


Introduction, notions et concepts : le décryptage des enjeux



**Croissance et extension des données – ce que les données font aux bibliothèques !
Journée d'étude co-organisée par Médiadix (CRFCB) et l'URFIST de Paris, 08/12/2017**

Gildas Illien, directeur des bibliothèques et de la documentation
Muséum national d'Histoire naturelle

Data Blues?

J'me sens
largué.



<http://www.intothescreeen.com/2012/10/photoshoot-joseph-gordon-levitt-flaunt-michael-muller.html>



Data Brainwashing?

Plus j'essaie,
moins
j'comprends.

Dataliénation?

Je ne me sens plus tout à fait moi-même dans la start-up nation.

N'ayez plus peur : vous êtes bibliothécaires,
donc vous êtes nés dans les données



<https://wifflegif.com/tags/7077-library-gifs>

Nous sommes en 2021, c'est la pause café à Mediadix : (lieux communs et malentendus)

Elle me dit :

- « Je taffe au data center de la CUB, je m'occupe du hub. Trop de la balle. »
- « Depuis que je suis data librarian, je me challenge sur du big data . C'est juste ouf»
- « Mon job, c'est faire des DMP* pour les Data-scientists de ma fac. »

* *Data Management Plan*

Je comprends :

- Elle est catalogueuse à la bibliothèque de Bordeaux. Son travail l'épanouit.
- Plus trop le temps de faire de belles notices en Unimarc, elle nettoie et enrichit le catalogue avec des scripts pour gérer des imports et des exports de masse; ça la change.
- Elle passe des heures à expliquer à ses chercheurs comment faire des tableurs sous Excel et comment nommer et sauvegarder systématiquement leurs fichiers plutôt que de saturer le bureau de leur PC avec des docs qui s'appellent tous « BERNARD.doc »

Petit lexique apéritif à l'usage du biblio-hipster

En anglais

- Data (textual, digital, raw...)
- Metadata
- Linked data
- Open data
- Personal data / data privacy
- Big data
- Research data
- Data management plan
- Text and data mining (TDM)
- Data visualisation
- Data Hub/Lake
- Data center/repository
- Data librarian
- Data architect
- Data Officer

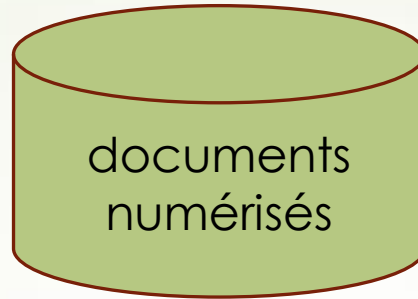
En français

- Données (textuelles, numériques, brutes...)
- Métadonnées
- Données liées
- Données ouvertes
- Données personnelles / protection des...
- Mégadonnées
- Données de la recherche
- Plan de gestion de données
- Fouille de textes et de données
- Visualisation de données
- Système de gestion de données
- Centre/entrepôt de données
- Catalogueur/ gestionnaire de données?
- Architecte de données : responsable SI?
- Directeur des données?



Cartographie simplifiée

Link



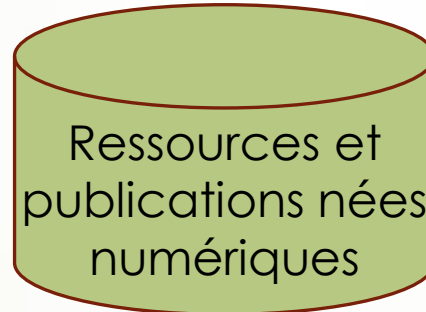
Bibliothèques numériques



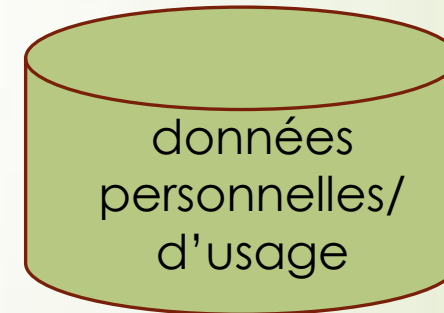
SIGB, enquêtes, outils de gestion...



Catalogues



Portails et outils de découverte



SIGB, outils web...

Open



Divers entrepôts, éditeurs...

Mine

C'est pas si compliqué...

(BnF) Catalogue général Espace pe

l'ours et la bibliothèque ? Tout Recherche avancée AUTEURS A-Z SUJETS A-Z PÉRIODIQUES CC

Accueil > Liste de notices > Notice bibliographique

Notice bibliographique

Notice Au format public

Auteur(s) : [Pastoreau, Michel \(1947-....\)](#). Auteur du texte

Titre conventionnel : [Conférence. 2006-01-04]

Titre(s) : L'ours médiéval [Enregistrement sonore] : histoire d'un roi déchu : conférence du 4 janvier 2006 / Michel Pastoreau, aut. ; Michel Pastoreau, participant

Édition : Reproduction numérisée

Publication : Paris : Bibliothèque nationale de France, 2006

Description matérielle : 2 disques compacts enregistrables (1 h 17 min 39 s, 16 m)

Collection : Conférences de la Bibliothèque nationale de France
Autour de l'exposition "Bestiaire médiéval"

Lien à la collection : [Conférences de la Bibliothèque nationale de France](#)

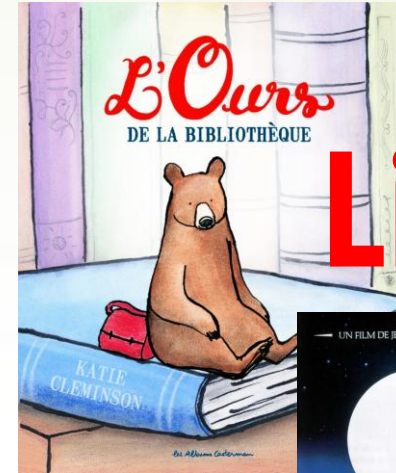
Note(s) : Enr. public
Paris : prod. BnF, P 2006

Enregistrement : (France) Paris, BnF, 20060104

Reproduction : Numérisation de l'enregistrement réalisé par la BnF

Participant(s) : [Pastoreau, Michel \(1947-....\)](#). Participant

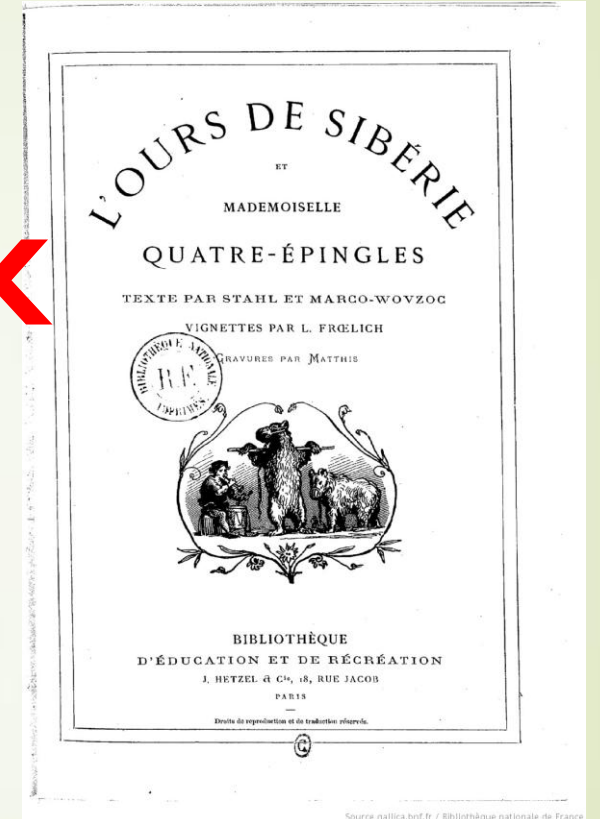
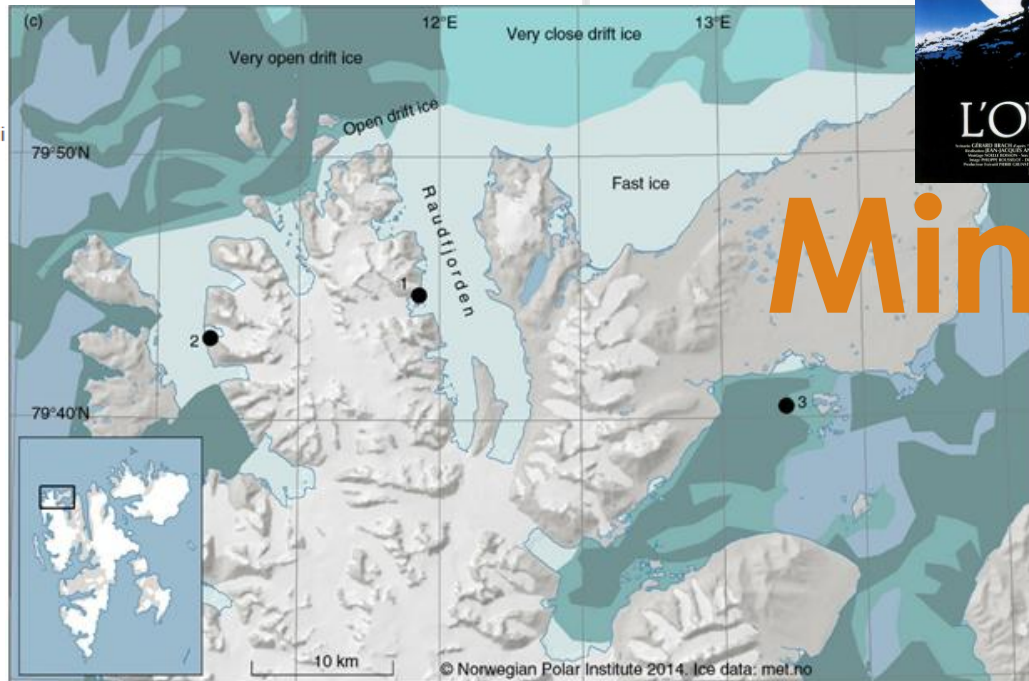
Open



Link



Mine





(1) Des clés pour comprendre le LOD (Linked Open Data)

- **Champ d'application** : les métadonnées, les documents numérisés, les données de gestion et d'activité (horaires, fréquentation, transactions...)
- **Enjeux / intentions** : mettre à disposition du plus grand nombre, humains et machines, faciliter l'interopérabilité, faciliter la réutilisation, mutualiser, s'ouvrir à de nouveaux partenaires, inventer de nouveaux services
- **Leviers** : les standards du web (sémantique), l'engagement des opérateurs nationaux (ABES, BnF...), la transition bibliographique, la numérisation des collections, l'incitation des pouvoirs publics à l'Etat comme dans les collectivités territoriales et au niveau européen
- **Cadre juridique** : de la licence ouverte de l'Etat (Etalab) à la Loi Lemaire pour la République numérique et à la loi Valter
- **Points de vigilance** : documents soumis au droit d'auteur, données personnelles et métadonnées « sociales », risques de se voir re-facturer ce que l'on a ouvert




Focus : les nouveautés juridiques en matière d'ouverture des données

- **Avant la Loi Lemaire** : principe de réutilisation des données publiques, mais avec un régime dérogatoire pour les données culturelles, d'enseignement et de recherche ; l'ouverture revêtait un caractère obligatoire pour les administrations centrales mais relevait du libre choix au niveau local.
- **Avec la loi Lemaire** : Les administrations sont tenues de mettre en ligne et de rendre librement réutilisables : les bases de données qu'elles produisent ; les informations essentielles qu'elles détiennent, les données dont la publication présente un intérêt économique, social, sanitaire ou environnemental.
- = consécration d'un **principe général de libre réutilisation et de gratuité des données** (Loi Valter).
- Pour les données de la recherche : lorsqu'elles sont issues d'une activité financée à 50% sur fonds publics et ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été rendues publiques par le chercheur, leur réutilisation est libre.
- L'ouverture par défaut devient la règle pour les projets financés dans le cadre de H2020.

Demeure une exception possible pour les collections numérisées

- La réutilisation peut également donner lieu au versement d'une redevance lorsqu'elle porte sur des informations issues des opérations de numérisation des fonds et des collections des bibliothèques, y compris des bibliothèques universitaires, des musées et des archives et, le cas échéant, sur des informations qui y sont associées lorsque ces dernières sont commercialisées conjointement. Le produit total du montant de cette redevance, évalué sur une période comptable appropriée, ne dépasse pas le montant total des coûts de collecte, de production, de mise à disposition ou de diffusion, de conservation de leurs informations et d'acquisition des droits de propriété intellectuelle.
- Voir loi Valter :
- « Lorsqu'un tel droit est accordé, la période d'exclusivité ne peut dépasser dix ans. Le bien-fondé de l'octroi d'un droit d'exclusivité fait l'objet d'un réexamen périodique au moins tous les trois ans. »
- « Lorsqu'un droit d'exclusivité est accordé pour les besoins de la numérisation de ressources culturelles, la période d'exclusivité peut, par dérogation, être supérieure à dix ans, sans dépasser quinze ans. Elle doit faire l'objet d'un réexamen au cours de la onzième année et ensuite, le cas échéant, lors de la treizième année



(2) Des clés pour comprendre les enjeux de la « data privacy »

- **Champ d'application** : certaines métadonnées et surtout les données usagers
- **Enjeux/ intentions** : garantir la protection de la vie privée des personnes dans le contexte du Big Data et de l'ubérisation de l'économie numérique
- **Leviers et cadre juridique** : des déclarations à la CNIL au nouveau règlement européen des données personnelles
- « Lorsque les documents comportent des données à caractère personnel, ils ne peuvent être rendus publics qu'après avoir fait l'objet d'un traitement afin de rendre impossible l'identification des personnes concernées »
- **Points de vigilance** : quelles données personnelles dans nos catalogues et nos bibliothèques numériques?

(3) Des clés pour comprendre la gestion des données de la recherche

- **Champ d'application** : l'édition électronique scientifique, les jeux de données primaires et secondaires associées aux publications
- **Enjeux/intentions** : repositionner les bibliothèques en appui au cycle complet de la recherche (et pas seulement la fourniture de documentation), faciliter la collecte et le libre accès au produit de la recherche (Open Access) et permettre l'exploitation et l'interprétation des données à grande échelle (TDM : « The right to mine is the right to read ») au moyen d'outils automatiques d'analyse, de calcul, de visualisation
- **Leviers** : mouvement en faveur de l'Open Access, HAL, Humanités et corpus numériques, obligation d'élaborer des plans de gestion de données (« data management plans »), développement des services à la recherche dans les BU, enjeu de l'évaluation de la recherche et bibliométrie, projets européens qui exigent un DMP
- **Cadre juridique** : Loi Lemaire, règlements européens
- **Points de vigilance** : différences culturelles recherche-documentation, stratégie des grands éditeurs scientifiques commerciaux (privatisation des données de recherche), transformation des activités et des organisations de travail



Parler et penser « données », c'est porter un regard différent et élargi sur les mêmes objets

- ▶ Savoir se distancer des OUTILS et des SYSTEMES pour comprendre les PROCESSUS
- ▶ Savoir se distancer des LANGAGES et des FORMATS et parler LA LANGUE COMMUNE
- ▶ Savoir se distancer du DOCUMENT pour voir l'INFORMATION qui est dedans
- ▶ Savoir se distancer de sa COMMUNAUTE METIER
- ▶ Bref, savoir s'extraire du contexte de production et de diffusion propre à la bibliothèque pour revenir aux fondamentaux :
 - ▶ L'utilisateur et son contexte
 - ▶ L'environnement, les partenaires et concurrents de la bibliothèque
 - ▶ La valeur ajoutée première de la bibliothèque : maîtrise de la structuration normalisée de l'information dans un domaine de qualification donné dans un but de diffusion, de partage et de médiation de la connaissance
- ▶ En intégrant trois facteurs de changement fondamentaux : la dématérialisation, la vitesse et le volume.



Nos atouts pour réussir dans la donnée: nos défauts font nos qualités

- Une profession **homogène** dans sa formation et sa pratique professionnelle
- **L'habitude d'appliquer des normes**, des standards et d'utiliser des référentiels pour structurer et qualifier l'information plutôt que de réinventer la roue à chaque projet
- Une profession **qui s'est acculturée très tôt à l'informatique** (traitements de données, connaissance des SI, gestion de projets applicatifs...)
- Un rapport (généralement) **agnostique aux contenus**, aux disciplines, aux genres...et une grande polyvalence là où d'autres se spécialisent forcément : notre spécialité, c'est l'esprit de système et la vision des processus et des flux
- Une capacité à penser les données **dans l'usage comme dans la durée**
- Une **capacité collective d'adaptation**... fortement sollicitée aujourd'hui.

