

Cyberlibrary

Valentin Maynadié

Mémoire DNSEP

Mémoire DNSEP, 2020
Atelier Communication graphique
Haute école des arts du Rhin

La numérisation des documents, conservation & diffusion. Les limites et les problématiques soulevées par les bibliothèques numériques.

Valentin Maynadié



Cyberlibrary

Mémoire DNSEP, 2020
Sous la direction d'Anne Laforet
Atelier Communication graphique
Haute école des arts du Rhin

Valentin Maynadié

Notes aux lecteurs :

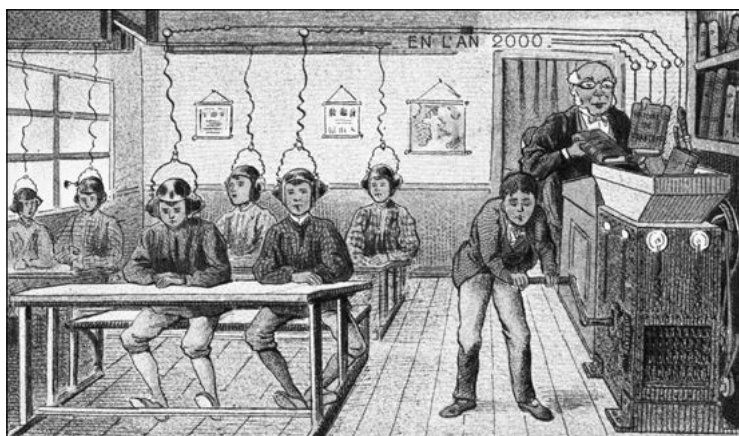
1. Les adresses des sites internet citées ont été consultées et vérifiées à la date du 21 septembre 2019.

2. Les termes techniques rencontrés pendant la lecture sont munis d'un exposant [00] renvoyant au glossaire.

[#0] Introduction	7
[#1] Vous avez dit numérisation ?	13
[1.0] Contexte de deux matérialités	15
[2.0] Du physique au numérique	17
[3.0] Pourquoi numériser ?	20
[4.0] Quoi numériser ?	26
[5.0] Comment numériser ?	28
[6.0] Du numérique au physique	37
[#2] Trois scénarios pour la bibliothèque numérique	41
[1.0] Apparition des bibliothèques numériques	43
[2.0] Transformation des métiers & techniques	45
[3.0] Changement d'Economie	46
[4.0] Changement sur l'accès au patrimoine	56
[5.0] Participatif	62
[#3] Conclusion	71
[#A] Annexe	79
[#A1] Glossaire	81
[#A2] Formats	89
[#A3] Bibliographie	93

[#0]

Introduction



En 1910 l'artiste Français Villemard dessine une série de cartes postales présentant sa vision futuriste de l'an 2000.

Sur celle-ci, on y voit une classe en plein apprentissage. Villemard imagine que le médium papier par l'intermédiaire d'une machine est converti et transmis électriquement pour être diffusé par un autre médium audio à un ensemble d'étudiants.

L'information devient flux.

Nous vivons aujourd'hui dans une période fascinante de l'histoire de l'édition et des bibliothèques.

Grâce au développement des technologies numériques et des réseaux interconnectés, le monde de l'information a explosé: les vecteurs du savoir matérialisés qu'étaient les écrits et les livres ont vécu une brusque révolution quand est apparue la possibilité de construire des bibliothèques d'un genre nouveau, reposant sur de nouvelles règles.

Aujourd'hui, par la numérisation, nous transformons la réalité physique des documents en flux d'information binaire et la stockons dans d'immenses bibliothèques électriques et magnétiques qui, grâce à la toile internet, deviennent accessibles à travers un écran, porté dans notre poche, quand et où on le veut y compris au bout du monde.

Je fais partie de la première génération née avec internet. Très jeune, j'ai eu accès à l'ordinateur de mon père sur son bureau. Au moment où j'apprenais à lire, la toute première chose que j'ai faite sur un écran a été de consulter sur CD l'encyclopédie Hachette 2000. Derrière un gros écran cathodique depuis les combles, je pouvais consul-

ter une multitude de choses qui me passionnaient alors : les pompiers, les avions, New York ou encore les planètes... je découvrais un monde bien plus passionnant que celui de mes cours de CP... avec la liberté de m'y mouvoir selon mes envies de savoir en savoir.

Collégien plus tard, je n'allais pratiquement jamais dans des bibliothèques et CDI. Et encore aujourd'hui il m'arrive très rarement de m'informer à travers elles.

Mais les outils formidables que m'offraient l'internet ont été ma porte sur la culture universelle, et ont construit la presque totalité de mon savoir.

Je suis né dans une période charnière entre médium physique et numérique.

Pour autant, mes études m'ont amené ces dernières années à concevoir des documents imprimés. Mais, à l'heure où le numérique devient le support de communication dominant, les livres que je conçois maintenant finissent généralement scannés et publiés sur internet.

Est-ce donc la finalité d'un livre ou d'une information que d'être visible ?

Toutes ces interrogations ont progressivement constitué le terrain de recherches qui m'intéressait : la question sur la numérisation et le futur de nos livres, et plus précisément celui de la numérisation de masse pour les bibliothèques numériques existantes et à prévoir.

Voyant dans internet un espace utopique pour beaucoup de sujets, en quoi sera-t-il bénéfique qu'il change celui des bibliothèques ?

Au premier abord, j'y ai vu une forme de libération de l'accès à la culture, telle que je la désirais : libre, partagé et prolifique.

Mais mes recherches, on le verra, m'ont vite amené à nuancer mon propos initial.

À travers le mémoire que je vous propose, j'envisage d'analyser dans un premier temps les mécanismes de numérisation du support physique et les immenses nou-

velles possibilités qu'amène ce nouveau support.

Nous verrons ensuite quels changements profonds apportent au monde du livre les bibliothèques numériques, à travers les exemples que nous fournissent les trois acteurs majeurs déjà existants : Google Books, Gallica (BNF) et enfin Open Library, et les questions qu'ils pourraient engendrer.

À travers ce tableau brossé, nous essayerons ainsi de trouver réponse à ma problématique : la numérisation des documents. Conservation & diffusion, les limites et les problématiques soulevées par les bibliothèques numériques.

En écho au sujet de mon mémoire, son contenu a été nourri majoritairement de recherches collectées en ligne : il est le fruit d'internet.

De ce fait il sera rendu accessible pour tous dès que possible sur la bibliothèque numérique open Library, afin qu'il alimente un jour les recherches d'autres personnes.

[Chapitre #1]

**Vous avez dit
numérisation ?**



Tablette d'argile
mésopotamienne, IIIe
millénaire avant J.-C.
Paris, musée du Louvre
mariellebrie.com (2019)

[1.0] **Contexte de deux matérialités**

Les sociétés ont eu depuis leurs origines besoin de moyens pour établir leur communication et en garder trace. Elles ont donc créé des supports à cette fin et les ont de plus en plus élaborés au fil du temps. C'est au tout dernier stade de cette évolution que se place aujourd'hui la numérisation.

En termes de support, la numérisation opère une vraie révolution de nature. Parce qu'elle fait office de passage entre deux matérialités : la physique, tangible et palpable, et le numérique, binaire et électronique.

Pour bien appréhender ce concept il est nécessaire de recontextualiser les fondements et les mécaniques propres à ses deux structures foncièrement différentes.

Physique

Les supports de communication et de stockage d'informations ont vu le jour il y a 6 000 ans, quand la sédentarisation des humains a nécessité la décharge de leur mémoire (leur savoir) sur des supports externes solides, pérennes et transmissibles.

Avec l'apparition de l'écriture, le développement de nombreuses techniques et supports s'accéléra sous diverses formes : hiéroglyphes inscrits à l'encre sur papyrus aux bords du Nil, signes sculptés sur des tablettes d'argile en Mésopotamie, gravures sur bois en Chine orientale... jusqu'aux lettres de l'alphabet grec frappées par plomb encre sur papier par l'allemand Gutenberg, précurseur de l'imprimerie moderne.

De 1500 à aujourd'hui, le papier et les techniques de l'imprimerie vont se répandre vite et partout à travers le monde, par leur capacité à multiplier le contenu écrit plus aisément qu'auparavant, et participer grandement à la démocratisation du support « livre ». À lui seul en effet, grâce à la superposition d'une multitude de feuilles reliées entre elles, il permet la diffusion et la conservation de contenus, textes et images très variés, dans un volume assez compact, manuellement manipulable et consultable.

À noter. J'utilise et j'utiliserais dans un souci de compréhension le terme « physique ». Le terme le plus approprié est « analogique »^[2] mais à mon sens il peut porter à confusion par rapport au début de l'informatique qui repose aussi sur des appareils analogiques.

Le numérique lui aussi est fondé sur des infrastructures physiques, mais je parle ici du contenu qui ne se concrétise que sous forme d'informations stockées électromagnétiquement donc non physiques.

Numérique

L'informatique, elle, a vu le jour en plein contexte de globalisation industrielle vers le milieu du XX^e siècle. Sa conception est attribuée en très grande partie à des ingénieurs et mathématiciens occidentaux se trouvant face à un important besoin de calculs.



En 1960, l'IBM 7000 est le premier ordinateur à base de transistor.

ahmedhergli.bs.com

(2019)

On se rappelle que les premiers ordinateurs analogiques^[2] ressemblaient à de grosses calculatrices « programmables » de la taille d'une maison.

Mais très vite, au-delà des capacités de calcul, on a développé une utilisation intéressante: le stockage de l'information et des savoirs au sens large, puis leurs transmissions avec l'apparition du stockage de masse et des réseaux en temps réel.

Les fondements de l'électronique et des

réseaux informatiques sont tous basés sur les mathématiques (et leur algèbre) que l'on peut qualifier d'universelles. Sur un fonctionnement unifié, et par des protocoles et des codes communs, l'ensemble des machines du monde parle le même langage et peut communiquer et se comprendre mutuellement.

De ce fait, en théorie, le fonctionnement de chaque terminal^[28] informatique étant le même, tout contenu est lisible et compréhensible par toutes les machines.

[2.0] **Du physique au numérique**

Passer de l'un à l'autre, c'est traduire dans une nouvelle structure.

Le mot traduction provient du verbe traduire, dont l'origine est le verbe latin traducere: « faire passer », faire passer d'une structure à l'autre.

On retrouve cette idée du passage d'un état à un autre dans le préfixe latin « trans » utilisé dans de nombreuses sciences qu'il est intéressant de noter.

- En biologie, la Transcription est le mécanisme qui permet de recopier les données des gènes et de recréer de la matière biologique grâce au code génétique.

- En philosophie, la Transcendance indique l'idée de dépassement ou franchissement d'une barrière, comme son propre corps ou une réalité.

- En informatique, le Transcodage se définit par une transformation d'un code (généralement informatique) en un autre code différent.

Ces concepts expriment parfaitement bien l'idée de la numérisation que nous allons aborder: une information dépassant son état physique pour devenir une information numérique peut être abordée comme un flux et non comme un objet unique et tangible.

De façon plus abordable, sur Wikipédia elle est communément définie comme: « la conversion des informations d'un support (texte, image, audio, vidéo) ou d'un signal

électrique en données numériques que des dispositifs informatiques ou d'électronique numérique pourront traiter. »

Cette définition montre qu'il existe des sources variées et nombreuses à numériser... des signaux sonores analogiques^[2], des films, voire des objets 3D! Mais nous ne nous intéresserons ici qu'aux documents, aux textes et images qui font leurs contenus.

C'est le scanner qui est à l'origine de cette possibilité. Son apparition fait suite au développement des ordinateurs modernes à transistors basés sur le signal électrique, ainsi que des premières cellules photoélectriques. Dès 1957 le premier scan numérique fut effectué grâce à un système de tambour couplé en tandem à un ordinateur, mais ce système qui ne pouvait numériser que des pages volantes feuille à feuille ne fut guère développé et sera réservé surtout à la recherche.

Le scanner à plat permettant de poser à plat des livres ouverts apparaît quant à lui en 1970, attendant pour se démocratiser réellement le début des années quatre-vingt-dix avec le développement des ordinateurs personnels et de l'internet.

L'appareil photo numérique lui aussi grandement utilisé pour la numérisation, attendra 1980 pour voir ses premiers prototypes mis au point, mais son utilisation réelle dans les champs de la numérisation ne se développera qu'au début des années 2000.

Beaucoup de foyers aujourd'hui possèdent une imprimante avec scanner intégré, ou encore des smartphones pouvant faire office d'appareil photo, mais leurs utilisations se résument souvent à des besoins de reproductions de documents administratifs ou familiaux.

Pour ce qu'y est de la numérisation de masse que nous allons aborder, celle du traitement de l'information et de son stockage, l'utilisation de scanners rapides et automatisés devient plus que nécessaire au vu de la tâche à accomplir.



Le premier scan
numérique en 1957.
computerhistory.org (2019)



Scanner de livres haute vitesse

BFS-Auto, Ishikawa
Senoo Laboratory 2016
cnetfrance.fr (2019)

Un exemple: Google s'est fait remarquer avec des machines qui peuvent numériser des livres à la vitesse de 250 pages par minute, ces dernières équipées d'ailleurs d'appareils photos plus rapides que les scanners à vitre que l'on connaît.

[3.0] Pourquoi numériser ?

Dans l'univers des sciences sociales nous entrons dans une ère dite de « société de l'information », qui fait suite à la société industrielle. Cette désignation fait cas d'un état dans lequel les technologies de la communication et de l'information ont un rôle central, leurs utilisations sont devenues quasi indissociables de nos métiers et nos vies, voire essentielles.

Une grande partie du contenu écrit produit par l'homme du XXI^e siècle naît désormais sur ordinateur, et n'est la plupart du temps consultée que sur ce dernier, des tweets aux mails, des articles, aux tableaux Excel...

*« Tous les deux jours, l'humanité produit autant d'informations que ce qu'elle a généré depuis l'aube de la civilisation jusqu'en 2003. Plus de 90 % des données disponibles aujourd'hui ont été produites ces 2 dernières années »**.

Nicolas Martin Pour
l'émission La méthode
scientifique, France
culture 06/09/2017

De l'économie à l'industrie, et de la science au social, nombre de pans de notre société seraient totalement tributaires de ces technologies de l'information et de la communication.

Mais attention, il ne faut pas négliger la proportion de textes qui est ensuite matérialisée sous forme papier, par exemple en livres ou en brochures publicitaires.

Il s'agit d'ailleurs souvent pour le livre d'un contenu différent. Dans bien des cas on ne retrouve pas d'équivalent numérique. Cela est sûrement dû au système économique de l'édition papier qui semble rester encore beaucoup

plus confortable et viable pour un éditeur.

De ce fait nous verrons que la numérisation travaille non seulement sur un contenu historique, mais aussi sur celui actuel.

Elle prend une place majeure dans le champ des bibliothèques et plus largement dans le domaine du livre qui est au centre de notre intérêt à travers ce mémoire. En effet ces lieux qui sont un espace concentré de savoir sous forme imprimée, soucieux de répondre à ses objectifs de conservation et de diffusion, vont s'articuler parfaitement avec les possibilités du numérique.

Depuis la mythique bibliothèque d'Alexandrie qui conservait en son sein la plus grande collection de documents écrits du monde antique, les bibliothèques ont pris la responsabilité de la sauvegarde du patrimoine documentaire écrit.

Ce fantasme d'accumulation infini dans un même lieu de tous les textes écrits, étaient évidemment un projet voué à l'échec tant la masse de documents produits était déjà énorme.

2000 ans après, l'arrivée des stockages numériques a permis un gain de place exceptionnel : 10 000 livres dans quelques centimètres carrés, plus facilement duplicables et transportables ! Il semblerait qu'elles aient trouvé l'outil de leur mythique utopie : des armoires de rangement à l'infini dans une place finie.

Mais on voit aussi que cette nouvelle matérialité n'empêche pas une cohabitation avec le physique, elle permet dans beaucoup de cas de se faire en double amenant une solution quand on ne peut garder l'original faute de place.

Wikipédia nous éclaire aussi sur d'autres objectifs avec cette liste assez fournie :

- Préserver et protéger des documents contre les risques d'altération (l'acidité par exemple)
- Archiver des documents originaux en vue d'un gain de place, et les dupliquer sans risque de dégradation pour les mettre à disposition du public

- Permettre au public de consulter et d'accéder à des documents anciens et/ou rares
- Donner accès à la connaissance à distance dans une perspective de communication (bibliothèques électroniques en ligne, projet « Bibusages » sur Internet, etc.)

Focus sur deux objectifs qui selon moi apportent une avancée particulièrement intéressante :

[3.1] **Pérenniser et protéger de la destruction**

Il existe de nombreux exemples au cours de l'histoire de l'humanité où celle-ci a été confrontée à la destruction massive d'un patrimoine culturel et notamment de documents et de livres.

Parmi eux, les incendies de bibliothèques ont été particulièrement ravageurs, avec une liste très longue commençant par le triste destin de la mythique bibliothèque d'Alexandrie qui brûla entièrement au début de notre ère grégorienne. Ou encore les cas de destructions volontaires de la mémoire lors de guerres ou de changements idéologico-politiques. On citera évidemment les sinistres autodafés de l'Inquisition en Espagne en 1481 ou de 1933 en Allemagne sous le régime Nazi où de grands rassemblements politiques invitaient les citoyens à brûler des livres sur leurs places publiques.

Mais l'altération naturelle du livre peut être aussi à l'origine de sa destruction progressive: un casse-tête pour les conservateurs qui se battent avec toutes sortes de bactéries, champignons et autres altérations temporelles dues à la manipulation et à la lumière.

Par leur préciosité, aujourd'hui de nombreux documents ne sont plus physiquement accessibles, ou seulement sur autorisation spéciale, ce qui rend leur étude compliquée.



La Carte Pisane abîmée
par l'humidité, 1300
gallica.bnf.fr (2019)

On voit alors dans le stockage numérique parallèle le moyen efficace de sauvegarder une « copie » de chaque document ou livre, au cas où un drame arriverait... mais aussi afin de permettre une consultation assurant l'intégrité physique des documents.

Pour autant, cette sauvegarde n'est pas destinée à tout le contenu d'un fond de livres.

La sauvegarde implique aussi un choix dans ce qui doit être conservé ou non, c'est le travail de l'archiviste. En effet le coût et le temps nécessaires à numériser une collection est important et tout ne vaut peut-être pas le coup d'être conservé sous forme numérique.

Il faut aussi noter que malgré toutes les précautions prises, les supports dits numériques ne sont pas parfaits et peuvent souffrir eux aussi des mêmes problématiques inhérentes à leur physicalité. Un disque dur^[14] est avant tout un concentré de mécanique fragile, mais nous aborderons cette problématique plus en détail plus loin.

[3.2] **Rendre toujours plus accessible**

L'accessibilité numérique peut se définir comme la mise en œuvre de ressources numériques à disposition de tout individu consultant depuis un terminal^[28] informatique.

C'est là l'avantage fondamental du support numérique : grâce aux réseaux comme le web^[31] on peut y accéder depuis n'importe quel endroit du monde, en une seconde, jour et nuit.

De plus, tout fichier ou document peut être théoriquement dupliqué, multiplié, et consulté simultanément par une infinité de personnes.

(Du moins si les droits qui s'appliquent le permettent, comme par exemple les œuvres tom-

bées dans le domaine public.)

La mise en ligne des premières bibliothèques numériques a vu le jour au cours des années quatre-vingt-dix, au même moment que la démocratisation de l'informatique se développait.

En France, le projet Gallica mené par la Bibliothèque Nationale de France (BNF) fut précurseur en la matière et ouvrait les portes d'un site internet réunissant déjà des milliers d'ouvrages numérisés dès 1997. Le projet s'est donné les moyens de devenir aujourd'hui l'une des plus grandes bases de données d'ouvrages au monde, accessible librement et gratuitement sur internet.

Une démarche qui s'inscrit ici dans une dynamique patrimoniale de diffusion du savoir. Chercheurs, étudiants ou simples lecteurs et curieux, peuvent en quelques clics accéder à des documents de la bibliothèque se situant à des milliers de kilomètres d'eux.

[4.0] Quoi numériser?

[4.1] Contenu et contenant

Lorsqu'on aborde les questions de retranscription d'une information (ici à travers la numérisation), il est important de se demander d'abord à quel besoin répondrait la version numérique d'un document, et de cette question en découlera une deuxième: qu'est-ce qui donne valeur à un document?

En effet on peut dissocier deux éléments au sein même d'un document imprimé: le contenu et le contenant.

Le contenu serait le texte, c'est-à-dire tout simplement ce qu'il y a d'écrit, et le contenant serait quant à lui le support, l'encre, la forme du texte, sa mise en page...



Bible de Gutenberg
ibookbinding.com (2019)

Prenons l'exemple de la célèbre bible de Gutenberg^[5] imprimée en 1454:

Le texte de la Bible en latin fait dans ce cas office de contenu et le livre l'objet en soi, à travers sa reliure, sa mise en page, ses illustrations et surtout son histoire représente le contenant.

Un théologiste trouvera une pertinence pour ses recherches dans le contenu tandis qu'un chercheur travaillant sur l'apparition de l'impression trouvera sûrement plus d'intérêt dans le contenant. Mais comme nous allons le voir, les bibliothèques proposent justement deux approches du document: le mode Texte et le mode Image.

[4.2] Le choix

La numérisation peut représenter des coûts importants, autant que du temps, et de ce fait tout ne peut pas être numérisé. Il est alors important

de se poser la question de décider ce que l'on devra numériser ou non.

Tout travail d'archivage ou de sauvegarde physique demande de faire des concessions dans ce que l'on garde accessible facilement, ce que l'on entrepose dans des réserves, ou ce que l'on détruit.

Le numérique ne déroge pas à la règle.

Certes, en créant le doublon de tout document dans un rangement informatique infiniment réduit, la numérisation apporte une nouvelle possibilité qui en fait un outil d'archive très performant, mais qui pourrait peut-être vite devenir à son tour surchargé.

La problématique du choix demeure. Tout dépendra finalement de la nature du document à conserver.

Si nous considérons deux raisons totalement différentes qui motiveraient une numérisation, on comprend que la question du choix appelle une réponse différente selon cette nature :

- Premier exemple : numériser un livre historique fragile et précieux afin de figer une copie dans le temps, le rendre accessible et de ce fait protéger l'original, oblige à l'évidence à tout intégralement garder.

- Deuxième exemple : numériser un document administratif dans la mémoire d'une organisation publique ou privée, dépendra de l'importance qu'on lui aura attribuée. En règle générale, le choix [garder/détruire] se fait simplement entre un administratif quotidien jugé superflu (les notes de service) et un administratif de fonctionnement dont on créera systématiquement un doublon numérique.

[5.0] **Comment numériser ?**

[5.1] **Les techniques d'acquisition**

Dans leur très grande majorité, les numérisations d'ouvrages s'effectuent selon deux moyens : l'appareil photo numérique ou le scanner à plat. Dans les deux cas on utilise une même mécanique : le balayage d'une surface par un capteur. L'utilisation de l'un ou l'autre de ces capteurs dits « photographiques » a pour fonction de convertir un rayonnement lumineux en signal électrique. Par balayage de la surface du document il réceptionne la couleur renvoyée par une source lumineuse projetée généralement artificiellement contre le document et en définit sa couleur qu'il convertit en valeurs mathématiques.

On appelle ce processus « l'échantillonnage », car en effet la numérisation constitue un compromis en termes d'information récoltée et de « poids » possible du fichier.

Tous les détails ne pouvant pas être stockés, la fréquence d'échantillonnage d'une zone représente sa précision, qui est définie par deux valeurs principales :

- le nombre de pixels par pouce, PPP (points par pouce), qui donnera la définition de l'image finale.
- le nombre de couleurs possible pour chaque pixel (la profondeur) noté en bit^[7].

Par exemple, pour un format JPEG^[B] (généralement il s'agit de 8 bits, soit 254 valeurs pour chacune des 3 couleurs rouge, vert et bleu, donc 16 millions de couleurs). Une image brute (RAW) d'un appareil photo professionnel en 14 bits représente quant à elle 4 milliards de couleurs potentielles, soit incroyablement plus que ce que l'humain peut percevoir.

Arrêt du temps

L'action de la numérisation crée un doublon numérique du document à une temporalité donnée. La photographie numérique fige dans le temps cette version, car sa singularité est qu'un fichier ne s'altère pas avec le temps. Le papier et l'encre sont remplacés par des valeurs de bits. Les zéros, les uns et zéros ne vieillissent pas.

Soit ils marchent, et ils sont justes. Soit ils sont cassés, corrompus, et ils sont illisibles.

Image / texte

La numérisation génère un fichier de type Image, généralement brut (RAW, TIFF^[E]) ou parfois JPEG^[B], PNG^[B] ou directement PDF^[D]. Mais cette version ne constitue pas spécialement une finalité pour le fichier.

En effet un certain programme dit de Reconnaissance Optique de Caractères (OCR)^[26] permet d'automatiser la récupération de texte dans l'image d'un texte imprimé et de le sauvegarder dans un fichier qui pourra être exploité dans un traitement de texte.

Cette action nécessite une interprétation des caractères et une transcription sur l'un des systèmes de codage de caractères informatiques. Il s'agit d'une norme qui associe un jeu de caractères abstraits (comme les alphabets) à un code numérique.

Ce système constitue la deuxième forme courante de document numérisé : le mode Texte.

La plus ancienne est l'ASCII^[1], très peu inclusive avec 128 caractères, suffisante pour du texte anglais mais trop limitée pour d'autres langues (par exemple pour le français, elle ne prend pas en compte les accents).

De ce fait L'Unicode^[30] a été développé: ce standard ISO^[19] est constitué d'un répertoire de 137929 caractères, couvrant une centaine d'écritures. Mais lui aussi ne peut aller jusqu'à couvrir la richesse des langues du monde entier.

Dimension 3D/2D

Un des problèmes innés des scanners et appareils photos est que par essence ils ne retranscrivent qu'une prise de vue à plat, en 2 dimensions: largeur et hauteur. Alors que n'importe quel élément physique ou document en possède 3: hauteur, largeur et profondeur. Même une feuille possède une épaisseur, certes de quelques microns, mais bien réelle.

Il est alors d'autant plus compliqué de retranscrire la forme d'un livre qui possède de multiples faces (couverture, quatrième de couverture, reliure, tranche, dos etc.)

Certes, des prises de vues latérales peuvent permettre une compréhension de la forme, mais elles résulteraient d'une interprétation du photographe. Le scan 3D constitue une piste technique intéressante, plus appropriée pour la conservation de la matérialité et de la forme du modèle. D'autant mieux qu'avec l'apparition des casques et autres accessoires de réalité augmentée il serait possible un jour de retrouver virtuellement une manipulation quasi-physique du document ou du livre.

[5.2] Les moyens de conservation

Stockage

Il existe un grand nombre de supports de stockage numérique. Chacun apporte un compromis entre fiabilité, espace et prix. Pour nos problématiques de stockage de documents, on peut les diviser en deux catégories:

Ceux optimisés par des serveurs pour un accès continu à l'information. Et ceux faits pour du stockage externe non soumis à un besoin d'accès courant.

En effet lorsqu'on veut accéder en ligne à du contenu, on fait appel à un serveur. Ce dernier est généralement constitué d'un ou plusieurs importants disques de stockage.

On retrouve les disques durs (HDD), très répandus pour le stockage de masse avec des capacités jusqu'à 16 téraoctets soit environ la capacité de stockage de 250 000 mémoires au format texte par exemple. Cela en fait un rapport espace-prix imbattable.

Sont apparus depuis quelques temps des supports dits de mémoire flash (SSD) connus pour avoir une vitesse de lecture très élevée et une fiabilité accrue par rapport aux HDD classiques. Leur capacité peut atteindre 20 téraoctets sur les modèles les plus coûteux. Mais c'est là aussi leur point négatif. Ils coûtent très cher, ce qui ne leur confère pas une grande notoriété dans le monde des serveurs de stockage.

Ensuite vient la deuxième catégorie, celle qui concerne les stockages qui ne nécessitent pas un accès permanent. On y retrouve les disques optiques tels que CD ou DVD-Rom plus connus pour le stockage de fichiers audio ou vidéo pouvant aller jusqu'à 20 gigaoctets.

Enfin viennent les cartouches magnétiques de type LTO qui sont constituées de bandes magnétiques. Très connues dans le monde de l'archive, elles sont cependant de moins en moins utilisées au profit des disques durs. Mais leurs performances n'en restent pas moins supérieures: elles sont plus fiables, et peuvent stocker jusqu'à 220 To sur une seule cartouche, pour un prix souvent faible.

Redondance

Quels que soient ces supports, ils ne sont pas très fiables seuls car leur durée de vie est extrêmement aléatoire et tourne généralement autour de 10 ans.

C'est pour cela que l'utilisation de doublons, voire de triplons, est la plupart du temps mise en place sur des serveurs. Eux-mêmes se situant parfois dans deux endroits physiques distincts. Cette redondance permet théoriquement une pérennisation parfaite des fichiers. Mais elle n'est que partielle, car elle nécessite une intervention humaine incessante afin de remplacer et recopier les disques durs vieillissants, endommagés ou corrompus.

Taille du support

Un des grands avantages du numérique est la taille de ce que le support peut contenir: un disque dur^[14] récent peut théoriquement stocker la numérisation de millions de livres, et il est aujourd'hui théoriquement possible de transporter dans un sac à dos les 200 To de la collection de 20 millions de documents Archive.org!

Dépendance à la société

Les stockages numériques de la simple clef USB jusqu'aux serveurs et centres de données dépendent de ressources importantes pour fonctionner et donc d'une stabilité économique et d'un environnement stable.

Ils sont souvent tributaires d'un approvisionnement électrique, d'une main-d'œuvre qualifiée et payée, et surtout d'un approvisionnement en supports, comme par exemple ces disques durs de rechange, fabriqué quasiment exclusivement dans les pays d'Asie, notamment la Thaïlande. À vrai dire, c'est un problème que les biblio-

thèques physiques ne rencontrent pas aussi fortement qu'on l'imaginerait car les livres, étant de la matière morte, n'ont en principe pas besoin d'intervention humaine s'ils sont conservés dans un endroit protégé.

Cohabitation

Comme abordé plus haut, les livres rencontrent cependant d'autres problèmes : altérations par usage fréquent, destructions volontaires, etc.

En fin de compte aucun de ces deux systèmes n'est parfait. Ils répondent tous deux à une problématique propre, mais ensemble ils constituent une chaîne plus solide de conservation. Ils seront sûrement amenés à cohabiter pendant encore très longtemps.

On doit toujours se rappeler que conserver un document électronique est à proprement parler impossible. Seule la capacité de le reproduire peut-être préservée. Le support numérique ne dure pas. Il mute. Il migre.

[5.3] Consultation

Pour pouvoir accéder au contenu stocké, un terminal^[28] informatique est toujours nécessaire. Et cela implique deux choses :

Une compatibilité de logiciel (software)

L'ordinateur doit reconnaître et pouvoir lire le fichier, donc la forme de stockage doit être compatible avec lui.

L'utilisation de programmes spécifiques avec des codecs à jour ou d'anciens codecs de déchiffrement est nécessaire, car les formats évoluent inexorablement. Par exemple, le format d'image portable pixmap (ppm) développé en 1988 par IBM^[18] n'est aujourd'hui plus supporté par la plupart des logiciels.

Une compatibilité de matériel (hardware)

L'ordinateur doit physiquement pouvoir recevoir le support dans son lecteur.

Comme par exemple des branchements IDE^[21] pour les disques durs déjà en voie de disparition, ou bien le lecteur de CD / DVD qui disparaît petit à petit de tous nos ordinateurs portables.

En principe, et pour le moment, le problème ne se pose pas. En effet la numérisation massive a débuté il n'y a qu'une quinzaine d'années et les formats utilisés à cette époque (JPEG^[B], PDF^[D], txt^[C], Epub^[A]) n'ont pas eu le temps d'évoluer ou si peu, et des conversions (transcodages) peuvent au besoin être encore effectuées.

Interface

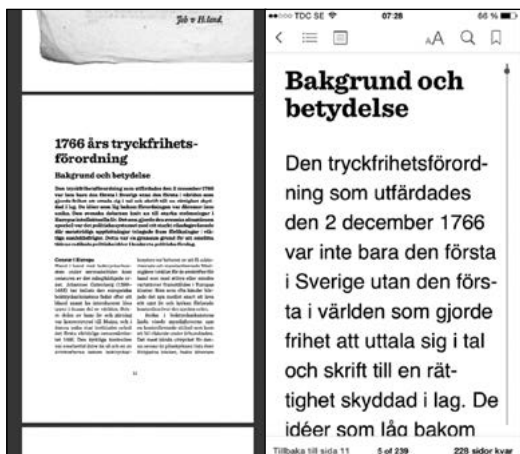
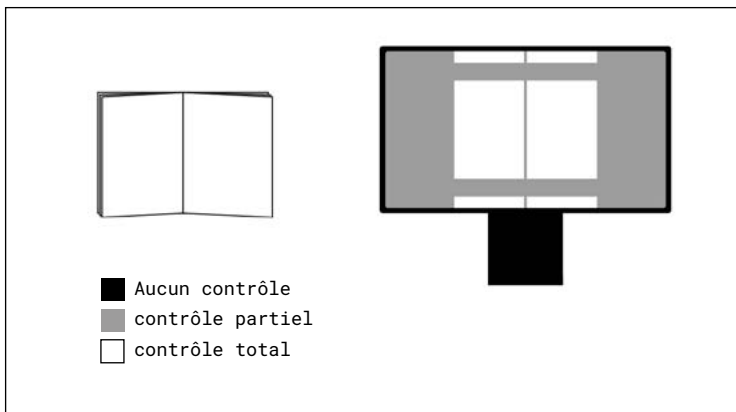
Les évolutions technologiques ont mené au développement d'une multitude de terminaux^[28] aux formes et tailles d'écran différentes.

D'abord ordinateur de bureau, il a ensuite été portable, et il tient maintenant dans notre poche avec les tablettes et les smartphones.

Il peut aller jusqu'à un écran noir et blanc à cristaux comme pour les liseuses, ou incurvé, et même pour certains pliables !

Contrairement à un livre dont la consultation ne demande aucun appareil, le fichier numérique est interprété par la machine du lecteur. C'est ce qu'on nomme une « interface de consultation ». Par sa pluralité et sa non-uniformité, son expérience varie énormément d'un support à l'autre.

Quand un éditeur produit une édition il est assuré du résultat utilisateur prévu, car c'est lui qui choisit le format, le papier, la forme, la typographie... toute la matérialité.



Epub zoomable &
 PDF fixe (sur iPhone)
upload.wikimedia.org (2019)

Avec un contenu numérique, on ne peut avoir que très peu de contrôle sur l'interface qu'aura le lecteur.

Certains formats numériques pourront même changer la mise en page du livre!

Les fichiers de lecture ePub (mode texte) illustrent bien ce propos car même la mise en page (échelle), la police de caractère utilisée et bien d'autres choses peuvent être modifiées depuis le terminal^[28].

Mais les formats de type image comme les PDF^[D] par exemple gardent un certain contrôle du moins sur la forme du contenu.

[6.0] Du numérique au physique

[6.1] Réversibilités

Quel barème pourrait distinguer une bonne d'une mauvaise sauvegarde par numérisation?

En cas de destruction ou d'altération d'un document, il serait en effet très intéressant de pouvoir le recréer au plus près de l'original.

La question de la réversibilité entre alors en jeu.

Dans notre cas d'étude, c'est le fait de pouvoir avoir accès à assez d'informations pour avoir la possibilité de la recréer, ou surtout de mieux la concrétiser et contextualiser a posteriori, ce qui est le cas pour les chercheurs lors de recherches par exemple.

Informations techniques comme : type de papier, construction des cahiers, reliure, technique d'impression, encre, etc. Autant d'éléments propres à l'histoire même du document.

[6.2] Re-Physicalisation

A peine la numérisation de masse a-t-elle été mise en place mondialement qu'aussitôt un besoin de re-physicalisation est redevenu une nécessité.

Une méthode déjà bien employée est le fac-similé^[16]. Il s'agit d'une copie ou d'une reproduction d'un livre souvent fragile ou rare, qui se veut être aussi ressemblante que possible à l'identique original. Ils sont utilisés, par exemple, par les étudiants ou les chercheurs pour des recherches sur une source à laquelle l'accès physique ne leur est pas permis.

Avec la démocratisation des imprimantes personnelles, imprimantes 3D, et fab-labs pour ne citer qu'eux, on voit clairement que le « tout numérique » est loin d'être une fin en soi.

Dans Post Digital Print, Alessandro Ludovico souligne la fausse rumeur de la mort du papier face à l'édition numérique qu'on entend beaucoup et partout depuis une vingtaine d'années déjà! Et pourtant, l'édition papier ne s'est jamais aussi bien portée et a su en fin de compte cohabiter avec le livre numérique.

Alors comme le propose Alessandro Ludovico : comment l'analogique ^[2] et le numérique vont-ils coexister dans l'ère post-numérique?

Une question que beaucoup d'intéressés, designers, graphistes, bibliothécaires, archivistes et futurs acteurs du monde de l'édition tentent de répondre.

En France les initiatives et projets dits du « Web To Print » s'attellent à donner des outils simples pour repasser de l'écran au papier.

On peut par exemple citer le collectif de recherche Pré-Post-Print, qui se présente comme un laboratoire et groupe de recherche autour des systèmes de publication libres et alternatifs.

[Chapitre #2]

Trois scénarios pour la bibliothèque numérique

[1.0] Apparition des bibliothèques numériques

Tâchons avant toute chose de définir plus précisément ce qu'est une bibliothèque avant qu'elle ne soit une bibliothèque numérique.

Selon l'encyclopédie une bibliothèque est un lieu au sein duquel est conservée et lue une collection organisée de livres. Certaines peuvent proposer aussi d'autres types de documents patrimoniaux ayant une valeur historique importante. Les bibliothèques sont souvent publiques. En plus de répondre à des problématiques de conservation, elles jouent un rôle important dans la diffusion des patrimoines culturels. Certaines d'entre elles, municipales et universitaires par exemple, vont jusqu'à autoriser le prêt de leurs documents gratuitement.

L'expression bibliothèque numérique quant à elle, est une expression technique qui consacre la jonction entre deux mondes différents : celui des bibliothèques et celui de l'informatique lié au système de diffusion permis par les réseaux (internet, intranet^[20], web^[30]).

Les objectifs d'une bibliothèque numérique sont par extension les mêmes que ceux d'une bibliothèque classique mais vont devoir être atteints au moyen des médiums nouveaux présentant des contraintes techniques différentes.

L'expression « bibliothèque numérique » reste cependant imprécise et peut faire l'objet de nombreuses interprétations.

Comme nous allons le voir, les nouvelles plateformes^[25] en ligne (par exemple celles issues du secteur privé) peuvent servir d'autres intérêts que ceux historiques des bibliothèques publiques, en diversifiant considérablement leurs offres.

Peut-on dans ces cas encore parler de bibliothèques?

Les premières bibliothèques dites numériques sont apparues assez tôt, avec l'usage dès 1957 du premier « scan » numérique et, surtout à partir de 1970, autour des premiers scanners à plat sur lesquels on pouvait dorénavant poser des livres et des cahiers sans en séparer les feuilles.

De cette avancée technique sont nés de nombreux projets, dont on retiendra le projet Gutenberg de 1971 pré-curseur en la matière.

Il est porté par Michael Hart qui, grâce à l'accès à un ordinateur central auquel il avait accès pendant ses études à l'Université de l'Illinois, chercha à optimiser son temps sur la machine coûteuse non pas en temps de calculs, mais pour du stockage.

C'est dans ce dernier qu'il voit l'intérêt de la machine. Il choisit d'ailleurs comme premier document numérisé la saisie d'une copie de la Déclaration d'Indépendance des États-Unis et l'envoya à tous les utilisateurs du réseau informatique de l'université, comme exemple de son concept.

C'est véritablement au début des années 1990 avec la révolution de la « toile » internet que ces bibliothèques d'un nouveau genre vont émerger et prendre une ampleur sans précédent.

En effet la possibilité de transmettre et consulter ces documents sur requête, via un réseau commun (internet), chamboule totalement notre rapport aux sources du savoir.

À partir de 1990 jusqu'aujourd'hui de nombreux projets de bibliothèques numériques vont fleurir dans le monde. Elles sont institutionnelles comme par exemple Gallica, la bibliothèque numérique de la Bibliothèque Nationale de France, Europeana, la plate-forme numérique européenne. Ou bien le fruit d'entreprises privées comme Google avec le projet Google Book, L'Open Library de la fondation Internet Archive.

Il faut préciser que ces nouveaux portails culturels qui ont changé nos pratiques à travers de nouvelles dimensions, ont dû s'articuler et cohabiter avec nos bibliothèques classiques et leur fonds existants. Finalement ces bibliothèques numériques, restent bien souvent dépendantes de ces contemporaines qui leur apportent les fonds d'ouvrage à numériser.

Se retrouvant face à ce nouveau défi technique incontournable, il a fallu bien sûr faire appel à de nouvelles compétences devenues indispensables.

[2.0] Transformation des métiers & techniques

[2.1] Métiers



Magasins BNF site
François Mitterrand
bnf.fr (2019)

Jusqu'à présent, les professionnels de la mémoire c'est-à-dire les archivistes, conservateurs et bibliothécaires, opéraient en diffusion directe en rassemblant, indexant, concevant et diffusant manuellement le savoir au sein des institutions.

La consultation s'effectuant directement sur le lieu même de leur stockage, l'interaction était directe avec le professionnel. Par exemple, à la Bibliothèque Nationale de France, la consultation d'un document particulier devait suivre de nombreuses étapes nécessitant un personnel important : rédaction du formulaire de consultation, autorisations, recherches, préparation du document, acheminement par TAD^[29] jusqu'à la consultation proprement dite.

Avec l'apparition des bibliothèques numériques et leurs nouveaux modes de diffusion, les professionnels ne sont plus uniquement des conservateurs de biens publics, mais deviennent

d'importants fournisseurs et diffuseurs de biens culturels et patrimoniaux répondant à un nombre grandissant de demandeurs. En attestent les 43 400 visites par jour sur le site internet gallica.com contre 5 000 entrées sur le site François Mitterrand de la BNF.

On comprend que ce changement implique l'arrivée de nombreux nouveaux corps de métiers comme ceux liés à la programmation, particulièrement importante et centrale à tous les échelons de ce processus. C'est un nouveau terrain qui s'ouvre en grand, avec la possibilité d'y créer nombre de nouveaux métiers, correspondant à des développements techniques qui vont en s'accéléralant.

[2.2] Techniques

Le numérique implique la mise en place d'infrastructures complexes nécessitant des algorithmes^[3] de classements poussés pour permettre d'affiner au maximum d'éventuelles recherches dans cette masse d'informations et de documents

L'organisation d'un rayonnage en bibliothèque se suffisait la plupart du temps d'une classification sous forme de cote comme celle de Dewey^[9] et d'un catalogage unique pour permettre une recherche et un accès suffisamment efficace. Tout en considérant qu'un classement bien compris et bien fait ne pouvait partir que du postulat qu'un livre ne pouvant être qu'à un seul endroit à la fois.

Mais les possibilités intrinsèques au numérique permettent maintenant l'indexation multiple d'un seul et même document : c'est une révolution. Ainsi le point d'entrée courant des utilisateurs sur ces nouvelles bibliothèques passe inévita-

blement par les moteurs de recherche sous la forme de requête à un index, moteurs nourris par un ensemble de mécanismes qui se développent comme ce qui suit.

Les métadonnées : comme l'utilisation du vocabulaire standardisé de métadonnées « Dublin Core^[11] » qui constitue des informations utiles à l'intention du moteur de recherche.

Les hyperliens : qui permettent de naviguer par liens transversaux entre les différents ouvrages qui se référencent entre eux.

Les recommandations : proposées vers d'autres ouvrages pertinents que la recherche va affilier à chaque résultat.

Les « Recherches plein texte » : qui consistent à examiner chaque mot d'un document afin de pouvoir effectuer une recherche dans le corps du texte.

Les mécanismes de ce qu'on appelle le « web sémantique^[32] » permettent entre autres une structure par croisement des contenus et métadonnées les concernant.

En bref pour être consultable un document ne peut exister seul, il doit être accompagné de nombre de métadonnées relatives à son contenu pour pouvoir être indexé et donc être accessible. Des techniques de plus en plus poussées qui nécessitent donc de nouveaux acteurs opérateurs très qualifiés.

[3.0] **Changement d'Économie**

En économie, principalement depuis l'industrialisation, la valeur d'un bien marchand repose pour l'essentiel sur sa nature de bien matériel. Elle se définit par un ensemble de notions bien tangibles comme sa composition, son coût de production, sa rareté, son unicité...

Pour parler concrètement on peut prendre l'exemple d'un litre de pétrole depuis son extraction d'une plateforme pétrolière jusqu'à sa commercialisation. Son prix est défini par l'ensemble des coûts que représentent les ressources humaines et naturelles nécessaires.

On comprend que pour tout bien, la notion de ressource limitée est centrale car c'est elle qui justifie pour beaucoup la valeur marchande.

C'est sa physicalité inhérente qui le contraint à cette réalité.

A contrario les biens stockés numériquement ne sont pas tangibles et se désignent comme « biens immatériels ». Théoriquement ils peuvent être dupliqués et ainsi être accessibles par plusieurs personnes à la fois.

Par exemple, si l'on pouvait dupliquer un litre d'essence comme on le fait pour un fichier numérique, Shell fermerait assez vite boutique!

Cette problématique est d'autant plus importante dans les champs des œuvres de l'esprit (livres, photos, musiques, films etc.) par le fait que leur valeur historique ne repose pas uniquement sur leur physicalité, mais aussi sur la propriété intellectuelle qu'elles représentent.

On voit que notre modèle de propriété devient difficilement applicable sur ce nouveau support et nécessite donc de repenser l'économie.

Certaines grandes entreprises américaines de la Silicone Valley^[27] excellent justement dans ce domaine comme Google par exemple à travers son projet Google Books.

[3.1] Google Book

2004 marque l'arrivée d'une entreprise privée dans le monde de la numérisation et de la diffusion en ligne : Google présente son projet Google Books imaginé dès son origine comme une bibliothèque numérique.

Lors du lancement de Google Books, le géant de Mountain View annonçait :

Laurent Brosset,
Google Livres et la
numérisation, 2016 51 p.
dumas.ccsd.cnrs.fr (2019)

« L'objectif du projet est simple : aider à conserver la prééminence du livre et des bibliothèques au sein d'une culture toujours plus centrée sur l'internet en plaçant leur information au sein du monde en ligne. Nous espérons que cela conduise plus de monde vers leur bibliothèque locale, que cela conduise à la numérisation des fonds de quelques-unes des plus grandes institutions de recherche, que cela rende tout ceci accessible. »

Cette déclaration fait écho à l'idéologie revendiquée par Google depuis ses débuts encore lisible sur leurs pages web about google :

« Notre objectif est d'organiser les informations à l'échelle mondiale pour les rendre accessibles et utiles à tous ».
On ne saurait mieux dire !

Bien sûr, derrière cette façade flamboyante, on aurait lieu de s'attendre à trouver un projet 100 % philanthropique^[23]. Mais en vérité, ce projet ne porte-il pas aussi et surtout toute une dynamique économique à l'inverse des des bibliothèques historiquement publiques, attachées à leur sacro-saint but non lucratif ?

Dès lors, quel impact risque-t-on d'avoir si des motivations économiques contrôlent de manière universelle et puissante l'accès à la « connaissance » ?

En quelques années Google Books s'est imposé comme un des leaders mondiaux dans plusieurs domaines liés à

l'archivage et à la distribution de livre numérique. En 2019 le site a revendiqué plus de 25 millions de livres numérisés, répertoriés et accessibles, ce qui en fait à l'heure actuelle la plus grande bibliothèque numérique du monde.

De ce fait, son développement est source de nombreuses interrogations et inquiétudes. Parmi elles, on retrouve des questions sur des enjeux politiques ou culturels internationaux, celles très sensibles qui concernent les nouvelles formes de contrôle de l'information. Sans oublier à l'avenir le sort de la conservation et de la pérennisation des patrimoines qui historiquement étaient gérés par les bibliothèques publiques.

[3.1.2] **Politique & Économie**

Économiquement Google, à l'instar des formes industrielles standard connues jusqu'à présent, représente une nouvelle forme de système économique que l'on nomme « le capitalisme cognitif », appelé aussi « l'économie du savoir ».

Derrière ce mot barbare introduit par Fritz Machlup en 1962 dans son livre « The production and distribution of knowledge in the US » se cache un concept nouveau de mutation économique : l'Information devient une valeur centrale dans la capitalisation.

Google depuis son origine en 1998 développe tous ses systèmes autour de l'information. La plupart des services qu'elle propose étant gratuits, l'entreprise s'attelle alors à accumuler des informations et de la matière connaissance (données) dont elle peut tirer bénéfice par le biais, par exemple, de la publicité ciblée « en conduite assistée déguisée » sur son moteur de recherche.

Elle transforme son service gratuit en monétisation des actes de ses utilisateurs.

Google

Livres

Ma bibliothèque

Rechercher dans ma bibliothèque **Rechercher**

NOUVELLE ÉTAGÈRE

Ma bibliothèque

Mes livres sur Google Play (9)

Favoris (0)

En cours de lecture (0)

À lire (0)

Lus (0)

Livres pour vous

Mon historique

Acheter des livres sur Google Play

Parcourez la librairie en ligne la plus vaste au monde et commencez dès aujourd'hui votre lecture sur le Web, votre tablette, votre téléphone ou un lecteur d'e-books.

[Accéder à Google Play »](#)

Mes livres sur Google Play - Privée Plus »

Candide
Voltaire

Candide, ou l'Optimisme
Voltaire, Lac...

Temple Run, R...
Temple Run

Achetés - Privée Plus »

Candide
Voltaire

À la recherche d...
Marcel Proust

De la Terre à la ...
Jules Verne

Évalués - Public Plus »

Cette étagère ne contient pas encore de livres. En savoir plus.

Consultés - Privée Plus »

Molière

Home page Google Books,
ma bibliothèque
books.google.fr (2019)

Le service Google Book ne déroge pas à la règle. Derrière ce projet se cachent de nombreux dessins autres que la conservation de façade dont se réclame l'entreprise Google. Aujourd'hui, Google Book contient une réserve impressionnante de livres libres de droits et de sous-droits d'auteur, et les diffuse à une échelle mondiale. Souvent gratuitement, quelquefois non.

Mais nous retrouvons ce que nous avons vu précédemment: pour le cas d'un document libre de droit, par essence gratuit et multipliable, sa valeur financière numérique est nulle.

Conscient de cet écueil dès 2010, Google a vite proposé son propre service de vente d'E-book^[15] avec des liens d'affiliation pour un achat physique du livre.

Enfin de compte la firme en revient là au modèle économique de l'unicité monnayable du produit!

[3.1.3] **Comment?**

Google, contrairement à d'autres projets de bibliothèque numérique, ne possède pas de fonds d'ouvrages.

Sa réussite se base sur de nombreux partenariats avec des bibliothèques institutionnelles.

Ces dernières doivent dans un souci de préservation et de diffusion passer aujourd'hui par une numérisation de leur fonds. Malheureusement, faute de moyens financiers ou techniques, elles sont amenées à travailler avec Google qui, elle, possède ces moyens et ces compétences.

Dans l'idée, cette collaboration n'est pas spécialement un mal. D'un côté la bibliothèque va fournir généralement gratuitement son fond d'ouvrages et de l'autre, Google lui procure en même temps qu'une sauvegarde sérieuse, un accès très élargi à l'intégralité de son contenu ainsi numérisé.

Mais ce partenariat privé - public pose cependant une question morale : elle peut positionner un service public en dépendance d'une société privée. Or, la culture a été majoritairement gérée pour le moment par l'État, avec la garantie d'indépendance et de neutralité commerciale.

Aucune fluctuation économique, quelle qu'elle soit, ne devrait pouvoir mettre en danger un patrimoine public aussi précieux que l'histoire et le savoir.

Sommes-nous vraiment en mesure de savoir comment une entreprise, a fortiori étrangère, liée à des contraintes financières, guidée par des investisseurs et dépendante d'une maison mère... va évoluer dans le futur ?

[3.1.4] **Motivation et intérêts**

On peut analyser facilement tous les « intérêts » de cette plateforme^[25] Google Play Books.

Aujourd'hui Google est une entreprise privée qui cherche avant tout le profit. L'évolution de la plate-forme en ligne a laissé place à des liens d'affiliation vers son service Google Play Books, qui propose des livres à l'achat au format Epub^[A]. On peut de ce fait y voir certainement une forme de concurrence déloyale.

En 2012 Google ouvre sa plate-forme de vente d'E-Book^[15], Google Play Books, sur le modèle de son store d'applications déjà très célèbre Google Play.

Dans cette démarche, elle affine les livres qui étaient en consultation libre scannée sur Google Book vers sa « boutique » de livre payants !

À l'aide d'un format Epub^[A] soumis à des DRM^[11] encapsulés, elle s'assure un contrôle du produit grâce à des systèmes verrouillant la possibilité de partager le fichier du livre.



Guerre et paix, leo
tolstoy E-Book,
Google Play Livre (2019)



Affiliation commerciale
books.google.fr (2019)

Par la suite elle intègre même Google Play Books à Google Play: aujourd'hui une grande partie des smartphones Android vendus dans le monde est de ce fait fournie par défaut avec Play Book pré-installé.

Ainsi, assise sur des services utilisés quotidiennement par la majorité du monde (Android, google - recherche etc.), aidée par sa position dominante, elle crée des portes d'entrée vers sa bibliothèque numérique, qui devient de facto incontournable.

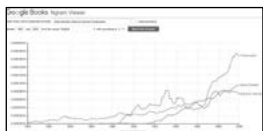
Affiliation commerciale

Pour concrétiser un système de monétisation, les affiliations commerciales Google proposent des liens à côté de certains livres, permettant l'achat physique sur des plateformes^[25] d'e-commerce.

Malheureusement dans ces partenariats commerciaux, les petits revendeurs ne sont pas proposés, et seuls les mastodontes de la vente en ligne sont présents. Nous sommes encore en présence d'une forme de concurrence déloyale.

Traitement textuel de masse

L'immense nombre d'ouvrages que contient Google Books représente aussi un fond de connaissances et de données textuelles propice aux opérations de calculs et aux traitements algorithmiques, utiles à l'amélioration d'autres services de l'entreprise. Ainsi, l'énorme corpus multilingue extrait grâce au système d'OCR^[26] constitue une aubaine pour améliorer des outils tel que Google Traduction.



Comparaison de l'utilisation des nom Albert Einstein, Sherlock Holmes, Frankenstein dans la littérature depuis 1800
books.google.com/ngrams
 (2019)

On peut retrouver pareil « siphonnage légal » pour alimenter les énormes bases de données nécessaires au développement de futurs systèmes tel que ceux dans les laboratoires de Google X.

Un puissant outil en ligne permet par exemple de comparer l'utilisation de mots à travers les siècles recensés dans des ouvrages numérisés. Des outils qui peuvent être très pertinents pour des chercheurs!

[3.1.5] Conclusion

Que cela soit pour un simple lecteur d'e-book^[15], un étudiant, ou encore un éminent chercheur, on ne peut nier l'expertise et la facilité déconcertante que procurent les outils de Google. Elle arrive à elle seule à couvrir un champ impressionnant entre tous les points d'entrée et de sortie de son projet.

Enfin de compte Google Books est difficile à définir, il ne peut s'apparenter au seul terme de bibliothèque numérique. Il en est même bien loin, ne serait-ce par sa volonté finale de vendre des livres.

En fait Google books se base sur l'expertise multifacettes de son système global, une interconnectivité de ses différents services, une multitude de fonctionnalités à tous les niveaux qui, d'une manière ou d'une autre, sait avec pertinence rendre service à tous nos besoins de savoir, de comprendre et d'agir.

Mais son futur incertain ne peut apporter une véritable réponse à long terme aux enjeux de conservation sûre et de diffusion libre du patrimoine culturel qui reste, rappelons-le, la vraie raison d'être des bibliothèques, numériques ou pas.

[4.0] **Changement sur l'accès au patrimoine**

La numérisation, qui permet de reproduire des documents sans dégradation, autant qu'on en souhaite en nombre, et de les diffuser à la vitesse du temps réel, semble offrir un terrain idéal à leur circulation.

Via Internet, cet accès massif à du contenu a théoriquement pour effet d'assurer l'égalité d'accès à la culture. Il est en effet possible de rentrer en quelques clics dans nombre de grandes bibliothèques et de consulter des fonds d'archives rares, normalement réservés à une élite de chercheurs, et cela depuis tous les bouts du monde. Certain droit universel trouve ici une concrétisation évidente : le droit de tous à accéder à l'information, affirmé d'ailleurs dans la déclaration des Nation Unies.

D'autre part, sa numérisation sert le patrimoine culturel en permettant sa préservation : bien que le support numérique ne puisse constituer une sécurité entièrement avérée, il permet en tout cas sa consultation sur écran sans manipulation de l'original.

Sur la question de l'accessibilité physique, il existe en France par exemple de nombreuses bibliothèques publiques de proximité : les bibliothèques municipales, intercommunales, départementales, ou bien encore celles d'enseignement comme les bibliothèques universitaires. Selon leur taille, malgré la richesse affichée de leurs collections, elles ne peuvent contenir qu'une toute petite partie du patrimoine français. Surtout quand l'on parle d'ouvrages rares, fragiles et parfois uniques.

Cette tâche en France n'est quasiment réservée qu'à la Bibliothèque Nationale de France située dans la capitale, dont la majeure partie des collections historiques est réservée à une élite de chercheurs autorisés.

On comprend aisément que la numérisation de ces collections et sa mise en ligne participe grandement au dynamisme d'ouverture et d'égalité d'accès au riche patrimoine français.

Le projet Gallica de la BNF est un bon exemple de prise en considération de cette problématique particulière qui a amené très tôt à concevoir une hybridation physique/numérique.

[4.1] **Gallica**

Gallica est la bibliothèque numérique de la Bibliothèque Nationale de France.

Sa création fait suite au projet de créer une « bibliothèque d'un genre nouveau », selon le vœu de François Mitterrand en 1988. Elle fut une des pionnières dans le milieu jalousement protégé des bibliothèques historiques à afficher la vocation de numériser et rendre accessible de façon libre une grande partie des documents de sa grande bibliothèque mère, la BNF. Gallica demanda 10 ans d'étude et de mise au point pour une mise en ligne fin 1997.

On y retrouve majoritairement des livres, mais aussi des manuscrits, de la presse et revues, des cartes, estampes, photographies, partitions, monnaies, médailles et des documents sonores et vidéos.

Cette plate-forme numérique a été conçue pour être une bibliothèque de référence notamment pour les chercheurs souhaitant exploiter les potentialités du numérique, proposant ainsi une large sélection d'œuvres à un accès pluridisciplinaire, le tout reconstitué selon une charte documentaire très détaillée.

[4.1.1] **Politique**

En tant que projet de la Bibliothèque Nationale elle ne sert pas un but financier, sa mission ne cache aucun besoin de rentabilité, elle est financée par le budget de l'État directement.

Par son statut institutionnel elle est soumise à une certaine charte patrimoniale et administrative, et demeure sous la tutelle du Ministère de

la Culture. Un décret du 3 janvier 1994 nous clarifie ses objectifs :

« - (...) collecter, cataloguer, conserver et enrichir dans tous les champs de la connaissance, le patrimoine national dont elle a la garde, en particulier le patrimoine de langue française ou relative à la civilisation française. » (...) d'assurer l'accès du plus grand nombre aux collections, sous réserve des secrets protégés par la loi, dans des conditions conformes à la législation sur la propriété intellectuelle et compatible avec la conservation de ces collections ».

<http://comitehistoire.bnf.fr/dictionnaire-bnf/biblioth%C3%A8que-nationale-france>

Gallica jouit donc des mêmes engagements que la BNF, ce qui l'oblige à s'assurer d'une grande pluralité représentative dans les choix de son contenu.

L'ensemble du contenu numérisé est accessible sans inscription et sans limitation via un portail numérique.

L'utilisateur a la possibilité de consulter le document directement sur le site ou bien de le télécharger dans différents formats de lecture libre, en image ou en texte PDF^[D], Epub^[A], texte (txt^[C]).

[4.1.2] **Comment ?**

La majeure partie de ses ressources de contenu vient de son propre fond d'archives.

En effet depuis 1537 la BNF remplit le rôle du dispositif de dépôt légal en France, c'est-à-dire qu'un exemplaire de chaque document (livre, revue, magazine...) imprimé en France (comptant un ISBN^[22]) doit lui être remis, ce qui la dote d'une quasi-totale inclusivité du panel culturel imprimé Français.

Mais elle nourrit aussi ses collections avec des dons, des legs et des achats.

BnF Gallica 🔍 ☰ FR ▼

TOUTES NOS SÉLECTIONS PAR TYPES DE DOCUMENTS PAR THÉMATIQUES PAR AIRES GÉOGRAPHIQUES BLOG

LES ACTUALITÉS
de la Bibliothèque numérique

5 830 837 DOCUMENTS EN LIGNE

RENSEIGNAGE

“
Quel est donc ce crime contre nature de tenir une moitié du genre humain dans une éternelle enfance ?”
George Sand, Léïla

À LA UNE

Jean Douchet
Intelligente passionnée et pédagogue, Jean Douchet fit découvrir le cinéma à la jeunesse grâce à la télévision scolaire des années 1960-1970. En 2018, Gallica lui consacrait un billet.

Essentiels de la philosophie
Retrouvez les philosophes français du 16^e au 20^e siècle et accédez rapidement aux grands textes classiques dans leur édition de référence.

LA PÉTITE BOTE

LA SÉLECTION

En l'an 2000
Illustrations futuristes réalisées entre 1895 et 1910, les images de la série "En l'an 2000" sont désormais consultables dans Gallica !

Droit(s) et jardin
Suite au colloque "Droit(s) et jardin" tenu à la BnF, Gallica retrace l'histoire des jardins, de la communauté des maîtres jardiniers de Paris au Moyen Âge à la création des jardins ouverts à la fin du 19^e siècle.

LA SÉLECTION

Presses et revues **Livres** **Manuscrits** **Objets**

DÉCOUVRIR DÉCOUVRIR DÉCOUVRIR DÉCOUVRIR

Home page BnF Gallica

gallica.bnf.fr/accueil/fr (2019)

La numérisation se fait soit en interne sur le site de la Bibliothèque François Mitterrand à Paris, soit par le biais de prestataires externes. Ce qui lui confère une certaine maîtrise de la chaîne de production.

[4.1.3] Rôle et objectifs

Gallica a les moyens de remplir de nombreux objectifs pour la BNF.



Manuscrits de la bibliothèque de l'Arsenal année 1772

gallica.bnf.fr/ark:/12148/bv1b53057297s?rk=2145
(2019)



Recherche par types de documents

gallica.bnf.fr/accueil
(2019)

Parmi eux on peut tout d'abord évoquer son aptitude exceptionnelle à la conservation. Comme on l'a vu plus haut, le support numérique apporte une version supplémentaire et décentralisée d'un document, sa consultation par le biais de la plateforme^[25] n'altère plus le document original.

En effet des actions de consultations répétées sur des documents fragiles peuvent gravement réduire sa durée de vie.

Tout document, même très bien conservé, étant voué à l'altération inéductible du temps, sa copie pourra ainsi constituer une trace de son existence pour le futur.

D'autre part, leur mise en ligne permet une diffusion et une mise en connaissance et en valeur des documents « bien gardés » du patrimoine.

Comme la BNF organise régulièrement des expositions pour mettre en avant certaines parties de ses collections, la page d'accueil de Gallica est très éditorialisée.

Les points d'entrée sont multiples: on peut accéder à une collection simplement par le moteur de recherche, mais

aussi par des sélections, par type de documents, par thématiques, par aires géographiques. Ce qui constitue un système avantageux pour un chercheur ou un étudiant.

Par exemple, en cherchant par thématique, il pourra trouver un ensemble de contenus relatifs à son sujet.

[4.1.4] **Conclusion**

À l'évidence, les contenus culturels doivent tenir une place de premier plan.

La France, par le biais du travail des bibliothèques et de ses conservateurs, dispose d'une grande richesse culturelle, il est important de la diffuser.

Sa bibliothèque numérique Gallica est un porteur majeur des enjeux culturels français avec des conséquences bénéfiques dans nombre de facettes de la société, de la recherche, du droit à l'information, et finalement de son rayonnement économique.

Il est incontestable que ce projet a su doter la France d'une souveraineté forte sur son patrimoine. Au moment où nos sociétés de plus en plus tournées sur l'information doivent tenir bon face au projet tentaculaire Google Book venu d'outre-Atlantique.

[5.0] **Participatif**

Un autre aspect particulier de cette mutation est l'aspect participatif que permet la « dématérialisation^[10] »

Le numérique entraîne le passage d'une mémoire de stock, forcément bornée par les espaces physiques de conservation, à une mémoire de flux, extensible dans l'espace et le temps, dynamique, que tout le monde peut alimenter.

Concrètement, cela signifie un sérieux changement dans la construction d'une bibliothèque.

Dans le cas d'une bibliothèque numérique par exemple, le contenu peut être en lui-même rassemblé et/ou numérisé par une communauté elle-même.

L'utilisateur n'est plus uniquement en bout de chaîne comme simple consultant mais bien acteur d'un projet patrimonial. De la rédaction d'une fiche de lecture par un amateur, jusqu'au don d'une collection entière numérisé par une bibliothèque publique, tout le monde apporte sa pierre à l'édifice.

On peut rattacher cette volonté à l'idéologie de la culture libre sur internet. Il s'agit d'un mouvement existant avant internet, mais amplifié par ce dernier pour défendre et promouvoir la liberté de distribuer et de modifier des œuvres de l'esprit sous la forme d'œuvres libres d'accès et de droits et copyrights. Il prône en premier lieu l'accessibilité universelle à « des biens communs informationnels » comme par exemple les productions tombées dans le domaine public grâce aux nouveaux moyens numériques.

L'essence même de ces derniers est en effet propice à ces concepts, la valeur de la connaissance pouvant être multipliée à l'infini et profiter à tous et à toutes pour leur émancipation intellectuelle.

Cette citation de Laurent Chemla pour Owni.fr* illustre parfaitement ce propos :

<http://owni.fr/2012/12/12/internet-cest-un-truc-de-hippies/index.html>

« Conçu en pleine période Flower Power par des barbus libertaires, Internet n'a jamais perdu – malgré les tentatives de récupération politiques et commerciales – son esprit profondément lié au partage ».

Quand il parle de partage, il fait référence au partage multidirectionnel commun.

L'encyclopédie Wikipédia est l'exemple type de cette idéologie. Depuis sa création sous forme de fondation en 2001, ce sont tous les membres de la communauté planétaire des internautes qui alimentent, vérifient et améliorent un contenu encyclopédique partagé qui grossit de manière exponentielle en toute liberté sur le web^[30].

Entièrement financé à partir de dons internes, avec seulement 350 employés, il s'agit pourtant du 5^e site le plus visité au monde !

En effet la création d'une plate-forme participative en ligne demande un investissement matériel et humain bien différent et souvent moins important qu'une bibliothèque classique, dans le sens où une partie des livres n'ont pas besoin d'être acquis et/ou numérisés en interne.

Théoriquement seul le développement de la plate-forme^[25] et son hébergement sont à la charge de l'organisation.

On peut dire que les valeurs portées par ces concepts se retrouvent parfaitement dans la philosophie d'une bibliothèque numérique libre. Il est assez évident que devait apparaître un de tel projet de fondation autour de la conservation du patrimoine.

[5.1] Le cas Open Library

Open Library est un projet né de la fondation Américaine Internet Archive.

Hormis son activité principale d'archivage de l'internet au travers de sa « wayback machine » initiée en 1997, Internet Archive comptabilise de nombreuses collections : vidéos, films, logiciels, images, audios et surtout livres.

Aaron Swartz

Est un informaticien, écrivain, militant politique et hacktivist américain, il a participé notamment à la création de l'Open Library

Elle possède un des plus grand catalogues de livres numérisés au monde sous le nom de domaine openlibrary.org

Brewster Kahleest

Est un informaticien américain, Il cofonde en 1996 Internet Archive, dont il demeure le dirigeant, ainsi que Internet Memory.

Ce projet a débuté en 2007, développé entre autres par Brewster Kahle et Aaron Swartz, deux fervents activistes défenseur de la « culture libre ».

Open Library est une fondation. Elle est financée majoritairement par des dons d'utilisateurs.

[5.1.2] Politique et économie.

Comme abordé plus haut, son statut de fondation lui apporte une certaine indépendance financière dans le sens où elle ne dépend pas des fluctuations du marché que peut subir une entité privée. Seule la confiance que lui accordent les utilisateurs maintient son existence.

Des campagnes de dons sont souvent mises en place : par exemple #savebook fin octobre 2019 a pu proposer de parrainer l'achat et la numérisation d'un livre par Open Library.

D'autre part elle n'est pas non plus affiliée à un État. Ce deuxième point lui permet de ne pas être soumis aux obligations patrimoniales ou culturelles d'un pays en particulier.

À noter qu'étant née américaine, nombre de ses

The screenshot displays the Open Library homepage with the following elements:

- Header:** Includes the Open Library logo, a navigation menu (☰), a search bar with "Rechercher" and a magnifying glass icon, and a "S'enregistrer" button. Below the header are "Naviguer" and "Mes livres" dropdown menus.
- Browse by Subject:** A horizontal carousel with four categories: "Art" (45,291 Books), "Fantasy" (11,160 Books), "Biographies" (7,304 Books), and "Science" (49,260 Books). Each category has a representative icon and a left/right arrow.
- Books to Sponsor:** A carousel of four book covers with "Sponsor" buttons below them:
 - "THE SELLOUT" by PAUL BEATTY
 - "THE BODY KEEPS THE SCORE" by BRAIN, BIOD, AND BODY IN THE WILLEN OF TRAUMA
 - "The Art of Translation" by Theodore Savory
 - "the END of PLENTY" by JOEL K. BOURNE JR.
- Classic Books:** A row of four classic book covers:
 - "THÉÂTRE DE J. RACINE."
 - "NANA"
 - "LECRECE BOGGIA MARIE TEDOR ANGELO"
 - "Hall of Fame"

Home page BNF Gallica
gallica.bnf.fr/accueil/fr
2019

utilisateurs sont quand même anglophones, et de culture occidentale.

Cette indépendance est certes un compromis, mais elle lui assure une forme de neutralité, notamment éditoriale et son lexique est à l'image de ses utilisateurs: sans frontière, assez hétéroclite...



Seveurs dans la nouvelle-Bibliothèque d'Alexandrie, Égypte (photo 2002)

<https://archive.org/about/bibalex.php>
(2019)

Dans cette optique d'indépendance et de décentralisation, une copie (backup) de l'intégralité de ses serveurs est gardée en disponibilité à la Bibliotheca Alexandrina en Égypte, où elle fait référence à la légendaire bibliothèque d'Alexandrie^[4] du monde antique qui rassemblait tout le savoir du bassin méditerranéen.

[5.1.3] **Comment ?**

Son contenu provient de sources très hétéroclites.

La plus grosse partie du fond de livres a été fournie par plus de 150 bibliothèques et éditeurs majoritairement américains. Mais elle provient aussi de dons en tous genres.

On peut noter par exemple une grande campagne organisée par certains membres de la communauté en 2010, intitulée « Libération des Livres ». Presque 1 million de livres libres de droits ont été téléchargés sur la plate-forme Google Book pour être à nouveau rechargés sur archive.org*, afin qu'ils puissent être débarrassés des restrictions d'usage imposées par Google et regagner ainsi pleinement le domaine public.

Ensuite comme abordé plus haut, des dons peuvent être effectués par la communauté sous

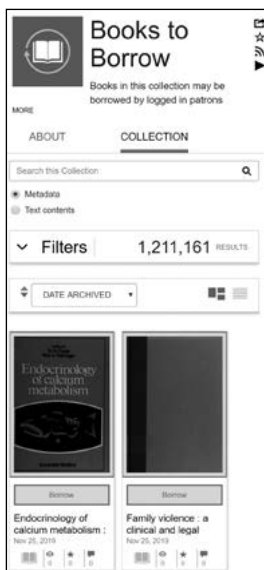
forme de fichiers numériques ou en livres physiques à condition que les droits d'auteur soient respectés.

En fin de compte elle ne possède qu'un petit budget qui ne lui permet pas une grande numérisation en interne, et de ce fait il en découle une pluralité de forme, de contenu, d'origine géographique. Entraînant un travail titanesque pour la communauté, obligée à trier tout cet ensemble.

[5.1.3] Service

Les services de consultation sont divisés en 2 blocs :

Tout d'abord OPEN l'Open Library qui propose



Home page Book to Borrow
<https://archive.org/details/inlibrary?sort=-publicdate>
 2019

uniquement des livres appartenant au domaine public. Ils sont consultables en ligne, imprimables sans limite, et téléchargeables dans les formats les plus répandus, que ce soit en image ou en texte : PDF^[D], DAISY's, ePub^[A] et texte en .txt^[C]

Ensuite vient Books to Borrow, qui joue le rôle d'un système de prêt de livres semblable à ce que l'on connaît dans des bibliothèques publiques. C'est-à-dire que gratuitement, pour une durée maximum de 2 semaines, vous pouvez emprunter un E-book^[15], cette action ayant pour effet de le rendre indisponible pour un autre utilisateur. C'est sur ce concept que repose le flou juridique qui permet à Open Library pour fournir librement du contenu tamponné « sous droit d'auteur ». Le livre, possédé physiquement, n'est que prêté dans sa version électronique ! Le contenu est soumis à des DRM^[11] (Gestion des droits

numériques), est alors encapsulé dans un format ePub^[A], lisible en local, restreint dans le temps et empêchant une copie frauduleuse.

5.1.4] Conclusion

Cette forme de bibliothèque d'un nouveau genre rendu possible par la démocratisation d'internet constitue une formidable ouverture égalitaire à la culture. De surcroît, son statut de fondation confère à cette Open Library une légitimité évidente attachée à ses devoirs de conservation et d'accessibilité au plus grand nombre.

D'autant que son indépendance renforce sa stabilité et donc sa pérennité dans le temps.

Malgré son budget restreint et ses problèmes liés aux droits d'auteur, force est de constater qu'elle remplit parfaitement son rôle de bibliothèque numérique dans un certain respect et éthique de cette tâche.

[#3]

Conclusion

Ce mémoire a pour ambition de mettre en perspective les mécanismes de la numérisation et les bouleversements qu'ils ont généré dans le monde des bibliothèques.

Dans le premier chapitre nous avons tenté de rendre compte du processus de numérisation de l'information et des documents qui est à l'œuvre depuis le développement de l'informatique à la fin du XX^e siècle dans un but de conservation, d'archivage et de diffusion. Mais, en analysant leur impact respectif, nous nous sommes aussi rendu compte de certaines de leurs limites. En effet cette nouvelle structure de mémoire, par sa complexité de mise en œuvre, fait de lui un médium fragile et beaucoup plus dépendant de systèmes électroniques en perpétuelle évolution.

Puis, dans le second chapitre, nous avons voulu analyser les nouveaux enjeux des grands projets de numérisation de masse à travers trois des scénarios que représentent les bibliothèques numériques déjà existantes : Google Books de l'entreprise Google, Gallica de la Bibliothèque Nationale de France et enfin l'Open Library de l'organisme à but non lucratif Internet Archive.

L'étude de chacun de ces exemples a fait apparaître une vision hétéroclite sur les questions de conception et de définition de bibliothèques numériques, au point qu'on peut voir parfois se créer entre elles une sorte de complémentarité.

En effet, Google books se lance dans le projet avec la force de frappe typique que l'on connaît de sa maison mère Google : en se tournant vers une approche ouvertement marchande qui n'hésite pas à prendre le parti

de renvoyer ses utilisateurs vers sa librairie Google Play Books et ses partenaires de vente.

Mais elle permet aussi, grâce à ses algorithmes^[3], ses puissants moteurs de recherche et l'interconnectivité de ses différents services une importante possibilité de recherche individuelle et de travail textuel de masse.

Gallica elle, est davantage tournée vers des objectifs qui restent ceux des bibliothèques patrimoniales conventionnelles : elle cherche avant tout à conserver et garantir un accès public au savoir et à préserver un ancien, précieux et considérable patrimoine majoritairement français. Disons que par convention elle s'appuie sur une mise en valeur des documents par une éditorialisation de son portail.

Open Library enfin, représente un 3ème cas de figure qui vient trouver sa place dans les dynamiques d'égalité et de liberté d'accès à l'information ouvertes par internet : elle est le fruit d'une idéologie où le partage, l'accès libre et gratuit à l'information et la participation sont les pierres angulaires d'une bibliothèque née numérique. Avec un budget bien plus limité que ceux de Google et de la BNF, elle a tout de même pu se développer et rassembler aujourd'hui près de 20 millions de documents en accès libre. De plus, avec Books-to-Borrow elle se positionne comme une bibliothèque de prêt à l'image d'une bibliothèque publique et trouve par ce moyen une solution au problème des droits d'auteur appliqués aux livres.

Trois aventures permises par le numérique, ce nouveau médium plein de promesses.

Mais apporte-t-il une réponse suffisante aux objectifs de pérennisation et de diffusion d'une bibliothèque telle que nous l'avons définie ?

Numériser un document ou un livre, c'est lui créer un double sur un médium où le physique devient flux.

Premier intérêt : Une fois numérisé, le document peut être multiplié aisément et à l'infini.

Ce qui induit un deuxième avantage : la conservation patrimoniale de son contenu dans le temps assurée par cette multiplication. L'apparition de l'impression avait déjà permis d'accélérer ce processus. Avec le support numérique immatériel cette singularité est encore amplifiée.

Mais ses caractéristiques particulières sont aussi la faiblesse de la numérisation : pour être accessible sur simple requête à un serveur central, un contenu sous forme de flux nécessite une infrastructure importante, complexe et fragile. La consultation se fait indirectement et nécessite un décodage de langages informatiques. Les systèmes de stockage et les terminaux de consultation doivent se maintenir perpétuellement et mutuellement à jour pour pouvoir en bout de course afficher le contenu sur l'écran de l'utilisateur.

On peut en conclure que la seule solution viable de préservation passe par la cohabitation des deux supports : le physique et le numérique.

On remarque d'ailleurs que seule la BNF est soumise à des enjeux patrimoniaux et propose au sein de cette même institution les deux versions d'un document.

La diffusion est elle aussi bouleversée.

Comme nous l'avons vu, l'apparition des réseaux informatiques a changé radicalement dans nos sociétés notre rapport à l'information. Face au phénomène, les bibliothèques devaient trouver une réponse suffisante au défi culturel de diffusion qu'elles ont la tâche d'accomplir. Elles l'ont fait à une échelle démultipliée.

Le nombre réduit de certains ouvrages, parfois unique, réduit drastiquement la possibilité de pouvoir y accéder aisément. Une bibliothèque de quartier ne peut proposer un contenu aussi riche que celui des réserves de la BNF

Avec Gallica par exemple, le plus fragile des manuscrits peut être désormais consulté par des millions de personnes simultanément à travers le monde. Il s'agit là d'une vraie avancée sociétale pour l'accès à la culture.

Quel devenir pour les bibliothèques numériques ?

Il est encore trop tôt pour évaluer la fonction qu'elles occuperont dans les décennies futures, mais on peut tenter d'imaginer comment cette cohabitation évoluera à travers le temps.

Le Tout-Numérique ne semble pas être une réponse totalement convaincante pour le moment, aussi bien en matière de pérennisation que de diffusion. Pour cette dernière par exemple, alors qu'on envisageait encore il y a quelques années une démocratisation de masse de la lecture sur écran à travers par exemple les liseuses numériques, l'attachement des gens au livre physique est encore une réalité fortement majoritaire.

Pendant ce temps, les bibliothèques numériques ont surtout profité aux étudiants ou aux chercheurs, répondant à un vrai besoin d'accès à du contenu spécifique depuis n'importe où.

Il n'en demeure pas moins qu'elles constituent une base textuelle colossale, une base qui n'a pas tardé de profiter à la naissance d'une nouvelle forme de technologie : le big data^[6] et les traitements automatisés d'informations : l'intelligence artificielle^[17].

Même si son application semble encore aujourd'hui très abstraite, il y a fort à parier qu'elle constituera un pilier majeur dans une société toujours plus tournée vers l'information. On observe déjà depuis plusieurs années son utilisation dans de nombreux champs d'application, mais qu'en est-il dans celui des bibliothèques ?

Par sa capacité à traiter « intelligemment » et rapidement de grandes bases de données, elle apporte de nouvelles

possibilités d'accès à l'information. À l'image du moteur de recherche Google déjà dopé à l'IA^[17], nous pouvons d'ores et déjà imaginer que des « robots » parviendront à lire nos livres à notre place pour en extraire l'information la plus pertinente pour nous.

Mais un livre ne relève pas uniquement de l'information, mais d'un cheminement de la pensée, de ce fait quelle place occupera-t-il en tant que tel dans un monde où l'information instantanée prend une place de plus en plus prépondérante ?

Toujours est-il qu'avec des bibliothèques plateformes^[25] contenant maintenant plusieurs millions d'ouvrages numérisés, l'IA^[17] deviendra probablement une solution incontournable pour explorer leurs gigantesques « rayonnages ».

C'est là un pari ambitieux. Avec chaque année toujours plus de livres édités dans le monde, la bibliothèque classique trouvera très vite sa limite physique. Elle ne peut pas tout traiter, stocker et diffuser. Un nouveau monde numérique s'est ouvert à une échelle sans précédent.

Il s'agit là d'une mutation de taille autant que de nature, étant donné les moyens informatiques nécessaires mais qui représente une vraie chance pour la conservation du patrimoine et une révolution pour la diffusion du savoir.

[#A]

Annexe

[#A1]

Glossaire

[1] **ASCII**

De l'anglais: American Standard Code for Information Interchange, en français: Code américain normalisé pour l'échange d'information. C'est une norme informatique de codage de caractères, apparue dans les années 1960. C'est la norme de codage de caractères la plus influente à ce jour. ASCII définit 128 codes à 7 bits.

wikipedia.org

[2] **Analogique**

Se dit de systèmes, dispositifs ou procédés qui représentent, traitent ou transmettent des données sous la forme de variations continues d'une grandeur physique (par opposition à numérique).

Larousse

[3] **Algorithmes**

Un algorithme est une suite finie et non ambiguë d'opérations ou d'instructions permettant de résoudre une classe de problèmes.

wikipedia.org (16.11.2019)

[4] **Bibliothèque d'Alexandrie**

La bibliothèque d'Alexandrie a été la plus grande bibliothèque du monde antique. Construite à la fin du IV^e siècle av. J.-C., elle a été définitivement détruite, probablement au moment de la prise de la ville par les Arabes en 641 apr. J.-C.

vikidia.org (20.11.2019)

[5] **Bible de Gutenberg**

Appelée aussi Bible latine à quarante-deux lignes, elle est le premier livre imprimé en Europe à l'aide de caractères mobiles.

[6] **Big data**

Big data est un anglicisme appartenant au domaine de l'informatique et signifiant littéralement «grosses données». Cette locution désigne les mégadonnées, c'est-à-dire l'ensemble des données informatiques à très grande échelle. Le big data est en général trop volumineux pour être traité avec des outils classiques.

linternaute.fr

[7] Bit

Le bit est l'unité la plus simple dans un système de numération en informatique, ne pouvant prendre que deux valeurs, désignées le plus souvent par les chiffres 0 et 1

wikipedia.org (10.10.2019)

[8] Born digital

Littéralement en français : né digital. Il s'agit d'une expression désignant le contenu créé sur ordinateur. Par opposition par exemple à l'ensemble du contenu créé avant la révolution informatique.

[9] Classification Dewey

La classification décimale de Dewey est un système visant à classer l'ensemble du fonds documentaire d'une bibliothèque, développé en 1876 par Melvil Dewey, un bibliographe américain.

wikipedia.org (13.11.2019)

[10] Dématérialisation

La dématérialisation est le remplacement pour une organisation, de ses supports d'informations matériels (souvent en papier) par des fichiers informatiques et des ordinateurs.

[11] Dublin Core

Le Dublin Core est un format descriptif simple et générique créé en 1995 à Dublin (Ohio) par OCLC (Online Computer Library Center) et le NCSA (National Center for Super-computing Applications).

Aussi connu sous l'acronyme DCMI, il s'apparente à un format de balisage (type XML ou XHTML) mais a été spécialement conçu pour décrire des documents de manière simple et standardisée. Il très utilisé dans la bibliothèque numérique aux fins d'enrichir de métadonnées un document et servir les moteurs de recherche (et des systèmes du web sémantique par exemple).

Le Dublin Core prévoit 15 éléments, tous facultatifs et tous repérables, qui portent sur la description

wikipedia.org, dublincore.org

[12] DRM

La gestion des droits numériques ou la gestion numérique des restrictions, en anglais « digital right management », ou encore les mesures techniques de protection, ont pour objectif de contrôler l'utilisation qui est faite des œuvres numériques.

wikipedia.org (08.10.2019)

[13] Deep Learning

Littéralement dit en français, l'apprentissage profond est un ensemble de méthodes informatiques d'apprentissage automatique, basé sur des mécanismes neuronaux.

[14] Disque dur (HDD)

Un disque dur est une mémoire de masse à disque tournant magnétique utilisée principalement dans les ordinateurs.

[15] E-Book

Au sens strict, un e-book, ou livre électronique, est un ouvrage (livre ou magazine) lisible sur un outil informatique, ordinateur, tablette tactile (comme l'iPad d'Apple) ou appareil dédié appelé liseuse ou e-reader.

futura-sciences.com

[16] Fac-similé

Reproduction physique fidèle ou copie d'un livre, manuscrit, dessin ou œuvre d'art originale, généralement par procédé photographique.

[17] IA

L'intelligence artificielle est l'imitation de l'intelligence humaine à l'aide de procédés informatiques.

linternaute.fr

[18] IBM

International Business Machines Corporation, connue sous le sigle IBM, est une société multinationale américaine présente dans les domaines du matériel informatique, du logiciel et des services informatiques.

wikipedia.org (09.11.2019)

[19] ISO (standard)

L'ISO un standard de norme établie par consensus et approuvé par un organisme reconnu, qui fournit, pour des usages communs et répétés, des règles, des lignes directrices ou des caractéristiques pour des activités ou leurs résultats, garantissant un niveau d'ordre optimal dans un contexte donné.

ooreka.fr

[20] Intranet

Désigne un réseau informatique privé utilisé par les employés d'une même entreprise ou de toute autre entité organisationnelle et qui utilise les mêmes protocoles qu'Internet.

wikipedia.org

[21] IDE (Branchements)

IDE est un mode de raccordement de périphériques en parallèle, notamment utilisé pour le branchement de disque dur interne. Il est de plus en plus remplacé par la norme SATA.

**[22] ISBN
(International Standard Book Number)**

En français : Numéro international normalisé du livre est un numéro internationalement reconnu, créé en 1970, identifiant de manière unique chaque édition de chaque livre publié

wikipedia.org (01.11.2019)

[23] Philanthropie

Sentiment qui pousse les hommes à venir en aide aux autres ; amour de l'humanité.

larousse.fr

[24] RVB

Acronyme de Rouge Vert Bleu, c'est le système de codage et d'affichage des images en informatique : l'ordinateur reconstitue l'ensemble des couleurs par addition de luminosité de ses 3 couleurs primaires

[25] Plateforme

Service Web qui met à la disposition de l'utilisateur un ensemble de programmes permettant la diffusion et le partage d'information, de contenus multimédias, de fonctionnalités propres aux médias sociaux ainsi que des applications tierces.

Office québécois de la langue française

[26] OCR (Reconnaissance optique de caractères)

Technique qui, à partir d'un procédé optique, permet à un système informatique de lire et de stocker de façon automatique du texte dactylographié, imprimé ou manuscrit sans qu'on ait à retaper ce dernier.

futura-sciences.com

[27] Silicon Valley

Est le nom donné à une vallée située aux États-Unis d'Amérique, en Californie, au sud de San Francisco, qui regroupe de très nombreuses entreprises de haute technologie et a été le berceau de la révolution électronique et informatique.

vikidia.org (13.11.2019)

[28] Terminal

En informatique, un terminal est une variété de périphérique réseau placé à l'extrémité d'un nœud. Le terminal est un point d'accès de communication entre l'homme et la machine, ou un réseau de machine. Par exemple un ordinateur, une tablette ou un smartphone sont tous les trois des terminaux.

wikipedia.org (14.11.2019)

[29] TAD

Système de transport sur rail des livres au sein du site François Mitterrand de la BNF.

[30] Unicode

L'Unicode désigne un système de codage utilisé pour l'échange de contenus à l'échelle internationale, dans différentes langues, en faisant fi de la langue, de la plateforme et du logiciel utilisé pour cet échange.

journaldunet.fr (14.11.2019)

[31] Web

Le World Wide Web, est un système hypertexte public fonctionnant sur Internet. Le Web permet de consulter, avec un navigateur, des pages accessibles sur des sites. L'image de la toile d'araignée vient des hyperliens qui lient les pages web entre elles.

wikipedia.org (14.11.2019)

[32] Web sémantique

Appelé aussi Web 2.0 le Web sémantique désigne un ensemble de technologies visant à rendre le contenu web accessible, utilisable et compréhensible non seulement par l'homme, mais aussi par la machine. Grâce à un système de métadonnées formelles, utilisant notamment la famille de langages développés par le W3C.

[#A2]

Formats

[A] EPUB

Pour « Electronic Publication ». C'est un format standardisé de livre numérique, créée en 2007. Il est open source et est fondé sur du XML.

Il est modulable sur sa mise en page, et le texte est ajusté en fonction du dispositif d'affichage et du choix du lecteur (à contrario du PDF ou les fichiers image de type jpeg). Il a été conçu fondamentalement pour la lecture de livres. Il est totalement tourné vers le contenu, qui est soit né digital (born digital^[8]) ou alors extrait d'une numérisation grâce à L'OCR^[26].

Il est majoritairement utilisé pour la lecture de contenu sur liseuses. Il est aussi proposé en vente ou avec la distribution d'E-book^[15].

[B] Format de type Bitmap (JPEG PNG)

Développé à l'origine par Microsoft et IBM^[18] pour leurs systèmes, il s'agit de Formats d'images matricielles encodées majoritairement en 8 bits, soit 255 valeurs de Rouge, Vert et Bleu^[24].

Les formats de cette famille ont la particularité d'être invariables et justes sur la conservation du contenu de l'image dans sa forme originale qui ne subit pas d'interprétation de la part du logiciel de lecture.

Dans les champs de l'archive il est beaucoup utilisé comme format de lecture. Il est en effet léger donc optimal pour la consultation web^[31] et, de plus, lisible nativement sur la plupart des terminaux. Son seul défaut est qu'il ne conserve pas beaucoup d'information dans l'image contrairement à son père, le TIFF.

[C] TXT (texte brut)

Le texte brut TXT représente seulement les caractères contenus, sans leur apparence, ni image, ni format. Ils sont encodés parfois en ASCII, mais majoritairement aujourd'hui en Unicode^[30]. C'est un format léger qui représente parfaitement le concept de contenu brut.

[D] PDF

Pour « Portable Document Format ». C'est un format de document standardisé crée par Adobe en 1992.

La spécificité du PDF est de préserver la mise en page d'un document mais aussi ses éléments source de manière distincte tel que la police de caractère ou les images qui sont encapsulées à l'intérieur du document.

Il ne subit en théorie aucune interprétation du système de lecture. Il peut même dans ses dernières versions être interactif et contenir des formulaires, champs de texte etc. Il s'est imposé comme format d'échange et d'archivage international. Et remplace le format PostScript aussi développé par adobe.

Le PDF est à l'origine utilisé pour le born digital^[8] (né digital), notamment en préresse.

Et il possède aussi l'avantage de pouvoir contenir en lui-même plusieurs pages sous un même fichier (type bitmap) ainsi que des métadonnées riches, ce qu'il en fait un format de prédilection pour le stockage et l'archivage du livre (multiples pages) pour lequel il s'est imposé.

[E] TIFF

Le format TIFF a été développé par Adobe suite à son rachat à Microsoft. Il s'agit d'un format bitmap, mais contrairement aux Jpeg et PNG, il peut contenir beaucoup plus d'information et n'est nativement presque pas compressé. Chaque pixel peut être encodé en 32 bits soit environ 4 milliards de valeurs possibles pour le Rouge Vert Bleu^[24].

Dans les champs de la numérisation, beaucoup de scanners fournissent nativement ce format brut, ce qui en fait un fichier de travail et de retouche intéressant avant publication.

[#A3]

Bibliographie

Livres & mémoires

Brosset, Laurent. (2016) Google Livres et la numérisation : quels impacts pour les bibliothèques numériques ? Mémoire, Université de Grenoble Alpes, France.

Cotier, Nicolas. (2018) Vers la pluralité des mémoires, Mémoire, École supérieure des arts et industries graphiques, Paris France.

Laforet, Anne (2011) Le Net Art au musée. Stratégies de conservation des œuvres en ligne, édition LP, Questions Theoriques, France.

Ludovico, Alessandro (2016) Post-Digital Print, La mutation de l'édition depuis 1894, Édition B42, France.

Ory-Lavollée, Bruno. (2002) La diffusion numérique du patrimoine, dimension de la politique culturelle, Rapport à Mme la ministre de la Culture et de la Communication, France

Articles web

À propos, BNF Gallica
<https://gallica.bnf.fr/edit/und/a-propos>

About Open Library (28.10.2019)
<https://openlibrary.org/help/faq/about>

Achiver, après ? Chabin, Marie-Anne Djakarta, (2007)
<http://www.marieannechabin.fr/archiver-et-apres/3-du-papier-au-numerique/>

Big data : des chiffres et des chiffres, Martin, Nicolas, France culture, (06/09/2017).
<https://www.franceculture.fr/emissions/la-methode-scientifique/big-data-des-chiffres-et-des-chiffres>

Des « Robins des Bois » libèrent les livres de Google Book sur Internet Archive, calimaq, – S.I.Lex –, (8.10.2010)
<https://scinfolex.com/2010/10/08/des-robins-des-bois-liberent-les-livres-de-google-book-sur-internet-archive/>

Dublin Core, Bibliothèque nationale de France
<https://www.bnf.fr/fr/dublin-core>
Gallica contre Google Books: le match!, Barthe, Emmanuel, [precisement.org](http://www.precisement.org) (14.02.2011)
<http://www.precisement.org/blog/Gallica-2-contre-Google-Books-le.html>

Genèse et description des bibliothèques numériques, Da Sylva, Lyne, Erudit (11.10.2013)
<https://www.erudit.org/fr/revues/document-2013-v59-n3-documentation0849/1018843ar/>

Les bibliothèques à l'ère du numérique, Eduscol
<https://eduscol.education.fr/numerique/dossier/archives/bibliotheques-numeriques/les-bibliotheques-a-l-ere-du-numerique>

Petites révisions d'été: CCD ou CMOS? Le débat éternel..., Labarbere, Bruno, Les numériques, (30.07.2017)
<https://cutt.ly/oeS87on>

Richard Stallman: refuser l'e-book, pour protéger nos libertés Solum, Clément, ActuaLité.com (08.06.2011)
<https://cutt.ly/qel6Deq>

Vidéo

C'est pas sorcier, France 3, Bibliothèque Nationale: La mémoire à livre ouvert, YouTube (23.05.2013) <https://www.youtube.com/watch?v=teU8tAQo3CE>

Synfonia, BNF /// numérisation de masse, YouTube (26.6.2012) <https://www.youtube.com/watch?v=bKmBm7Ry-GM>

Je tiens à remercier particulièrement ma tutrice Anne Laforet pour son super travail de suivi et de relecture. Mais aussi mon père et ma mère, Isabelle et Phœbé pour leurs corrections et leur soutien.

Il y a 10 ans pourtant personne ne l'aurait cru, mais aujourd'hui, j'ai fini d'écrire un mémoire.

Valentin Maynadié

Mémoire de DNSEP, 2019
Atelier de Communication graphique
Haute école des arts du Rhin

Tuteur :

Anne Laforet

Conception graphique :

Valentin Maynadié

Police de caractère :

Theinhardt
Roboto mono

Impression :

Imprimerie Launay
Paris décembre 2019