Journée d'études URFIST Bordeaux / Médiaquitaine

Catalogues de bibliothèques : quel avenir ?

Emmanuelle Bermès Bibliothèque nationale de France 9 novembre 2010

Postulat de base : les catalogues de bibliothèque ONT un avenir.

???

Souvenez-vous, Il y a quelques années...



Le catalogue est mort : vive le catalogue !



Qu'est-ce qu'un catalogue ?

Un outil de recherche documentaire ?

 Un outil informatique de production de données?

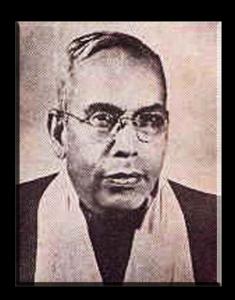
Un outil de gestion d'une collection ?

Fondements théoriques des catalogues



- Charles Ammi Cutter (1837-1903, USA)
 Cutter, Charles A. Rules for a Dictionary Catalog. 4th ed., rewritten. Washington,
 D.C.: Government printing office, 1904
- Shiyali Ramamrita Ranganathan (1892-1972, Inde) Ranganathan, Shiyali Ramamrita. The five laws of library science. Madras: Madras Library Association; London: Edward Goldston, 1931 Numérisé et disponible en ligne: http://dlist.sir.arizona.edu/1220>.
- Seymour Lubetzky (1898-2003, USA) Lubetzky, Seymour. Principles of Cataloguing: Final Report. Phase I, Descriptive Cataloging. - Los Angeles, Calif.: University of California, Institute of Library Research, 1969
- ➤ Elaine Svenonius (1933-...., USA)
 Svenonius, Elaine. The Intellectual Foundation of Information Organization. Cambridge, Mass.: MIT Press, 2000

- Impact des 5 lois de Ranganathan (1931) sur le catalogue
 - Les livres sont faits pour être utilisés : exhaustivité du catalogage
 - * À chacun son livre, À chaque livre son lecteur : éviter le bruit et le silence
 - * Il faut épargner le temps du lecteur : Le lecteur ne doit pas avoir à demander à consulter un document pour savoir s'il répond ou non à ses besoins ; c'est au catalogue de s'en charger. Le lecteur ne devrait pas avoir à mener de longues recherches bibliographiques avant d'interroger le catalogue.
 - * La bibliothèque est un organisme en croissance : enrichissement du catalogue et des liens entre les anciens documents et les nouveaux.

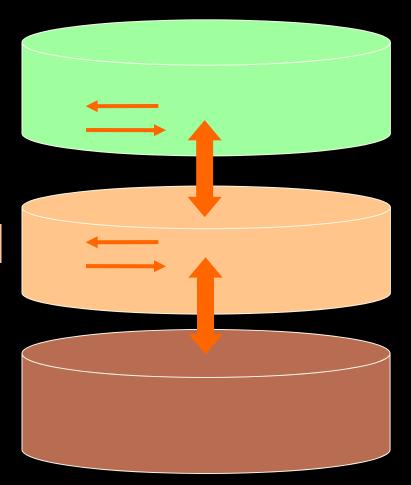


Structure des catalogues actuels

Fichier d'autorité

Fichier bibliographique

Données d'exemplaire



John Ford ou John Ford ? Frédéric Dard ou San Antonio ? IDHEC ou FEMIS ? Story-boards ou Scénarimages ?

Le Grand Bleu version longue ou version courte?
En DVD ou en VHS?
Avec ou sans bonus?

Il est en bon état ? Quelle cote dois-je demander pour le consulter ?

« Trouver »

- Idée qu'il est nécessaire de répondre à un besoin, une question précise
 - Des documents dont on connaît le titre, l'auteur, le sujet ou d'autres caractéristiques
- Problématique :
 - imaginer en amont toutes les caractéristiques qui peuvent être demandées par un utilisateur
 - les rendre identifiables (sans ambigüité)
- Définition de la notion de points d'accès
- Qui débouche dans les catalogues informatisés sur les fichiers d'autorité

« Localiser / colocaliser »

- Le catalogue comme un outil local
 - Est-ce que la bibliothèque X possède cet ouvrage ?
 - Que possède la bibliothèque X sur le même sujet / par le même auteur etc. ?
- Utilisation du catalogue comme un outil de gestion
 - Où se trouve ce document ?
 - Est-il en bon état ?
 - Est-il disponible ?
 - Données d'exemplaires, données locales

« Choisir »

- Le catalogue, une collection de descriptions
- Choisir c'est...
 - confirmer que ce qui est décrit correspond à ce qu'on cherche
 - ou bien, établir une distinction entre des entités qui présentent des caractéristiques analogues
 - Certaines informations sont purement descriptives

Enjeux du catalogue en 2010

 Intégration dans un écosystème global, celui du Web

Économie du catalogage

Que change le Web?

Règne du bruit



Règne du plein texte

Règne du document primaire



Zoom sur Le lecteur



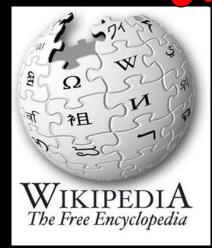




Sur le Web, le lecteur ressemble plutôt à ça...



Où est









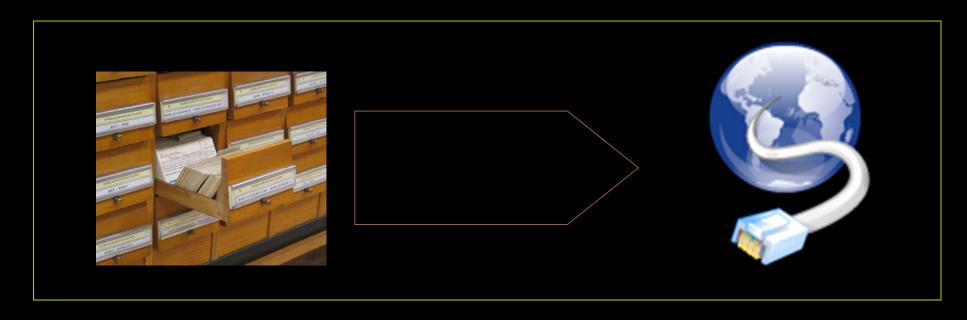
Pour être visible sur le Web, la bibliothèque doit...

- Mettre à disposition des données...
- Dans l'écosystème du Web :
 - Indexables par les moteurs de recherches
 - Reliées avec d'autres données existantes
 - Sur le passage des utilisateurs



Problématique : mettre les ressources documentaires sur le Web

- Pas un nouveau protocole...
- Ni un Web service... Ni une API...



Juste... LE WEB

Opac Web - 1998

Le web profond ou web invisible (en anglais deep web) est la partie du web accessible en ligne mais non indexée par des moteurs de recherche classificatés généralistes.

Consultation de la base

Connexion directe à la base

Contenu de BN-OPALE

Consultation de la base

Le catalogue BN-OPALE de la Bibliothèque nationale de France est accessible via Internet.

Attention:

- La consultation s'effectue avec le protocole telnet Vt100; vous devez disposer d'un émulateur Telnet et configuence...).
- 2. <u>Consultez notre aide</u> en cas de problème de connexion
- 3. Les bases sont indisponibles à certaines heures





Des silos de données cachées

- Des silos de données inaccessibles pour l'architecture du Web (pas d'URL, pas de liens)
 - Protocoles spécifiques : Z39.50, SRU-SRW, OAI-PMH
 - Web services, API Web



http://www.flickr.com/photo s/library_of_congress/217 8285893

Côté utilisateurs :

- * Revoir les règles de catalogage
- * (ISBD/AACR -> RDA)
- ✗ Prendre acte du modèle FRBR
- ✗ Rénover les interfaces (Web 2.0)



Il faut réconcilier la bibliothèque et le Web



Côté machines:

- Revoir la syntaxe
- ➤ Prendre acte du modèle FRBR
- Adopter l'architecture du Web
- * vers le Web sémantique

Réconcilier la bibliothèque et le Web

Étape 1.
Revoir les règles de catalogage
Prendre acte des FRBR

Forces et faiblesses des formats MARC



Les forces :

- Finement structuré
- Sépare le fond et la forme
- Les faiblesses
 - Rigide
 - Pas de modèle

1998 : le Modèle FRBR

Functional requirement for Bbiliographic records

Spécifications fonctionnelles des notices bibliographiques

FRBR ou les pièges du langage

J'ai perdu mon document

un objet matériel

• J'ai distribué le document à tous les participants

un ensemble d'objets matériels partageant des caractéristiques communes en termes de contenu, d'aspect physique et de circonstances de production

• Le document est disponible en PDF, HTML, sur papier

un contenu textuel précisément identifié par delà la diversité de ses supports

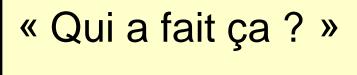
• Il existe trois versions du document

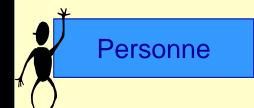
une pure abstraction qui n'existe qu'à travers un vague lien intellectuel entre divers contenus textuels

FRBR : des entités...

Création, vue à un niveau abstrait par delà les différentes versions qui en existent Œuvre Exemple : Les Travailleurs de la mer Version sous laquelle une Œuvre est disponible Expression Exemple : traduction anglaise des *Travailleurs de la mer* Publication (pour simplifier) **Manifestation** Exemple: Toilers of the Sea, G. Routledge & sons, 1886 Exemplaire matériel Item Exemple: ouvrage coté 16-Y2-7542

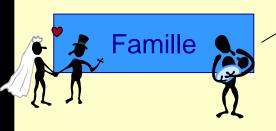
FRBR : des entités...





Individu, pseudonyme, ou pseudonyme collectif

Exemple: Victor Hugo



Au moins deux personnes, quelle que soit la nature du lien familial

Exemple : Famille Brontë

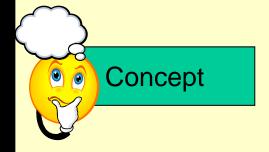


Organisme ou groupement ayant un nom et pouvant agir comme une entité

Exemple : Société des amis de Victor Hugo

FRBR : des entités...

« C'est à quel sujet ? »



Exemple: Transports maritimes



Exemple : la Mary Celeste



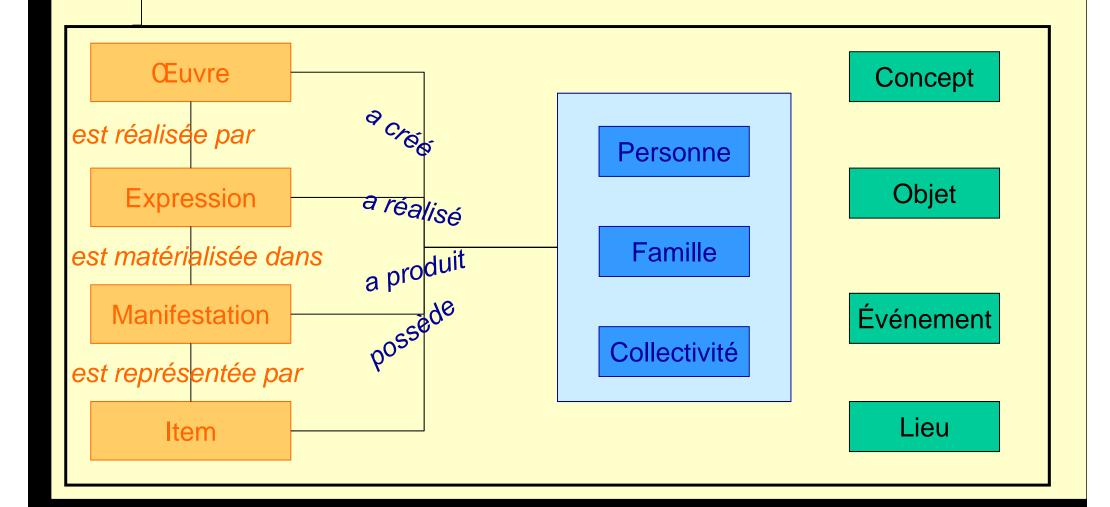
Exemple : Bataille de Waterloo

Exemple: Guernesey

FRBR : des relations...

3. Relations entre les 4 premières entités et toutes les autres

est au sujet de



FRBR : ... mais aussi des « tâches utilisateur »...

• FRBR : tâches utilisateur sur les données bibliographiques :

• trouver

• les ressources correspondant à une recherche

• identifier

- confirmer que la ressource trouvée correspond bien à la recherche
- distinguer entre des ressources analogues

• <u>sélectionner</u>

• entre des ressources analogues, choisir celle qui correspond à ses besoins

• <u>obtenir</u>

• avoir accès à une ressource

Et aussi...

- FRAD: functional requirements for authority data
- FRSAD: functional requirements for subject authority data

L'évènement de l'été

Juin 2010 : publication officielle de RDA

RDA = Resource description and access

« RDA - the new, unified standard for resource description and access, designed for the digital world and an expanding universe of metadata users »

Qu'est-ce que RDA?

- Nouveau code de catalogage destiné à remplacer les règles de catalogage anglo-américaines AACR (AACR = Anglo-American Cataloguing Rules)
 - élaboré par et pour une communauté particulière : les bibliothèques anglo-saxonnes
 - piloté par le Joint Steering Committee for Development of RDA (JSC)
 - composé de représentants de quatre pays : Australie, Canada, Etats-Unis, Royaume-Uni

MAIS

- utilisation supra-nationale de fait
- volonté de devenir un code à vocation internationale

Qu'est-ce que RDA?

- Code de catalogage conçu pour adapter les catalogues de bibliothèques à l'environnement actuel et en particulier au web
 - documents décrits > ressources numériques
 - notices produites lors du catalogage >
 métadonnées qui intéressent les utilisateurs du web
 - outils de catalogage
 - catalogues informatisés > évolution des SIGB vers une structure compatible avec les pratiques du web
 - règles > RDA Toolkit = site web payant

Une révolution dans le catalogage ?

- RDA introduit une nouvelle approche du catalogage
 - « penser autrement » l'information bibliographique
 - à partir des modèles FRBR / FRAD
 - à partir des tâches des utilisateurs
 - « créer autrement » l'information bibliographique
 - vers une nouvelle structure des catalogues
 - une notice par entité FRBR
 - importance croissante et généralisation des notices d'autorité
 - vers de nouvelles méthodes de catalogage
 - dérivation/récupération de métadonnées existantes
 - prise en compte de l'informatisation généralisée des catalogues et des outils de catalogage

Nouvelle démarche de catalogage

- Fondée sur une référence étroite aux modèles de l'information bibliographique
 - modèle FRBR pour les notices bibliographiques
 - modèle FRAD pour les données d'autorité
- Fondée sur la prise en compte des besoins des utilisateurs
 - →associer chaque élément de données
 - aux entités FRBR / FRAD
 - aux tâches des utilisateurs du catalogue

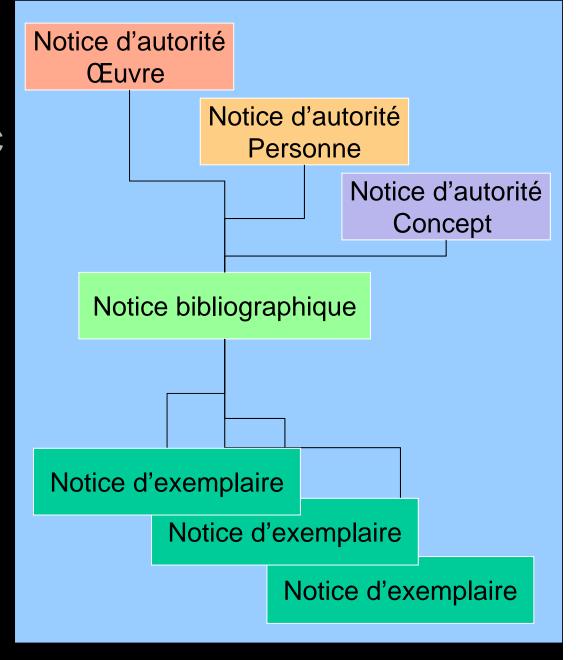
Organisation générale de RDA

- 10 sections réparties en 2 grands groupes
 - décrire les attributs des entités
 - décrire les relations entre les entités
- Chaque section contient
 - des règles générales
 - un chapitre pour chaque entité FRBR
- Chaque chapitre est associé à une (des) tâche(s) des utilisateurs du catalogue exemples:
 - 2. Identifier les manifestations et les exemplaires → TROUVER / IDENTIFIER
 - 3. Décrire les supports → CHOISIR

Structure des catalogues

Période de transition (SIGB existants)

- Évolution des formats MARC pour rendre compte des attributs FRBR / FRAD
- Structure actuelle maintenue
 - notices bibliographiques autonomes
 - → « description composite »
 - notices bibliographiques et notices d'autorité peuvent être liées (scénario 2) ou non (scénario 3)

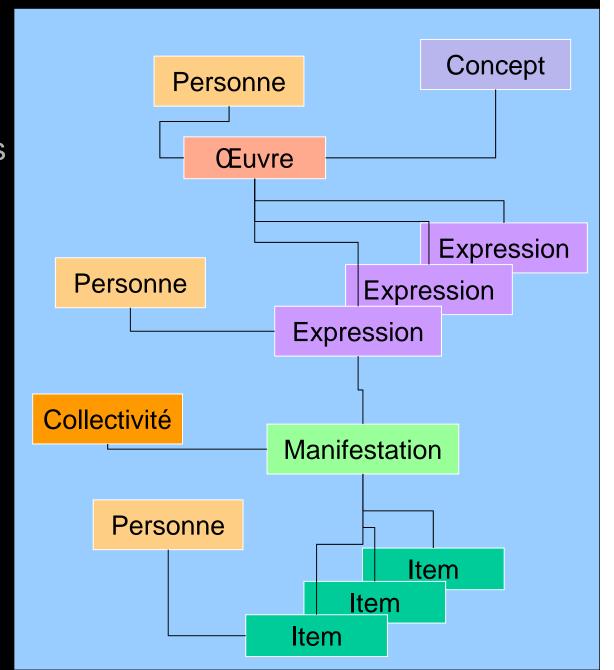


Structure des catalogues

Mise en œuvre complète

(nouvelle génération de systèmes)

- Des notices liées entre elles pour les différentes entités
 - Œuvre
 - Expression
 - Manifestation
 - Item
 - Personne
 - Collectivité
 - Famille
 - Concept
 - etc.

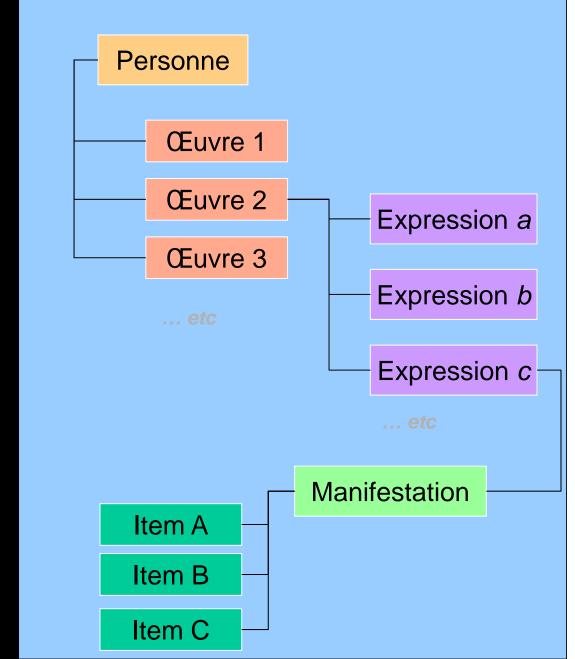


Structure des catalogues

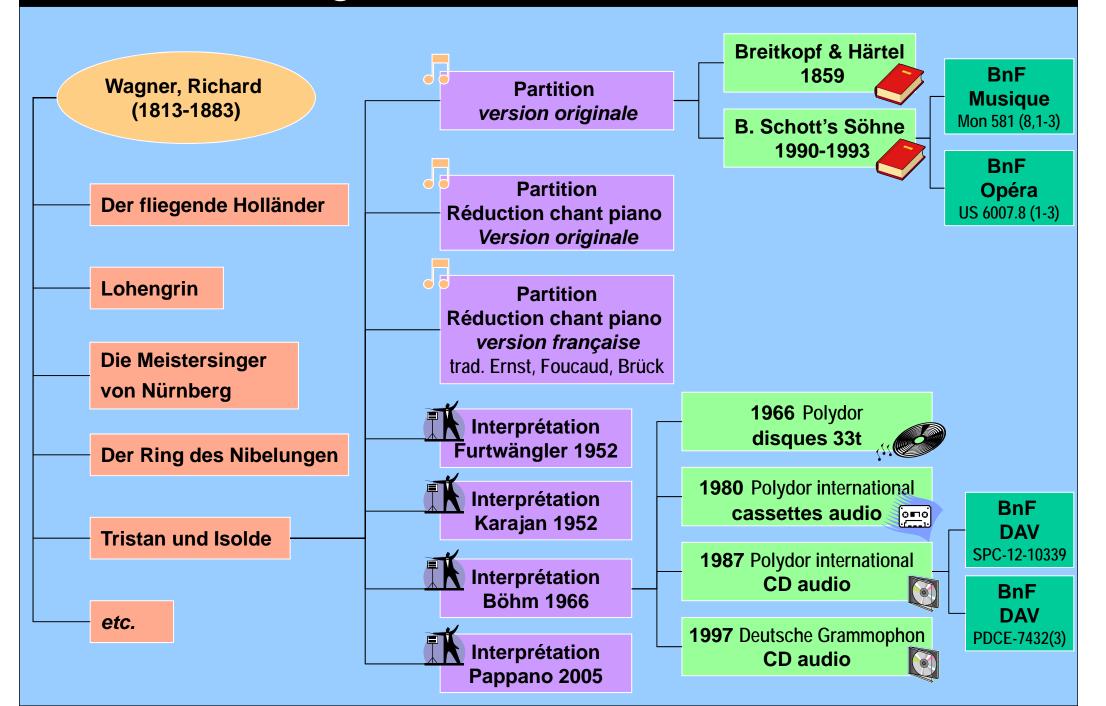
Mise en œuvre complète

(nouvelle génération de systèmes)

- Affichage selon le modèle FRBR
 - toutes les œuvres associées à une personne
 - toutes les expressions associées à cette œuvre
 - toutes les manifestations de la même expression
 - les œuvres ou expressions en relation



Affichage des résultats d'une recherche



Trois scénarios d'implémentation

Scénario 3

- Maintien de la structure actuelle des catalogues
- Établissement de notices d'autorité **non liées** aux notices bibliographiques pour les œuvres et les agents

• Scénario 2

- Maintien de la structure actuelle des catalogues
- Établissement de notices d'autorité **liées** aux notices bibliographiques pour les œuvres et les agents

Scénario 1

- Nouvelle structure des catalogues proche des bases de données relationnelles / orientées-objet
- Une « notice » par entité FRBR



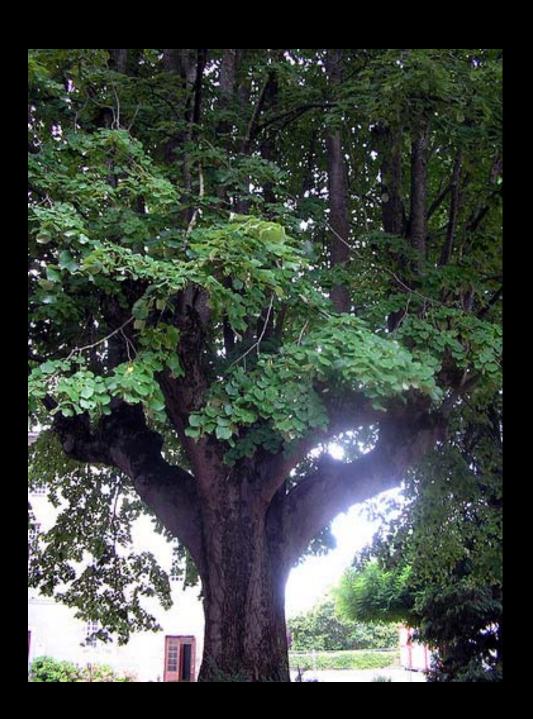
Réconcilier la bibliothèque et le Web

Étape 2.
Revoir la syntaxe
Adopter l'architecture du Web

Que change XML?

- Passage des formats MARC vers XML
 - en agissant sur la structure : passage de ISO 27.09
 à MarcXchange
 - En agissant sur la syntaxe : MODS
- En dehors de la sphère MARC :
 - Dublin Core,
 - * EAD,
 - * etc.





Que change XML?

- Force
 - Une syntaxe plus souple
 - Granularité
 - Structuration adaptée au contexte
 - * facilite les conversions
- Faiblesse
 - Toujours le même modèle (modèle de l'arbre)
 - * 1 notice = 1 document

Les données



Faire exploser la notice ...

Le Web sémantique, qu'est-ce que c'est?

Web sémantique

 un ensemble de technologies & de standards développés par le W3C pour construire le Web de données

Web de données

 une extension du Web permettant de relier non pas des documents (pages HTML) mais les données elles-mêmes, et de les rendre exploitables par des machines

OWL

RDF

SPARQL

SKOS



RDFS

N3

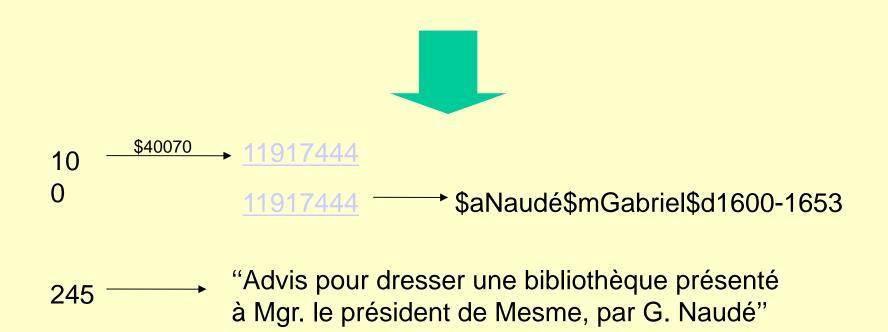
URI

HTTP



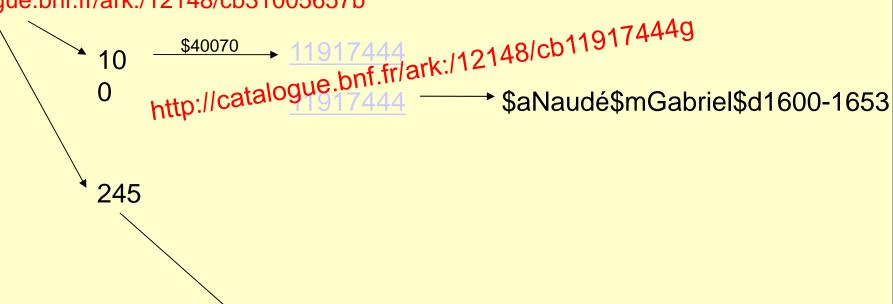
1. Structurer l'information

100 \$3<u>11917444</u> \$w 0 b.....\$aNaudé\$mGabriel\$d1600-1653\$40070
245 1 \$aAdvis pour dresser une bibliothèque présenté à Mgr. le président de Mesme, par G. Naudé,...\$dTexte imprimé



2. Nommer les ressources avec des identifiants Web (URI)

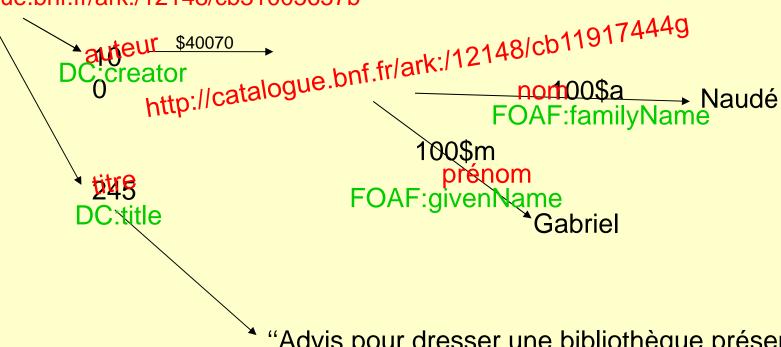
http://catalogue.bnf.fr/ark:/12148/cb31005657b



"Advis pour dresser une bibliothèque présenté à Mgr. le président de Mesme, par G. Naudé"

3. Définir les relations entre les ressources... en utilisant des vocabulaires normalisés

http://catalogue.bnf.fr/ark:/12148/cb31005657b



"Advis pour dresser une bibliothèque présenté à Mgr. le président de Mesme, par G. Naudé"

4. Nommer les relations avec des URIs http://catalogue.bnf.fr/ark:/12148/cb31005657b http://catalogue.bnf.fr/ark:/12148/cb11917444g http://purl.prg/dre/tarms/creator http://xmlns.con/http:/ Naudé http://purl.org/cc/terms/title http://xmlns.com/foaf/0.1/familyName **≺**Gabriel "Advis pour dresser une bibliothèque présenté à Mgr. le président de Mesme, par G. Naudé" rce description Framework

5. Exprimer le tout avec une syntaxe normalisée

@prefix dc: < http://purl.org/dc/terms/>

@prefix foaf: < http://xmlns.com/foaf/0.1/>

http://catalogue.bnf.fr/ark:/12148/cb31005657b dc:creator http://catalogue.bnf.fr/ark:/12148/cb11917444g.

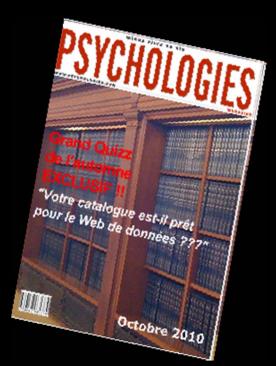
http://catalogue.bnf.fr/ark:/12148/cb11917444g foaf:familyName "Naudé";

foaf:givenName "Gabriel".

http://catalogue.bnf.fr/ark:/12148/cb31005657b dc:title "Advis pour dresser une bibliothèque présenté à Mgr. le président de Mesme, par G. Naudé".

Formalisme RDF - Quel intérêt ?

- Représenter l'information sous forme de graphe
 - Un réseau global d'informations
 - reliées
- Faire éclater la notice
 - Des données et pas des « documents »
 - Plus de souplesse pour manipuler, sélectionner...
- Quelques inconvénients quand même (en termes de modèle)
 - Complexité pour représenter certaines déclarations, les périodes, les provenances, l'absence d'information, globalement, tout ce qui nécessiterait un 4^e élément dans le triplet! Comme par ex. :
 - Jacques Chirac a été président de la République française de 1995 à 2007
 - Ce livre a été publié en 1981 par l'IGN à Paris et par la Chambre d'Agriculture de l'Indre à Blois
 - La date de publication de ce livre est inconnue mais elle doit se situer entre 1917 et 1923 selon la source X
 - Etc.





Identifiants URI

Vocabulaires normalisés

Syntaxe RDF

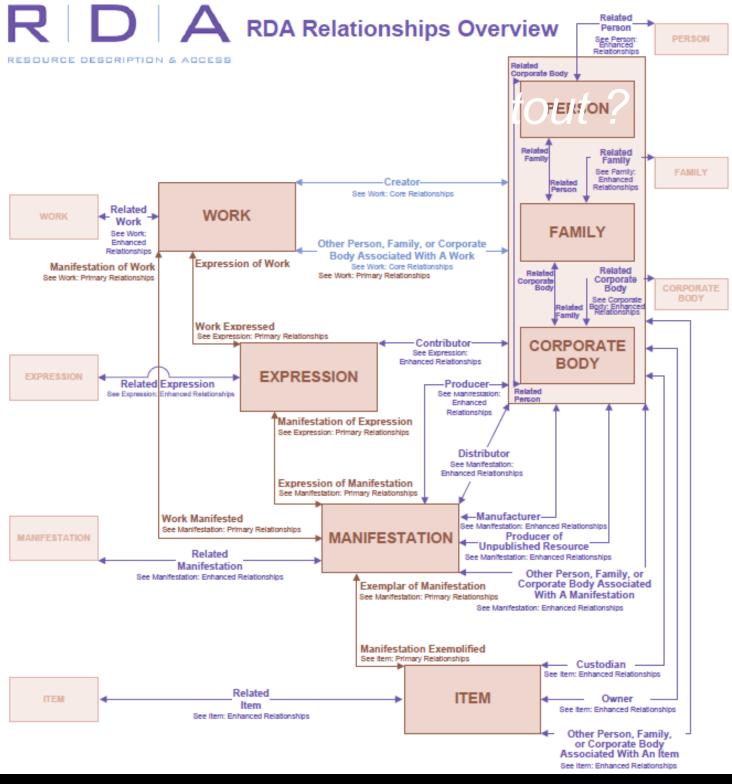
Résultat : Bravo ! Votre catalogue est presput prespout le Web de données. Il ne vous reste reste firm e modèle de données adapté pour le converti en CF.

Définir un modèle pour le Web de données

- Déterminer les classes et les propriétés
- Leur attribuer des URI
- Se concerter avec sa communauté
- Faire valider par d'autres communautés
- Utiliser, tester, réviser...



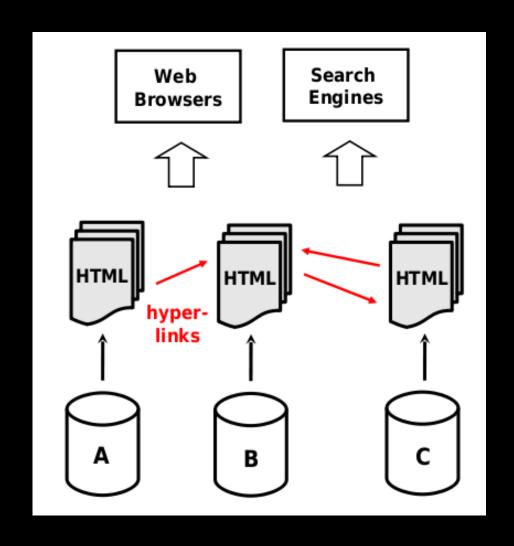
- Prépondérance des liens entre les différentes entités
- Fournit une vue globale (pas de séparation biblio/autorités)



D'un Web de documents...

Principes du Web 1.0

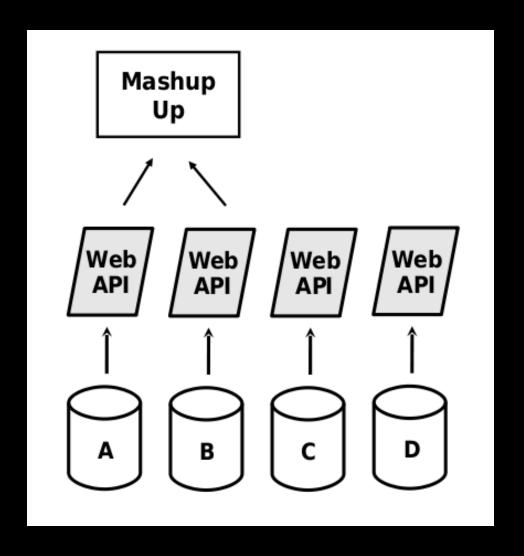
- Les documents (HTML) contiennent des liens entre eux
- » Le contenu n'est pas structuré
- On ne peut pas faire de requêtes
- On ne sait pas l'exploiter avec des machines



...à un Web d'applications...

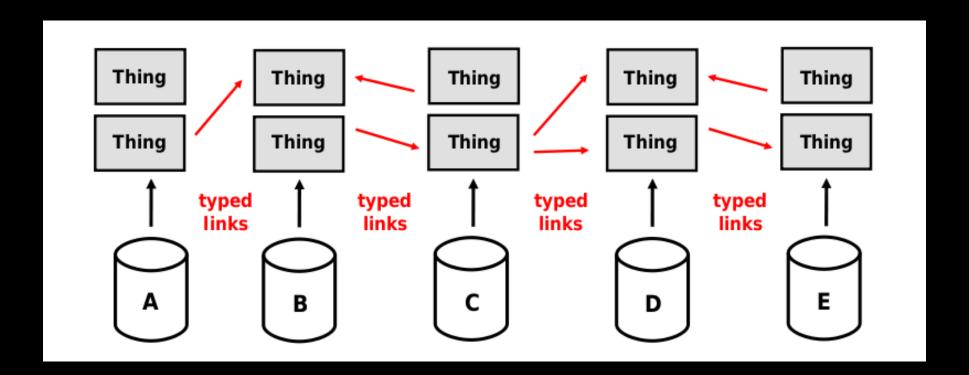
Principes du Web 2.0

- » Les APIs permettent d'exposer les données structurées
- » ... donc de créer de nouvelles applications
- » Mais chaque API est propriétaire
- » Chaque API n'est valable que pour un « silo » de données
- On ne peut pas faire de liens entre les données d'un silo à l'autre



... à un Web de données

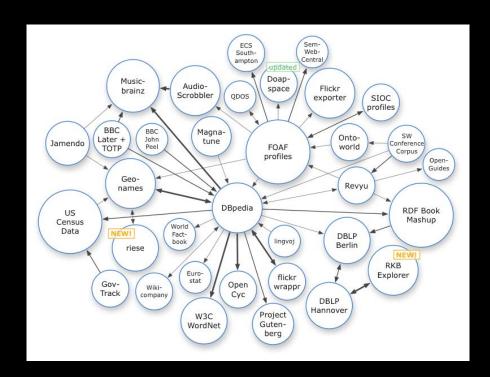
- » Le Web de données est un espace documentaire unifié, global, commun
- On peut faire des liens d'une donnée à une autre
- Les données ne sont plus isolées dans des silos
- » Les APIs propriétaires sont remplacées par des standards ouverts (RDF, SPARQL, HTTP...)



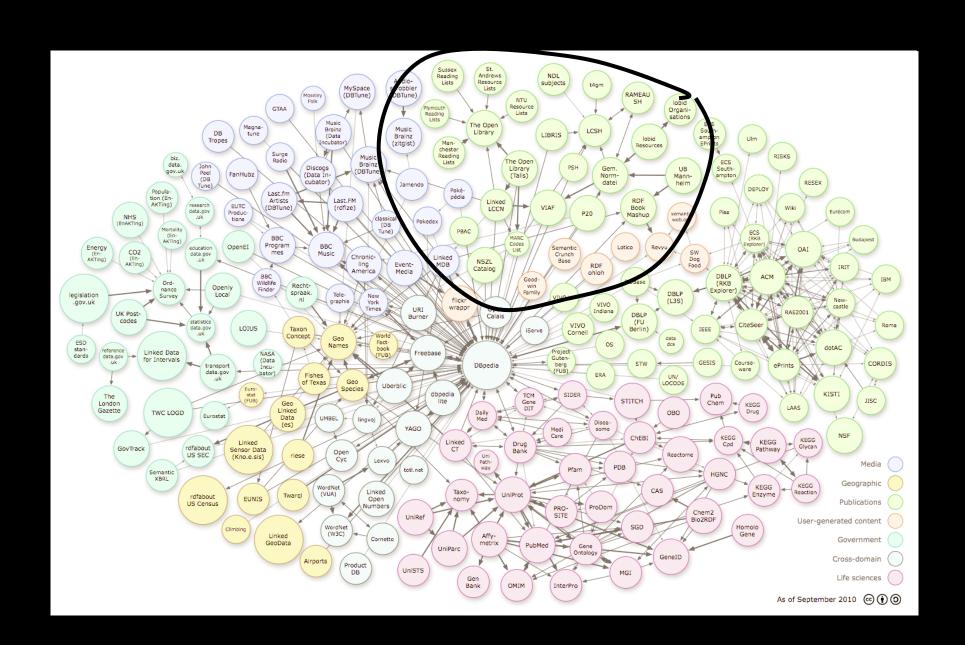
Principes du « Linked Data »



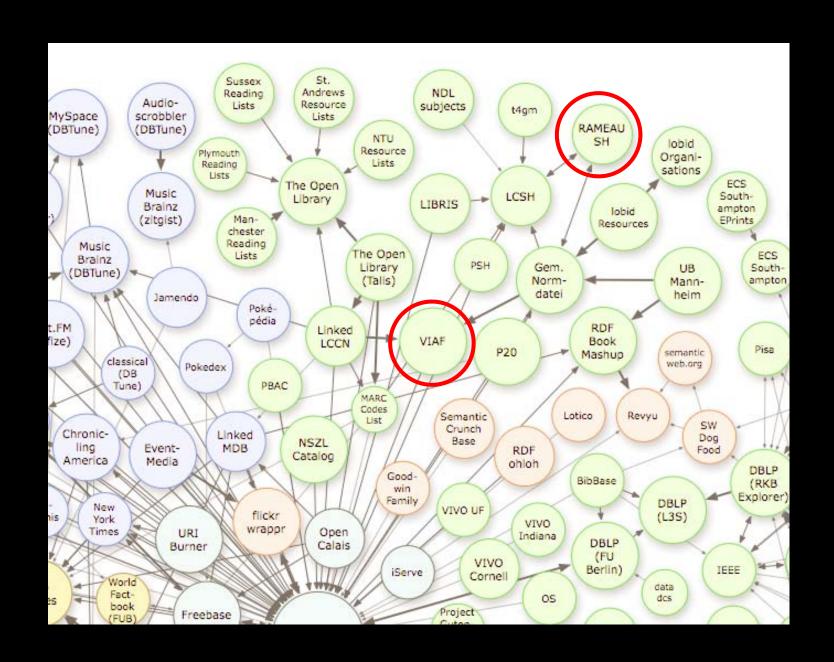
- Villiser des URIs pour nommer les ressources
- ★ Utiliser des URIs accessibles via HTTP
- X Quand quelqu'un demande une URI, lui renvoyer des données utiles en RDF
- Proposer des triplets faisant le lien vers d'autres URI, afin de faire le lien avec d'autres ressources



Les bibliothèques dans le Web de données aujourd'hui



Les bibliothèques dans le Web de données aujourd'hui

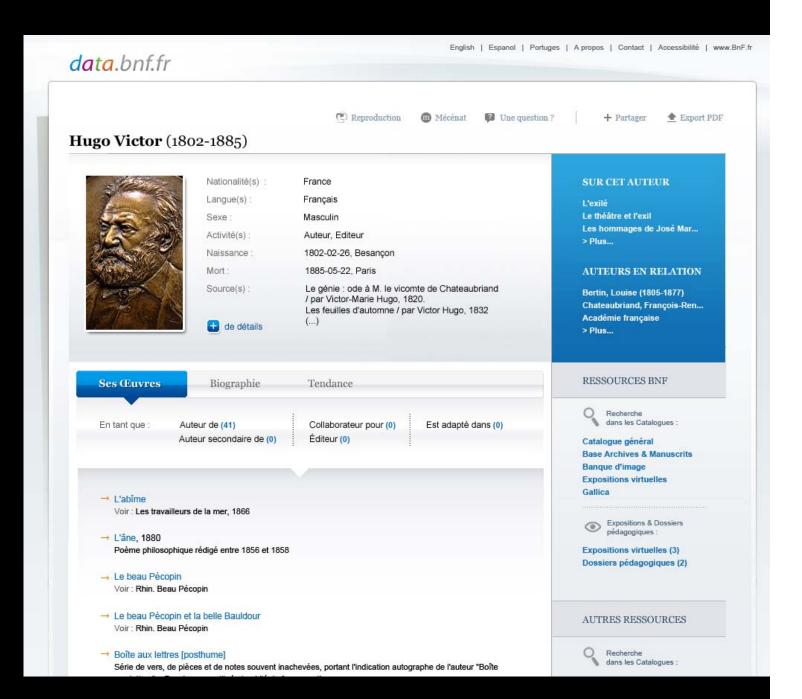


Réconcilier la bibliothèque et le Web

Étape 3. Rénover les interfaces (à suivre...)

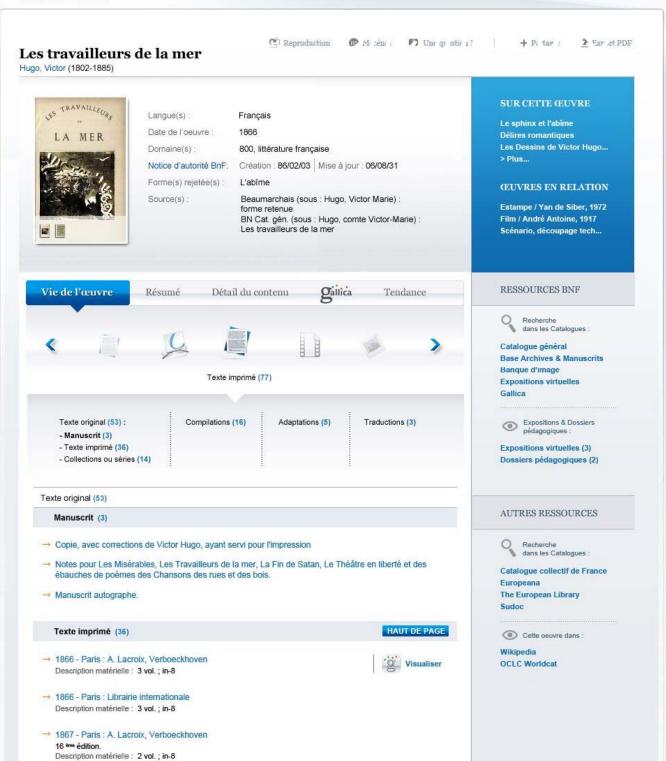
Une interface

Exemple de page « auteur »





Exemple de page« œuvre »



Conclusion

Bon. Alors, cet avenir des catalogues ??????

Quelques questions pour lancer le débat...

- Va-t-on adopter RDA en France ?
- Quand les règles de catalogage vont-elles changer?
- Pourra-t-on encore être catalogueur et ne pas savoir ce qu'est FRBR ?
- Les catalogueurs devront-ils faire du XML ? Du RDF ?
- Combien ça va coûter ?
- Dois-je demander à mon SIGB de faire du Web de données ?

Quelles perspectives en France?

- Pas de code de catalogage
- Règles de catalogage éclatées
 - par type de ressources sur le modèle des anciens ISBD
 - par type d'entité pour les points d'accès
 - ... et désormais « anciennes »
 - obsolescence accrue avec la publication de l'ISBD intégré
- Besoin d'actualiser les règles
 - pour suivre les évolutions au niveau international
 - pour répondre aux besoins des catalogues d'aujourd'hui

Adopter RDA en France?

- Quel intérêt ?
 - Cohérence avec les grands catalogues internationaux
 - Structuration des catalogues adaptée aux besoins actuels des bibliothèques (Web sémantique) avec le scénario 1
- Quels obstacles ?
 - Changement profond des règles de catalogage
 - perte de la cohérence des catalogues
 - Abandon de la référence prioritaire aux recommandations de l'IFLA (ISBD, etc.)
 - Aucune garantie aujourd'hui sur la mise en place du scénario 1 par les SIGB

Adopter RDA en France?

- Remplacer les normes AFNOR par RDA?
- Pourquoi se poser la question aujourd'hui?
 - peu de forces à consacrer à la normalisation du catalogage en France
 - promotion de RDA par l'IFLA
 - RDA = futur code international de catalogage ?
 - perspectives offertes par RDA
 - rénovation en profondeur des règles de catalogage
 - mutation des catalogues
 - ouverture vers le web sémantique

Une réflexion nationale

- Réflexion menée dans le cadre de l'AFNOR
 - Groupe de travail technique
 - début des travaux : septembre 2010
 - Groupe de travail stratégique
- Représentation de la BnF dans les deux groupes de travail

Décision fin 2011

Les 3 âges des catalogues de bibliothèques sur le Web

Données Web

Données de bibliothèques

Opac Web: seulement un accès de surface à un silo de données invisible Données Web

Données de bibliothèques

Next gen'
Catalogue:
fonctionnalités Web
2.0 + APIs pour
rendre les données
visibles

Données Web

Données de bibliothèques

Web Catalogue: catalogue catalogue entièrement conforme à l'architecture du Web

1990's

2000's

2010's?

