaux types de supports mécaniques ont fait du numérique le principal support de la mémoire. Supprimant, au passage la problématique du stockage volumineux des documents matériels il est alors logique de s'interroger sur la capacité de stockage de ces nouveaux supports. Depuis longtemps déjà les ingénieurs rêvent d'une mémoire universelle<sup>3</sup>, rapide d'accès et faible en énergie, capable de puissante capacité de stockage, mais est-ce vraiment envisageable?

## Qu'est-ce qu'une mémoire?

Pour penser les mémoires de demain il est intéressant de se référer à des mémoires déjà existantes sur les modèles construits par la nature. On observe ainsi de nombreuses analogies entre la mémoire artificielle que l'homme développe et la mémoire biologique dont il est doté. Si on analyse notre propre mémoire, on se rend compte que plusieurs modèles théoriques existent. On considère dans tous les cas qu'elle est composée de différents sous-systèmes auxquels sont associées des

---

<sup>2</sup> Luca PERNIOLA, « Les mémoires du futur » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p92, Novembre 2013, 188p

<sup>3</sup> *Ibid.*

fonctions. Dans les années 60, émergent les premiers modèles d'interprétation. On distinguait alors globalement la mémoire sensorielle, la mémoire à court et à long terme. La première retient la majorité de l'information issue des différents sens de perception humaine. La mémoire à court terme, dont la capacité d'utilisation se limitera à quelques dizaines de secondes et qui nous permet de nous souvenir, par exemple, du début de notre phrase. Enfin pour retenir toutes ces informations nous sommes dotés d'une mémoire à long terme capable de la conserver longtemps et en importantes quantités.

Ce modèle de mémoire s'est enrichi avec les années en différentes sous catégories<sup>4</sup>. Il semble donc que l'évolution nous ait conduits à adopter cette structure afin d'être à même de percevoir et de traiter rapidement une information tout en acquérant un savoir. Cette structuration de la mémoire dans notre cerveau ne fonctionne pas sur le modèle d'une mémoire universelle, la combinaison de mémoires spécialisées semble donc plus pertinente.

Les technologies de la mémoire actuelle fonctionnent déjà sur ces combinaisons en associant dans une machine deux systèmes de mémoire. Comme pour le cerveau humain, chaque mémoire à ses propres capacités et fonction, par exemple conserver ou non l'information sans alimentation électrique.

D'un côté on trouve les mémoires de travail, dites mémoires vives ou que l'on appelle également la RAM (Random Access Memory). Principale mémoire du système, c'est elle qui permet de stocker de façon temporaire les données nécessaires à l'exécution d'un programme. Cela signifie qu'en l'absence d'une source d'énergie électrique, l'information qui y est enregistrée disparaît en quelques dizaines de millisecondes. Lors de la mise hors tension de l'appareil ou lors d'une coupure de courant par exemple les informations contenues dans la mémoire de travail vont être effacées. Elles ont donc un temps de mémoire très court et sont dites volatiles.

---

<sup>4</sup> Luca PERNIOLA, « Les mémoires du futur » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p93, Novembre 2013, 188p

D'un autre côté nous avons des mémoires de stockage, le disque dur par exemple, au prix d'un temps d'accès plus long ils peuvent enregistrer d'énormes quantités d'information. Lorsqu'on associe ces deux types de mémoire il est possible de construire ce système que l'on utilise au quotidien pour exécuter des opérations en peu de temps donc tout en pouvant les conserver.

En plus des deux mémoires précédents cités ils existent d'autres types de mémoire, comme la mémoire morte, qui va stocker l'information et la conserver même hors tension, on l'appelle ROM (Read-Only Memory) que l'on pourrait traduire par « *en lecture seule* ».

Ou encore la Mémoire flash. Cette mémoire, nous y reviendrons combine le fonctionnement des deux mémoires précédemment citées. Elle possède la non-volatilité des mémoires mortes mais reste accessible en lecture ou en écriture par contre avec un temps d'accès à peine plus long que celui d'une mémoire vive.

On peut imaginer la hiérarchie des mémoires comme une pyramide, au sommet on y trouve les mémoires de travail avec des capacités de stockage faible mais un temps d'accès très court, puis plus on descend dans la pyramide plus le volume d'information stockable augmente et le temps d'accès diminue. Tout en bas on retrouve donc les mémoires de stockage.

### *Quels fonctionnements ?*

Au plus près du processeur, on distingue deux types de mémoires de travail : les mémoires statiques (ou SRAM pour Static Random Access Memory) et les mémoires dynamiques (ou DRAM pour Dynamic Random Access Memory)<sup>5</sup>. Les premières occupant le sommet de la pyramide.

---

<sup>5</sup> Luca PERNIOLA, « Les mémoires du futur » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p94, Novembre 2013, 188p

Elles sont directement construites sur la puce de silicium du microprocesseur, car elles sont constituées de transistors, comme ce dernier. Leur temps d'accès est de quelques nanosecondes, mais le volume de stockage est très faible.

Les mémoires dynamiques, juste en dessous dans la pyramide des mémoires, ont une capacité de stockage plus importante ; elles sont en général placées sur des puces indépendantes, dont le temps d'accès est compris entre 10 et 100 nanosecondes. Ces mémoires de travail sont utilisées pour effectuer les opérations de calcul des microprocesseurs. Des mémoires volatiles qui perdent les données. Dans une mémoire dynamique, le bit, unité de 0 et 1 qui permet d'encoder de l'information, est enregistré par un système composé d'un transistor et d'un condensateur. Ces composants sont gravés à côté de cette puce en silicium du microprocesseur. Le transistor va alors agir comme un interrupteur qui autorise, s'il est ouvert, la lecture du bit par un courant électrique. Physiquement, l'état du bit est défini par la charge électrique (l'accumulation d'électrons) du condensateur. Si le condensateur est chargé, le bit vaut 1, s'il est déchargé, le bit vaut 0.

Le problème des condensateurs c'est qu'ils retiennent mal les électrons, qui fuient dans le silicium dont sont composés les éléments. Pour cette raison, la mémoire dynamique est volatile. Il est donc nécessaire de réécrire le bit d'information régulièrement, toutes les quelques dizaines de millisecondes, pour éviter que le condensateur ne se décharge trop et que le bit change d'état. Ainsi, sans alimentation électrique, l'information disparaît de la mémoire.

Dans les premières mémoires dynamiques des années 1960, les transistors et les condensateurs étaient accolés. La miniaturisation a conduit à réduire la taille des composants, mais aussi à changer leur placement relatif. Les condensateurs sont maintenant construits au-dessus des transistors.<sup>6</sup> De plus, la forme des condensateurs a été modifiée pour réduire un peu plus la consommation électrique, ils ressemblent désormais à des petits tubes très long. Avec un rapport de la longueur du condensateur sur la largeur de sa base qui peut atteindre jusqu'à 50, et une densité élevée de condensateurs, les fabricants ont dû adapter les autres

---

<sup>6</sup>Christophe GAUTHIER, "Comment ça marche : la mémoire Flash", Mars 2005 (<http://www.01net.com/editorial/294823/comment-ca-marche-la-memoire-flash/>, consulté le 15 Juin 2014)

composants en produisant des cavités longues et étroites tout en gérant les problèmes de fuites.

Mais la limite de miniaturisation semble atteinte pour ces dispositifs, dont la densité de stockage (quatre gigaoctets) augmente difficilement. Ainsi, les ingénieurs travaillent en priorité sur deux aspects relatifs aux mémoires dynamiques : l'augmentation de la densité d'information et l'économie d'énergie, obtenue en supprimant la volatilité.

### *L'arrivée de la mémoire flash*

Les « *mémoires flash* » sont justement un type de mémoire non volatile. Mais elles ne peuvent malheureusement pas remplacer les mémoires dynamiques. On pourrait les situer plus bas dans la pyramide des mémoires car elles ont un temps d'accès d'une dizaine de microsecondes, plus que les mémoires dynamiques. Depuis leur développement les mémoires flash ont été miniaturisées.

Elles atteignent actuellement des capacités de stockage de 128 gigaoctets. Mise au point en 1980 par le Professeur Fujio Masuoka et l'un de ses collègues Shoji Ariizumi, employés de Toshiba, la création de la mémoire flash avait pour but de réduire le temps d'écriture et de suppression et d'offrir une forte capacité de stockage tout en réduisant la taille. La technologie est dévoilée 4 ans plus tard à l'International Electron Devices Meeting (IEDM) de San Francisco<sup>7</sup>. Toshiba innovera dès 1987 en lançant la mémoire flash NAND et un plus tard c'est Intel qui mise sur le potentiel de cette technologie en créant la mémoire flash NOR. La taille de ces mémoires ont a progressivement réduit et aujourd'hui, la mémoire flash NAND est gravée à 30 nm et la mémoire flash NOR gravée à 45 nm.

Au niveau du fonctionnement, dans ces mémoires flash, chaque bit est représenté par l'état ouvert ou fermé d'un transistor. Un réservoir au sein du transistor peut accumuler une certaine quantité d'électrons qui créent un champ électrique. Celui-

---

<sup>7</sup> Article Wikipédia : [http://fr.wikipedia.org/wiki/Mémoire\\_flash](http://fr.wikipedia.org/wiki/Mémoire_flash)

ci laisse passer ou bloque le courant qui traverse le transistor. Les mémoires flash sont non volatiles, car les électrons restent bloqués dans le réservoir même en l'absence d'une source extérieure qui permet dans les mémoires dynamiques de rafraîchir les données.

Pour bloquer ces électrons sans courant électrique le réservoir a été intercalé entre deux couches d'oxydes qui l'isolent du substrat environnant. Son fonctionnement réside dans l'utilisation d'une cellule de base, un transistor MOS, possédant une grille flottante qui va piéger les électrons ce qui va permettre le stockage de l'information. Pour que les électrons traversent cet oxyde il existe deux solutions, soit on injecte des électrons chauds soit on obtient un effet tunnel en appliquant une haute tension sur la grille de contrôle. Cependant, pour conserver ces propriétés isolantes, ces couches d'oxydes ne peuvent être soumises à une trop grande miniaturisation qui affaiblirait ses capacités de blocage et augmenterait le risque d'erreurs d'écriture et de lecture. En moyenne on estime que la mémoire flash peut supporter en fonction de la qualité de l'oxyde utilisé jusqu'à 100 000 écritures et effacement. L'avantage de ce type de mémoire réside dans l'absence d'éléments mécaniques, en conséquence elle résiste très bien aux chocs.

## **Les progrès de la mémoire informatique**

Récemment des entreprises comme Samsung se sont lancées dans la production de puce mémoire flash en 3D<sup>8</sup>, l'objectif étant d'améliorer les capacités de ces puces en multipliant les couches. Aujourd'hui les capacités varient entre quelques Mo et 1 To. Les constructeurs Sun et Micron envisagent aussi d'améliorer la durée de vie de ces supports en permettant plus de réécriture sur le support.

Il s'agit donc pour les fournisseurs de mémoire, les industriels et leurs ingénieurs de travailler sur l'amélioration de ces capacités, comme leur résistance et leurs durées de vie, le temps d'accès pouvant être compris entre la nanoseconde et la milliseconde.

---

<sup>8</sup> Bruno Cormier, « Samsung SSD 850 Pro : de la mémoire flash 3D très rapide et ultra-durable », Juillet 2014 (<http://www.01net.com/editorial/622914/samsung-ssd-850-pro-memoire-flash-3d-tres-rapide-et-ultra-durable/>, consulté le 2 Juillet 2014)

Chaque jour de nouvelles mémoires sont développées, des mémoires qui pourront même disposer des caractéristiques des mémoires de travail et des mémoires de stockage permettant par exemple de réduire les dépenses énergétiques qu'elles engendrent. Les pistes s'orientent plutôt vers de nouveaux dispositifs qui permettraient de palier aux performances des mémoires dynamiques et des mémoires flash.

C'est par exemple le cas avec la mise au point d'une résistance électrique à la place du condensateur des mémoires dynamiques. Cette modification va permettre de faire varier le bit correspondant en fonction d'une résistance faible ou élevée. Cette mémoire devient alors non volatile. On peut citer dans ce principe de mémoire à résistance variable l'exemple des mémoires magnétiques et les mémoires à changement de phase, moins énergivores et présentant des capacités de stockage sensiblement similaires à celle des mémoires dynamiques.

Les mémoires non volatiles restent pour l'instant les plus rentables et les moins énergivores. Ces mémoires dites « *statiques* », avec leurs temps d'accès plus long, ne remplaceront pas tout de suite les mémoires volatiles dont l'ultra rapidité est indispensable au fonctionnement du microprocesseur. Les mémoires flash du type de la clé USB, des cartes mémoires et aujourd'hui des disques SSD sont adaptées à un stockage d'environ 5 ans et dont le volume se limite à 256Go. Petite et légère elle s'adapte à de nombreuses utilisations et servira souvent à dépanner pour le transport de donnée.

Depuis quelques années pourtant les disques durs de certains appareils sont remplacés par des dispositifs plus ingénieux disposant de mémoires SSD (Solid State Drive) par exemple. Équipés de ce type de disque les ordinateurs sont débarrassés d'éléments mécaniques et donc plus rapides. Les solid-state drive, SSD ou disque électronique pour les québécois. Le matériel est en réalité constitué de mémoire à semi-conducteurs à l'état solide et non plus sur un support magnétique en rotation rapide comme les disques durs que nous connaissons au quotidien.

Plus solides que des disques durs classiques, ils sont dotés généralement, notamment depuis 2003, d'un plateau en verre. Grâce à cette particularité ils sont plus résistants aux vibrations et aux chocs, de plus leurs performances en consommation et en débit sont nettement supérieures à celle d'un disque dur mécanique. Seul le prix était encore près de dix fois plus chères en 2012 que celui de ses prédécesseurs. La solution adoptée alors pour les ordinateurs du quotidien est d'utiliser un SSD pour stocker le système et un disque dur pour les données. La différence de taille est également observable avec l'illustration suivante présentant à gauche un disque dur classique de 160Go et à droite un disque SSD de 250 Go.



**Figure 1 : Comparaison d'un disque dur classique et d'un disque dur SSD**

### *Les mémoires magnétiques*

Les disques durs de nos machines sont souvent lents, c'est en raison du temps nécessaire à la tête de lecture qui doit d'abord se positionner au bon endroit par un déplacement du bras mécanique et par une rotation du disque. On peut considérer les disques durs occupant le bas de la pyramide des mémoires<sup>9</sup>. En revanche leur prix est bas et leurs coûts de stockage supérieurs à ceux de la mémoire flash.

---

<sup>9</sup>Luca PERNIOLA, « Les mémoires du futur » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p92-99, Novembre 2013, 188p

Contrôlée par l'aimantation des éléments, la résistance qui code le bit est contrôlée par des éléments dont est formé le composant. Avec un temps d'écriture de l'ordre de la nanoseconde avec toutefois un temps un peu plus long pour lire une information. Les mémoires magnétiques ont été créées dans les années 50 mais leurs utilisations se limitaient alors à des domaines bien spécifiques comme l'aérospatial de par leurs résistances aux rayons cosmiques par exemple. Mais depuis une dizaine d'années les progrès technologiques relance l'intérêt pour cette technologie. On trouve par exemple aujourd'hui des mémoires magnétiques ayant des capacités de stockage de 64 mégaoctets.

Récemment des résultats de recherche ont été publiés par l'Université de York en Angleterre et diffusés sur le web par Tom's Hardware<sup>10</sup>, concernant des disques à laser pulsé annonçant plus de 200 Gb/s pour une densité supérieure aux disques actuels. Le disque dur est composé d'une tête magnétique fonctionnant sur un plateau ferromagnétique qui induit un sens 0 ou 1 en quelques nano secondes.

La collaboration a réuni des chercheurs venant de Suisse, de Russie, du Japon, des Pays-Bas et même d'Espagne qui expliquent qu'il suffirait de chauffer un court instant ( moins 5 pico secondes) le plateau composé alors de fer et de gadolinium avec un laser pour inscrire ou modifier la polarité ( 0 ou 1 ) inscrite sur celui-ci. Ce système sera développé pendant la prochaine décennie le temps de perfectionner le processus et d'industrialiser la technique.

Ce type d'innovation n'est pas sans rappeler l'annonce, dans la revue Science, qu'avait fait IBM sur certains de ses travaux portant sur le bit magnétique. L'article explique en outre que les chercheurs auraient réussi dans seulement 12 atomes à maintenir stocker à basse température près d'un bit de données magnétiques.

Le web journal Tom's Hardware<sup>11</sup> a souhaité en savoir plus et a contacté le géant pour comprendre un peu plus le fonctionnement du "plus petit bit magnétique". Le bit fonctionne donc selon leurs investigations à 1 °K (-272,15 °C) et commencera à montrer des signes d'instabilité à partir de 5 °K (-268,15 °C), lui procurant une forte résistance aux aléas du stockage. Il est actuellement nécessaire pour stocker

---

<sup>10</sup> LECLERCQ Denis, « Le disque dur à laser pulsé, prochaine révolution du stockage ? », Février 2012 (<http://www.pcworld.fr/stockage/actualites,le-disque-dur-laser-pulse-presque-finalise,524649,1.htm?comments=1#comments>, consulté le 15 Mars 2014)

<sup>11</sup> Ibid.

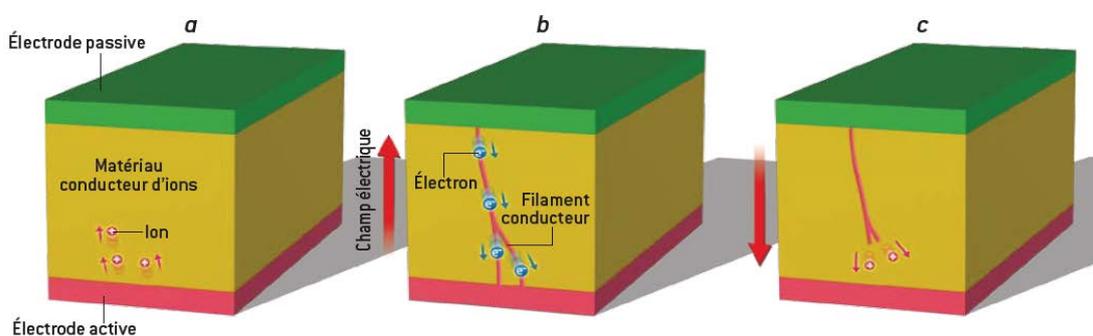
un bit de données sur un disque dur magnétique, pour le même bit, à température normale, on devrait pouvoir stocker sur à peine 150 atomes.

Pour l'instant il n'est pas envisageable de construire de façon industrielle ces supports, la précision nécessaire serait trop coûteuse pour une production à la chaîne. Mais cela reste possible dans un futur proche multipliant par 100 les capacités de stockage actuelles.

### *Les nouvelles mémoires*

D'autres solutions pourraient voir le jour dans les années à venir<sup>12</sup>, deux pistes sont étudiées actuellement par les industriels. Les mémoires à oxydes (OxRAM ou Oxide-based Random Access Memory). Dispositifs récents mais les études se multiplient et les essais aussi.

On note notamment que les sociétés Sandisk et Toshiba ont déjà annoncé en 2013 l'arrivée d'une mémoire à oxyde de 32 gigaoctet.



**Figure 2 : Fonctionnement des mémoires à Oxyde PERNIOLA Luca, « Les mémoires du futur » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, Novembre 2013, pp. 92-99.**

Concrètement on code l'information via une résistance qui crée à l'aide d'un stimulus électrique, un filament permettant de conduire ou non les électrodes.

Observé depuis les années 60, ce phénomène électrique qui permet la formation du filament, qu'on appelle le claquage de l'oxyde, n'est utilisé que depuis 2004.

<sup>12</sup> PERNIOLA Luca, « Les mémoires du futur » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, Novembre 2013, pp. 92-99.

L'objectif est désormais d'améliorer le procédé pour optimiser les performances de cette mémoire.

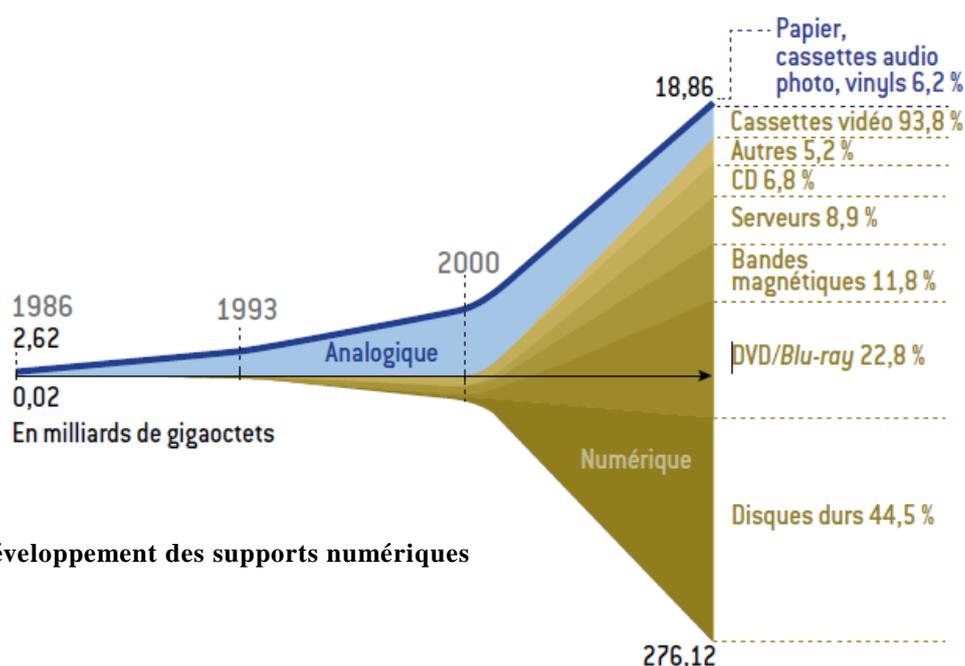
Les progrès sont si rapides dans ce domaine qu'il est difficile de donner les caractéristiques de ces mémoires. Le temps d'écriture varie de la nanoseconde à la milliseconde suivant les dispositifs. Par ailleurs, les techniques de fabrication, surtout pour les mémoires à oxydes, sont relativement courantes, et leur coût de production est très compétitif. En revanche, les phénomènes microscopiques mis en jeu restent à mieux comprendre.

### *Les supports actuels : entre applications et réflexions*

Outre les disques durs, il existe encore la possibilité de stocker sur des supports externes les données.

Sans le lecteur, le support n'est rien, les exemples d'obsolescence se multiplient et aujourd'hui on se tourne vers des solutions plus durables, le nombre de documents générés est lui aussi multiplié par l'arrivée d'un nombre immense de moyens d'enregistrer.

Ce schéma permet de mettre en évidence les supports les plus utilisés actuellement.



**Figure 3 : Développement des supports numériques**

On crée plus donc on veut stocker plus, fini la photo qui devait être parfaite et dont la pellicule devait être soigneusement développée par un professionnel. La fonction a été reléguée à des imprimantes ou stocker sur un CD au fond d'un tiroir dans les meilleurs des cas. Et pourtant ces supports ont facilité grandement l'organisation des archives ces dernières années. Formats de boîtier universel, standardisation des lecteurs, capacités de stockage et de conservation intéressantes, en bref une opérabilité matérielle qui tend à la normalisation. Ces supports évoluent en même temps que les besoins, la pérennité des données est devenue un enjeu majeur émergent, une préoccupation latente qui génère beaucoup de travaux de recherches. Les résultats de ces recherches ne sont pas si éloignés de nous que cela et déjà les prochaines générations de supports de l'information sont annoncées.

## **Graver l'information**

### *Les disques :*

Après 30 ans d'utilisation le CD est encore utilisé ponctuellement pour des petits travaux de sauvegarde et l'on s'inscrit inconsciemment dans une logique d'archivage. Le CD gravé n'est généralement pas réinscriptible et finalement pour des copies quotidiennes et personnelles ce support se positionne plutôt bien dans le classement.

Dernièrement, avec l'arrivée de nouvelles méthodes de compression audio ayant donné naissance au MP3 par exemple, des lecteurs audio pouvant lire ces formats sur un CD-R(W) ont été développés. Permettant de stocker alors de 4,4113 à 1114 fois<sup>13</sup> plus de musique que sur un CD audio avec cependant une dégradation plus ou moins perceptible de la qualité sonore en fonction du débit auquel le disque / le morceau a été compressé.

Le CD a cependant rapidement des limites quand il s'agit d'archiver de grande quantité de données où l'on atteindra facilement la limite de 650Mo. La vitesse n'est pas non plus son fort lorsqu'il s'agit de temps d'accès mais pour la lecture de

---

<sup>13</sup> Article Wikipédia :[http://fr.wikipedia.org/wiki/Disque\\_compact](http://fr.wikipedia.org/wiki/Disque_compact)

films compressés les débits peuvent être très élevés en matière de lecture continue de fichier. Il n'empêche qu'en 2013, le CD audio est en progressivement en baisse.

On trouve également sur le marché du disque le DVD-ROM qui existe depuis 15 ans mais a tout de même mis du temps à s'imposer dans les usages. Et pourtant depuis 5 ans la plupart des ordinateurs sont vendus avec d'un lecteur DVD, et même combiné à un graveur.

C'est en 1995 que le DVD voit le jour<sup>14</sup> et que, progressivement il va s'imposer face à la traditionnelle cassette VHS.

Non seulement d'un point de vue facilité d'utilisation, on oublie le rembobinage, on peut accéder à un moment précis du film, qui est par ailleurs chapitré et on a même accès à des bonus.

La qualité de l'image est nettement supérieure et on peut stocker jusqu'à 7 fois plus de données. Le prix est abordable et stable, les lecteurs sont plus robustes et le gain d'espace, sans être au cœur du succès du support, n'en reste pas moins un avantage pour le stockage en archive.

Un système de zones a été conçu par les industriels pour encadrer l'exportation des disques dans le monde. C'est pour eux une façon de maintenir un contrôle non seulement sur le prix de vente mais aussi sur les dates de sortie.

Deux systèmes étaient également censés en empêcher la copie : le système CSS interdisant la copie de données d'un DVD vers un ordinateur, et le système Macrovision, interdisant la copie d'un DVD vers un magnétoscope.

Ces systèmes de protection ont été vite cassés et on trouve aujourd'hui une masse de logiciels permettant d'extraire un DVD vers un PC et de le recopier soit au format original soit au format VCD ou DivX qui autorise de le copier sur un CD classique (avec néanmoins une dégradation légère de la qualité). La plupart des lecteurs DVD récents sont « compatibles DivX », ce qui signifie qu'ils peuvent lire des DVD vidéo aussi bien que des CD ou DVD contenant des vidéos compressées au format DivX (bien entendu un DVD peut contenir plusieurs films DivX).

---

<sup>14</sup> Article Wikipédia : <http://fr.wikipedia.org/wiki/DVD>

### *La standardisation :*

Les dimensions sont identiques à celle du CD avec un diamètre maximum de 12cm et un diamètre de trou central de 1,5cm.

Il est également possible de personnaliser la face de présentation du DVD, au départ avec des kits d'autocollants imprimables à apposer sur le disque, mais certaines imprimantes permettant une impression directement sur le disque à la condition que le disque dispose d'une surface imprimable.

Les DVD ont adopté un format de stockage standard<sup>15</sup>, des boîtiers identiques, quelqu'en soit le contenu film ou jeux vidéos. Ils sont à court terme plus résistants que ceux des CD qui avaient tendance à craquer et se fendre dans les transports. Cependant la notion de court terme est effectivement de mise de par la composition des boîtiers, le plastique vieilli mal. Physiquement le boîtier DVD mesure en moyenne 19cm de hauteur par 13,6cm environ 1,4cm d'épaisseur. La hauteur et la largeur correspondent aux boîtiers VHS, il est possible d'insérer un CD dans un boîtier DVD ce qui implique de nombreuses facilités de stockage. Le DVD, partageant les mêmes dimensions que le CD va exploiter le même support. Il existe cependant des différences techniques concernant la densité et la capacité de stockage et enregistrement mais aussi la façon dont ces procédés sont sauvegardés.

Avec une capacité de 4.5 Go, le DVD peut convenir pour l'archivage de documents professionnels voir familiaux mais ne conviendra pas pour stocker une vidéothèque. La fiabilité, reste la même que celle du CD-ROM, la fragilité et le maintien de bonnes conditions de conservation peuvent être gênants, sans parler de son indexation qui peut s'avérer problématique. Le DVD offre une bonne vitesse de lecture et d'écriture mais son temps d'accès, à l'instar du CD reste long et la gravure nécessite toujours l'utilisation d'un logiciel dédié.

Aujourd'hui l'évolution la plus connue est le disque Blu-ray mais en réalité il y'a deux formats en compétition. Le format Blu-ray et le HD DVD, 25 Go pour le

---

<sup>15</sup> Article Wikipédia : <http://fr.wikipedia.org/wiki/DVD>

premier 15 Go pour le second. Toshiba abandonne progressivement le projet en 2008 et met fin à son développement permettant l'essor du Blu-Ray.

On note aussi l'apparition bien que non disponibles pour l'instant dans le commerce, de format comme le HVD (Holographic Versatile Disc ou Disque holographique polyvalent) avec une capacité de plusieurs téraoctets.

### ***La nouvelle génération, le Blu-Ray:***

Dévoilé en octobre 2000, il faudra attendre avril 2003 pour voir le premier lecteur prototype sortir au Japon, le nom de Blu-ray<sup>16</sup> vient en réalité du type de rayon laser utilisé qui s'approche du bleu, « *Blue* » signifiant bleu et « *ray* » signifiant laser. Il adopte les mêmes dimensions standard que ces versions précédentes, mais compte tenu de sa technologie le Blu-ray est plus vulnérable aux rayures.

Annoncée en mars 2004, l'introduction de TDK dans la fondation Blu-ray s'accompagna d'indications qui permettaient l'amélioration de l'avenir du disque. Une technologie consistant à rajouter une couche de polymère très fine et très dure que l'on appelle DURABIS, permettait au support de mieux résister aux rayures. Les constructeurs Sony et Panasonic développèrent rapidement des technologies de protection similaire. D'ailleurs la fabrication du Blu-ray exige le passage d'un test de résistante aux rayures, première cause de dégradation du support. Le lecteur utilisé fonctionne avec une diode laser possédant une longueur d'onde de 405 nm, d'une couleur bleue, qui permet de lire et d'écrire les données. En comparaison les CD et les DVD utilisent quant à eux des lasers infrarouges et rouges à respectivement 780 nm et 650/635 nm.

Cette réduction de taille permet aussi de stocker davantage de données sur un disque qui pourtant fait la même taille en combinant un système de double lentille et en réduisant l'épaisseur du disque, limitant ainsi certains effets optiques, on améliore considérablement la qualité de lecture. Plus les pistes vont être rapprochées entre elles plus l'encodage des données est efficaces et peut permettre de stocker plus de données.

---

<sup>16</sup>Article Wikipédia : [http://fr.wikipedia.org/wiki/Disque\\_Blu-ray](http://fr.wikipedia.org/wiki/Disque_Blu-ray)

Différentes études sont faites pour permettre aux institutions archivistiques de s'y retrouver parmi lesquelles nous pouvons citer à titre d'exemple le Rapport sur la Qualité des Disques Blu-ray enregistrables pour l'archivage des données numériques effectuées pour le Ministère de la Culture et de la Communication au sein de la Direction générale des patrimoines. Ce dernier compare justement les capacités de résistance d'un Blu-Ray enregistrable, un BD-R, avec celle du DVD. Les recommandations de cette étude établissent un bilan satisfaisant du BD-R, faisant de lui un support prometteur pour l'archivage des données numériques<sup>17</sup>.

### *Les supports de demain*

Déjà de nouvelles améliorations pour le Blu-Ray sont en cours d'élaboration comme la création de l'Archival Disc. Né d'une alliance entre Panasonic et Sony<sup>18</sup>, c'est un support prometteur pour l'archivage. Annoncé début 2014, il pourrait être disponible dès la fin 2015. Il semble que les deux sociétés commercialiseront le format optique sous leurs marques respectives<sup>19</sup>.

Le principe est de simplement l'écart entre les pistes de 0,32 à 0,225 micromètre ce qui permet en outre de conserver la même technologie utilisée pour les Blu-ray. La possibilité d'être gravé sur les deux faces lui permettra de stocker davantage et tout cela sans changer de lecteur. Sony a annoncé que le disque sera dans premier temps conçu pour héberger jusqu'à 300 gigaoctet de données, puis 500 et prévoit même une troisième version capable de stocker jusqu'à un téraoctet de données. C'est un support envisagé comme une solution pour l'archivage des données de niveau professionnel<sup>20</sup>. L'intérêt principal de ces nouveaux Blu-ray réside dans la volonté de ces fabricants de viser un public d'entreprises et non plus de particuliers. Sony et Panasonic espèrent s'imposer auprès des acteurs de

---

<sup>17</sup> Jean-Michel LAMBERT, Jacques PERDEREAU, « Qualité des disques Blu-ray enregistrables pour l'archivage des données numérique », Novembre 2011 (<http://www.lne.fr/publications/recherche/Etude-qualite-bd-r.pdf>, consulté le 15 Février 2014)

<sup>18</sup>Sébastien TRIBOT, « L'archival Disc: le successeur du Blue Ray », 13 Juin 2014 ([http://www.techniques-ingenieur.fr/actualite/high-tech-thematique\\_193/l-archival-disc-le-successeur-du-blue-ray-article\\_231715/](http://www.techniques-ingenieur.fr/actualite/high-tech-thematique_193/l-archival-disc-le-successeur-du-blue-ray-article_231715/), consulté le 10 Janvier 2014)

<sup>19</sup>

<sup>20</sup>Sébastien TRIBOT, « L'archival Disc: le successeur du Blue Ray », 13 Juin 2014 ([http://www.techniques-ingenieur.fr/actualite/high-tech-thematique\\_193/l-archival-disc-le-successeur-du-blue-ray-article\\_231715/](http://www.techniques-ingenieur.fr/actualite/high-tech-thematique_193/l-archival-disc-le-successeur-du-blue-ray-article_231715/), consulté le 10 Janvier 2014)

l'archivage professionnel de données, qu'il s'agisse d'institutions ou d'entreprises mais visent aussi des hébergeurs de données comme outils de maintenance des données.

### *L'avenir de la gravure:*

Aujourd'hui la gravure va plus loin dans le processus de gravure et de nouveaux matériaux sont à l'essai. On peut évoquer le cas de support en quartz. Composé de nanostructures de quartz fondu avec une durée de vie plus d'un million d'années. C'est en référence au super-héros Superman que les chercheurs ont décidés de baptisé ce support: « *cristal mémoire de Superman* »<sup>21</sup>. L'histoire de Krypton, planète futuriste dont est originaire le héros, y étant stockée dans des cristaux.

Dès 2009<sup>22</sup>, la société Hitachi avait stocké des données sur du quartz mais à ce moment-là l'écriture et la lecture restaient les problèmes principaux. Gravant de minuscules points pour stocker des bits elle peut aujourd'hui en graver plus de 100 à la fois ce qui améliore considérablement le temps d'écriture.

Au début de l'année 2012, l'entreprise Hitachi avait annoncé la fabrication de support de stockage basée sur cette même technologie. Les laboratoires de la firme en collaboration avec ceux de l'Université de Kyoto, avait trouvé une façon de graver des données sur un bloc de quartz avec un laser. Les tests conduits par le constructeur lors d'essai de vieillissement accéléré promettent une résistance de près de 2h lors d'une exposition à une température de 2000°C et une durée de vie de centaines de millions d'années.

Toutefois le porte-parole précise tout de même qu'il faudra encore un peu de temps, jusqu'en 2015 au moins pour mettre ce produit sur le marché, le temps de finir de tester les capacités du produit. Ce temps prévoit également la mise en

---

<sup>21</sup>Frédéric BORDAGE, « Le quartz, support de stockage d'avenir », Novembre 2012 (<http://www.greenit.fr/article/materiel/le-quartz-support-de-stockage-davenir-4626>, consulté le 15 Mars 2014)

<sup>22</sup> Frédéric BORDAGE, « Le quartz, support de stockage d'avenir », Novembre 2012 (<http://www.greenit.fr/article/materiel/le-quartz-support-de-stockage-davenir-4626>, consulté le 15 Mars 2014)

place d'un système de transfert des données des entreprises vers des laboratoires en charge de les encoder. Aujourd'hui le seul bémol de cette technologie reste comme souvent le support de lecture, les microscopes optiques nécessaires à la lecture de ce support et leurs disponibilités dans le temps.

Récemment les universités de Southampton et d'Eindhoven ont testé un laser à impulsions ultra-rapides et ont réussi à stocker sur du verre de synthèse composé de quartz un fichier texte de 300Ko. Pour cela le codage est réalisé en 5 dimensions, comprenant, la hauteur et l'orientation de la gravure, s'ajoutent trois dimensions pour ces nanostructures. Les données sont également codées sur trois couches séparées d'environ 5 micromètres par des points. Les trois dimensions vont permettre de modifier la façon dont la lumière se déplace à travers le verre, ainsi en modifiant la polarisation on peut le lire à l'aide d'un microscope optique et d'un polariseur. En fait, il utilise la même technique que sur des lunettes de soleil pour gérer la lumière.

Avec des estimations de stockage de 360To selon les chercheurs, des résistances thermiques jusqu'à 1000°C et avec une durée de vie de plus d'un million d'années, ce support pourrait bien apporter une réponse aux problèmes de stockage de demain.

Le directeur des équipes de l'université de Southampton, Jingyu Zhang insiste d'ailleurs sur le fait que ce procédé de stockage pourrait être d'une grande utilité pour les entreprises disposant de grands volumes d'archives sans avoir à faire des sauvegardes tous les 5 à 10 ans comme l'obsolescence des supports l'oblige aujourd'hui.

Le marché du quartz est donc en train de démarrer et Hitachi a déjà annoncé une commercialisation sous quelques années visant justement des grandes entreprises ou des données culturelles plutôt que le grand public selon Tomiko Kinoshita, porte-parole du laboratoire de recherche de la firme japonaise.<sup>23</sup>

Les limites de ces évolutions sont encore floues et les capacités qu'il sera possible d'atteindre encore difficiles à estimer. Mais il n'est pas impossible de nommer ces

---

23

limites, il s'agit aujourd'hui d'améliorer les capacités de stockage, la vitesse de lecture et d'écriture, et la consommation d'énergie.

## DES LIMITES PHYSIQUES

### L'obsolescence programmée

On entend par obsolescence programmée la planification volontaire de la durée de vie normale pour un produit. Cette méthode permet aux constructeurs de pousser à une consommation extrême. L'un des exemples récents concerne les trois premières générations d'iPod qui ne disposaient pas d'une batterie amovible, sa durée de vie était ainsi limitée à une vingtaine de mois, ils ont évité de justesse un procès en dédommageant les clients. Récemment les problématiques soulevées lors de ces affaires ont amené le Sénat à débattre de ces effets néfastes sur l'environnement comme sur le pouvoir d'achat. En Mars 2013, Jean-Vincent Placé, Sénateur et Président du Groupe écologiste du Sénat recommande différentes propositions pour lutter contre ces méthodes. L'obsolescence programmée est pour lui une aberration et une impasse face aux enjeux économiques et environnementaux.

Auteur d'une proposition de loi sur le sujet<sup>24</sup>, il préconise la définition d'un «*délit d'obsolescence programmée*» pour permettre des actions en justice. Il propose d'allonger la durée minimum du délai de garantie des produits de 6 mois actuellement à 2 ans. Un système «*d'accès aux pièces détachées pendant 10 ans pour la réparation*» et la mise en place d'un bonus/malus sur l'éco-contribution pour valoriser la durée de vie des produits ou encore la mise en place de notices de réparation.

L'allongement de 6 mois à 2 ans du délai d'utilisation de la garantie des produits inciterait les industriels à concevoir des produits durables. Le 13 Février 2014 la nouvelle loi sur la consommation, la loi Hamon, prévoit en effet deux mesures pour lutter contre l'obsolescence programmée.

---

<sup>24</sup>AFP, « La guerre contre «l'obsolescence programmée» est déclarée », Avril 2013 ([http://www.liberation.fr/economie/2013/04/23/la-guerre-contre-l-obsolescence-programmee-est-declaree\\_898359](http://www.liberation.fr/economie/2013/04/23/la-guerre-contre-l-obsolescence-programmee-est-declaree_898359))

### ***Première mesure :***

Le délai de six mois après achat durant lequel le consommateur devait prouver que le dommage s'inverse. C'est désormais aux fabricants de prouver que c'est la faute du consommateur et ce dans les deux ans suivant l'achat de l'appareil.

### ***Deuxième mesure :***

Le fabricant devra désormais indiquer à ses consommateurs la disponibilité des pièces détachées nécessaires à la réparation du produit. Mais la loi ne prévoit pas l'accès à la notice de réparation recommandé par le sénateur.

En revanche le fabricant devra obligatoirement fournir les pièces détachées, aux vendeurs professionnels ou réparateurs (agrés ou non), dans un délai maximum de deux mois.

## **La durée de vie des supports en archive**

Les nouveaux supports ne sont pas encore arrivés en archives et les supports actuels restent peu convainquant dans leurs durées de vie se dégradant spontanément en moyenne au bout de dix ans<sup>25</sup> demandant la mise en place des migrations permanentes et indispensables. La durée des supports peut-être très limitée et même dans de bonnes conditions de conservation le cycle de validité est d'une durée de 2 à 5 ans.

Au départ, les disques durs se sont imposés comme une solution plus durable que les bandes magnétiques car plus performant. Mais la problématique de l'archivage des données a poussé les centres d'archivage à se retourner vers ces solutions magnétiques qui peuvent conserver de façon passive les données sans consommation d'énergie. En contrepartie leurs durées de vie restent limitées et la migration devient la seule solution à la conservation.

Concernant les autres supports comme les mémoires flash que nous avons évoqués, ils disposent également d'une durée de vie estimée à cinq ans. Les disques optiques

---

<sup>25</sup> "Archivage pérenne de données numériques", 8 et 9 octobre 2013 à Paris (UPMC) organisé par le GIS SPADON et le GDR Verres ([http://www.ustverre.fr/site/images/Journees/Workshop\\_GDR\\_Archivage.pdf](http://www.ustverre.fr/site/images/Journees/Workshop_GDR_Archivage.pdf), consulté le 12 Juin 2014)

numériques comme les DVD, les CD et aujourd'hui les Blu-ray pressés ont la meilleure durée de vie, on l'estime entre 20 et 30 ans. Certes ils disposent d'un temps d'accès plus long mais s'imposent néanmoins comme le support le plus fiable. Exception faite des CD enregistrés qui perdent très rapidement leurs données, parfois en moins d'une année.

L'amélioration de la durée de vie en archive passe donc nécessairement par un rafraîchissement du support en menant soit une politique de migration soit d'émulation du support. Comme le DVD avant lui, le CD est très présent dans les archives, surtout pour les enregistrements sonores, selon les constructeurs la longévité d'un CD pourrait être de 75 à 200 ans selon sa fabrication. Avec des tests de vieillissement les CD enregistrables gravés sont plus fragiles que les CD pressés.<sup>26</sup> En laboratoire comme le National Media Lab Test des CD enregistrables résistent en moyenne 5 ans et les pressés de 10 à 25 ans.

Concrètement dans des conditions «*naturelles*», dans le centre d'archivage, il semble que le vernis protecteur de la couche métallique entraîne parfois l'oxydation du disque pour les plus vieux CD conservés, les rendant illisibles plus rapidement.

## **L'altération du matériel**

Pour les disques durs, le risque majeur est la panne mécanique mais des erreurs dans les transferts peuvent aussi rendre l'information illisible ou défectueuse. Les transistors des mémoires flash peuvent être déchargés à cause des rayons cosmiques. Dans la détérioration de ces supports plusieurs facteurs entrent en jeu et il est parfois difficile de les dissocier.

La pellicule de plastique qui recouvre les CDs et les DVD est très sensible à la chaleur et aux surexpositions ultraviolettes de la lumière du Soleil. De façon générale les tests de ces supports ont démontré une forte sensibilité aux réactions d'oxydation par l'eau mais aussi par l'air. C'est d'autant plus vrai pour les supports d'optique utilisables.

---

<sup>26</sup>LALOË Franck, SPITZ Erich « La quête d'un support numérique durable » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p78-82, Novembre 2013, 188p

Pour mieux comprendre et étudier le vieillissement de ces supports plusieurs groupements se sont créés. C'est par exemple le cas du Groupement d'Intérêt Scientifique (GIS) qui réunit près de six organismes de recherche. Ces membres sont à titre d'illustration appelés à participer aux comités de normalisation ISO, AFNOR ou encore AES et contribue à faire évoluer les normes en ce domaine. L'objectif principal c'est avant tout l'amélioration des informations enregistrées sur les Disques Optiques Numériques (DON). L'idée est de pouvoir proposer différentes solutions et méthodes d'archivage satisfaisantes. Le GIS-DON a été renommé début 2012 le GIS-SPADON pour Groupement d'Intérêt Scientifique sur les Supports Pérennes d'Archivage des Données Numériques.

Les résultats de ces réflexions et expérimentation permettent de guider les acteurs du domaine dans des solutions pérennes sur plusieurs décennies ou plusieurs siècles contribuant efficacement à la sauvegarde du patrimoine mémoriel.

L'une de ces expériences a souhaité se pencher justement sur les supports du type DVD et a comparé au DVD Interdisciplinaire la Société Syylex<sup>27</sup> d'autres DVD-classiques. Ce modèle en particulier permet de graver directement dans le verre et n'utilise pas de couche ou de polycarbonate actuellement le point faible des DVD actuel. Au vue de plusieurs expérimentations, il semble que les facteurs d'altération physico-chimiques soient trop multiples pour admettre une validité expérimentale. Pour disposer de donnée il a donc fallu exposer plusieurs DVD à des températures et des taux d'humidité élevés. Lorsque le taux de données erronées dépasse les limites normalisées ou est totalement illisible on considère le DVD mort. Le détail de cette expérimentation est précisé dans l'Annexe 1.

Pour générer ces conditions sévères les DVD sont placés dans une enceinte climatique pendant au maximum 1000 heures par tranches de 250 heures. Soumis à des températures de 90°C avec une humidité relative de 85 %. On démontre alors que DVD GlassMasterDisca résiste nettement mieux que les autres, après près de 1000 heures d'exposition sa lecture ne présente pas d'altération significative alors que les DVD-R tiennent au maximum 250 heures. Il présente un bon potentiel pour

---

<sup>27</sup> Cf : Annexe 1

l'archivage numérique pérenne et pourtant il a fallu réfléchir davantage les tests et les procédures de comparaison pour arriver à ces recommandations.

L'altération des nouveaux médias demande des recherches poussées et des expérimentations. Cela engendre parfois des études techniques contradictoires à ce que prévoit le constructeurs, elles sont donc vitales à la bonne conduite des projets d'archivages.

# L'INFORMATION DEVIENT IMMATERIEL

## Les formats de l'information et leurs évolutions

Pour conserver les documents il est nécessaire de les transformer en données qui seront traduites en bit. Plusieurs solutions existent et des standards ont été mis en place<sup>28</sup>, par exemple concernant le texte, on peut le coder en se référant à la table ASCII pour les caractères latins. Ce format ne prenant pas en compte tous les autres systèmes d'écriture, UNICODE a été développé pour combler ces manques. Dans certains cas des entreprises se basent sur leurs propres schémas de codage pour les caractères comme IBM avec EBCDIC ou Apple construit sur Intel avec un ASCII étendu.

Pour l'archivage il est donc nécessaire de pouvoir effectuer une correspondance entre ces différents systèmes pour maintenir l'intégrité de l'information. Cette manœuvre peut s'avérer difficile surtout lorsqu'il s'agit d'information multilingue par exemple. On pourra alors encapsuler le texte dans une structure logique en codant les langages de façon indépendante des environnements informatiques.

Si l'on prend l'exemple des langages XML ou SGML ou TEX, indépendant du logiciel et du matériel, ils ne sont pourtant utilisés que dans certains domaines comme les éditeurs médicaux, techniques et scientifiques.

Il est donc nécessaire de s'interroger sur les caractéristiques de ces formats d'encodage pour mettre en place une conservation à long terme. Si l'on prend par exemple un document créé sur Word 97, il pourra être conservé pour une certaine durée, on peut alors supposer qu'il sera facile de lire le document tant que le logiciel existe et qu'il effectuera, au besoin, lui-même la conversion. Pourtant rien ne garantit que cette conversion fonctionne qu'aucune perte d'information ne soit faite.

---

<sup>28</sup> BEAUDRY Guylaine, BOISMENU Gérard, « Expertise technique et organisationnelle ; L'archivage », Février 2001 (<http://revues.enssib.fr/titre/6tecnic/5archivage/> consulté le 25 mai 2014)

La mise en place d'un format d'encodage universel permettrait de rendre pérenne le document numérique et c'est davantage le cas avec les formats XML ou SGML qui sont reconnus comme stables. Par exemple le XML n'est pas rattaché à un logiciel particulier et son format libre, en open Access lui garantit cette stabilité et donc une meilleure conservation de l'intégrité du document sur le long terme. Il peut être représenté avec le jeu de caractère proposé par ASCII et donc le format actuellement le plus universel.

Le HTML est une solution, un sous ensemble simplifié de SGML, utilisé comme une base de l'encodage du web et qui garantit, là encore sa pérennité. Malgré l'utilisation de format standard, le document est souvent édité via une application qui le publie dans un format propriétaire. Le secteur de l'édition électronique est bien souvent encore dominé par des codages et des mises en forme propriétaire dont la conversion entrainera peut-être des pertes de données.

La problématique des images soulève aussi des questions. La résolution nécessaire au maintien de l'intégrité de l'image est souvent élevée et difficilement compatible avec les impératifs financiers de l'archivage. Cela peut amener à l'élaboration d'algorithmes qui compresseront l'image en entraînant des pertes d'informations comme avec la norme JPEG. Résultat l'objet diffusé pourra être légèrement différent de l'objet créé. Si les modifications sont considérées comme des modifications de contenu intellectuel il peut être nécessaire d'obtenir une autorisation par le détenteur de la propriété intellectuelle et on parlera de transfert de responsabilité de conservation.

L'évolution de ces formats est donc un autre paramètre important pour la conservation sur de nouveaux supports. De nouveaux langages sont régulièrement élaborés comme lors de la conférence WWDC avec l'annonce par Apple d'un nouveau langage de programmation baptisé Swift<sup>29</sup>.

---

<sup>29</sup> Antoine CROCHET- DAMAIS, "Swift : Apple dévoile le remplaçant d'Objective-C", Juin 2014, (<http://www.journaldunet.com/developpeur/outils/apple-swift-et-sdk-ios8-0614.shtml>, consulté le 2 Juillet 2014)

Ce langage moderne, conçu pour les applications Ios, c'est la mise en place d'une syntaxe plus claire et plus accessible à l'utilisateur lambda. Ce langage devrait ainsi être plus simple à apprendre par les débutants et les lignes de codes à produire sont réduites par 4 parfois par rapport à d'autre langage. La nouveauté c'est aussi la possibilité de voir le résultat d'une modification immédiatement sans avoir besoin de recompiler le code.

## **Les solutions pour préserver l'intégrité d'un document**

Pouvoir maintenir l'intégrité et l'authenticité du document électronique c'est l'enjeu majeur de l'archivage. Et au-delà des problèmes matériels, il y a les risques de pratique malveillante comme l'espionnage industriel et le vol d'informations. Des solutions existent comme le cryptage. Il s'agit d'associer au document une clé qui permet de décrypter un document protégé. Le texte est ainsi codé et illisible et seul le détenteur de la clé pourra le rendre visible. Mais un problème se pose, comment assurer la pérennité de ce système si par exemple l'éditeur de clé venait à cesser ses activités ?

D'autres solutions sont alors envisagées comme le développement à l'initiative du W3C et l'aide un groupe de travail, le Digital Signature Initiative. L'objectif est de mettre au point une syntaxe XML qui représenterait la signature des ressources internet. Elles permettraient de maintenir l'intégrité des données et l'authentification et d'établir un « *Web of trust* ».

Pour garantir l'intégrité d'un document numérique qui se définit par sa nature (fixité, référence, provenance et contexte) il faut que ces informations accompagnent le document tout au long de son cycle de vie, depuis sa mise à disposition par l'auteur, sa publication et sa distribution et enfin son archivage.

## **Les défis de la duplication**

Il est alors important de dupliquer sans erreur et la migration doit s'attacher à conserver l'intégrité de l'information numérique, elle inclut effectivement une migration de support mais ne va pas s'appuyer sur une copie exacte de ces données.

Elle va modifier le format de ces objets numériques et en conservant son contenu intellectuel.

Chaque changement d'environnement informatique a donc des implications légales et la quantité d'information numériques augmentant considérablement chaque jour il est urgent de mettre en place des infrastructures au niveau national et international en charge d'organiser ces opérations de migration. Cela passe souvent inaperçu aux yeux des utilisateurs, pour eux la copie d'un document numérique ne sous-entend pas nécessairement des erreurs.

Si aucune lettre ne s'échappe d'un document texte on ne peut affirmer la même chose pour les supports analogiques. Malgré des précautions chaque copie va introduire un peu de « bruit » qui progressivement s'amplifie au fil des recopies jusqu'à être inexploitable.

Si on arrive à dupliquer aujourd'hui sans erreur c'est avant tout grâce au travail dans les années 1960 de deux mathématiciens américains Irving Reed et Gustave Solomon<sup>30</sup>. Ils réussirent à mettre au point des codes de corrections d'erreurs numériques en sur-échantillonnant l'information. En donnant une redondance à ces informations la machine peut vérifier si elle n'a pas fait d'erreur. Ces algorithmes permettent de reconstituer l'information mais soulèvent le problème de la préservation à long terme des données.

La méthode d'archivage durable la moins onéreuse aujourd'hui consiste à graver des DVD en verre<sup>31</sup>. Enfin pour palier à la rapidité d'évolution des technologies on peut noter l'émergence de l'émulation. Il s'agit de recréer l'environnement informatique d'une plate-forme, déjà très utilisée pour jouer à de vieux jeux en ligne, cette technologie pourrait s'adapter à la lecture de vieux CD nécessitant des systèmes d'exploitation obsolètes. Cette solution tient aussi son potentiel car elle est peu coûteuse face aux sommes allouées à la migration de données.

---

<sup>30</sup> LALOË Franck, SPITZ Erich « La quête d'un support numérique durable » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p78-82, Novembre 2013, 188p

<sup>31</sup> Ibid.

# LES STRATEGIES ARCHIVISTIQUES

---

## REDEFINIR LE DOCUMENT

### Du papier au numérique

Il est plus complexe de s'attacher à définir le contenu intellectuel d'un document numérique, le concept est complexe et on peut définir plus niveaux d'abstraction. Si l'on prend le niveau le plus bas, l'information numérique n'est qu'une succession de 0 et 1, les bits qui en créant des chaînes se distingue les uns des autres créant un langage compréhensible par une machine. Il s'agit donc de conserver la succession de bits dans le bon ordre et de mettre en place un système chargé de vérifier cet ordre. Mais cela limite et ne rend pas compte des problèmes de formats, de la structure de la donnée sans parler des plates formes matérielles ou logicielles. Dans un sens on a délégué une partie de notre contrôle à la machine, une autonomie profondément différente de celle conférée par le document matériel. La décontextualisation du document est une opération qui modifie la nature du document la rendant difficile à définir. Le milieu de vie du document n'est plus lié à des limites physiques mais est codé sous différentes formes et accessible via des interfaces de lectures.

Selon les propos de Robert<sup>32</sup> la dématérialisation se situerait davantage dans la dissociation des activités de création en jeu. Selon Gardey<sup>33</sup> on peut découper le lieu de la frappe de celle de l'affichage et de la mémorisation, on distingue alors une similitude avec l'ordinateur avec l'espace de saisie (le clavier), d'affichage (l'écran) et d'enregistrement (la mémoire). Les trois activités restent séparées lorsqu'on utilise des systèmes électroniques de lecture, comme lorsqu'on écrit un SMS.

---

<sup>32</sup>Pascal ROBERT, Nathalie PINÈDE, « Communication & Organisation », n° 42, p191-202, 2012, 280 p  
( <http://communicationorganisation.revues.org/3948> , consulté le 25 Mars 2014)

<sup>33</sup> D Gardey., Ecrire, compter, classer : comment une révolution de papier a transformé les sociétés contemporaines, 1800-1940, Paris, La Découverte, 2008.

Deux types de document sont alors distingués, celui qui contient les trois fonctions comme un document matériel et l'autre qui articule dans un système les trois fonctions. Les données étant alors sous forme codée, binaire, on a une rupture qui se crée entre le support et le message. L'utilisateur perd son autonomie à accéder au document et l'insère même dans un processus de contrôle, implicite ou explicite. Les problèmes de formats et de supports mettent en avant les enjeux sociaux de la numérisation des données. L'objet numérique perd son apparence, son support matériel qui donne toujours la possibilité de définir un contexte, une histoire, aussi réduite soit-elle.

## **La sélection de l'archive, un processus intellectuel**

La sélection des documents à archiver ou non est un facteur essentiel dans le développement d'une politique d'archivage, aujourd'hui les procédures ont évolué et peuvent varier d'un pays à l'autre. C'est une démarche universelle qui place l'archiviste dans une position délicate où ces choix engagent l'avenir des futures recherches historiques<sup>34</sup> et dans un sens ce que les prochaines générations connaîtront du passé. Quand bien même nous aurions les moyens techniques et surtout économiques de stocker toute l'information, tout n'est pas à conserver de façon pérenne. Il y'a donc un aspect matériel mais aussi intellectuel dans le processus d'évaluation du document. La possibilité de mise en relation, la maniabilité et surtout le degré d'agrégation de l'information font du support numérique un médium bien particulier<sup>35</sup>. Stocker moins mais stocker mieux revient à réfléchir aux nouvelles façons de trier les archives.

L'évaluation va ainsi reposer sur la valeur de témoignage, liée à la place qu'il occupe dans la hiérarchie du créateur, et sur la valeur de l'information, établi à partir de critères non absolus comme le temps, le lieu, l'individu.

---

<sup>34</sup>Eric KETELAAR, « (Dé) Construire l'archive Matériaux pour l'histoire de notre temps » ; « Matériaux pour l'histoire de notre temps », N° 82, p65-70, 2006, p138, (<http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2-page-65.htm>, consulté le 15 mai 2014)

<sup>35</sup> Rosine CLEYET-MICHAUD, « La sélection dans les archives contemporaines », 2001 (<http://www.archivesdefrance.culture.gouv.fr/gerer/records-management-et-collecte/principes/>, consulté le 15 Mars 2014)

Une bonne évaluation est conditionnée par l'intervention de l'archiviste dans des processus comme celui de la création jusqu'à la veille de l'obsolescence de ces supports et formats. Même si l'on tend à appliquer les principes de l'archivage dans le numérique, la place de l'archiviste est modifiée fortement.

Ce type d'archive contemporaine participe à la réflexion dans le domaine du tri et les procédures mise en place depuis 2004 sont issues de ce type de réflexion. Mais concrètement un texte réglementaire prend du temps, jusqu'à deux ans de procédure sont nécessaires. On peut également se dire que les procédures de tri sont les mêmes pour les archives numériques, que les principes devraient être identiques mais ce support virtuel change en réalité beaucoup de choses.

Néanmoins, chaque année des instructions sont distribuées pour les archives papier mais aussi pour les archives électroniques et constituent un corpus méthodologique qui permet aussi une meilleure transparence de ces politiques pour le citoyen. Elles couvrent des champs variés et parfois non couverts comme le traitement et la conservation des archives des délégations du CNRS et des archives des unités de recherche et de service traitées dans l'instruction du 15 janvier 2007 alors que le travail de réflexion avait commencé depuis juin 2005 dans le but d'améliorer la collecte de ces archives publiques. Dans certains cas elles peuvent reprendre une instruction antérieure et faire des reprises plus en profondeur. Les processus de sélection sont en eux-mêmes des processus créatifs car chaque classement nécessite un choix, une catégorisation de l'information.

Ces classifications pratiquées par les archivistes apportent aussi une information le contexte original dans lequel l'information a été collectée. C'est une stratification du classement dont l'analyse archéologique permet de mettre en avant les méthodes mais aussi les supports utilisés.

## **La nature du produit numérique**

L'un des fondements<sup>36</sup>, c'est de définir le document lui-même, il faut le décrire et associer des informations au document numérique qui permettront, malgré les évolutions technologiques, de créer comme une petite carte de visite pour chacun. On utilise pour cela des métadonnées qui s'organisent autour de 4 aspects essentiels qui composent la notion d'intégrité du document. Assurant une bonne conservation et de facilité pour son exploitation.<sup>37</sup>

### ***Le Contexte :***

La conservation d'un lien dans un document pose problème car l'hébergement du document ne comprend pas l'hébergement de tous les documents annexes et cités. Doit-on conserver la page web en même temps que le document pour assurer sa pérennité ? L'archivage du web est ainsi une problématique émergente dans ce contexte. L'environnement matériel et logiciel utilisé pour sa création peut engendrer des configurations spécifiques ou non. L'information numérique doit donc être associée à un contexte technique, pour son utilisation, les documents auxquels il est relié et son interface de lecture.

### ***La Provenance :***

La provenance est une notion légèrement différente et un des concepts centraux de l'archivage moderne. Il est préconisé d'enregistrer l'objet et l'historique de ses modifications internes ou techniques qui accompagnent sa vie. Cet aspect est connecté à celui de la fixité. C'est un des concepts centraux de l'archivage moderne : on enregistre l'origine de l'objet ainsi que tout l'historique de sa conservation et des environnements techniques qui ont accompagné sa vie. Cette chaîne de conservation permet de connaître le contexte, point important de la garantie de l'intégrité, et ainsi de s'assurer de l'authenticité de l'objet.

---

<sup>36</sup> LUPOVICI, Catherine, « Les stratégies de gestion et de conservation préventive des documents électroniques », Bulletin des Bibliothèques de France, N°4, p43-54, Avril 2000 (<http://bbf.enssib.fr/consulter/bbf-2000-04-0043-004>, consulté le 20 novembre 2013)

<sup>37</sup> LUPOVICI, Catherine, « Les stratégies de gestion et de conservation préventive des documents électroniques », Bulletin des Bibliothèques de France, N°4, p43-54, Avril 2000 (<http://bbf.enssib.fr/consulter/bbf-2000-04-0043-004>, consulté le 20 novembre 2013)

### ***La Fixité :***

Pour garantir la fixité on « *marque* » les versions des objets, cela peut se faire via un filigrane électronique. Ces méthodes de marquages sont en plein développement et seront normalisées dans les années à venir. Soit on met en place un suivi de l'évolution du document, en d'autres termes on enregistre l'historique des modifications soit on effectue une prise de vue instantanée de l'intégralité de la base de données à des moments donnés.

### ***La Référence :***

Mais la préservation du document et de ces stades évolutifs ne suffit pas à le rendre intègre, pour cela il faut qu'il dispose en plus d'une référence univoque. On dispose de plusieurs solutions combinables ou non telles les bibliographies ou les catalogues, mais dans un espace numérique le lien doit être aussi actif quel que soit la localisation du document sur le réseau.

Les systèmes traditionnels se sont déjà adaptés en ajoutant par exemple au format MARC, qui sert à décrire les documents, le champ 856 qui permet d'inclure un lien. L'autre moyen pour obtenir des références c'est d'inclure ces références à la source d'une partie distincte de l'information numérique, par exemple le format TEI (Text Encoding Initiative) va contenir les informations descriptives dans l'en-tête du document, c'est le même principe qu'avec les métadonnées du format HTML. Les URL (Uniform Resource Locators) qui servent à encodés en HTML les pages web vont servir de code de localisation de l'objet numérique sur le réseau Internet. Mais même avec ces solutions, des problèmes se posent, dans le premier cas l'URL n'est pas forcément persistant et pour la seconde les informations codées ne sont pas toujours de qualité suffisante.

## DES APPROCHES DE CONSERVATION EN EVOLUTION

### Fonctionnement actuel, quels projets?

Des projets de pérennité voient le jour, comme celui d'Internet Archives qui utilise dans la banlieue californienne les archives du XXème siècle en cas de catastrophe numérique. Chaque semaine plus de 20.000 nouveaux livres arrivent dans cet entrepôt, souvent des dons d'universités ou de bibliothèques. A l'initiative de ce projet Brewster Kahle qui fait notamment fortune en vendant en 1999 une société de traitement de données à Amazon.<sup>38</sup>

*«Nous devons conserver le passé à mesure que nous inventons le futur. Si la bibliothèque d'Alexandrie avait fait une copie de chaque livre et l'avait envoyée en Inde ou en Chine, nous aurions conservé le travail d'Aristote, les œuvres d'Euripide. Une seule copie à un seul endroit ne suffit pas.»* déclare-t-il.

### *Le Cloud*

Les documents se divisaient avant l'ère du numérique en deux catégories, ceux qu'il fallait stocker et ceux qu'il fallait garder accessibles. Aujourd'hui on veut **en** conserver un accès rapide aux documents même si celui-ci est archivé. C'est une véritable problématique pour les entreprises et les institutions et même si les dispositifs existant sont nombreux, peu s'adaptent aux besoins de ces structures.

Elles se tournent alors vers l'hébergement massif sur des serveurs. Nécessairement en louant des espaces ou en mettant en place leurs propres salles serveurs, des solutions moins coûteuses que l'entretien d'entrepôts d'archivages.

Ces espaces virtuels de stockage sont appelés le Cloud, les documents peuvent y être suivis, marqués, tracés et retrouvés<sup>39</sup> mais pas seulement. Ce type d'hébergement permet de dématérialiser les données mais aussi des applications. Concrètement, c'est un ensemble de matériel raccordé en réseau qui fournit un service que les individus peuvent exploiter depuis n'importe quelle localisation dans le monde. Le nuage est le symbole utilisé pour représenter l'Internet dans les

---

<sup>38</sup>Christelle DI PIETRO, « Une arche pour sauver les livres imprimés en Californie » Mars 2012 <http://www.enssib.fr/en/breves/2012/03/05/une-arche-pour-sauver-les-livres-imprimés-en-californie>

<sup>39</sup>Article Wikipédia : [http://fr.wikipedia.org/wiki/Cloud\\_computing](http://fr.wikipedia.org/wiki/Cloud_computing)

diagrammes des réseaux informatiques, d'où l'origine du nom. C'est le basculement d'une tendance, on obtient de la puissance non plus grâce à son propre ordinateur mais en utilisant les ressources de machines distantes mises à disposition par un prestataire.

Dès les années 2000 apparaissent les premiers hébergeurs Web capables de stocker les applications dans leurs locaux et les premiers outils collaboratifs comme les CRM ou plutôt simplement les boîtes de messageries se développent, c'est ce qu'on appelle le web 2.0. La généralisation de l'accès à Internet a permis le développement massif du Cloud. La baisse des coûts des supports de stockage du type disque dur et l'amélioration des technologies permettent aux hébergeurs de proposer des tarifs de plus en plus abordables.

La mise en réseau de ces données entraîne l'implication dans leurs processus de gestion d'une multiplicité d'intervenants comme le fournisseur internet ou le prestataire d'hébergement. On considère aujourd'hui les entreprises comme Amazon ou Google comme les leaders de l'émergence du Cloud, suivies par les grands éditeurs de logiciel comme Microsoft et Oracle qui ont rapidement mis en place également des offres de Cloud computing.

L'accroissement de ce système entraîne le développement des centres de données (data-center en anglais) et entraîne l'agrandissement des infrastructures et des technologies comme la bande passante ou l'achat de nouveaux serveurs pour faire face à la demande. Par conséquent un comparatif s'impose avant de faire son choix. En prenant bien en compte l'implication des coûts directs et cachés que peut engendrer le Cloud. Sachant que les durées d'engagements varient en moyenne de 2 à 3 ans selon les prestataires. . A court terme, ce fonctionnement permet aux particuliers comme aux entreprises de réaliser des économies mais sur le long terme c'est une autre histoire et le coût peut s'avérer supérieur en comparaison d'un hébergement en interne en fonction du mode d'utilisation et de la durée de vie de l'application, sans parler de l'impact écologique de l'expansion des data-center. La

qualité du réseau et de la connexion deviennent essentielles pour le client sachant qu'aucun fournisseur ne peut garantir une disponibilité à 100 %.

Plusieurs inconvénients demeurent comme les risques liés à la sécurité car les connexions passent par un réseau et peuvent être interceptées et utilisées par un tiers. L'éventualité que le prestataire ait à cesser ses activités doit engendrer la mise en place de clause de restitution des données et la localisation géographique prise en compte. La législation étant différente d'un pays à l'autre, la confidentialité des données n'est pas la même partout, des lois différentes vont s'appliquer en fonction du pays où se trouve le data-center.

Pour Richard Stallman<sup>40</sup>, à l'origine du projet GNU, le Cloud « *est un piège* » car les utilisateurs perdent le contrôle de leurs applications. C'est pour lui un concept publicitaire sans intérêt, rejoignant les critiques exprimées par Larry Ellison, fondateur d'Oracle, selon lequel il s'agit d'un phénomène de mode.

Malgré ces solutions il est évident que des augmentations des besoins de stockage vont nécessiter le développement de technologies de conservation afin de réduire les coûts de migration et de les rendre plus supportables économiquement. A titre d'illustration 5 000 euros c'est le coût annuel moyen de la préservation à long terme d'un téraoctet de données numérisées.<sup>41</sup>

La mission fondamentale des institutions comme les bibliothèques et plus particulièrement les bibliothèques de recherches c'est la préservation de ce patrimoine documentaire en maintenant une accessibilité et une visibilité de ces données. C'est ainsi que des politiques documentaires spécifiques ont été définies par les bibliothèques, déterminant des priorités parmi les natures de données à préserver. Toutefois, la question de la préservation à long terme des données numériques reste largement ignorée des porteurs de projets de numérisation. Et l'ampleur des coûts fait peser un risque fort sur la pérennité des résultats de ces projets.

---

<sup>40</sup> Article Wikipédia : [http://fr.wikipedia.org/wiki/Cloud\\_computing](http://fr.wikipedia.org/wiki/Cloud_computing)

<sup>41</sup> BLIN Frédéric, « La numérisation des bibliothèques » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p48-51, Novembre 2013, 188p

## *Les nouvelles stratégies en Bibliothèques*

Depuis plusieurs années déjà des organismes se sont intéressés à la préservation des informations produites par les institutions et l'attention s'est très rapidement portée vers la détérioration des supports qui étaient alors souvent magnétiques. Dans les années 90, les archivistes mettent en place un rafraîchissement du support en créant des copies ou en migrant vers d'autres supports. Mais cette solution n'est efficace que lorsqu'on encode les données sous un format qui pourra demeurer indépendant des supports et de leurs évolutions successives. Il faudra tout de même que le logiciel, qui aura permis la création du code, soit maintenu ou mis à jour en assurant la comptabilité avec les précédentes versions. Des projets sont effectivement en place comme la mise au point de LOCKSS (Lots of Copies Keeps Stuff Safe) qui permet aux bibliothèques d'héberger des publications par exemple sur les disques durs locaux et de confronter ces versions à celle en ligne pour éviter une détérioration des contenus. C'est une approche un peu récente très différente des stratégies de sauvegarde classique. Les stratégies classiques de conservation vont donc en fait avoir deux approches, passive et active.

La stratégie passive est celle pratiquée depuis toujours c'est la conservation sur des supports lisibles, des parchemins à l'analogique le fonctionnement reste le même. Elle demande tout de même un investissement initial, comme l'achat du support.

La stratégie active consiste à inscrire sur un support durable les données et à les stocker dans un endroit adapté. On met alors en place une veille des supports et de leurs évolutions qui permet de prévoir les migrations futures. L'investissement nécessaire n'est pas aussi fort que pour la stratégie passive car la plupart des données vont être enregistrées dans des centres de données. Mais avec ce système le propriétaire de l'information est obligé de faire confiance à un prestataire.

Les récents cas comme l'affaire Snowden participent à la sensibilisation sur ce sujet, mettant en lumière des lois comme le « *USA Patriot Act* » qui permet notamment à la NSA (National Security Agency), d'avoir un accès illimité aux données gérées ou hébergées par les acteurs du Cloud américains et leurs filiales étrangères, sans en informer leurs clients.

Les utilisateurs sont alors à la recherche de plus en plus de transparence sur ces législations et de la protection de leurs informations, plus particulièrement lorsqu'il s'agit d'hébergement européen. Pour certains elles sont plus à l'abri dans un coffre-fort sur place que dans des data-center à des dizaines de kilomètres. C'est surtout le coût qui est déterminant. Mettre en place une stratégie de conservation pérenne peut s'avérer onéreux, à titre d'exemple un DVD de cinq giga-octets gravés dans le verre peut coûter jusqu'à 150 euros, mais le lieu de stockage sera moins cher donc plus économique.

Dans la stratégie active certains coûts vont être abordables comme le coût en temps et matériel d'une migration, mais il faut ajouter à cela le salaire des informaticiens, de locaux et du matériel adéquat et sa consommation électrique.

Contrairement au coût de la stratégie passive ces coûts augmentent chaque année de façon linéaire, c'est en termes de délais que l'une ou l'autre des stratégies pourra être pertinente ou non. Certains secteurs connaissent en effet des difficultés financières comme les hôpitaux dont les impératifs d'archivages, en durée notamment, sont fort, au minimum 20 ans<sup>42</sup>, donc coûteux et problématiques.

Pourtant les enjeux économiques ne sont pas forcément incompatibles avec des enjeux qui préoccupent de plus en plus d'acteurs comme la protection du patrimoine historique que représentent les archives. Prenons le cas des archives africaines, dont le patrimoine audiovisuel est d'un grand intérêt historique mais malheureusement souvent dans un état de dégradation très avancé. Pourtant elles détiennent un fort potentiel économique.

Pour exploiter le potentiel commercial un projet a vu le jour en 2014 baptisé « *Capital numérique* ». Lancé par l'Organisation internationale de la Francophonie avec pas moins de 7 partenaires africains et européens, ce projet vise à mettre sur le marché du numérique des œuvres numérisées, issues du Fond Francophone de production audiovisuelle du Sud et couvrant près de 25 ans, avec l'aide la BNF. Ces programmes remis en circulation sont et seront accessibles via un portail de

---

<sup>42</sup> BLIN Frédéric, « La numérisation des bibliothèques » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, Novembre 2013, pp. 48-51.

vidéo à la demande «africafilms.tv» grâce à l'un des premiers fournisseurs de programme d'Afrique francophone « Côte Ouest Audiovisuel ». L'intérêt de ce genre de pratique c'est que l'on constate bien que la logique patrimoniale et la logique économique ne s'opposent pas forcément dans le domaine du numérique. Au contraire elles fonctionnent l'une avec l'autre et permettent de préserver de façon presque autonome la mémoire collective.

Au milieu de cette émergence de projets on trouve Google. En 2004 l'entreprise lance son projet Google Books qui vise à numériser et diffuser l'ensemble des livres du monde. Comme le souligne Frédéric Blin, « *Pour la première fois, une entreprise privée disposait du potentiel de contrôle sur le patrimoine documentaire mondial, concurrençant directement les bibliothèques.* »

Ces dernières ont donc dû réagir et se positionner, comme la BNF Bibliothèque nationale de France. Sa volonté forte d'affirmer sa légitimité, comme gardien du patrimoine français informationnel, explique sa volonté de constituer une masse documentaire en passant d'une logique de numérisation sélective à une logique de numérisation systématique.

## **Des outils de conception**

### ***ORI-OAI, le chaînon manquant***

Conçu pour gérer des archives institutionnelles<sup>43</sup>, ORI-OAI est un outil qui permet de valoriser tous types de ressources constituant le patrimoine numérique des universités. Plusieurs projets sont ainsi en cours dans plusieurs universités et sont tous basés sur ce modèle, l'objectif est de tester la mise en place de système d'archives pour adapter les méthodes de préservation. La finalité serait de pouvoir organiser une normalisation autour des métadonnées de gestion techniques des documents numériques.

Cette valorisation du patrimoine numérique telle qu'elle est décrite souvent dans ces politiques de conservation c'est l'ensemble des profits dont pourra bénéficier

---

<sup>43</sup> <http://www.ori-oai.org/>

le projet et la masse de connaissances produite, comme avec le travail des chercheurs et les enseignants. Concrètement on crée une archive institutionnelle qui regroupera tous les documents importants et donc à conserver sur le long terme mais dont on a besoin au quotidien, répondant ainsi aux usages de ces institutions.

Satisfaisant la problématique de normalisation et préservation, le logiciel permet aussi de mettre en place un réseau inter-universités avec des portails qui, là encore, permettent d'étendre considérablement l'exploitation des données tout en gérant les paramètres d'accès à ces informations.

Concrètement cet outil permet de faire le lien avec les autres outils déjà en place en bibliothèque, et il articule même son fonctionnement sur ces outils tout en les valorisant. Ces possibilités d'échanger facilitent le travail collaboratif comme pour le référencement et l'indexation toujours en respectant des standards. ORI-OAI est donc généralement utilisé pour cataloguer les productions numériques de son établissement ou pour agréger des ressources d'autres universitaires.

S'il est capable de s'adapter si bien au besoin particulier de chaque structure c'est parce que son fonctionnement repose sur huit composants paramétrables, optionnels et indépendants, donc pouvant être activés ou non. L'une d'elle c'est justement l'interaction avec les autres composants qui, très souple, permet notamment de communiquer avec les logiciels extérieurs. Une bibliothèque pourra, si elle le souhaite, utiliser son propre interface de recherche et donc ne pas utiliser l'option *ORI-OAIsearch* mais de faire tout de même un renvoi vers la fonctionnalité d'index de recherche *ORI-OAI-indexing*.

Début 2011 par exemple, l'Université de Lorraine a lancé une plate-forme basée sur ORI-OAI pour faciliter la diffusion de ses thèses et mémoires en version numérique. Son nom est PETALE (Publications et travaux académiques de Lorraine). Les documents étaient déjà disponibles en ligne mais les besoins d'une plateforme unique se sont imposés. Depuis, sa création a permis la diffusion commune pour les quatre universités de Lorraine. En permettant une large diffusion de ses productions scientifiques, sur les moteurs de recherche plus

particulièrement, les universitaires lorrains bénéficient d'une forte valorisation de leur travail et d'une nette visibilité.

PETALE s'inscrit aussi largement dans le mouvement de l'Open Access sur lequel nous reviendrons un peu plus loin, en accordant un accès libre à l'intégralité des thèses et des mémoires.

### *Open Archival Information System<sup>44</sup>*

Pour encadrer ces projets d'archivage numérique et leurs adaptations sur de nouveaux supports, des normes sont également mise en place. L'Open Archival Information System ou OAIS (Système ouvert d'archivage d'information) est un modèle conceptuel à l'initiative du Consultative Committee for Space Data Systems, et a été enregistré comme norme ISO sous la référence 14721:2003, elle est destinée à la gestion, à l'archivage et à la préservation à long terme de documents numériques. Le long terme est défini comme suffisamment long pour être soumis à l'impact des évolutions technologiques.

Le modèle OAIS constitue donc une référence décrivant dans les grandes lignes les fonctions, les responsabilités et l'organisation d'un système qui voudrait préserver de l'information, en particulier des données numériques, sur le long terme, pour en garantir l'accès à une communauté d'utilisateurs identifiés.

Le modèle OAIS définit le fonctionnement de l'archive, c'est-à-dire les différentes entités fonctionnelles et la façon dont elles s'articulent entre elles. Les différentes entités sont :

Les entrées : reçoivent les paquets d'information à verser et les transmettent au stockage ;

Le stockage : stocke et sauvegarde les paquets d'information archivés ;

La gestion des données : met à disposition toutes les informations utiles au fonctionnement de l'archive ;

L'administration : pilote le système ;

---

<sup>44</sup> Article Wikipédia : [http://fr.wikipedia.org/wiki/Open\\_Archival\\_Information\\_System](http://fr.wikipedia.org/wiki/Open_Archival_Information_System)

La planification de la préservation : assure une veille technologique et propose les évolutions et les stratégies pour prévenir l'obsolescence ;

L'accès : communique les paquets d'information diffusés aux utilisateurs.

Chacune de ces entités est associée à des rôles et des fonctions et doit communiquer avec les autres sous la forme de flux de données.

## **Exemple du fonctionnement de la plate-forme PAC**

Il existe cependant d'autres projets en bibliothèque que l'on peut prendre en exemple et qui ont à cœur de s'adapter aux nouveaux supports de l'information.

CEDARS (CURL Exemplars in Digital Archives) est un projet qui nous vient du Royaume-Uni à l'initiative du CURL, le Consortium of University Research Libraries et subventionné dans le cadre du programme eLib (Electronic Libraries Programme). Ici aussi c'est une association de plusieurs bibliothèques et notamment Oxford, qui ont permis de mettre en place une expérience d'archivage pérenne pour le patrimoine numérique. Ce projet explore les relations avec les éditeurs sur le droit d'archiver notamment sur l'aspect Management.

Le CINES s'est vu confier en 2004 par le ministère de l'Enseignement supérieur et de la recherche la mise en place d'un service dédié à l'archivage pérenne des documents numérique du patrimoine scientifique<sup>45</sup>.

A plusieurs reprises ces décisions ont été renforcées notamment en 2006, où le CINES se voit également confier la mission de conservation des thèses électroniques françaises. Mais aussi en 2007 la gestion de l'archivage du programme Persée et plus récemment en mars 2014, la modification du décret de 1999 du Centre informatique national de l'enseignement supérieur confirme la charge au CINES de mettre en place un archivage pérenne.

---

<sup>45</sup> <https://www.cines.fr/archivage/une-plateforme-en-production/des-reponses-techniques/>

Dès 2004 la conception d'une plate-forme d'archivage au Cines commence et trois ans plus tard une première version (PAC v1.0) est mise en service. L'objectif est d'intégrer STAR, une plate-forme d'archivage développée auparavant par l'ABES (Agence des Bibliothèques de l'Enseignement Supérieur) qui sert au dépôt et au référencement des thèses électroniques. Mais cette première version peine à gérer des grands volumes de stockage et dès 2008 PAC v2.0 voit le jour. Basé cette fois sur le logiciel Arcsys associé à un module conçu en interne en java et utilisant une base de données en MySQL. Tous les documents reconnus comme d'intérêt national peuvent y être hébergés. Son approche est toujours généraliste pour éviter le traitement au cas par cas en mutualisant la plate-forme pour tous les projets d'archives. Conforme à la norme OAIS, PAC s'attache à respecter les pratiques et les normes actuelles de l'archivage. Aujourd'hui encore la plate-forme reste à l'écoute des évolutions et s'adapte régulièrement

Un nouveau projet a également vu le jour en 2011, le projet EUDAT<sup>46</sup>, prévu pour trois ans. Il s'agit cette fois d'un projet européen, à l'initiative des communautés scientifiques, pour répondre aux besoins des chercheurs à l'information. Dans ce but 25 partenaires se sont réunis et s'attachent à recenser les besoins et exigences des communautés de recherche. Le but est de pouvoir sécuriser et déployer une infrastructure persistante mais surtout commune des données numériques, la « *Collaborative Data Infrastructure* » (CDI). Le CINES fait partie des organismes mondialement reconnus dans le domaine de l'archivage pérenne et fait figure d'exemple dans ses projets.

L'application EUDAT permettra de fournir des services communs comme une réplique sécurisée des données, le transfert vers des centres de traitement ou encore un catalogue de métadonnées

La stratégie d'EUDAT est de faire travailler des acteurs différents avec les utilisateurs finaux pour élaborer un service au plus proche des besoins.

---

<sup>46</sup> <https://www.cines.fr/archivage/le-projet-europeen-eudat/>

Le 28-30 Octobre 2013 la seconde conférence EUDAT s'est tenue à Rome et a permis de réaffirmer la vision de EUDAT<sup>47</sup>, en effet Kimmo Koski, Directeur Général du Centre SCC-IT, coordinateur du projet Science & EUDAT, a résumé cette stratégie «*La science est mondiale et les e-infrastructures et les services connexes devraient donc l'être également. [...] Les services doivent être axés sur l'utilisateur, nous avons besoin d'établir une confiance entre les chercheurs et les fournisseurs d'e-infrastructures.*».

### ***Fonctionnement des échanges entre CINES et producteur/service versant***<sup>48</sup>

Les installations matérielles sont surveillées par des experts qui, en parallèle de leurs activités de veilleurs technologiques, effectuent des contrôles d'intégrité sur le matériel et les technologies utilisés, et des contrôles métiers qui comparent l'emprunte effectuée à l'archivage et celle issue des sauvegardes. Il n'est jamais possible de supprimer tous les risques et il faut donc les évaluer en détail. L'objectif est d'évaluer et de gérer ces différents scénarii pour être le plus apte à les maîtriser. Il s'agit donc d'évaluation. Il faut évaluer le risque acceptable et mettre en place des solutions adaptées. Les risques possibles concernent généralement un logiciel ou matériel manquant, la perte du contenu d'un fichier et évidemment la détérioration du support.

Pour évaluer ces risques, le service PAC dispose d'une méthodologie concernant l'audit de dépôt d'archive électronique et y à former ses équipes, DRAMBORA.

Cette étude a permis d'identifier au sein du projet 38 risques potentiels, détaillant leurs probabilités et le temps d'impacts et a permis de mettre en place en 2009 un « *Plan de gestion des risques* » a été mis en place et est remis à jour chaque semestre.

Mais d'autres problématiques stratégiques sont soulevées avec ce travail, le CINES priorise le traitement préventif des risques alors que d'autres structures comme la Bibliothèque nationale de France priorise les curatifs.

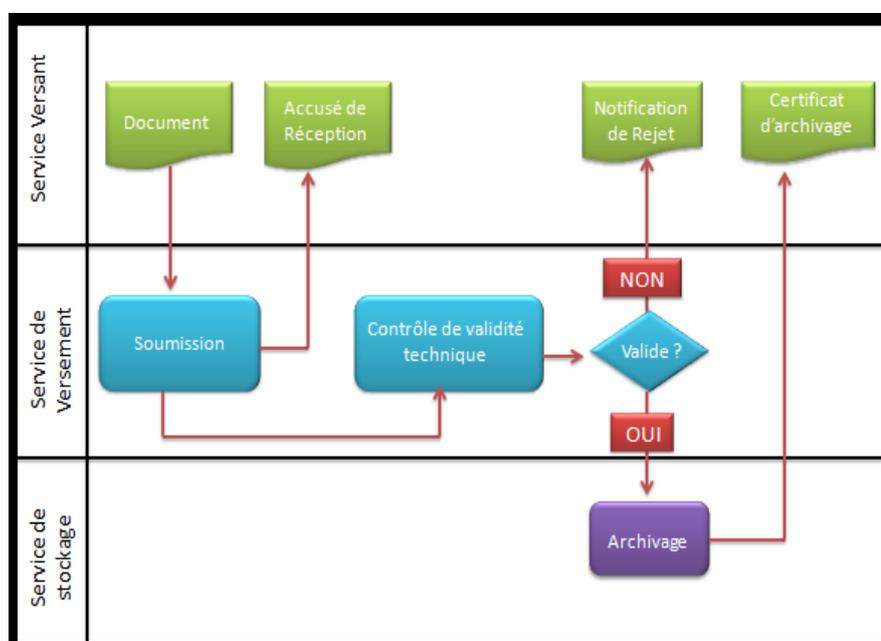
---

<sup>47</sup> <http://www.eudat.eu/2nd-conference>

<sup>48</sup> <https://www.cines.fr/archivage/une-plateforme-en-production/des-reponses-techniques/>

### ***Fonctionnement concret de PAC:***

Concrètement lorsqu'un service versant contact le service PAC du CINES, on élabore tout d'abord un plan de transfert qui permet de définir tous les composants du projet. On doit identifier les données à pérenniser, en estimer la volumétrie et les formats choisis, se conformer à la norme légale mais aussi des aspects plus humains avec la formation du personnel du service versant. Techniquement le service verse alors ses données via le réseau ou sur un support amovible et placé dans un répertoire, c'est ce qu'on appelle un SIP (paquet d'information soumis).



Une fois transférée PAC procède à une série de contrôles de validité technique des données transférées comme la conformité des formats. Si le contrôle est positif, le versement se fait dans un serveur de stockage. Un certificat d'archivage est alors automatiquement transmis avec un identifiant unique et pérenne ainsi que les métadonnées. Si le contrôle est négatif, une notification est envoyée au service versant et le paquet d'archives retourné, pour informer de cette non-conformité.

La pérennisation d'un document ne dépend pas uniquement de questions de stratégie, d'organisation ou de format de fichier, il s'agit de décrire suffisamment pour assurer sa traçabilité. Les métadonnées ont donc un rôle important surtout

lorsque l'on veut traiter des projets d'archives très différents et pérenniser sa localisation. PAC associe du coup plusieurs jeux de métadonnées. Elles s'inscrivent dans une stratégie où les documents archivés doivent fournir suffisamment d'information sur leurs contextes si l'on veut garder leur contenu accessible.

Directement inspiré de l'OAIS qui associe un PDI, des informations de pérennisation du document, à l'archiva, PAC associe des informations sur le projet d'archives en lui-même. Concrètement ces métadonnées sont regroupées dans un fichier XML nommée PDDL (Project Preservation Description Information), et archivé lui aussi dans la plate-forme.

L'ensemble de ces métadonnées est organisé sous forme d'un fichier sip.xml qui permet à PAC de vérifier lors du versement la validité de ces données. Une fois validé le fichier est converti en aip.xml à qui l'on rajoute alors des métadonnées pour l'archivage à proprement parlé, à savoir un identifiant unique et surtout pérenne, sa date d'archivage, les empreintes numériques des fichiers et le fameux identifiant du projet d'archives. Ces fichiers aip.xml vont être conservés au même emplacement que le fichier qu'il accompagne et répertorié dans une base de données pour faciliter son indexation.

La particularité de PAC c'est qu'elle s'attache à perfectionner son système de métadonnées et pour cela le CINES a mis en place un système pour préserver ces spécificités : les métadonnées métiers.

Dans cet objectif le CINES incite les services versant à joindre aux documents à archiver un fichier de métadonnées « *métier* » s'il existe. Ces métadonnées sont transmises sous format XML mais doivent indiquer dans une DTD ou un schéma XSD la structure du fichier et la définition des balises.

## **CADRE, LIMITES ET RECOMMANDATIONS POUR LE NUMERIQUE :**

### **Le contexte législatif et réglementaire**

Les institutions techniques et scientifiques produisent depuis près de quarante ans des données numériques en raison de la complexité des calculs informatiques produits, mais les administrations et les entreprises restent souvent fidèles au papier pour des nombreux aspects.

Et même si les vingt dernières années ont vu augmenter l'informatisation des procédures et que dans l'administration les bases de données se sont considérablement développées et sophistiquées le papier reste souvent unificateur.

Il a toutefois été nécessaire d'encadrer ces processus de numérisation. Plusieurs lois sont alors créées comme la loi du 13 mars 2000<sup>49</sup> qui marque un bouleversement dans le traitement du numérique, et portent sur l'adaptation des droits de la preuve et l'applique au numérique. Cette loi permet ainsi de mettre fin à la domination du papier comme preuve et modifie l'article associé 1316-1 du code civil pour permettre au numérique de bénéficier de la même valeur probatoire.

C'est seulement à partir de 2004 que l'archivage numérique va être, en France, inscrit sur le plan stratégique de l'administration électronique. Cette prise de conscience était un enjeu majeur des acteurs de l'archivage, tant pour l'objet que pour la profession d'ailleurs. Trois axes sont définis et portés par les archives de France et la Direction générale de la modernisation de l'État.

Tout d'abord donner une impulsion à la mise en place de plate-forme d'archivage, sensibiliser les principaux acteurs et la mise en place de référentiels. Des actions sont ainsi menées depuis plusieurs années en direction de ces réseaux avec des formations de mieux en mieux structurées et qui s'améliorent chaque année.

---

<sup>49</sup> Françoise BANAT-BERGER, « Les archives et la révolution numérique : Le Débat », n° 158, p70-82, 2010, 192p  
([http://www.cairn.info/resume.php?ID\\_ARTICLE=DEBA\\_158\\_0070](http://www.cairn.info/resume.php?ID_ARTICLE=DEBA_158_0070), consulté le 27 novembre 2013)

Deux ans plus tard en 2006, deux référentiels sont élaborés, le premier définit le cadre générique d'une politique d'archivage sécurisée appartenant au secteur public sous la tutelle de l'Agence Nationale de la sécurité des systèmes d'information (ANSSI) et le second pour faciliter les échanges entre un service d'archives et ses différents partenaires en les structurant davantage afin de réduire les ruptures de charges entre ces acteurs.

### ***Les normes:***

Issue de ces lois, la norme française NF Z42-013 sur l'archivage électronique à récemment été transposée en norme internationale par l'ISO. La norme ISO 14641-1, porte sur les différentes spécifications pour la conception et l'exploitation des systèmes informatiques utilisés en archivage électronique. Elle s'adresse aux entreprises et organismes qui souhaitent archiver des documents numériques et qui surtout souhaitent assurer la finalité et l'intégrité pour le compte de leurs responsables.

Cette norme permet ainsi aux entreprises du monde entier de prendre connaissance des spécifications relatives aux mesures techniques et organisationnelles dans le cadre d'un plan d'archivage Ce plan encadre l'enregistrement, l'archivage, la consultation mais aussi la communication de documents numériques en vue de leur conservation et de la préservation de leurs intégrités

## **La préconisation de l'Open Access**

Depuis quelques années on a vu l'émergence des entrepôts d'archives ouvertes pour les publications scientifiques, cela a induit un bouleversement dans les processus de communication à plusieurs niveaux. Le principe des archives ouvertes consiste en le dépôt par les auteurs d'articles en version pré-print, voire en version post-print quand l'éditeur l'y autorise. Plus les articles sont diffusés et accessibles à travers le monde, plus ils ont de chance d'être lus, et donc d'être cités, permettant alors à la revue source d'obtenir un

important facteur d'impact et d'attirer ainsi de bons auteurs. La réputation d'un chercheur ou d'un organisme de recherche est bien souvent évaluée par l'impact que ses publications auront.

Le document numérique, en tant qu'objet informationnels sont de nature hybride ce qui entraîne aujourd'hui des problèmes dans la diffusion des résultats. Pourtant la forte croissance de ces entrepôts témoigne d'une augmentation de l'information disponible qui pose la question de l'information pertinente. Ce phénomène engendre un mouvement et une mutation des archives ouvertes.

Ces ressources constituent en effet un réservoir potentiel immense de savoir et permettrait de développer davantage l'apprentissage à distance, bien que le système d'archives ouvert soit encore peu développé en France, les mentalités et les programmes évoluent rapidement.

Plusieurs voies existent, le premier est l'Open Access Green qu'on appelle aussi « *auto-archivage* ». Dans ce modèle, l'auteur, ou une personne qui agirait en son nom, va déposer l'article dans un répertoire dit ouvert, sur un site personnel ou d'une institution. En France l'exemple est HAL, mis en place par le CNRS. Souvent les revues, même sous abonnements vont coexister avec ces archives ouvertes et beaucoup d'éditeurs autorisent le dépôt de ces articles dans une version pré-print avec ou sans embargo.

La seconde solution c'est l'Open Access Hybrid, seuls certains articles d'une revue sous abonnement pourront en fonction du choix de l'auteur être disponibles en libre accès. Contre un paiement des frais de publication, les éditeurs peuvent proposer ce genre d'option, c'était au départ une solution pour basculer progressivement d'une revue sans abonnement à une revue en total open Access.

Nous pouvons prendre l'exemple du développement de l'open Access gold, des revues en libre accès, dont le payeur est soit l'auteur physique, soit une personne morale. Généralement sous forme électronique, il existe parfois des possibilités d'impression à la demande.

Le 17 juillet 2012, la Commission Européenne confirme la volonté de basculer dans le Libre Accès en publiant une recommandation Horizon 2020 préconisant la publication en accès ouvert des travaux de recherche financés par des fonds publics d'ici à 2020.

Néanmoins le travail des publishers, ces personnes responsables de la mise en page des articles, est encore vu par beaucoup comme primordial à la bonne diffusion du savoir scientifique. Permettant une sélection mais aussi une valorisation des données exposées par les auteurs l'éditeur se charge ainsi de vérifier la pertinence de l'article, notamment en le soumettant à des pairs qui seraient aptes de juger objectivement le travail proposé ce qu'on appelle le peer reviewing. Il optimise la lecture du texte en adaptant la typographique, la charte graphique, et corrige même les éventuelles fautes d'orthographe ou de syntaxe, facilitant le travail des archivistes par la suite.

L'Union européenne encourage donc vivement les états membres à adopter des politiques publiques en faveur du libre accès aux publications scientifiques et les gouvernements sont libres de favoriser la voie dorée de la publication en open Access ou la voie verte des archives ouvertes.

Pour que le modèle Open Access Gold marche bien et s'impose de façon permanente il faut tout d'abord garantir la qualité scientifique du résultat final pour encourager les éditeurs, les chercheurs et leurs institutions à investir dans cette solution de façon massive. La présence dans les marchés de plusieurs éditeurs prédateurs qui font de plus en plus d'affaires au détriment des auteurs qui, très souvent de façon naïve, décident de publier avec eux, représente en fait une menace considérable qui risque de compromettre la crédibilité du modèle Gold et de le rendre un investissement infructueux à long terme.

Quant aux chercheurs il faut éviter que le passage à ce nouveau système ne favorise pas trop ceux qui sont rattachés aux institutions les plus riches, indépendamment de la qualité des produits des leurs recherches. A l'heure

actuelle un chercheur « *pauvre* » peut soumettre un article à une revue traditionnelle prestigieuse qui pourra être publié sans aucune contribution économique de sa part tout simplement s'il passe le processus de sélection. Les coûts de publication des articles sont couverts grâce aux abonnements donc de fait la publication de l'article de notre chercheur ne disposant pas de moyens est financée par les abonnements payés à la revue par des institutions « *riches* ». Dans le nouveau système c'est au chercheur même de contribuer aux coûts de la publication. Ce système pourrait empêcher plusieurs chercheurs capables mais sans trop de moyens pour faire connaître les résultats de leurs recherches. Les résultats de ces recherches sont d'un grand intérêt pour l'archivage et l'utilisation de l'Open Access permet une diffusion large et donc une sauvegarde plus facile à répartir sur des disques regroupés en serveur.

Il faut donc s'assurer que l'Open Access ne soit pas incompatible avec la liberté de la recherche. Le passage au modèle Open Access Gold ne peut se faire que de façon très graduelle et doit donc nécessairement être accompagné d'une politique sérieuse d'évaluation des produits de la recherche.

## **Les risques du libre échange**

Le développement de structures privées comme Google ou Facebook met en péril la gestion étatique de données de la recherche ou des données sensibles ou relevant du moins de la vie privée de l'individu. L'étude des conditions d'utilisation d'un site comme Facebook tout d'abord permet de prendre conscience de cette problématique.

*« Partage de votre contenu et de vos informations »*

*Le contenu et les informations que vous publiez sur Facebook vous appartiennent, et vous pouvez contrôler la façon dont nous partageons votre contenu, grâce aux paramètres de confidentialité et des applications. En outre :*

1. Pour le contenu protégé par les droits de propriété intellectuelle, comme les photos ou vidéos (propriété intellectuelle), vous nous donnez spécifiquement la permission suivante, conformément à vos paramètres de confidentialité et des applications : vous nous accordez une licence non-exclusive, transférable, sous-licenciable, sans redevance et mondiale pour l'utilisation des contenus de propriété intellectuelle que vous publiez sur Facebook ou en relation avec Facebook (licence de propriété intellectuelle). Cette licence de propriété intellectuelle se termine lorsque vous supprimez vos contenus de propriété intellectuelle ou votre compte, sauf si votre compte est partagé avec d'autres personnes qui ne l'ont pas supprimé.

2. Lorsque vous supprimez votre contenu de propriété intellectuelle, ce contenu est supprimé d'une manière similaire au vidage de corbeille sur un ordinateur. Cependant, vous comprenez que les contenus supprimés peuvent persister dans des copies de sauvegarde pendant un certain temps (mais qu'ils ne sont pas disponibles). »<sup>50</sup>

Concrètement, on donne à Facebook le droit d'utiliser gratuitement toutes les œuvres créées par ses utilisateurs. D'autres sites comme Instagram ou Twitter appliquent la même politique. Chez Google on retrouve le même fonctionnement. L'auteur reste donc propriétaire en un sens mais cède ses droits d'exploitation gratuitement à ces entreprises. C'est ce que Pascal Robert appelle l'« *auto-gestionnarisation de la société* »<sup>51</sup>, l'utilisateur donnant lui-même les informations nécessaires à sa « traçabilité ». Selon lui, ce phénomène enclenche le processus par lequel la mémoire est de moins en moins envisagée comme une question politique que comme une question économique. Il y a une véritable monétisation de la donnée et dans sa logique Facebook et Google deviennent des entreprises de flux mémoriel à forte valeur ajoutée. La politique de Google d'étendre ses applications dans des projets comme Google image bascule son statut de moteur de recherche à celui d'une « *mémoire vive* » et développe une « *véritable politique*

---

<sup>50</sup> <https://www.facebook.com/terms.php?locale=FR>

<sup>51</sup> Pascal ROBERT, Nathalie PINÈDE, « Communication & Organisation », n° 42, p191-202, 2012, 280 p  
( <http://communicationorganisation.revues.org/3948> , consulté le 25 Mars 2014)

*culturelle* ». Ce nouveau positionnement l'amène à être en concurrence directe avec l'État.

Toujours d'après Robert<sup>52</sup>, si Facebook est une entreprise de flux c'est avant tout parce que ce qui passe par son site est en un sens éphémère tout en permettant d'accumuler des informations sur l'utilisateur. Ces pratiques ne sont pas illégales en soit et sont encadrées par l'article L. 122-7 de la propriété intellectuelle qui dispose que «Le droit de représentation et le droit de reproduction sont cessibles à titre gratuit ou à titre onéreux», et ajoute que «*L'auteur est libre de mettre ses œuvres gratuitement à la disposition du public, sous réserve des droits des éventuels coauteurs et de ceux des tiers ainsi que dans le respect des conventions qu'il a conclues*». En revanche toute cession de droit est encadrée par l'article L.131-3 du code de la propriété intellectuelle qui précise qu'une durée limite doit être fixée sans quoi, en l'absence de cette mention, ou si celle-ci est jugée trop imprécise la cession est nulle et toute reproduction sera considérée comme une contrefaçon.

Dans les licences citées précédemment la durée n'est justement pas stipulée. L'article L.131-3 précise également que les modes de reproduction doivent être indiqués, là encore ce n'est pas le cas dans les licences que nous avons regardées. Il est alors possible de mener une action en justice mais même avec la loi de son côté l'utilisateur reste en pratique souvent dépouillé de ses droits et ne maîtrise que difficilement sa propriété privée numérique.

Récemment, en particulier dans le cadre de l'affaire Edward Snowden aux États-Unis, la presse a souligné que divers gouvernements utilisent l'analyse de données privées à des niveaux surprenants. La principale raison invoquée pour ce type d'utilisations est la lutte contre le terrorisme. Mais le système PRISM de la NSA<sup>53</sup> (National Security Agency, l'agence américaine de

---

<sup>52</sup>Ibid.

<sup>53</sup> Serge ABITEBOUL, Pierre SENELLART, « Un déluge de données » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p32-35, Novembre 2013, 188p

sécurité) et les systèmes équivalents des autres États sont aussi utilisés pour l'espionnage, notamment industriel. Ils peuvent l'être pour surveiller les opposants politiques. Le contrôle et l'analyse des données sont certainement les nouvelles armes parmi les plus inquiétantes des dictatures et des gouvernements totalitaires.

Récemment encore, Edward Snowden a dévoilé un document secret rédigé par un membre de la NSA<sup>54</sup> qui explique comment l'agence pouvait traquer l'e-mail privé et les comptes Facebook des administrateurs système avant de pouvoir pirater leurs ordinateurs et ainsi avoir accès aux réseaux qu'ils contrôlent.

En infiltrant les ordinateurs des administrateurs système travaillant pour des compagnies étrangères de téléphone et d'Internet, la NSA pouvait avoir accès aux appels et courriels qui circulent sur ces réseaux. Un exemple lorsque Der Spiegel a révélé l'an passé que les réseaux des ingénieurs de Belgacom en Belgique avaient été victime de ce type d'intrusion par l'agence britannique de communication GCHQ.

L'appropriation de ce contenu constitue l'un des versants de cette problématique, l'autre versant est ce que Robert<sup>55</sup> appelle «*l'auto-dévoilement de la personnalité*», l'étude et l'observation des comportements amenant à cette «*transparence nouvelle*» de l'identité de l'individu. Elle permet aux entreprises d'avoir un certain pouvoir qui revenait auparavant uniquement à l'Etat. Les enjeux de la dématérialisation sur des serveurs en réseaux sont donc la préservation des droits qui protègent l'identité numérique. Pourtant l'arrivée de ces stratégies de libre échange, de libre circulation de l'information peut aussi avoir ses avantages comme l'arrivée de l'open

---

<sup>54</sup> Ryan GALLAGHER, Peter MAASS, « Inside the NSA's Secret Efforts to Hunt and Hack System Administrators », Mars 2014 ( <https://firstlook.org/theintercept/article/2014/03/20/inside-nsa-secret-efforts-hunt-hack-system-administrators/>, consulté le 15 Août 2014)

<sup>55</sup> Pascal ROBERT, Nathalie PINÈDE, « Communication & Organisation », n° 42, p191-202, 2012, 280 p ( <http://communicationorganisation.revues.org/3948> , consulté le 25 Mars 2014)

Access dans les pays démocratiques qui pourrait permettre aux journalistes de partager plus librement des informations.

# L'AVENIR DE L'INFORMATION

---

## LES ENJEUX EMERGENTS DES NOUVEAUX SUPPORTS

### L'évolution de l'information

Nous l'avons vu, la masse d'information produite ne cesse d'augmenter et sa préservation sur de nouveaux supports engendre des conséquences importantes sur la mémoire sociale. La diversité des supports étudiés peut laisser entendre que de nombreuses solutions existent et pourtant ces systèmes ne fonctionnent que sur une faible quantité de principes matériels. L'information même est devenue différente, massive et parfois incontrôlable. Comme nous l'avons vu la dissociation du contenu et de son support ont entraîné la disparition progressive de la notion de document original qui permettait de fonder les principes même d'archivage. Le numérique et les enjeux de pérennisation s'attachent désormais à englober l'archivage du web. Singulier et inédit ce type d'archivage va aujourd'hui poser de nombreuses questions et amène les archivistes à revoir leurs méthodologies. Le document a pris en profondeur, pouvant être relié à d'autre, entraînant les problèmes de pérennité des liens auxquels il renvoie. Stocker sur un support matériel, aussi performant soit-il, ne permet pas de pérenniser les liens qui se créent entre les documents, il est donc indispensable aujourd'hui d'envisager l'information sous un angle nouveau. Le document, selon Pascal Robert<sup>56</sup> est ainsi « *sur-stocké* » mais impliqué dans un flux, le rendant « multiple et insaisissable ».

Or pour construire l'histoire il sera nécessaire de pouvoir retrouver ces connexions entre les documents. Des documents qui autrefois étaient recontextualisables par leurs formats, leurs structures et leurs nomenclatures et on pouvait avec des études visuelles et matérielles mettre en lumière ces connexions. Un historien ne se satisfait pas du contenu mais apporte autant d'attention au contenant, au niveau de langage, aux circonstances historiques, une approche globale qui selon Éric

---

<sup>56</sup>Pascal ROBERT, Nathalie PINÈDE, « Communication & Organisation », n° 42, p191-202, 2012, 280 p  
( <http://communicationorganisation.revues.org/3948> , consulté le 25 Mars 2014)

Ketelaar<sup>57</sup> apportait également un apport à l'anthropologie, il cite notamment Ann Stoler qui met en avant le « *tournant archivistique* » que subit la profession et qui permet de réfléchir sur la rédaction, l'appréhension même du processus d'archive. En d'autres termes l'étude de l'étude des archives.

## La «logistication» du document numériques<sup>58</sup>

Le document s'enrichit aussi grâce à des modes comme le mode révision. Dans de nombreuses structures, le mode révision des logiciels type suite Office comprenant des logiciels de traitement de texte ou de tableur par exemple, sont apparus. Ces systèmes permettent d'annoter les documents et sont un exemple de l'enrichissement qu'autorise désormais le numérique. Cette fonction permet de capter une immédiateté qui apporte un outil de contextualisation supplémentaire qui n'existait pas avant l'arrivée de l'informatique. C'est cet aspect qui permet ce que Pascal Robert<sup>59</sup> qualifie de « *mise en miroir du document* ». La persistance du document ne dépend plus du nombre de copies mais de la capacité à extraire le document qui sera archivé. Mais cette persistance reste étroitement liée à ses copies qui permettent la sauvegarde par comparaison et donc on en revient à la problématique des supports sur lesquels ils sont fixés. La pérennité des supports reste donc un facteur majeur car c'est d'eux que dépendra l'accessibilité au produit et donc à la capacité de le reproduire et de le diffuser.

Le document s'inscrivant désormais dans une logique de flux et disposant d'une nature instable, l'archivage devient un processus automatique informatique dont il faut paramétrer les algorithmes de sélection et de protection. L'archive élargit sa fonction de stockage à une fonction d'informatisation du document. Pascal Robert ira plus loin en évoquant avec la « *logistication du document* ». Les exemples que nous avons vu illustrent bien ces notions, à travers la mise en place de procédures

---

<sup>57</sup>Eric KETELAAR, « (Dé) Construire l'archive Matériaux pour l'histoire de notre temps » ; « Matériaux pour l'histoire de notre temps », N° 82, p65-70, 2006, p138, (<http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2-page-65.htm>, consulté le 15 mai 2014)

<sup>58</sup> Pascal ROBERT, Nathalie PINÈDE, « Communication & Organisation », n° 42, p191-202, 2012, 280 p  
(<http://communicationorganisation.revues.org/3948>, consulté le 25 Mars 2014)

<sup>59</sup> Ibid.

de stockage mais aussi de protections de l'intégrité, on s'inscrit effectivement dans une tendance à la « *logistication du document* ». <sup>60</sup>

On peut considérer que cette logistique entraîne également ce que Robert appelle « *la logistique informationnelle documentarisée* » <sup>61</sup> La multiplication des échanges et leurs centralisations sur un seul canal d'échange, l'informatique, un seul langage, le bit, rend ce flux observable, quantifiable et analysable. Prenons par exemple les flux de consommation, que seule une étude basée sur de petit échantillon permettait de rendre exploitable, aujourd'hui ce sont des échantillons beaucoup plus importants, une visibilité presque mondiale peut être obtenue. Des exemples contradictoires se mettent en place parfois. Le piratage par exemple est un phénomène mesurable et souvent réprimandable dans beaucoup de pays et pourtant certaines productions utilisent parfois ces données pour se vanter sur les réseaux sociaux de ces audimats illégaux lorsque le dernier épisode de leurs séries est diffusé. A échelle plus individuel c'est tout le profil web de l'individu qui peut se dessiner et c'est cela qu'entend Robert avec sa « *documentarisation logistique* » mettant en avant une « traçabilité sur les réseaux sociaux » à l'initiative même de l'utilisateur.

L'information est dans une dualité plus il y a d'information plus l'incertitude d'un risque informationnel grandit. Pour établir un plan d'action face à ces risques, conséquences de l'incertitude, il faut davantage comprendre la nature et le périmètre d'influence de ces flux.

### ***Archives et propriété intellectuel***

D'un point de vue législation, le document d'archive ou conservé par un service à vocation patrimoniale peut être soumis à une protection intellectuelle qu'il faut distinguer de ce qui est inédit. Un document reconnu comme inédit ne pourra bénéficier

---

<sup>60</sup> Pascal ROBERT, Nathalie PINÈDE, « Communication & Organisation », n° 42, p191-202, 2012, 280 p  
( <http://communicationorganisation.revues.org/3948> , consulté le 25 Mars 2014)

<sup>61</sup> Ibid.

d'une communication qu'après autorisations de l'auteur ou des héritiers. Ce qui est permis c'est la copie privée et la reproduction à visé technique pour le dépôt légal ou pour la reproduction de conservation par les bibliothèques, les musées et les archives. Pour les données personnelles, d'une part elles ne peuvent être communiquées à des tiers sans l'accord de la personne concernée et d'autre part toute personne peut avoir accès à ces données personnelles et demander à les faire modifier. L'évolution du web, du point de vue de l'archiviste, permet une diffusion beaucoup plus large donc des risques beaucoup plus grands de dispersion d'où l'intérêt de faire un travail de sensibilisation du public.

Se pose alors la question de la responsabilité des institutions archivistiques et de leurs missions face à cette démultiplication du volume informationnel.

## **Nouveaux enjeux, nouvelles missions**

### *La responsabilité de la conservation et de l'archivage*

Selon les dispositions de la loi du dépôt légal les bibliothèques doivent assurer que toutes les publications de leurs pays sont récupérées, conservées puis diffusées. Mais plusieurs pays n'ont pas encore débattu au sujet des publications électroniques et une nouvelle problématique émerge.

Des rapports comme celui du Task Force on Archiving of Digital Information<sup>62</sup> suggèrent dans ses conclusions de faire des éditeurs les responsables de l'archivage et de la conservation de leurs publications. Cette conclusion s'appuie sur le fait que ces tâches sont complexes et demandent que les éditeurs soient les premiers responsables de l'archivage et de la conservation de leurs publications, réaffirmant le rôle des publishers que nous avons évoqués avec l'Open Access. Mais la mise en place des responsabilités au niveau international appelle à une organisation surtout dans le domaine de l'information publiée en réseau.

---

<sup>62</sup> Guylaine BEAUDRY, Gérard BOISMENU, « Expertise technique et organisationnelle », Février 2001 (<http://revues.enssib.fr/titre/6tecnica/5archivage/2formats.htm>),

Plusieurs institutions tentent donc de mettre en place des solutions pour garantir l'accès sur le long terme à l'information. Récemment, des pays comme le Danemark, la Finlande et la Norvège <sup>63</sup> commencent à modifier leur législation concernant le dépôt légal et essaient de s'adapter pour couvrir toutes les ressources électroniques publiées en ligne. D'autres pays emboîtent le pas sur ces démarches, souvent en collaboration et à l'initiative des bibliothèques nationales. Chaque situation est différente et on évalue en détail les caractéristiques de chaque pays, notamment concernant la situation de l'industrie et de l'édition électronique.

Prenons l'exemple de l'Allemagne et des Pays-Bas dans lesquels on trouve une forte présence de l'édition électronique internationale assurée par les éditeurs traditionnels du domaine, cela influence grandement les réflexions en faveur d'un dépôt volontaire des éditeurs traditionnels de ces documents qui seront ensuite pérennisés sur des supports. Plus particulièrement, au Pays Bas, la Bibliothèque Royale propose de devenir un archiveur officiel de l'éditeur tel qu'Elsevier ou Kluwer.

Le cas opposé c'est l'arrivée, sur le marché de l'édition électronique, de partenaires complètement nouveaux et sans rapport à l'édition traditionnelle comme en Australie, en Norvège, au Danemark ou en Suède. Cette situation modifie les modalités de collecte et de transfert vers les archives nationales.

Pourtant la Suède et la Finlande pratiquent la collecte systématique des ressources en ligne du domaine national. Chez les norvégiens, on demande aux éditeurs de déposer toute les publications et ce sont les éditeurs qui assurent la mise en forme. Au Danemark, les responsabilités sont partagées entre bibliothèque et éditeurs, ces derniers devant notifier à la Bibliothèque national les documents dans la base de l'éditeur. L'Australie a fait le choix de sélectionner les ressources web qui ne correspondent qu'à sa politique documentaire, définie et publique <sup>64</sup>.

---

<sup>63</sup> Catherine LUPOVICI, « Les stratégies de gestion et de conservation préventive des documents électroniques », Bulletin des Bibliothèques de France, Avril 2000, N°4, p 43-54 <http://www.enssib.fr/bbf/bbf-2000-4/07-lupovici.pdf>,

<sup>64</sup> Catherine LUPOVICI, « Les stratégies de gestion et de conservation préventive des documents électroniques », Bulletin des Bibliothèques de France, Avril 2000, N°4, p 43-54 <http://www.enssib.fr/bbf/bbf-2000-4/07-lupovici.pdf>,

En France, la loi du Dépôt légal date de 1992 et ses décrets d'application règlent l'obligation du dépôt pour les publications électroniques sur support depuis 1993. Malheureusement ce texte ne s'applique pas aux publications en ligne. Pour l'instant donc elles ne sont pas enregistrables par la Bibliothèque nationale de France au titre du dépôt légal. Il faudrait une extension de la loi *est nécessaire* pour résoudre ce problème et mettre en place une politique documentaire. Car c'est aussi grâce à ses amnésies que la mémoire se construit, l'élimination et la sélection de ce qui doit ou non être archivé(ou *archivé ?*) constituent en soit un phénomène positif<sup>65</sup> à partir duquel on peut aussi édifier son objet. Le tri reste ainsi souvent vu comme un point négatif, mal compris et peu explicite.

L'émergence de ces problématiques confirme que de plus en plus les archivistes sont non seulement gardiens de la mémoire collective mais sont devenus également membres actifs de l'édification et de la structuration de celle-ci. Surtout cette question appartient désormais aussi aux citoyens.

Émerge donc à l'international en plus des normes, des recommandations comme celle issue de la charte de l'Unesco adoptée par la Conférence général de 2003. Elle recommande en outre, la création de politique et de stratégies pour assurer la préservation du patrimoine numérique. Une enquête conduite en 2009 a cependant révélé que peu de pays en développement suivent ces dispositions. Pour remédier à cette méconnaissance de la Charte dans ces pays, l'Unesco a pris l'initiative d'organiser un débat international pour protéger le document numérique. En septembre 2012 s'est donc tenu à Vancouver au Canada le débat « *La Mémoire du monde à l'ère numérique : numérisation et conservation* ». L'objectif était de recueillir des propositions concrètes en vue de préserver l'avenir du document numérique. Deux conclusions principales sont tirées, il faut une politique de préservation authentique et pérenne à échelle mondiale et une continuité de l'information maintenue.

---

<sup>65</sup> Catherine LUPOVICI, « Les stratégies de gestion et de conservation préventive des documents électroniques », Bulletin des Bibliothèques de France, Avril 2000, N°4, p 43-54 <http://www.enssib.fr/bbf/bbf-2000-4/07-lupovici.pdf>,

Les enjeux sont la mise à disposition de documents fiables et conservés sur des supports stables permettant d'associer une politique de transparence sur les affaires publiques avec un système de libre accès, tout en préservant les droits du citoyen. A l'issue de ce débat les participants ont adopté un texte, la Déclaration de Vancouver UNESCO/UBC qui contient des recommandations spécifiques à l'intention de l'UNESCO et des acteurs de la sauvegarde dans les États membres.

D'autres recommandations ont ainsi été formulées en vue de la conservation pérenne des données. On peut notamment citer deux commissions, la première américaine la Commission on Preservation and Access, sous la tutelle du Council on Library and Information en collaboration avec le Research Librarians Group réunissant une équipe de travail chargée d'élaborer un projet de conservation nationale de l'information numérique et qui fait référence au rapport Preserving Digital Information. La deuxième est l'équivalent européen, la European Commission on Preservation and Access (Ecpa), qui vise à promouvoir ces travaux américains pour accroître la prise de conscience.

### *L'archivage du web : enjeux et défis*<sup>66</sup>

Les usages ont changé, tout être pensant est potentiellement générateur de contenu, un phénomène en expansion grâce à la mise en place de « *pure players* », ces hébergeurs au service d'UGC (User Generated Content, ou « contenu généré par les utilisateurs ») et les services vidéo en ligne. Les artefacts type Smartphone, tablette et autres outils sont autant de moyens de créer et de partager du contenu.

Plusieurs types de logiciels ont donc été mis au point pour permettre la conservation du web. Ces « *moissonneurs* » permettent d'effectuer des sauvegardes régulières en parcourant le Web.

Le pionnier en la matière est Brewster Kahle, dont nous avons déjà parlé et qui dès 1999 fonde Internet Archive. Cet ingénieur diplômé du MIT (Massachusetts

---

<sup>66</sup> Claude MUSSOU, « Et le web devint archive : enjeux et défis », juin 2012 (<http://www.ina-expert.com/e-dossier-de-l-audiovisuel-sciences-humaines-et-sociales-et-patrimoine-numerique/et-le-web-devint-archiver-enjeux-et-defis.html>, consulté le 25 Mai 2014)

Institute of Technology) a eu pour objectif de rendre accessible tout le web, un projet qui a notamment permis d'en faire émerger d'autres dans les institutions archivistiques.

C'est par exemple le cas de la BNF qui fonde en 2003 un consortium dédié à ce projet, l'IIPC (L'International Internet Preservation Consortium), qui regroupe aujourd'hui une quarantaine d'institutions comme l'INA, des bibliothèques nationales ou universitaires. L'ensemble de ces structures travaille en partenariat en fonction des moyens de chacun à la préservation du Web mondial.

L'INAA est notamment en charge, selon la Loi Dadvsi (Titre IV) de 2006, de la collecte et de la préservation des contenus à titre de Dépôt légal. L'objectif de cette mission est de pouvoir rendre accessible les documents et permettre l'élargissement des champs de recherche.

Pour répondre pleinement à ces obligations l'Ina entreprend, dès 2001, la collecte et l'archivage des sites de médias plus particulièrement des vidéos à la demande des grandes chaînes mais aussi des sites personnels consacrés à certains programmes qu'il s'agisse de radio ou vidéo. A titre d'illustration cette collecte présentait en avril 2012 plus de 13 milliards de version d'URL.

Mais sélectionner ces ensembles de sites a demandé une réflexion autour de la notion même d'archivage documentaire. Les algorithmes de sélection doivent sélectionner à partir des métadonnées qui imposent une redéfinition de l'approche documentaire Claude Mussou, ajoute que dans certains cas les opérations de « *catalographique, bibliographique, archivistique ou documentaire de structuration des fonds peuvent se révéler caduques et inopérantes* ».

La plupart des moteurs de recherche comme Google mettent en place ce qu'on appelle une stratégie de « *ranking* », l'ordre des résultats à une requête est défini en fonction de facteurs, et plusieurs méthodes permettent de rendre la recherche plus pertinente et souvent cette pratique fera remonter en priorité les liens commerciaux.

L'INA au contraire a fait le choix de pratiquer une indexation « *full text* » qui permet de garantir une neutralité des réponses. Claude Mussou définit d'ailleurs

cela comme un « *contrepoids à la substitution de la mémoire collective par les monopoles industriels du Web mondial* ». <sup>67</sup>

Concrètement les requêtes vont être relatives au nombre d'occurrence des mots recherchés mais aussi à leurs places dans la structure éditoriale du texte. Ces priorités sont parfois renégociées et discutées permettant une relative assurance de la neutralité du résultat.

Les institutions patrimoniales à travers ces procédés poursuivent leurs missions de pérennisation de l'information et des savoirs<sup>68</sup>. Les instabilités du Web et la fragilité des pages web font de cette démarche un enjeu majeur.

Le lien mort est un phénomène qui touche tout le monde mais est particulièrement gênant dans les domaines de recherches scientifiques qui sont très consommateurs de source d'information. Les archivistes du web ont donc une grande responsabilité dans ce rôle.

## **QUELS EFFETS SUR LA MEMOIRE SOCIALE ?**

### **L'information comme bien commun**

L'information suit le développement des réseaux, Internet appartient à tout le monde et en même temps à personne générant une richesse collective grâce à ces mises en communs des savoirs. Il devient un bien commun même s'il existe ce que Lionel Maurel qualifie de « *passager clandestin* »<sup>69</sup> ceux qui ne participent pas à l'évolution des biens et récupèrent le bénéfice des évolutions globales. On peut citer concrètement la construction dès les débuts du Net d'Imdb une grande banque de données sur le cinéma nourrie par le travail de milliers d'internautes et qui sera récupérée plus tard par le géant Amazon.

---

<sup>67</sup> Claude MUSSOU, «Et le web devint archive : enjeux et défis », juin 2012 (<http://www.ina-expert.com/e-dossier-de-l-audiovisuel-sciences-humaines-et-sociales-et-patrimoine-numerique/et-le-web-devint-archive-enjeux-et-defis.html>, consulté le 25 Mai 2014)

<sup>68</sup> Ibid.

<sup>69</sup> Lionel MAUREL, « Les communs de la connaissance : un autre regard sur l'information » ; « Documentaliste-Sciences de l'Information », Vol. 48, p48-59, 2011, (<http://cergy0051.lb.u-cergy.fr/revue-documentaliste-sciences-de-l-information-2011-3-page-48.htm>, consulté le 26 Mai 2014)

Des licences virales, on justement été créées dans le but d'éviter ces appropriations. En 1989, (General Public Licence), la plus connue est rédigée par Eben Moeglin et Richard Stallman pour permettre l'utilisation des logiciels libres mais qui impose de rendre public et accessible toute modification du code. La notion de bien commun va donc également être définie par cette organisation, cette « *gouvernance* ». <sup>70</sup>

Car sans gouvernance plusieurs risques existent, d'abord celui « *d'embouteillage* » en effet, la congestion de cette ressource peut entraîner des ralentissements. Le réseau Internet est conçu comme une gigantesque toile d'araignée, les serveurs privés ou publics communiquent entre eux et lorsque l'on tape une adresse internet dans un navigateur la connexion chemine de serveur en serveur jusqu'à l'information

Récemment par exemple, le 12 août 2014 <sup>71</sup>, Internet a subi une panne remarquée notamment dans toute l'Europe de l'Ouest et aux États-Unis. Pour des utilisateurs la panne est probablement parue discrète mais chez les fournisseurs la panne était sérieuse. A chaque point d'intersection on trouve un routeur qui sert d'aiguilleur en se servant de la « *table de routage* » commune et sur laquelle ils communiquent pour être à jour. On ajoute donc régulièrement à cette table des nouveaux chemins, et c'est ce qu'a fait le 12 août l'opérateur de télécommunications Verizon aux États-Unis.

Le problème est que de nombreux routeurs en fonctions sont un peu dépassés et ne peuvent lire que jusqu'à 512 000 routes différentes, l'arrivée des milliers de routes de Verizon c'était trop et ils ne pouvaient plus remplir leur rôle. Même si l'impact de cette panne est limitée elle témoigne bien du manque de gestion technique qui existe parfois avec les nouveaux supports de l'information. Dans un sens, chaque

---

<sup>70</sup> LE CROSNIER Hervé, ERTZSCHEID Olivier, PEUGEOT Valérie, MERCIER Silvère, BERTHAUD Christine, CHARNAY Daniel, MAUREL Lionel, « Vers les « communs de la connaissance » ; « Documentaliste-Sciences de l'Information », Vol. 48, p48-59, 2011, 80p ([http://www.cairn.info/zen.php?ID\\_ARTICLE=DOCSI\\_483\\_0048](http://www.cairn.info/zen.php?ID_ARTICLE=DOCSI_483_0048), consulté le 1er Juillet 2014 )

<sup>71</sup>Martin UNTERSINGER, "Devenu trop grand, Internet a subi des perturbations", Août 2014 ([http://www.lemonde.fr/pixels/article/2014/08/15/devenu-trop-grand-internet-est-tombe-en-panne\\_4472153\\_4408996.html](http://www.lemonde.fr/pixels/article/2014/08/15/devenu-trop-grand-internet-est-tombe-en-panne_4472153_4408996.html), Consulté le 25 Août 2014)

fournisseur et hébergeur internet, chaque acteur de cette toile est responsable de son petit bout de réseau et doit s'assurer que les capacités de leurs routeurs conviendront.

C'est un phénomène d'interdépendance qui nécessite une gouvernance car elle frôle parfois l'autogestion. Concrètement, pour Frédéric Dhieux, directeur de projets de BSO Network Solutions, par exemple, c'est à l'ANSSI, organisme de l'État chargé de la défense des réseaux français, de jouer un rôle de veilleur et qui devrait alerter sur ces questions car pour l'instant la seule option est de rappeler les responsabilités de chacun pour le bien de chacun dans le maintien d'Internet. Cette pente glissante pourrait inciter les acteurs de ces infrastructures à promouvoir des politiques « *malthusiennes* »<sup>72</sup> de réduction de l'accès, en définitive, payer plus pour internet. D'autres menaces existent comme interdire l'accès à certains liens, l'exemple de la Chine avec sa « *muraille électronique* ». Les coupures d'Internet en Égypte durant les émeutes de 2011 est un autre exemple, et à l'inverse c'est parfois pour protéger sa population que les réseaux sont coupés, comme avec les conflits en Irak récemment où les kurdes ont coupé l'accès à Facebook durant plusieurs heures car la propagande des djihadistes engendrait des mouvements de peur chez la population même très éloignée des zones de conflits.

Il est nécessaire de permettre une ouverture démocratique vers les savoirs tout en protégeant ces savoirs de la pollution, car il existe aussi au-delà de la privation d'un accès, le phénomène inverse avec le spam et les pollutions intellectuelles du type de celle employée en Irak avec des conséquences certaines.

---

<sup>72</sup> LE CROSNIER Hervé, ERTZSCHEID Olivier, PEUGEOT Valérie, MERCIER Silvère, BERTHAUD Christine, CHARNAY Daniel, MAUREL Lionel, « Vers les « communs de la connaissance » ; « Documentaliste-Sciences de l'Information », Vol. 48, p48-59, 2011, 80p ([http://www.cairn.info/zen.php?ID\\_ARTICLE=DOCSI\\_483\\_0048](http://www.cairn.info/zen.php?ID_ARTICLE=DOCSI_483_0048), consulté le 1er Juillet 2014 )

## Les risques éthiques

### « *Les incertitudes de la société du savoir* »<sup>73</sup>

Dans cette partie nous discuterons les « *incertitudes de la société du savoir* » telles que les décrit Sven Ove Hansson, professeur de philosophie à l'Institut royal de technologie de Stockholm. Ce chercheur mène essentiellement des recherches autour de la philosophie du risque et des articulations avec la dynamique des croyances et la théorie des valeurs. Dans cet article il évoque l'altération de la qualification de notre société, tantôt société du savoir, tantôt société du risque.

C'est ce que nous avons tâché de mettre en évidence précédemment. D'un côté la multiplication des moyens de communication et de production de l'information ont permis d'aboutir à ces appellations de « *société du savoir* » et de l'autre les risques mémoriels inquiètent les institutions et la perte de donnée nous rend vulnérable.

Hansson admet en un sens que ces deux nominations sont en fait valables pour qualifier notre société, en partie du moins. Il définit le savoir comme étant à la fois cognitif, reposant sur un système de croyance, et un système plus complexe d'appropriation des données contenues qui permettent de générer du savoir. Subjectif et objectif à la fois, le savoir permet de définir une « *croyance justifiée* ».

Les informations permettent de construire ce savoir et sont constituées de données qui doivent être contextualisées pour que l'information prenne un sens et soit « *assimilable* ». Hansson précise que ces distinctions sont nécessaires<sup>74</sup>. Il faut en plus disposer d'une forte certitude, car c'est le degré de confiance qu'on associe à l'information qui permet de la valider. Dans le cas inverse c'est l'incertitude qui s'installe, incluant dans les deux cas un état subjectif. C'est cette perception qui va

---

<sup>73</sup> Ove HANSSON SVEN, « Les incertitudes de la société du savoir », *Revue internationale des sciences sociales*, n° 171, p43–51, 2002, 192p (<http://www.cairn.info/revue-internationale-des-sciences-sociales-2002-1-p-43.htm>, consulté le 25 juin 2014)

<sup>74</sup> Ove HANSSON SVEN, « Les incertitudes de la société du savoir », *Revue internationale des sciences sociales*, n° 171, p43–51, 2002, 192p (<http://www.cairn.info/revue-internationale-des-sciences-sociales-2002-1-p-43.htm>, consulté le 25 juin 2014)

inclure la notion de risque. Le risque fonctionne sur la mise en jeu d'un facteur inconnu engendrant de l'incertitude que le niveau de savoir ne pourra dissiper. Avec l'arrivée du numérique des risques ont évolué et des risques d'un nouveau genre sont apparus.

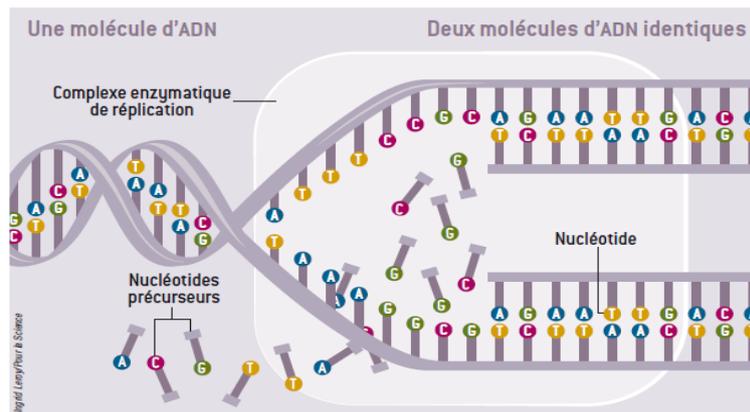
Récemment un projet<sup>75</sup> dirigé par Nick Goldman de EBI (European Bioinformatics Institute) de Hinxton au Royaume-Unis a permis de coder 154 des sonnets de Shakespeare dans de l'ADN. D'autres tests ont été réalisés en codant 26 secondes d'un clip audio du célèbre discours de Martin Luther King « *I have a dream* », une copie de travaux de recherche des scientifiques James Watson et Francis Crick et ils ont même ajouté une photo de l'institut des chercheurs avec un fichier descripteur, la table ASCII. Soit en définitive quatre types de document, un fichier texte en code ASCII pour les sonnets, un PDF pour l'article scientifique, un extrait sonore au format mp3, un document iconographique au format JPEG 2000.

« *Je pense que c'est une étape très importante* », explique le généticien moléculaire de la Harvard Medical School à Boston, George Church.

En 2012 il avait déjà de son côté codé son dernier livre dans seulement un milliardième de gramme d'ADN. Selon Sriram Kosuri, l'un des membres de cette équipe de recherche en biologie de synthèse on pourrait avec cette méthode stocker la totalité des informations dans le monde, soit 1,8 zettabits ( $1,8 \times 10^{21}$ ), dans seulement 4 grammes d'ADN.

---

<sup>75</sup> YONG Ed, « Synthetic double-helix faithfully stores Shakespeare's sonnets », Janvier 2013 (<http://www.nature.com/news/synthetic-double-helix-faithfully-stores-shakespeare-s-sonnets-1.12279>, consulté le 12 Février 2014)



1. LE MÉCANISME qui confère sa pérennité à l'information codée dans le génome est la réplication de l'ADN, orchestrée par un complexe protéique, le complexe enzymatique de réplication. Les deux brins de la molécule d'ADN sont séparés et, sur chaque brin, sont recrutés les nucléotides complémentaires de la séquence lue par le complexe [A et C sont respectivement les complémentaires de T et G]. On obtient ainsi deux nouvelles molécules d'ADN identiques.

**Figure 4 L'ADN, mémoire numérique du vivant Vincent Daubin, Simon Penel et Éric Tannier Pour la science**

L'ADN c'est avant tout de l'acide désoxyribonucléique, elle est composée de brins complémentaires composés de nucléotides dont l'agencement détermine la lecture des informations génétiques. On les organise selon quatre bases (T, G, A et C) que l'on pourrait faire correspondre à un système binaire en imaginant que le T et le G représente le 1 et le A et le C le 0). Lorsque l'on crée de l'ADN par synthèse il est possible d'écrire des informations en utilisant un système déjà utilisé en informatique.

Concrètement les scientifiques de Boston ont utilisé une imprimante chargée de graver les fragments d'ADN synthétisés sur une surface en verre et ont converti les 0 et 1 du code en séquences de nucléotides composées des quatre éléments. Pour retrouver le code dans l'extrait un fichier a été ajouté et il suffit d'un séquenceur ADN et d'un ordinateur pour lire les données.

Pour cette lecture, la chute du coût du séquençage de l'ADN permet d'envisager toute les pistes et encoder n'est pas compliqué. En revanche coder de longue séquence d'ADN peut coûter cher et les chercheurs ont pris le parti de coder sur des fragments courts avec seulement 150 paires de base.

Pour anticiper les éventuelles erreurs de lecture des séquences lorsque plusieurs séquences de nucléotides sont identiques, les chercheurs de l'équipe de Nick

Goldman ont décidé d'utiliser une base en 3 et utilisent le quatrième nucléotide comme opérateur de permutation. Le code est donc en réalité composé de 0, 1 et 2 qui correspondent à A, T et G ou C. Ainsi pour coder 22, que le premier chiffre est codé par A le second 2 sera codé par une correspondance nouvelle qui ne fait intervenir que les trois autres nucléotides. Pour limiter davantage les erreurs, les fragments ont été également placés de façon à se chevaucher. De cette façon, au début de chaque fragment on retrouve les informations qui indiquent sa position dans l'ensemble de la séquence.

La question du support s'est aussi posée, pour l'expérience ils ont utilisé un substrat en verre, figé l'ADN devient inaltérable pouvant résister à des variations de température extrême et lu par un algorithme informatique.

Les données ont été stockées sur 150 000 brins d'ADN qui a ensuite été lyophilisé et expédié en Allemagne par avion. Arrivé là-bas, il a fallu deux semaines pour le réhydrater et le séquencer. A l'exception d'un fragment de 25 nucléotides, l'intégralité des informations a pu être retrouvée.

L'équipe de Church, elle, n'a utilisé qu'un code simple où A et C représente zéro et G et T représente 1. Cette méthode présente le défaut de générer de longs segments avec la même lettre, plus difficile à comprendre pendant le séquençage et qui conduit parfois à des erreurs.

En raison de sa stabilité, de ses capacités de conservation et de vie, l'ADN a longtemps été considéré comme un potentiel support de stockage<sup>76</sup>. En effet, l'idée qu'il puisse stocker des suites de 0 et 1 est tout à fait cohérente avec sa structure.

Dès 1964 le physicien et officier russe Mikhail Neiman a vu dans ce système un moyen de miniaturiser les dispositifs de stockage. Dès 1990 l'expérience est mise en œuvre avec le développement du génie génétique, on réussit à produire des séquences courtes, comme un message secret annonçant le débarquement du 6 juin 1945, une petite comptine anglaise avec parole, musique et petite image ou encore la déclaration des droits de l'homme.

---

<sup>76</sup> DESSIMOZ Christophe « Stocker de l'information dans l'ADN » ; « BIG BANG numérique Les données massives changent-elles le monde ? Pour la Science, n° 433, p100-101, Novembre 2013, 188p

Grâce aux progrès des techniques de séquençage la recherche s'est accélérée et en 2010 Craig Venter et son équipe arrivent à synthétiser un génome complet avec un million de paires de bases à partir du génome d'une bactérie. Cet ADN a ensuite été injecté à une autre bactérie dont le génome propre avait préalablement été retiré. Cette expérience, certes fort coûteuse, a permis de démontrer qu'il était désormais possible de coder des informations avec de l'ADN avec une durée de fabrication raisonnable.

Néanmoins cette approche est différente de celle de Goldman qui code l'information dans un ADN synthétique fabriqué par une machine et non par une cellule de bactérie. Le fait que l'ADN soit un support universel permet de limiter les risques liés à l'obsolescence, depuis plus de trois milliards d'années c'est le support de l'information génétique des êtres vivants. Sous cette forme et tant qu'il y aura de la vie sur Terre il sera possible de développer une technique pour le lire.

Techniquement cette information devrait être capable d'être encore lisible dans des millénaires même dans des conditions de conservation extrême. Notons que sa lenteur dans le processus de consignation des données le destinerait déjà à une conservation pérenne de données mais dans l'avenir le procédé pourra être amélioré. Les capacités de stockage testées sont encore faibles avec environ 739Ko mais il devrait rapidement être possible de répéter l'opération à une échelle plus grande. Le procédé aussi reste cher, comptez 7000 euros par mégaoctet, ce qui là encore peut être amélioré.

Mais où stockerait-on cette ADN ? Il existe déjà des banques d'ADN comme la Réserve mondiale de semence du Svalbard, une enceinte souterraine réfrigérée au cœur d'une île norvégienne du Spitzberg mais n'importe quel endroit frais, sombre et sec pourrait suffire à le conserver sous forme lyophilisé.

L'informatique en envisageant ce type de support pourrait avoir des conséquences encore difficile à cerner. Les organismes habitués à récupérer l'ADN dans la nature par transfert horizontal pourraient voir apparaître de nouveau gène et l'informatique pourrait modifier le cours de l'évolution et engendrer d'importants risques éthiques.

## Les conséquences sur la mémoire sociale

Les processus d'archivage comme nous l'avons vu repose sur la sélection des documents à archiver pour supporter la mémoire humaine et sociale. Non seulement les échantillons relevés sont sélectionnés et donc ont déjà subi une première phase interprétative mais en plus la société qui interprète cet échantillon va influencer sa narration, sa perception. C'est souvent l'État qui décide le niveau d'intensité émotionnelle rattaché à cet échantillon et impose une « *politique de la mémoire* » (Derrida et Stiegler)<sup>77</sup>.

Il faut alors accepter qu'il existe de nombreuses significations, interprétations en croisant les contextes sociaux, historiques ou encore religieux, cette « *généalogie sémantique* »<sup>78</sup> en parallèle des narrations tacites. Et des données de plus en plus importantes doivent être associées au document numérique pour faciliter ces interprétations futures.

On passe ainsi comme le décrit Pascal Robert<sup>79</sup> : « *d'une mémoire sociale équipée ponctuellement par la technique à une technique qui devient la forme dominante de la mémoire sociale.* »

La mémoire sociale est connectée et donc potentiellement dé-connectable, soumise aux aléas matériels qui peuvent créer une véritable « *panne de mémoire* » engendrant un autre type de risque.

Même si le document numérique ne disparaît pas en tant que tel, si son support matériel n'est qu'un moyen de retranscrire une information à des lecteurs, il est tout de même pris dans un flux dont il est difficilement extractable.

---

<sup>77</sup> Pascal ROBERT, Nathalie PINÈDE, « Communication & Organisation », n° 42, p191-202, 2012, 280 p  
( <http://communicationorganisation.revues.org/3948> , consulté le 25 Mars 2014)

<sup>78</sup> Eric KETELAAR, « (Dé) Construire l'archive Matériaux pour l'histoire de notre temps » ; « Matériaux pour l'histoire de notre temps », N° 82, p65-70, 2006, p138, (<http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2-page-65.htm>, consulté le 15 mai 2014)

<sup>79</sup> Pascal ROBERT, Nathalie PINÈDE, « Communication & Organisation », n° 42, p191-202, 2012, 280 p  
( <http://communicationorganisation.revues.org/3948> , consulté le 25 Mars 2014)

Le droit à l'oubli par exemple est reconnu depuis mai 2014 lorsque la Cour de Justice européenne a rendu un arrêt reconnaissant un « *droit à l'oubli* », concrètement il s'agit du droit pour tout citoyen européen de demander à ce que ses données à caractère personnel n'apparaissent plus, si elles sont erronées ou ne sont plus pertinentes, dans les résultats de moteurs de recherche en ligne, sur Google notamment.

Cela veut-t-il dire pour autant que n'importe qui a la possibilité d'effacer ses traces en ligne ? La réponse est non, même après la publication de cet arrêt de la Cour de justice européenne car dans les faits, la suppression des données totales est moins facile qu'elle ne l'est avec des exemplaires matériels. Une fois le support matériel détruit, l'information n'existe plus, en informatique une sauvegarde a pu être faite et oubliée, non-indexée et donc perdue dans le réseau. Il est « *saisie et redéfinie par cette circulation* »<sup>80</sup> et vit désormais dans cet environnement. S'il perd sa capacité à être tangible et donc destructible il gagne dans sa capacité à être préservé mais peut être trop car la délégation de ces sauvegardes franchit peut être les limites des droits de l'individu.

Ces informations, utilisées à des mauvaises fins pourraient donc effectivement être un véritable risque pour notre société et ces modes de fonctionnement. Avant un document ne devenait archive que lorsqu'il était considéré comme un échantillon représentatif, une preuve, puis classé selon une procédure. Aujourd'hui le numérique fait que tout est créé et conservé au même instant faisant disparaître une hiérarchisation. Dans cette optique la bibliothèque du Congrès au États-Unis a pris la décision d'archiver l'intégralité des messages publiés sur Twitter et en France, on met en place un dépôt légal pour les publications internet à la BNF. En définitive sans sélection tout est mémoire et c'est la définition même de mémoire sociale qui va être soumise à une évolution. L'échantillonnage pour la recherche et l'étude de ces phénomènes s'avère plus complexe et redéfinit aussi la définition de l'archive. On dématérialise cette définition et on se rapproche de celle donnée

---

<sup>80</sup> Pascal ROBERT, Nathalie PINÈDE, « Communication & Organisation », n° 42, p191-202, 2012, 280 p  
( <http://communicationorganisation.revues.org/3948> , consulté le 25 Mars 2014)

Michel Foucault<sup>81</sup> c'est à dire « *la masse des choses dites dans une culture, conservées, valorisées, réutilisées, répétées et transformées* ».

Il faut donc sans arrêt repenser les mécanismes de tri et de hiérarchisation. Le retour en arrière n'est plus possible et le principal besoin émergent, avec l'arrivée de nouveaux supports d'archivage pérenne, serait donc non pas la mise en place d'une réglementation ou de normes fixes mais plutôt un système évolutif qui nécessitera toute la souplesse propre à la gestion d'un flux. Il s'agirait peut-être de tirer profit intelligemment des nouveaux moyens de stockage, la possibilité de pouvoir effectivement garder à disposition des quantités énormes d'informations va certes, créer un océan d'informations insondable par l'homme, mais également lui permettre d'avoir une vue plus large, une vision plus « *mondialisée* ». La mise en place de nouveaux dispositifs permet donc en un sens la création de nouvelle méthodologie de recherche et d'étude de notre histoire.

## **DONNER DU SENS : L'UTILISATION DE LA SEMIOTIQUE**

### **Enjeux de la capitalisation de l'information**

La capitalisation radicale de l'information sur des supports stables c'est aussi la possibilité de réaliser des opérations de penser/classer<sup>82</sup>, ces technologies intellectuelles<sup>83</sup>, comme le souligne Pascal Robert<sup>84</sup>, qui permettent cette décontextualisation.

Cette opération permet de restituer le document, de le découper et de lui associer des données universelles qui lui permettront de participer à une mise en réseau complexe et structurée de l'information. On pourrait qualifier cet ensemble de patrimoine intellectuel numérique, englobant les productions passées et les

---

<sup>81</sup> Ibid.

<sup>82</sup> Pascal Robert, « Qu'est-ce qu'une technologie intellectuelle », *Communication et langages*, n°123, 2000, p. 97-114

<sup>83</sup> Jack Goody, *La Raison graphique*, Paris, Minuit, 1979

<sup>84</sup> Pascal ROBERT, Nathalie PINÈDE, « *Communication & Organisation* », n° 42, p191-202, 2012, 280 p

( <http://communicationorganisation.revues.org/3948> , consulté le 25 Mars 2014)

productions à venir, on peut identifier plusieurs productions à valeurs patrimoniales. Le patrimoine intellectuel matériel numérisé, notamment ces fonds, dits patrimoniaux, se procurent une toute nouvelle dimension avec le numérique. Les ressources numériques acquises avec des partenariats externes et enfin la production intellectuelle interne. Les travaux des chercheurs comme les thèses ou les articles mais aussi ceux des étudiants, ou les travaux documentaires des bibliothécaires.

Le patrimoine numérique devient donc un enjeu à valeurs multiples. Dans un premier temps la recherche scientifique, nous l'avons vu se base sur des savoirs préalablement existants. Organisée rationnellement, l'accessibilité des connaissances est devenue essentielle à son exploitation. Il ya aussi une valeur économique, omniprésente, car la production de ces savoirs demande des financements, conditionnées souvent par la qualité des productions passés et surtout de sa perception dans ces milieux de recherche. La valeur communicationnelle est liée à cette perception, les ressources doivent être visibles, accessibles et surtout être portées par des équipes motivées et managées pour être investies dans des projets de valorisation.

Ces valeurs sont démultipliées par le caractère numérique. D'une part son accessibilité qui permet tout de même une gestion adaptée des droits d'accès avec pourtant une grande facilité de partage qui améliore considérablement les échanges entre les acteurs de la recherche. Les projets comme l'ORI-OAI que nous avons présentés sont des exemples qui illustrent parfaitement ces développements. La vision globale du patrimoine qu'autorise ce nouveau système permet d'accéder à la compréhension de ses évolutions. Cette dématérialisation permet une démultiplication des usages et donc une démultiplication du patrimoine intellectuel.

## **Le rôle de l'archiviste et l'analyse discursive**

Ces nouveaux usages des documents archivés, de ces savoirs pérennisés favorisés par l'arrivée des nouveaux supports donnent à l'archiviste un moyen de remonter le cours de création du document. La démarche d'enregistrement n'est plus passive,

automatique, au contraire elle demande un engagement différent, interactif, faisant de lui un « *créateur de valeur dans la chaîne documentaire* »<sup>85</sup>.

L'élaboration, la construction et l'utilisation des archives sont toujours plus ou moins influencées par des facteurs sociaux et culturels comme nous l'avons démontré. L'utilisation elle-même a varié et est aussi multiple que les acteurs qui l'approvisionnement. L'exploitation d'une archive est encadrée par des lois et des normes. Ce cadre législatif ne sera sensiblement pas le même pour un cabinet d'avocats ou pour des ingénieurs. Ces méthodologies archivistiques doivent donc associer la culture « *organisationnelle des concepteurs d'archive* »<sup>86</sup> dans ces multiples facettes (politiques, économiques, culturelles ou encore religieuses).

La contextualisation du document comme nous l'entendons par ces critères ne permet pas une signification complète qui se construit aussi par l'invisible, le non dévoilé, ce « *discours invisible des créateurs et « archiveurs* »<sup>87</sup>. Les utilisateurs, à la création d'un document, démarrent eux-mêmes ce processus archivistique. Ils partagent ensuite via le numérique et s'organisent en communauté, permettant d'établir des liens entre archives privées et public. C'est le révélateur des angoisses et des attentes de l'utilisateur qui reconstruit l'archive à sa manière, sélectionne, interprète et consomme en un sens l'information. Ces documents devenus presque communautaires, deviennent actifs comme nous l'avons vu, interconnectés et dépendant de ces réseaux demandant aux instances mémorielles de s'adapter à ces nouvelles structures informationnelles. C'est en effet à eux d'endosser cette mission nouvelle d'utilité sociale, continuité de leur fonction patrimoniale. Nous l'avons également démontré, le droit à l'oubli est une des initiatives récentes pour protéger ces utilisateurs dont cette soudaine implication massive les confronte à un risque. L'archiviste doit donc prendre en compte dans son processus de sélection la protection des données personnelles tout en sélectionnant des documents notables qui auront été revus, commentés et enrichis.

---

<sup>85</sup> Eric KETELAAR, « (Dé) Construire l'archive Matériaux pour l'histoire de notre temps » ; « Matériaux pour l'histoire de notre temps », N° 82, p65-70, 2006, p138, (<http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2-page-65.htm>, consulté le 15 mai 2014)

<sup>86</sup> Ibid.

<sup>87</sup> Eric KETELAAR, « (Dé) Construire l'archive Matériaux pour l'histoire de notre temps » ; « Matériaux pour l'histoire de notre temps », N° 82, p65-70, 2006, p138, (<http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2-page-65.htm>, consulté le 15 mai 2014)

*« Le témoignage physique peut nous en dire autant ou plus sur un document et son contexte que le contenu informatif lui-même »<sup>88</sup>*

---

L'analyse du contenu doit être faite à la lumière de ces enrichissements et l'archiviste doit pouvoir mettre en place une analyse sémantique qui permet une organisation de cette documentation plus efficace, pertinente et neutre.

L'archiviste aurait alors comme tâche d'appréhender non seulement le sens du document mais aussi les structures des discours, la profondeur du texte qui permettent la contextualisation selon des auteurs ou des méthodologies de rédaction.

Il se voit également confié la mission de mettre au point des solutions pour détecter à partir de grands corpus de textes des règles de grammaire ou des règles de traduction d'une langue en une autre. Il nous faut dans un premier temps revenir à l'opération de décontextualisation pratiquée par le document traditionnel. Rendu possible grâce à la dissociation matériel du contenant et du contenu.

## **Le développement d'une archive sémantique**

Depuis plusieurs années, l'objectif général est de faire évoluer le web, immense source de données, vers ce que l'on a donc appelé le web sémantique. L'objectif étant de développer un langage de métadonnées plus précis pour permettre aux machines d'identifier des documents plus précisément pour le compte des utilisateurs. La possibilité de collecter et de structurer les données fonctionne grâce à la mise en place de description formelle encadrée par le W3C23.

On pourrait rapprocher ces nouveaux aspects du web de la pratique de l'analyse discursive, étude du discours. D'après Pécheux<sup>89</sup> lorsqu'on analyse on suppose un regard subjectif sur la sémiotique des mots et c'est grâce à des processus sémantique que l'on peut établir un rapport.

---

<sup>88</sup> Heald, Carolyn. « Is There Room for Archives in the Postmodern World ? », p88-101, 1996, American Archivist- Cité par Eric KETELAAR dans : Ibid.

<sup>89</sup> Michel MELOT, Jean-Michel SALAÛN, Roger T PÉDAUQUE, « Le Document à la lumière du numérique: forme, texte, médium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité », 2007, Toulouse, Cepaduès éditions, 213 p.

En d'autres termes l'analyse discursive s'attache à faire émerger les structures permettant d'interpeller le lecteur, celui par laquelle un dialogue silencieux s'instaure entre le discours écrit et le lecteur lui permettant de réagir, songer. Le web sémantique tente lui aussi de se construire une science discursive, les métadonnées sont les prémices, les briques sur lesquelles le discours numérique et ses règles ce construisent. Conçu pour structurer en parallèle du document les informations utiles à son indexation, ses relations à d'autres documents ou encore pour décrire les conditions d'utilisations.

De nos jours, les pratiques documentaires et les outils à disposition convergent vers les technologies de ce web sémantique<sup>90</sup>.

L'objectif du web sémantique est d'améliorer les capacités de la machine à interpréter la demande de l'utilisateur. Beaucoup de manipulations lorsqu'on utilise le web sont encore à la charge de l'utilisateur car la recherche d'information est encore conçue pour être exploitée par un cerveau et une logique humaine

La solution réside dans la création de langage et d'outils permettant de représenter de façon formelle la connaissance pour que la machine puisse avoir accès au sens des documents.

Les acteurs de l'information ont déjà l'habitude de produire pour les images des descriptions textuelles et pour cela ils s'appuient sur des ressources terminologiques : dictionnaires de noms propres, listes d'autorité, schémas de classifications, thésaurus spécialisés ou généraux, etc. On peut alors dans une certaine mesure automatiser la production de ces métadonnées et améliorer la recherche d'image.

Les ontologies sont justement un ensemble structuré de termes et de concepts qui permettent d'appréhender le sens d'un champ d'informations, il s'agit de modéliser les connaissances d'un domaine pour faciliter l'accès des communautés qui y sont associées.

---

<sup>90</sup>Raphaël TRONCY, « Nouveaux outils et documents audiovisuels : les innovations du web sémantique » ; « Documentaliste-Sciences de l'Information », Vol. 42, p.392 – 404, 2005

([http://www.cairn.info/resume.php?ID\\_ARTICLE=DOCSI\\_426\\_0392](http://www.cairn.info/resume.php?ID_ARTICLE=DOCSI_426_0392), consulté le 30 novembre 2013)

Le langage RDF est utilisé pour décrire les ressources sémantiques et le langage OWL pour les ontologies du web sémantique. Recommandé par le W3C, le RDF (Resource Description Framework), utilise par exemple ce langage pour décrire les relations qu'ont des gens avec d'autres et les « choses » autour d'eux.

Petit à petit les technologies du web sémantique se développent et permettent une plus grande précision dans les requêtes et les réponses, l'ouverture des ressources permet également de rendre plus pertinent ces réponses. Le développement de ressources ouvertes serait donc d'un intérêt majeur et le bénéfice répercuté sur tous les utilisateurs. En effet si jusqu'au début du XXIème siècle on exprimait des théories scientifiques à partir des propositions basées de petits échantillons, aujourd'hui avec le développement de gigantesques volumes d'observations il y a fort à parier que ces propositions prennent de l'ampleur. L'utilisation d'ordinateurs puissant de calcul sera alors un moyen de mettre à l'épreuve des théories qui auront été « comprises » par l'ordinateur, pas seulement stockées mais davantage « apprises ». Les barrières de langue et de grammaire pourraient être résolues à l'aide d'un algorithme permettant une libre circulation des contenus. L'ordinateur selon Dowek<sup>91</sup> pourrait alors devenir un « prédicateur » et relégué les scientifiques à l'élaboration d'algorithmes complexes avec toutefois un contrôle et une expérimentation humaine de confirmation.

Nous pouvons dans ce contexte discuté de l'ouvrage de Roger T. Pédaque « *Le document à la lumière du numérique* ». En réalité c'est le nom collectif d'un réseau de scientifiques francophones qui travaillent dans des domaines divers d'expertises des SIB. Le choix d'un pseudonyme collectif avait pour objectif de montrer les mutations des auteurs et des idées, sans attribuer la parenté à un auteur précis.

---

<sup>91</sup> Gilles DOWEK, « L'explosion mémorielle change la donne » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p26-31, Novembre 2013, 188p

Une préface de Michel Melot décrit les procédés mis en œuvre dans cet ouvrage, il explique ainsi que Roger T. Pédauque<sup>92</sup> s'interroge sur la notion de document à l'heure du numérique. Le contexte actuel, l'intense documentation que produit notre société et la modernisation des supports entraînent un besoin de redéfinition de la notion même de document.

Pour Melot, la personnalité de Roger semble s'affirmer dans cet ouvrage qui regroupe en réalité 3 textes différents. Une introduction rédigée par Jean-Michel Salaün présente ainsi la réunion de cet ensemble et la construction de la réflexion transversale mise en œuvre pour l'élaboration de ces textes. Il y justifie la création du pseudonyme de Roger T. Pédauque qui permet de nombreuses lectures sans pour autant attacher à ces réflexions la pensée unique d'un auteur et ainsi nul n'est gardien, selon lui, de la « *pensée pédauquienne* ». A partir de ces trois textes il va tenter d'esquisser trois pistes, qu'il se défend d'être un programme de recherche, mais qui permet de se rendre compte des processus découlant de ces trois réflexions. La tridimensionnalité du document, l'empreinte cognitive des communautés et le paradoxe de Roger.

Le premier texte, Pédauque 1 tente d'expliquer les reformulations que vont impliquer le numérique et les variations de définition subi par le document. C'est à la fois un état des lieux et à la fois une prédiction sur l'avenir de cette définition. Il envisagera alors le document sous trois de ces aspects : forme, signe et médium. Il présentera chacun de ces aspects avec leurs spécialités, leurs évolutions puis en proposant plusieurs définitions et les questions qu'elles pourraient soulever.

Il y définit la notion de document comme forme, s'insérant dans un processus éducatif pour y accéder, l'école permettant d'acquérir les outils intellectuels nécessaires à la compréhension d'un système de signe. Les évolutions seront évidemment celles provoquées par les supports définis autrefois par « *document traditionnel = support + inscription.* »

---

<sup>92</sup> Michel MELOT, Jean-Michel SALAÛN, Roger T PÉDAUQUE, « Le Document à la lumière du numérique: forme, texte, médium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité », 2007, Toulouse, Cepaduès éditions, 213 p.

Ainsi la définition qu'il propose sera : « *Un document numérique est un ensemble de données organisées selon une structure* ». Aujourd'hui le support papier qui servait de structure est devenu le logiciel et l'inscription la donnée, donnant une nouvelle équation :

« *Document numérique = structures + données* »

Élargie à de nouvelles possibilités de composition avec le XML (structure) et ses feuilles de style. Celui-ci pouvant être appréhendé par l'homme. Ce qui donnerait comme équation du type :

« *Document XML = données structurées + mise en forme* »

Dans une seconde partie, il s'agira d'évoquer le document comme objet signifiant, le sens et non plus la forme. Le document correspond alors à l'équation « *document = inscription + sens* ». Il en donne une définition « *Un document numérique est un texte dont les éléments sont potentiellement analysables par un système de connaissance en vue de son exploitation par un lecteur compétent.* »<sup>93</sup>.

Il s'agira de traiter dans ce texte des différents aspects du web sémantique. Des apports possible par l'ajout de connaissance tout en limitant le bruit causé par un attachement trop fort au mot et non pas au sens, ou encore aux idées auxquelles il peut renvoyer. Le tissage d'un réseau entre les documents va apporter la possibilité d'une navigation entre document et la mise en place de métadonnées, de data, permet de réduire le bruit que pourrait causer une recherche mot à mot.

L'équation devenant alors « *Document web sémantique = texte informé + connaissances* » et deviendrait « *Document web sémantique = texte informé + ontologies.* »

Dans la seconde partie de l'ouvrage, Pédaïque 2, essaie de rattacher l'évolution des pratiques avec une communication plus humaine. La linguistique ne suffit pas à analyser le texte, il faut pour cela imaginer une structure indépendante de son contenu. Cette partie prône un retour aux fondamentaux avec une modification de

---

<sup>93</sup> Michel MELOT, Jean-Michel SALAÛN, Roger T PÉDAUQUE, « Le Document à la lumière du numérique: forme, texte, médium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité », 2007, Toulouse, Cepaduès éditions, 213 p.

la notion de texte dans un espace multimédia et d'analyser le rôle de la médiation humaine ou automatisée.

Le web sémantique est une notion conçue par Tim Berners Lee, il est perçu comme une extension du web qui devrait le transformer en un « *vaste espace d'échange de ressources entre êtres humaines et machines* ». Le but étant de décharger les utilisateurs d'une partie des tâches de recherche

La notion d'ontologie devient alors essentielle dans la mise en place d'une sémantique formelle. Pour cela il faut que les contributeurs du web respectent une infrastructure commune qu'on appelle le « *cake* » de Tom Berners-Lee, qui s'organise en couche de différents langages.

L'ouvrage s'intéresse aussi à la dimension intellectuelle<sup>94</sup> du document. Cette analyse est donc faite en deux voies qui rejoignent les autres dimensions documentaires repérées par ailleurs. Il apparaît que si un groupe, un pays ou un individu produit des documents avec des caractéristiques propres dans leurs productions, leurs discussions, alors Salaün envisage la possibilité d'une empreinte cognitive. Autant d'informations qui offrent des possibilités inédites derrière de mise en valeur du document et son exploitation.

Enfin la troisième partie est un article collectif écrit RTP-DOC qui justement défend comme hypothèse un processus inverse où ce n'est pas la modernisation des supports qui ferait évoluer le document mais celle de la société qui modifie notre façon de le traiter. C'est ce que Salaün va traiter dans sa partie Le Paradoxe de Roger. Il met en avant la *redocumentarisation* en insistant sur les relations entre les transformations documentaires et les changements sociaux. L'article suggère un découpage par fonctions, par territoires et s'attache à mettre en évidence la recomposition documentaire comme source et conséquence du jeu des protagonistes.

---

<sup>94</sup> Michel MELOT, Jean-Michel SALAÜN, Roger T PÉDAUQUE, « Le Document à la lumière du numérique: forme, texte, médium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité », 2007, Toulouse, Cepaduès éditions, 213 p.

Mais l'idée principale est que cette notion est construite sur un paradoxe, le paradoxe de « Roger » que Salaün se propose d'évoquer ainsi « *le Web favorise conjointement deux mouvements opposés : le développement d'échanges spontanés (conversations) et leur fixation sur support public, pérenne et documenté.* »<sup>95</sup>. La conclusion de ces trois textes permet d'établir un programme de recherche.

---

<sup>95</sup> Michel MELOT, Jean-Michel SALAÜN, Roger T PÉDAUQUE, « Le Document à la lumière du numérique: forme, texte, médium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité », 2007, Toulouse, Cépaduès éditions, 213 p.

## CONCLUSION

---

Nous l'avons vu, l'évolution rapide des supports et l'émergence de multiples créateurs d'informations plongent les institutions dans une course effrénée au développement de nouvelles stratégies. La place de ces communautés génératrices et utilisatrices de ces savoirs est ce bien commun.

Ce type de transformation technique ne s'invite pas que dans la science mais dans le culturel. C'est ainsi que la place de ces productions communes<sup>96</sup> forme cette richesse. Ces capacités que nous avons développées et que nous développons encore de collecter, référencer, stocker et partager engage dans une démarche de construction du passé. Les historiens de demain auront un travail différent de celui d'aujourd'hui. La visibilité généalogique prendra une toute nouvelle ampleur. En définitive le web vient transformer ce qui était éphémère en proto-document<sup>97</sup>.

Il ne s'agit donc plus du résultat de stratégies d'industriels ou de groupes de passionnés mais davantage d'un mouvement structurel qui, posé dans l'architecture du réseau dépassera la volonté des acteurs. En d'autres termes la mise en place d'un cadre fixe de production de l'information permettrait à cette dernière de trouver une autonomie de création à travers ses utilisateurs. Le résultat souhaité et mis en avant dans l'ouvrage de Pédauque c'est justement la mise en place de méthodes génériques qui permettront une compréhension améliorée et un développement plus cohérent d'outils de production, d'échange et de manipulation du document. La réflexion autour de ces thèmes est désormais ouverte et encadrée

---

<sup>96</sup> LE CROSNIER Hervé, ERTZSCHEID Olivier, PEUGEOT Valérie, MERCIER Silvère, BERTHAUD Christine, CHARNAY Daniel, MAUREL Lionel, « Vers les « communs de la connaissance » ; « Documentaliste-Sciences de l'Information », Vol. 48, p48-59, 2011, 80p ([http://www.cairn.info/zen.php?ID\\_ARTICLE=DOCSI\\_483\\_0048](http://www.cairn.info/zen.php?ID_ARTICLE=DOCSI_483_0048), consulté le 1er Juillet 2014)

<sup>97</sup> Gilles DOWEK, « L'explosion mémorielle change la donne » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p26-31, Novembre 2013, 188p



## *Sources*

### **AFNOR:**

<http://www.afnor.org/>

### **Archimag:**

<http://www.archimag.com>

### **CINES:**

<https://www.cines.fr>

### **EUDAT:**

<http://www.eudat.eu/2nd-conference>

### **Facebook:**

<https://www.facebook.com>

### **ORI-OAI:**

<http://www.ori-oai.org/>

### **UNESCO :**

<http://unesdoc.unesco.org>

### **Wikipédia :**

<http://fr.wikipedia.org>



## *Bibliographie*

### **Monographie :**

MELOT Michel, SALAÛN Jean-Michel, PÉDAUQUE Roger T, « Le Document à la lumière du numérique: forme, texte, médium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité », 2007, Toulouse, Cépaduès éditions, 213 p.

### **Articles :**

AFP, « La guerre contre «l'obsolescence programmée» est déclarée », Avril 2013 ([http://www.liberation.fr/economie/2013/04/23/la-guerre-contre-l-obsolescence-programmee-est-declaree\\_898359](http://www.liberation.fr/economie/2013/04/23/la-guerre-contre-l-obsolescence-programmee-est-declaree_898359), consulté le 25 Mars 2014)

ABITEBOUL Serge, SENELLART Pierre, « Un déluge de données » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p32-35, Novembre 2013, 188p

BANAT-BERGER Françoise, « Les archives et la révolution numérique : Le Débat », n° 158, p70-82, 2010, 192p ([http://www.cairn.info/resume.php?ID\\_ARTICLE=DEBA\\_158\\_0070](http://www.cairn.info/resume.php?ID_ARTICLE=DEBA_158_0070), consulté le 27 novembre 2013)

BLIN Frédéric, « La numérisation des bibliothèques » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p48-51, Novembre 2013, 188p

BEAUDRY Guylaine, BOISMENU Gérard, « Expertise technique et organisationnelle ; L'archivage », Février 2001

(<http://revues.enssib.fr/titre/6tecnic/5archivage/>, consulté le 25 mai 2014)

BORDAGE Frédéric, « Le quartz, support de stockage d'avenir », Novembre 2012  
(<http://www.greenit.fr/article/materiel/le-quartz-support-de-stockage-davenir-4626>, consulté le 15 Mars 2014)

CHAILLAN Franck « La mémoire biologique est-elle codée ? » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p110-117, Novembre 2013, 188p

CLEYET-MICHAUD Rosine, « La sélection dans les archives contemporaines », 2001 (<http://www.archivesdefrance.culture.gouv.fr/gerer/records-management-et-collecte/principes/>, consulté le 15 Mars 2014)

CORMIER Bruno, « Samsung SSD 850 Pro : de la mémoire flash 3D très rapide et ultra-durable », Juillet 2014 (<http://www.01net.com/editorial/622914/samsung-ssd-850-pro-memoire-flash-3d-tres-rapide-et-ultra-durable/>, consulté le 2 Juillet 2014)

CROCHET- DAMAIS Antoine, "Swift : Apple dévoile le remplaçant d'Objective-C", Juin 2014, (<http://www.journaldunet.com/developpeur/outils/apple-swift-et-sdk-ios8-0614.shtml>, consulté le 2 Juillet 2014)

DAUBIN Vincent, PENEL Simon, TANNIER Éric, « L'ADN mémoire numérique du vivant »; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p102-108, Novembre 2013, 188p

DAUMAS Francis, MASSOL Marion, «Au cœur d'un centre d'archivage» ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p43-45, Novembre 2013, 188p

DESSIMOZ Christophe « Stocker de l'information dans l'ADN » ; « BIG BANG numérique Les données massives changent-elles le monde ? Pour la Science, n° 433, p100-101, Novembre 2013, 188p

DI PIETRO Christelle, « Une arche pour sauver les livres imprimés en Californie » Mars 2012 (<http://www.enssib.fr/en/breves/2012/03/05/une-arche-pour-sauver-les-livres-imprimees-en-californie>, consulté le 6 Juin 2014)

DOWEK Gilles, « L'explosion mémorielle change la donne » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p26-31, Novembre 2013, 188p

KETELAAR Éric, « (Dé) Construire l'archive Matériaux pour l'histoire de notre temps » ; « Matériaux pour l'histoire de notre temps », N° 82, p65-70, 2006, p138, (<http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2-page-65.htm>, consulté le 15 mai 2014)

GALLAGHER Ryan, MAASS Peter, « Inside the NSA's Secret Efforts to Hunt and Hack System Administrators », Mars 2014 (<https://firstlook.org/theintercept/article/2014/03/20/inside-nsa-secret-efforts-hunt-hack-system-administrators/>, consulté le 15 Août 2014)

GAUTHIER Christophe, "Comment ça marche : la mémoire Flash", Mars 2005 (<http://www.01net.com/editorial/294823/comment-ca-marche-la-memoire-flash/>, consulté le 15 Juin 2014)

HANSSON SVEN Ove, « Les incertitudes de la société du savoir », Revue internationale des sciences sociales, n° 171, p43–51, 2002, 192p (<http://www.cairn.info/revue-internationale-des-sciences-sociales-2002-1-p-43.htm>, consulté le 25 juin 2014)

LECLERCQ Denis « Le disque dur à laser pulsé, prochaine révolution du stockage ? », Février 2012 (<http://www.pcworld.fr/stockage/actualites,le-disque-dur-laser-pulse-presque-finalise,524649,1.htm?comments=1#comments>, consulté le 15 Mars 2014)

LALOË Franck, SPITZ Erich « La quête d'un support numérique durable » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p78-82, Novembre 2013, 188p

LAMBERT Jean-Michel, PERDEREAU Jacques, « Recherche dans le cadre du Groupement d'Intérêt Scientifique sur les Supports Pérennes d'Archivage des Données Numériques », Juin 2013 (<http://www.lne.fr/publications/recherche/conservation-donnees-dvdr-juin-2013.pdf>, consulté le 25 mai 2014)

LAMBERT Jean-Michel, PERDEREAU Jacques, « Qualité des disques Blu-ray enregistrables pour l'archivage des données numérique », Novembre 2011 (<http://www.lne.fr/publications/recherche/Etude-qualite-bd-r.pdf>, consulté le 15 Février 2014)

LE CROSNIER Hervé, ERTZSCHEID Olivier, PEUGEOT Valérie, MERCIER Silvère, BERTHAUD Christine, CHARNAY Daniel, MAUREL Lionel, « Vers les « communs de la connaissance » ; « Documentaliste-Sciences de l'Information », Vol. 48, p48-59, 2011, 80p ([http://www.cairn.info/zen.php?ID\\_ARTICLE=DOCSI\\_483\\_0048](http://www.cairn.info/zen.php?ID_ARTICLE=DOCSI_483_0048), consulté le 1<sup>er</sup> Juillet 2014 )

LUPOVICI, Catherine, « Les stratégies de gestion et de conservation préventive des documents électroniques », Bulletin des Bibliothèques de France, N°4, p43-54, Avril 2000 (<http://bbf.enssib.fr/consulter/bbf-2000-04-0043-004>, consulté le 20 novembre 2013)

MAUREL Lionel, « Les communs de la connaissance : un autre regard sur l'information » ; « Documentaliste-Sciences de l'Information », Vol. 48, p48-59, 2011, (<http://cergy0051.lb.u-cergy.fr/revue-documentaliste-sciences-de-l-information-2011-3-page-48.htm>, consulté le 26 Mai 2014)

MUSSOU Claude, «Et le web devint archive : enjeux et défis », juin 2012 (<http://www.ina-expert.com/e-dossier-de-l-audiovisuel-sciences-humaines-et-sociales-et-patrimoine-numerique/et-le-web-devint-archive-enjeux-et-defis.html>, consulté le 25 Mai 2014)

NDUNGIDI Patrick, « Une nouvelle vie pour les archives audiovisuelles africaines », Mai 2014 ([http://www.huffingtonpost.fr/patrick-ndungidi/une-nouvelle-vie-pour-les-archives-audiovisuelles-africaines\\_b\\_5381426.html?utm\\_hp\\_ref=fr-culture](http://www.huffingtonpost.fr/patrick-ndungidi/une-nouvelle-vie-pour-les-archives-audiovisuelles-africaines_b_5381426.html?utm_hp_ref=fr-culture), consulté le 2 Juin 2014)

PERNIOLA Luca, « Les mémoires du futur » ; « BIG BANG numérique Les données massives changent-elles le monde ? », Pour la Science, n° 433, p92-99, Novembre 2013, 188p

RABUT Élisabeth « Que faites-vous de nos archives ? Massification, sélection, conservation » ; « Le Débat », n° 158, 2010 (<http://bbf.enssib.fr/consulter/bbf-2000-04-0043-004>, consulté le 10 Mars 2014)

ALAN Leo, «Writing the Book in DNA», Août 2012 (<http://hms.harvard.edu/news/writing-book-dna-8-16-12>, consulté le 6 Janvier 2014)

ROBERT Pascal, PINÈDE Nathalie, « Communication & Organisation », n° 42, p191-202, 2012, 280 p

(<http://communicationorganisation.revues.org/3948>, consulté le 25 Mars 2014)

TRIBOT Sébastien, « L'archival Disc: le successeur du Blue Ray », 13 Juin 2014 ([http://www.techniques-ingenieur.fr/actualite/high-tech-thematique\\_193/l-archival-disc-le-successeur-du-blue-ray-article\\_231715/](http://www.techniques-ingenieur.fr/actualite/high-tech-thematique_193/l-archival-disc-le-successeur-du-blue-ray-article_231715/), consulté 10 Janvier 2014)

TRONCY Raphaël, « Nouveaux outils et documents audiovisuels : les innovations du web sémantique » ; « Documentaliste-Sciences de l'Information », Vol. 42, p.392 – 404, 2005

([http://www.cairn.info/resume.php?ID\\_ARTICLE=DOCSI\\_426\\_0392](http://www.cairn.info/resume.php?ID_ARTICLE=DOCSI_426_0392), consulté le 30 novembre 2013)

UNTERSINGER Martin, « Devenu trop grand, Internet a subi des perturbations », Août 2014 ([http://www.lemonde.fr/pixels/article/2014/08/15/devenu-trop-grand-internet-est-tombe-en-panne\\_4472153\\_4408996.html](http://www.lemonde.fr/pixels/article/2014/08/15/devenu-trop-grand-internet-est-tombe-en-panne_4472153_4408996.html), Consulté le 25 Août 2014)

YONG Ed, « Synthetic double-helix faithfully stores Shakespeare's sonnets », Janvier 2013 (<http://www.nature.com/news/synthetic-double-helix-faithfully-stores-shakespeare-s-sonnets-1.12279>, consulté le 12 Février 2014)

### **Workshop:**

« Archivage pérenne de données numériques », 8 et 9 octobre 2013 à Paris organisé par le GIS-SPADON et le GDR Verres ([http://www.ustverre.fr/site/images/Journees/Workshop\\_GDR\\_Archivage.pdf](http://www.ustverre.fr/site/images/Journees/Workshop_GDR_Archivage.pdf), consulté le 12 Juin 2014)

## *Table des annexes*

**ANNEXE 1 : PERDEREAU JACQUES, « Recherche Sur La Conservation Des Archives Numériques Dans Le Cadre Du Gis-Don – Tenue Dans Le Temps Des Disques Enregistrables Dvd+ R Et Dvd En Verre Syylex À Température Et Hygrométrie Élevées. », JUIN 2012**

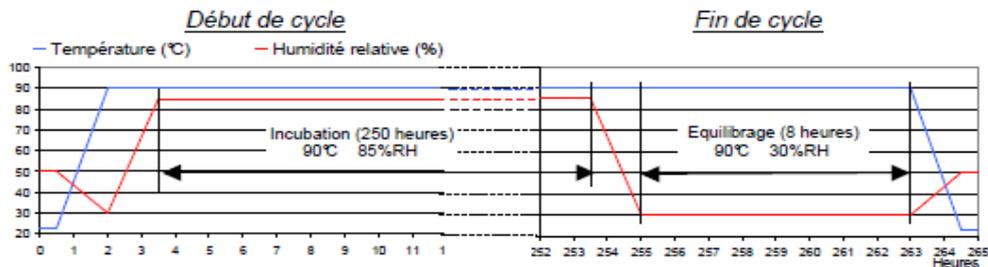
# ANNEXE 1 : RESULTAT DES ESSAIS DE VIEILLISSEMENT ACCELERE



- DVD sur support dans l'enceinte climatique Climats-

Afin de faire les analyses intermédiaires, on doit retirer les disques de l'enceinte. Pour cela on doit revenir progressivement aux conditions climatiques normales. En effet, afin d'éviter la formation de bulles d'eau dans le polycarbonate, il est nécessaire de « purger » le polycarbonate de l'eau qu'il a absorbée à température élevée. Pour cela, le taux d'humidité est d'abord diminué jusqu'à 30% tout en maintenant la température à 90°C. Un palier dit « d'équilibrage », d'une durée de 8 heures, est ensuite effectué, en maintenant toujours la température de 90°C. Les graphiques ci-dessous montrent l'évolution de la température et du degré hygrométrique au début et à la fin d'un cycle de 250 heures à 90°C et 85% d'humidité.

*Profil de Température-Hygrométrie pour une exposition de 250 heures à 90°C et 85%RH*



- Cycle de vieillissement accéléré -

## 5. RESULTATS DES ESSAIS DE VIEILLISSEMENT ACCELERE

Les résultats sont données ci-après par type de disque.

La fin de vie du DVD correspond à l'une des conditions suivantes

- PI8 (moyenne sur le disque) est supérieur à 280,
- ou présence d'erreurs incorrigibles POF,
- ou illisibilité partielle ou totale du DVD.

Il suffit qu'une de ces conditions soit satisfaite avec un des deux analyseurs pour que le disque soit considéré comme "mort".

## *Glossaire*

**AFNOR**<sup>98</sup>: L'Association française de normalisation est un organisme officiel français de normalisation. Créée en 1926, c'est elle qui représente la France auprès de l'Organisation internationale de normalisation (ISO), du Comité européen de normalisation (CEN).

**ASCII**<sup>99</sup>: American Standard Code for Information Interchange (Code américain normalisé pour l'échange d'information), c'est une norme de codage de caractères en informatique ancienne et connue pour sa forte influence sur les codages de caractères qui lui ont succédé.

**CNIL**: Commission Nationale de l'Informatique et des Libertés est l'institution en charge garant du respect de la vie privée et des libertés l'informatiques.

### **DÉPÔT LÉGAL**:<sup>100</sup>

Le dépôt légal peut être une imposition légale ou l'incitation des producteurs et/ou des diffuseurs de déposer dans une institution, comme la bibliothèque nationale du pays, un ou plusieurs exemplaires des documents qu'ils produisent dans le but d'assurer le contrôle bibliographique universel d'élaborer des bibliographies nationales.

**JPEG**<sup>101</sup>: Joint Photographic Experts Group JPEG (acronyme de Joint Photographic Experts Group) est un format d'enregistrement d'une image fixe pour une représentation numérique compressée

---

<sup>98</sup> Article Wikipédia : [http://fr.wikipedia.org/wiki/Association\\_française\\_de\\_normalisation](http://fr.wikipedia.org/wiki/Association_française_de_normalisation)

<sup>99</sup> Article Wikipédia : [http://fr.wikipedia.org/wiki/American\\_Standard\\_Code\\_for\\_Information\\_Interchange](http://fr.wikipedia.org/wiki/American_Standard_Code_for_Information_Interchange) Consulté le 23 Mai 2014

<sup>100</sup> Article Wikipédia: [http://fr.wikipedia.org/wiki/Dépôt\\_légal](http://fr.wikipedia.org/wiki/Dépôt_légal)

<sup>101</sup> Article Wikipédia : <http://fr.wikipedia.org/wiki/JPEG> Consulté le 23 Mai 2014

**TEX**<sup>102</sup>: Il s'agit d'un système logiciel de composition de documents particulièrement utilisé dans le milieu scientifique.

**UNICODE** Développée par le Consortium Unicode, il s'agit d'un standard informatique permettant l'échange de textes dans différentes langues, à un niveau mondial. Unicode 7.0 est la dernière version disponible depuis le 16 juin 2014

**XML**: Extensible Markup Language c'est un langage de programmation très utilisé pour l'échange d'information par Internet.

---

<sup>102</sup> Article Wikipédia : <http://fr.wikipedia.org/wiki/TeX> Consulté le 20 Août 2014

## *Table des illustrations*

FIGURE 1 : COMPARAISON D'UN DISQUE DUR CLASSIQUE ET D'UN DISQUE DUR SSD.....	20
FIGURE 2 : FONCTIONNEMENT DES MEMOIRES A OXYDE PERNIOLA LUCA, « LES MEMOIRES DU FUTUR » ; « BIG BANG NUMERIQUE LES DONNEES MASSIVES CHANGENT-ELLES LE MONDE ? », POUR LA SCIENCE, N° 433, NOVEMBRE 2013, PP. 92-99.....	22
FIGURE 3 : DEVELOPPEMENT DES SUPPORTS NUMERIQUES.....	23
FIGURE 5 L'ADN, MEMOIRE NUMERIQUE DU VIVANT VINCENT DAUBIN, SIMON PENEL ET ÉRIC TANNIER POUR LA SCIENCE.....	80



# Table des matières

<b>SIGLES ET ABREVIATIONS .....</b>	<b>7</b>
<b>INTRODUCTION.....</b>	<b>11</b>
<b>LES TECHNOLOGIES DE LA MEMOIRE.....</b>	<b>13</b>
<b>Les supports matériels .....</b>	<b>13</b>
<i>Qu'est ce qu'une mémoire? .....</i>	<i>13</i>
<i>Quels fonctionnements ? .....</i>	<i>15</i>
<i>L'arrivée de la mémoire flash .....</i>	<i>17</i>
<i>Les progrès de la mémoire informatique .....</i>	<i>18</i>
<i>Les mémoires magnétiques .....</i>	<i>20</i>
<i>Les nouvelles mémoires.....</i>	<i>22</i>
<i>Les supports actuels : entre applications et réflexions.....</i>	<i>23</i>
<i>Graver l'information .....</i>	<i>24</i>
<i>Les disques :.....</i>	<i>24</i>
<i>La standardisation :.....</i>	<i>26</i>
<i>La nouvelle génération, le Blu-Ray: .....</i>	<i>27</i>
<i>Les supports de demain.....</i>	<i>28</i>
<i>L'avenir de la gravure: .....</i>	<i>29</i>
<b>Des limites physiques .....</b>	<b>31</b>
<i>L'obsolescence programmée.....</i>	<i>31</i>
<i>La durée de vie des supports en archive.....</i>	<i>32</i>
<i>L'altération du matériel .....</i>	<i>33</i>
<b>L'information devient immatériel .....</b>	<b>36</b>
<i>Les formats de l'information et leurs évolutions.....</i>	<i>36</i>
<i>Les solutions pour préserver l'intégrité d'un document.....</i>	<i>38</i>
<i>Les défis de la duplication.....</i>	<i>38</i>

<b>LES STRATEGIES ARCHIVISTIQUES.....</b>	<b>40</b>
<b>Redéfinir le document.....</b>	<b>40</b>
<i>Du papier au numérique.....</i>	<i>40</i>
<i>La sélection de l'archive, un processus intellectuel.....</i>	<i>41</i>
<i>La nature du produit numérique.....</i>	<i>43</i>
<b>Des approches de conservation en évolution.....</b>	<b>45</b>
<i>Fonctionnement actuel, quels projets?.....</i>	<i>45</i>
Le Cloud.....	45
Les nouvelles stratégies en Bibliothèques.....	48
<i>Des outils de conception.....</i>	<i>50</i>
ORI-OAI, le chaînon manquant.....	50
Open Archival Information System.....	52
<i>Exemple du fonctionnement de la plate-forme PAC.....</i>	<i>53</i>
Fonctionnement des échanges entre CINES et producteur/service versant	55
Fonctionnement concret de PAC:.....	56
<b>Cadre, limites et recommandations pour le numérique :.....</b>	<b>58</b>
<i>Le contexte législatif et réglementaire.....</i>	<i>58</i>
Les normes:.....	59
<i>La préconisation de l'Open Access.....</i>	<i>59</i>
<i>Les risques du libre échange.....</i>	<i>62</i>
<b>L'AVENIR DE L'INFORMATION.....</b>	<b>67</b>
<b>Les enjeux émergents des nouveaux supports.....</b>	<b>67</b>
<i>L'évolution de l'information.....</i>	<i>67</i>
<i>La «logistication» du document numériques.....</i>	<i>68</i>
Archives et propriété intellectuel.....	69
<i>Nouveaux enjeux, nouvelles missions.....</i>	<i>70</i>
La responsabilité de la conservation et de l'archivage.....	70

L'archivage du web : enjeux et défis .....	73
<b>Quels effets sur la mémoire sociale ? .....</b>	<b>75</b>
<i>L'information comme bien commun.....</i>	75
<i>Les risques éthiques .....</i>	78
« Les incertitudes de la société du savoir » .....	78
<i>Les conséquences sur la mémoire sociale.....</i>	83
<b>Donner du sens : l'utilisation de la sémiotique .....</b>	<b>85</b>
<i>Enjeux de la capitalisation de l'information.....</i>	85
<i>Le rôle de l'archiviste et l'analyse discursive .....</i>	86
<i>Le développement d'une archive sémantique.....</i>	88
<b>CONCLUSION .....</b>	<b>95</b>
<b>SOURCES.....</b>	<b>97</b>
<b>BIBLIOGRAPHIE.....</b>	<b>99</b>
<b>TABLE DES ANNEXES.....</b>	<b>105</b>
<b>GLOSSAIRE.....</b>	<b>107</b>
<b>TABLE DES ILLUSTRATIONS.....</b>	<b>109</b>
<b>TABLE DES MATIERES.....</b>	<b>111</b>