## Éléments de définition :

* **Bibliothèques … : Numériques, Electroniques, Virtuelles.**

« *Concept organisationnel basé sur une alliance de l’informatisation de la bibliothèque, de l’accès aux télécommunications et d’une nouvelle gamme d’outils directement utilisables par le lecteur et le dispensant de recourir à un intermédiaire humain.* » Source : Corrado Pettenati.

* On apprend qu’une bibliothèque est une organisation au sens « structure » ou « institution » dans laquelle une activité qui est menée par des agents humains qui travaillent/gèrent des objets.
* L’informatisation de la bibliothèque est le passage des fichiers papiers à un système d’informations (associant des bases de données par auteurs ou éditeurs) pour constituer une fiche qui va décrire les documents qu’on a. L’intégration dans les bibliothèques d’un système d’information de type logiciel documentaire qui va aider à gérer globalement le fond documentaire (gestion des ouvrages, entrées sorties prêts et retours).
* L’accès aux télécommunications signifie l’accès à internet dans les années 90. Signifiant simplement l’accès au catalogue à distance.
* Ces 3 types ont pour point communs : d’être rattaché à une organisation de type bibliothèque, d’être un fond documentaire avec un accès à distance (grâce à une gamme d’outils mis à la disposition des usagers), et avec ces gammes d’outils on amène l’usager à utiliser des ressources à distances et à s’approprier aussi des capacités de recherche de l’information tout en autonomie.
* **Différences entre bibliothèque physique et virtuelle :** 
  + Selon des conditions d’accès aux documents (lecteur n’a plus à se déplacer)
  + Selon l’étendue ou nature des services offerts (aide à la lecture par exemple)
  + Cela ne correspond pas à l’ampleur ou couverture du fonds. Il existe de très petites bibliothèques numériques, proposant tout de même des ressources pertinentes pour l’usager !
* **D’après Gary Cleveland (IFLA) :** « *Organisations qui offrent des ressources, y compris en personnel, pour sélectionner, structurer, offrir un accès intellectuel, interpréter, distribuer et conserver l’intégrité de documents sous forme numérique. Une bibliothèque numérique garantit également un accès sur la durée aux œuvres électroniques dans le but d’être aisément et à un moindre coût disponible à un ou plusieurs publics spécifiques.* »

Cleveland met en avant que dans les bibliothèques numériques, il y a aussi du personnel et une continuité entre le modèle ancien et le modèle nouveau. Le modèle nouveau va consister à continuer les différentes tâches qui sont assignées à ces professionnels (sélectionner des ressources pertinentes, structurer les informations, faire en sorte qu’on puisse y accéder, envisager des possibilités d’accès différents et la diffusion/pérennité du fond documentaire). Toutes ces tâches sont réalisées au regard de la dimension instantanée et un peu trop foisonnante des informations, dans l’univers du web, pour être distinguées entre elles.

Donc peut-importe la tournure que prend l’apparition de nouveau système et les mises en place de nouveaux modèles ce n’est pas pour autant que les missions des bibliothèques vont disparaitre mais également il faut que les différents types de publics/usagers méritent de pouvoir choisir leur manière de renseignements adéquates. Il invite à prendre en compte tout l’aspect organisationnel d’une bibliothèque.

* « **Une collection numérique devient une bibliothèque numérique lorsqu’elle répond aux 4 conditions suivantes** :
  1. Elle peut être créée et produite dans un certain nombre d’endroits différents, mais elle doit être accessible en tant qu’entité unique. C’est-à-dire que nous pouvons avoir une interface « collection numérique » mais les éléments qui la composent peuvent être des ressources se trouvant à l’extérieur et à des endroits différents.
  2. Elle doit être organisée et indexée pour permettre un accès aussi facile que possible. Cette collection doit subir le processus d’analyse et de traitement documentaire/de l’information pour pouvoir identifier les auteurs de la source, ceux qui l’ont publié, etc.
  3. Elle doit être stockée et gérée de manière à avoir une existence assez longue après sa création. Il faut évidemment maintenir sa maintenance et sa gestion depuis sa création. Ce qui signifie qu’elle sera amenée à évoluer dans sa constitution et dans son interface.
  4. Elle doit trouver un équilibre entre le respect du droit d’auteur et la diffusion libre du savoir. Exemple des lois DADVSI, fin de l’exception culturelle et aux fins d’enseignement des droits de copie, droits de citation ? »

Avec la numérisation qui est un processus visant à transformer un document classique, de base en format analogique. Pour chaque ressource numérisée : spécifier l’utilisation autorisée (usage privé, enseignement et recherche, usage commercial …)

## Bibliothèque numérique?

Leur fonction première est évidemment d’apporter les savoirs et connaissances. Elles ont une meilleure diffusion mais tout en amenant à réfléchir sur de nouvelles formes de transmission.

* Années 90 : bibliothèques sur Internet
  + Mise en ligne des catalogues (**SUDOC** : base bibliographique faite par l’agence bibliographique française de l’enseignement supérieur pour trouver un document dans un lieu ou interface).
  + Puis documents numérisés
* Faciliter l’accès, améliorer la qualité des infos données et leur présentation
  + **Quantitatif** : supranational, dans une dimension internationale.
  + **Qualitatif** : recherche texte intégral ou non.
  + Mais ont toujours des limites, comme la barrière de la langue.
* Une bibliothèque patrimoniale, en mode image, pour une dimension historique.

## Intérêt des Bibliothèques numériques

1° **Confronter le lecteur avec de nouveaux codes** (écrit => multimédia). On passe d’une modalité communicationnelle qui est l’écrit à d’autres médias (comme le son, l’image) sans que l’usager n’est à fournir trop d’efforts. Permettre à l’usager de s’approprier des connaissances et savoirs en lien avec l’histoire et le patrimoine en lui soumettant un document avec de nouvelles modalité et en l’invitant à interagir avec cette nouvelle modalité.

2° **Réservoir de documents**. Ici, c’est la question de la traçabilité : il faut pouvoir faire un tableau de bord afin de prévenir les erreurs (pour qu’on puisse se retrouver), de suivre l’historique de la constitution de la bibliothèque et de retrouver les indications des systèmes et outils utilisées (car ils changent souvent pour survivre à leur temps).

3° **Conservation de documents**. On parle de patrimoine du temporaire car les documents numériques ont une durée de vie plus courte (exemple : couche chimique qui s’altère sur les cd/dvd).

4° **Organisation des connaissances** avec les logiques classificatoires ou la dictature du mot-clé dans une barre de recherche.

5° **Mise en ligne de services** (exemple : ask a librarian). On revient à la notion de médiation pour jouer l’intermédiaire et donner une réponse plus pertinente.

6° **Promotion de fonds patrimoniaux** avec la démocratisation de l’accès.

* + - * **Quels corpus ? Quelles collections ?**

Il existe 3 types de ressources pour les Bibliothèques Numériques :

* **Produites et mises en ligne par la bibliothèque elle-même en tant qu’institution** 
  + Ex : Gallica, BNF
  + Archives institutionnelles
* **Ressources commerciales auxquelles la bibliothèque est abonnée** 
  + Ex : périodiques scientifiques pour BU
* **Ressources extérieures mises à disposition de l’usager** 
  + Ex : liste de signets thématiques (des ressources repérer sur le web et qu’on a repéré comme des sources de qualité), etc …
    - * **Comment?**

**Numérisation :**

* Mode image (par scan)
* Mode texte (OCR = reconnaissance optique de caractère)

**Rétablir le lien avec le document matériel :**

* Métadonnées (Dublin Core), qui sont des fiches de références bibliographique sous format numérique et rattachées au document lui-même.

ATTENTION à la question des coûts (coûts de traitement, de matériel, de personnel, de virtualisation) !!

## Exemples

* **Gallica, 1997**. Une interface de base épurée mais suite à l’évolution, on remarque l’intégration d’évolutions technologiques comme la barre de recherche, puis un travail éditorial avec du texte et de l’image.
* **Bibliothèque municipale de Lyon**, à la fin des années 2000. On voit bien que dans les premières années, on retrouvait déjà le ‘catalogue’, et suite aux évolutions, on trouve encore une épuration de l’interface.
* **« Le guichet du Savoir »**, interface mettant en lien des professionnels et des usagers. On parle d’une médiation classique, communication, avec une médiation numérique, sur une interface numérique.
* **« Les Signets »** de la Bibliothèque nationale de France. Ce sont une organisation de connaissances à travers des thématiques, qui sont des liens hypertexte pour les relier avec d’autres interfaces. Mais ce service est de moins en moins utilisé, voir inaccessible aujourd’hui suite aux changements qu’a connu le Web après toutes ces années. C’est un travail se basant sur la manière du traitement documentaire.

## La course au contenu

Les questions à se poser sont : « Qui numérise ? Qui profite ? Qui sont les éditeurs ? D’où cela estparti ? »

* **D’après l’initiative du Budapest, on apprend que l’OAI (Open Access Initiative) :** « *Par "Open Access" nous entendons la mise à disposition gratuite sur Internet des articles scientifiques, autorisant tout utilisateur à lire, télécharger, copier, distribuer, imprimer, exploiter grâce à des outils de recherche, établir des liens hypertextes vers ces articles et matériaux, traiter automatiquement à des fins d'indexation, les réutiliser grâce à la mobilisation de leur contenus par des logiciels spécialisés ou à toute autre fin légale et ce sans autre barrière financière, juridique ou technique que celles qui sont intrinsèquement liées à l'accès sur Internet. Les seules contraintes pesant sur la reproduction et la distribution, et le seul impact des règles de la propriété intellectuelle en ce domaine, sont de donner aux auteurs un droit de contrôle sur l'intégrité de leurs œuvres dans le cadre de ces réutilisations libres et que leur qualité d'auteur soit clairement mentionnée et reconnue dans le cadre de celles-ci* ».

C’est un plaidoyer des scientifiques de l’époque qui souhaitent rendre accessible leurs travaux mais pas que l’industrie informatique se les approprie pour faire de l’argent sur de l’information, qui a été mise de base à disposition pour tous. On sort du cadre documentaire et on se dirige vers des enjeux de société, de positionnement sur le développement d’une société de l’information et d’une société numérique. Surtout, en opposition à un développement d’économie numérique qui vient se baser sur la circulation des informations et des savoirs.

* **OAI-PMH : L'OAI-PMH (Open Archives Initiative’s Protocol for Metadata Harvesting) ou protocole OAI :**
  + Facilite l’échange de données entre des fournisseurs de données et de service
  + Permet de créer, d’alimenter et de tenir à jour, par des procédures automatisées, des réservoirs d’enregistrements qui signalent, décrivent et rendent accessibles des documents, sans les dupliquer ni modifier leur localisation d’origine.
  + Visibilité accrue à des documents
  + Interopérabilité : collecte (moissonnage) les données descriptives de ressources et documents de tous types, accessibles sur l’Internet dans des entrepôts OAI.
  + Peut fonctionner en « surcouche » du Dublin Core
  + Sites de références sur le document : <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm>
* **Conclusion :** c’est une course contre la montre car celui qui acquiert le savoir peut développer beaucoup de choses derrière.
  + - * **Et pourquoi ??**

**NUMERISER** n’est pas tout car on se base sur des **OBJECTIFS** de partage, pour tout **PUBLIC** et quelques-uns en particuliers avec des **COLLECTIONS** diverses et variées (car elles vont concerner un certain nombre de thématiques, de spécialités et de disciplines) tout en proposant des **SERVICES** tout aussi divers et variés.

## Bibliothèque numérique & conservation

* **D’après Michel Fingerhut (directeur médiathèque IRCAM)***: « Avec le papier électronique, lisible sans souris, sans clavier, sans stylet (…) la numérithèque du futur ne possèderait plus alors que des livres blancs où chaque lecteur pourrait tout inscrire.* »

La problématique de la trace, de l’IN-scription avec l’intérêt et enjeu de la mémoire de la connaissance. Conserver des ouvrages n’est pas (que) conserver des écrits car il existe des bibliothèques sans livres. Or ces écrits numériques ont des problématiques de conservation spécifiques …

* **Les atouts du support numérique :** la **c**apacité de stockage, de plus en plus impressionnante et compacte plus le temps passe

Mais …

* **Les problèmes :** 
  + Altération des supports => obsolescence des supports et des dispositifs de lecture …
  + Evolution des formats:
    - Propriétaires ou non … multipliée par spécificités des médias numériques (image, son, vidéo, etc …) ==> exemple du PDF
    - Peu de chance que le phénomène ralentisse. Innovation technologique + stratégie industrielle conséquentes
* **D’après Alberto Manguel** : « *Au musée de Naples, on peut encore lire sur les cendres d’un papyrus sauvé de Pompéi. En revanche, beaucoup de mes disquettes vieilles de 5 ans sont déjà inutilisables. Les technologies modernes nous condamnent à nous occuper d’elles en permanence.* » in Internet, le cauchemar de Babel, L’express 28/03/05.
* On est tributaire avec cette technologie des modalités d’accès.

## L’information et la « Révolution volumétrique »

* Traitement de flux énormes d’informations
  + Exemple des volumes des collectes de l’Internet
* Nouvelles unités documentaires
  + Application de techniques archivistiques aux sites web
  + Qu’est-ce qu’un document ?
* Des techniques bibliothéconomiques industrielles
  + Traitements documentaires automatiques
  + Échantillonnage des contrôles qualité et alertes automatiques
  + Protocoles de recyclage des problèmes

# Bibliothèque numérique et/ou Googliothèque universelle ?

## **Les postulats de depart.**

Des transformations …

* Un changement d’environnement documentaire
  + L’augmentation exponentielle de l’offre documentaire, des ressources, des collections, etc.
  + La pénétration des infrastructures informatiques et de télécommunication
  + Une normalisation multisectorielle
* Au niveau des bibliothécaires
  + Une mutation de l’objet
  + Un rôle d’intermédiation, avec un renforcement
  + L’impact sur les fonctions des bibliothécaires
* **Slogan de Google « Organiser l’information du monde, la rendre accessible à tout l’univers »**

## Projet OCEAN

A l’origine du projet « ocean », Google passe un accord avec 16 grandes universités pour référencer les contenus de leurs "dépôts électroniques de documents". MIT (Massachussets Institute of Technology) et 16 autres universités (Cornell University, Hong Kong University of Science and Technology, University of Toronto et University of Washington) pour indexer et rendre systématiquement accessible les contenus des "archives numériques ouvertes" ("digital repositories") créées par ces institutions

## GOOGLE PRINT / BOOKS

Le 7 octobre 2004, Google Print voit le jour :

* **Editeur** : vous envoyez (gratuitement) un exemplaire de vos livres chez Google. Ils scannent tout ça (gratuitement) et l'index.
* **Recherche full text**
* **Feuilletage en ligne** de quelques pages de l'ouvrage (la table des matières)

**L’actualité du projet …**

Livres Hebdo, 20 Janvier :

* « Google aurait numérisé des centaines d’ouvrages français, sans aucune autorisation des maisons d’éditions. »
* « Les oeuvres "pillées" appartiendraient aux éditeurs Gallimard, Grasset, Hachette ou Fayard, et antérieures à 1970, les auteurs étant Albert Camus, Paul Valéry, André Malraux, André Gide ou André Breton. »
* « Le service juridique de Gallimard compte bien poursuivre Google, "nous allons réagir, c’est de la contrefaçon. Nous n’allons pas en rester là", des oeuvres impliquées et imprimés par Gallimard étant toujours soumis à des droits d’auteur. le SNE (Syndicat national de l’édition) risque lui aussi de se poursuivre Google. »

Google pourrait avoir lancé la numérisation des œuvres françaises avant que le projet de la BNE (Bibliothèque numérique européenne) ne prenne trop d’envergure...

**Ce que le débat Google apporte aux bibliothèques …**

On interroge simultanément les 3 facettes qui font, de manière constitutive, la part irréductible du livre dans son appréhension à savoir :

* + Son rapport avec l’individualité consciente qui le feuillette d’une part : *lecture fragmentaire*.
  + Son rapport avec l’inscription qu’il recueille et dont il est la trace d’autre part : *lecture numérique*
  + Le rapport de cette inscription à un héritage culturel partagé : *économie des savoirs du livre*.

**Le débat Google & les bibliothèques …**

* Un moteur de recherche (qui se double d’un modèle économique entre agence média et régie publicitaire) n’est NI une librairie, NI une bibliothèque, NI un éditeur.

MAIS …

* Le point commun de ces trois-là, le : CATALOGUE.
* Or Google s’attaque de manière sans équivalente dans l’histoire du livre et de la diffusion, de la conservation et de l’accès AU CATALOGUE.

**Aujourd’hui Google c’est :**

* Le catalogue
* Le fonds numérisé
* Le public captif + L’adéquation aux nouveaux usages
* Les outils de monétisation de services à l’usager (et aux partenaires : adwords, adsense, Payment Corp)
* L’infrastructure et la technologie de recherche

Conclusion : qu’est-ce qui l’empêche de devenir libraire ? Editeur ? Bibliothécaire ?

## Gallica

Gallica, fondée en Octobre 1997, contient 80 000 ouvrages numérisés (80% monographies, 20% revues). Les **PROBLEMES** sont les choix techniques de numérisation :

* Numérisation en mode image. Pas ou très peu d’OCRisation
* Recherche plein texte possible seulement sur table des matières
* Pas ou mal indexé par moteurs de recherche (pas d’URN fixe). Gallica travaille à intégrer protocole OAI)

La **SOLUTION** depuis 2005 est qu’une nouvelle numérisation en OCR avec rétro numérisation OCR des fonds existants.

## BNUE (Bib Num de l’UE)

Origine : « réplique » de JNJ à Google Print

**Avril-Mai 2005 :**

* 23 bibliothèques nationales signataires
* 6 états pour financement de numérisation (All, France, Italie, Espagne, Hongrie, Pologne)
* Soutien commission européenne (déblocage de 60 millions d’euros)
* 400 000 euros sur budget culture français pour 2006

**Objectifs :**

* « Bibliothèque des savoirs » orienté grand public (et non bibliothèque recherche)
* Interface de recherche commune aux contenus patrimoniaux ou sous droits
* OCRisation (recherche plein texte)
* Pertinence dans le choix des contenus non OCRisables
* Devenir un site portail
* Associer public/privé (= tirer enseignements de Google Print)
  + Mise en ligne de contenus sous droits, « à discrétion de chaque maison d’édition »)
  + Traduire : proposer des partenariats AVANT que Google n’ait constitué son catalogue en surfant sur les dernières « affaires » et « procès » sur l’opt-in

**2 maquettes sur des bases de test :**

* www.bnue.org (Thompson) … inaccessible
* www.bibnum.org (Isako) … minimaliste

**Calendrier (prévu) :**

* Ouverture au public (du portail et non des fonds) annoncée pour Juin 2006
* Mise en place archives européennes prévue d’ici 2008

**Apport français :**

* 300 000 à 400 000 documents (Provenance Gallica)
* + Nouveaux fonds numérisés.

**Le nerf de la guerre : une bataille perdue d’avance ?**

* Pour la numérisation (Réponse de JNJ à une question sur le « budget » au sénat) :

"3,5 millions d'euros pour la transposition de « Gallica » en mode texte et à son extension à quelques grands journaux ou périodiques, et 6 ou 7 millions d'euros (à raison de 100.000 euros par an à compter de 2007) pour la poursuite de la numérisation du fonds."

* Même en multipliant par les 22 bibliothèques européennes qui ont donné leur accord (de principe) … "capex" annuel de Google : 1 milliard de $.

## Une bibliothèque mondiale

Projet annoncé en décembre 2006 partenariat Unesco, bibliothèque du Congrès américain, d’une bibliothèque multilingue et multiculturelle