



# FRBRiser son catalogue en s'alignant avec les données de la BnF

## Pourquoi ? Comment ?

Sylvain Franceschi (Montpellier Méditerranée Métropole)  
Etienne Cavalié (BnF)

*Journée Systèmes & Données - 14/11/2017*



# Plan

- pourquoi s'aligner avec la BnF dans le cadre de la Transition bibliographique  
--> et pourquoi utiliser des ARK
- alignements :  
Stratégie, méthode, outils
- Ce qu'on peut faire aujourd'hui :
  - nettoyer ses données
  - interroger le SRU
  - recharger les ARK
- Impacts sur les équipes, charge de travail, montée en compétence  
Extraire, recharger
  - pour ceux qui nettoient
  - pour ceux qui alignent
- Partage du code et de la méthodologie :
  - SRU BnF ouvert au public
  - Transmission de la méthode et des outils



# Pourquoi les médiathèques de Montpellier rejoignent-elles la Transition bibliographique ?

Le besoin d'améliorer l'expérience de recherche

- perfectionner les fonctionnalités d'un catalogue agrégé
- prise en compte du comportement des usagers sur le web

La nécessité d'un catalogue fiable en contexte de mutualisation

- Poursuite de la dynamique de transferts de compétences
- Mutualisation avec bibliothèques spécialisées locales



# Pourquoi les médiathèques de Montpellier rejoignent-elles la Transition bibliographique ?

Pour manifester leur besoin d'un service bibliographique national rénové et capable de les aider à atteindre leurs objectifs de signalement



# Avec quel rôle ?

Dans le groupe Systèmes & Données depuis un an

- Rendre compte d'une expérience de réseau de lecture publique
- Proposer d'expérimenter les protocoles de transition
- Entamer avec ses éditeurs logiciels le dialogue sur les fonctionnalités attendues

# Pourquoi commencer par s'aligner avec la BnF ?

## Préparer la récupération des données FRBRisées

- Ne pas FRBRiser "dans son coin" ...
- Accepter le tempo de l'agence bibliographique nationale
- Se préparer à la récupération des métadonnées

## Expérimenter la répartition des rôles dans ce chantier

- Évaluer nos capacités techniques
- Anticiper la montée en compétence des équipes
- Définir le niveau d'implication à attendre de la part des établissements de lecture publique

# Comment ?

Le positionnement de la BnF

3-4 ans pour transformer ses données

- BnF : implémenter le modèle FRBR-LRM
- Bibliothèques : récupérer les identifiants BnF

Une opportunité : l'ouverture d'un SRU pour le catalogue de la BnF

# Le SRU de la BnF

<http://catalogue.bnf.fr/api/>

## Chercher via le SRU (Search Retrieve via URL) dans le Catalogue général de la BnF

Rechercher

Index	Relation	Valeur	Booléen
bib.anywhere	all		and
bib.author	all		and
bib.title	all		and
bib.subject	all		and
bib.doctype	all		and
bib.recordtype	all		and
bib.recordid	all		and
bib.ark	all		and
bib.isbn	all		and

Vue HTML



<http://catalogue.bnf.fr/ark:/12148/cb34592410j>

```
Auteur(s) : Gary Romain (1914-1980)
Titre conventionnel : [Gros-Câlin]français|
Titre(s) : Gros-Câlin [Texte imprimé] : roman / Émile Ajar
Publication : Paris : France loisirs, 1977
Impression : 53-Mayenne : impr. Floch
Description matérielle : 222 p. ; 22 cm
Numéros : ISBN 2-7242-0110-8 (Rel.) : 25 F
```



Vue XML

[http://catalogue.bnf.fr/api/SRU?version=1.2  
&operation=searchRetrieve  
&query=bib.ark%20any%20%22ark:/12148/cb34592410j%22](http://catalogue.bnf.fr/api/SRU?version=1.2&operation=searchRetrieve&query=bib.ark%20any%20%22ark:/12148/cb34592410j%22)

```
<?xml version="1.0" encoding="UTF-8" ?>
<record xmlns="http://www.loc.org/standards/sru" type="Bibliographic">
  <controlfield tag="001" value="FRBNF45924100000004" />
  <controlfield tag="009" value="http://catalogue.bnf.fr/ark:/12148/cb34592410j" />
  <datafield ind1="1" ind2="1" tag="100">
    <subfield code="a">Romain, Gary</subfield>
    <subfield code="d">1914-1980</subfield>
  </datafield>
  <datafield ind1="1" ind2="1" tag="101">
    <subfield code="a">Gros-Câlin</subfield>
    <subfield code="t">français</subfield>
  </datafield>
  <datafield ind1="1" ind2="1" tag="102">
    <subfield code="a">Gros-Câlin</subfield>
    <subfield code="t">roman</subfield>
  </datafield>
  <datafield ind1="1" ind2="1" tag="103">
    <subfield code="a">Ajar, Émile</subfield>
  </datafield>
  <datafield ind1="1" ind2="1" tag="104">
    <subfield code="a">Paris</subfield>
    <subfield code="c">France loisirs</subfield>
    <subfield code="d">1977</subfield>
  </datafield>
  <datafield ind1="1" ind2="1" tag="105">
    <subfield code="a">53-Mayenne</subfield>
    <subfield code="c">impr. Floch</subfield>
  </datafield>
  <datafield ind1="1" ind2="1" tag="106">
    <subfield code="a">222 p.</subfield>
    <subfield code="c">22 cm</subfield>
  </datafield>
  <datafield ind1="1" ind2="1" tag="107">
    <subfield code="a">ISBN 2-7242-0110-8</subfield>
    <subfield code="c">Rel.</subfield>
    <subfield code="d">25 F</subfield>
  </datafield>

```



# Le SRU de la BnF

Possibilité d'interroger automatiquement le catalogue de la BnF

Comment ?

- avec quels critères ?
- avec quel logiciel ?

# Extraction des données

N°	Source	ISBN	Titre	Auteur(s)	Date(100)
1/10000		2-84187-375-7	instruction pastorale de monseig	Fénelon Fran	1984
1/100001	FRBNF38850224		la moitié du geste	Noël Bernard	1982
1/1000017	ELC012743	9,78235E+12	fonds pierre caminade 1911 1998	Ville de Toul	2015
1/1000019			festival de musique à maguelone cathédrale d		2016
1/1000020	FRBNF42770998		Le mois de l'architecture du 9 ma	Direction Ré	2016
1/1000024			Le clergé basque rapports présen		1938
1/100003			Cioè Tragedie di Euripide intere	Euripide	1751
1/1000034			L'Armana adouba e publica de la		1964
1/1000039	OLCL74891		Le maréchal Lannes	Thoumas Ch	1991
1/1000043		2-86274-813-7	le Rhône par Louis Barron l'art l'hi	Barron Louis	1992
1/100009		978-2-9512512-8-1	podalyre à crotone	Audibert-Ca	2015
1/10001			le général Fournier Sarlovez le p	Delpech-Lab	1969
1/100010	FRBNF00124753		dictionnaire de la civilisation gre	Rachet M. F	1968
1/100015			la moitié du geste	Noël Bernard	1980
1/1000177			les petites pièces de Berlin comp	Centre choré	1988
1/1000179			Michele Iglesias peintures palai	Pyrénées-Or	1988
1/1000182			Paul Dardé 1888 1963 le monume		1988
1/1000183			histoire de la psychanalyse en al	Maison de H	1988

# Arbre de décisions

Présence d'un identifiant ?

- oui :
  - ISBN
  - EAN
  - FRBNF
  - ARK BnF
  - ELC
  - OCLC
  - EAN (not. CD/DVD)
- non :
  - Titre + auteur + date ?

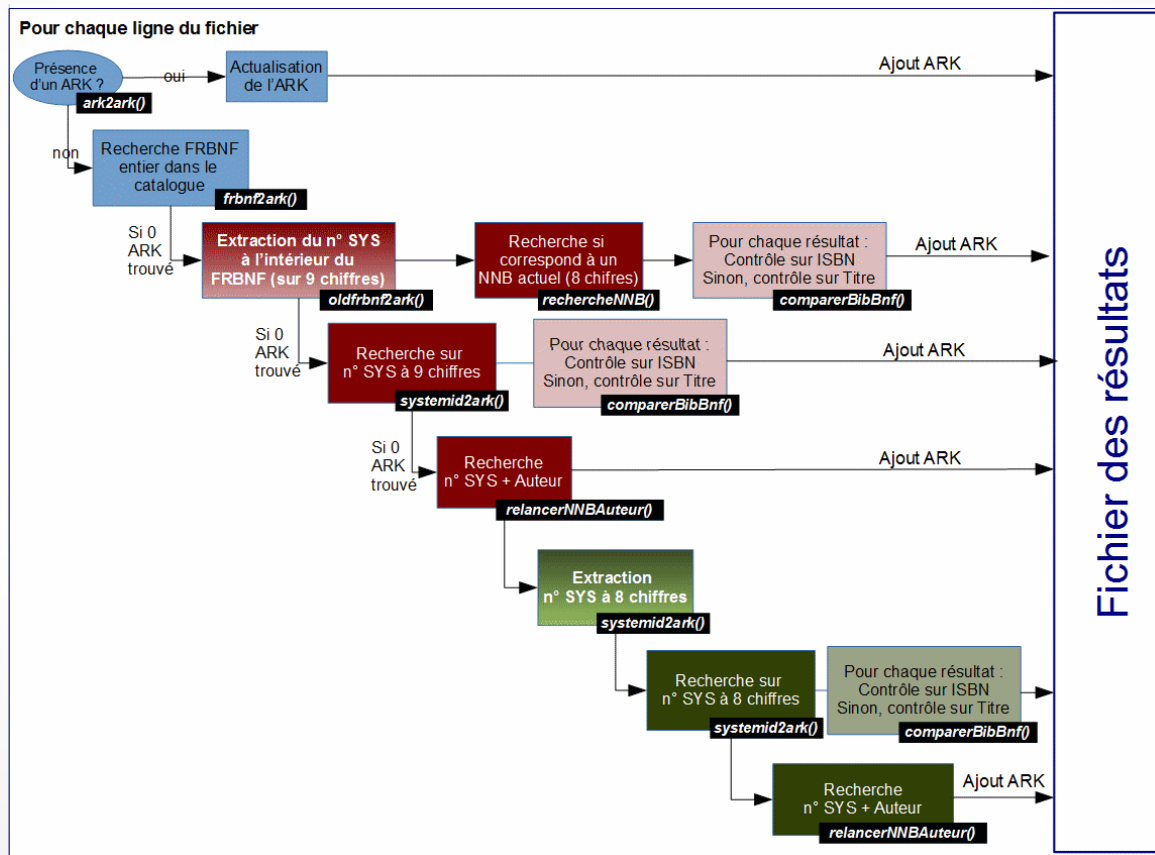
Quelles zones Unimarc ?

Quelle fréquence dans mon catalogue ?

Quelle précision ? -> quels contrôles ?

Quelle présence dans le catalogue BnF ?

# FRBNF



# ISBN

- Nettoyage de la zone 010\$a
- Requête sur ISBN dans le catalogue
- Si pas de résultat : requête sur l'EAN
- Si pas de résultat : conversion 10/13 ou 13/10
- Si pas de résultat : recherche dans le Sudoc
  - + Si résultat Sudoc : recherche alignement BnF

# Taux de recoupement

FRBNF : 98% des notices retrouvées

ISBN : > 90% des notices retrouvées  
(avec recours au Sudoc pour les réimpressions)

Titre - Auteur - Date : 60% des notices retrouvées

# A faire

Évaluer les résultats Titre-Auteur-Date  
0, 1, n résultats

Définir la méthodologie pour les CD/DVD

Et les autres types de ressources

# Quel outil de traitement automatique ?

## Expérimentations avec Open Refine

OPEN Refine Ensemble1\_MONISBN Permalink

Open... Export

Facet / Filter Undo / Redo 24

296317 rows Extensions: VIB-Bit

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 10 next > »

All	Id enreg.	ISBN	EAN nettoyé	ISSN	ISBN nettoyé	URL requête SRU	ISBN/EAN ?	Pb ISBN ?	EAN	préfixe EAN	OCLC	Orig
1.	1000	2-204-03065-1	9782204030656		2204030651	http://catalogueservice.bnf.fr/SRU?version=1.2&operation=searchRetrieve&query=ISBN%20all%20%222204030651%22&recordSchema=arkID	2		9782204030656	978	MO500310080701	BN
2.	1000000	978-2-35522-105-7	9782355221057		9782355221057	http://catalogueservice.bnf.fr/SRU?version=1.2&operation=searchRetrieve&query=ISBN%20all%20%229782355221057%22&recordSchema=arkID	9		9782355221057	978	ELC3178789	Electre
3.	1000001	978-2-35417-519-1	9782354175191		9782354175191	http://catalogueservice.bnf.fr/SRU?version=1.2&operation=searchRetrieve&query=ISBN%20all%20%229782354175191%22&recordSchema=arkID	9		9782354175191	978	ELC3292173	Electre
4.	1000002	978-2-267-02971-0	9782267029710		9782267029710	http://catalogueservice.bnf.fr/SRU?version=1.2&operation=searchRetrieve&query=ISBN%20all%20%229782267029710%22&recordSchema=arkID	9		9782267029710	978	ELC3276457	Electre
5.	1000003	978-2-7324-7731-2	9782732477312		9782732477312	http://catalogueservice.bnf.fr/SRU?version=1.2&operation=searchRetrieve&query=ISBN%20all%20%229782732477312%22&recordSchema=arkID	9		9782732477312	978	ELC3168802	Electre
6.	1000004	978-2-7588-0261-7	9782758802617		9782758802617	http://catalogueservice.bnf.fr/SRU?version=1.2&operation=searchRetrieve&query=ISBN%20all%20%229782758802617%22&recordSchema=arkID	9		9782758802617	978	ELC3263892	Electre
7.	1000005	978-2-9536657-6-5	9782953665765		9782953665765	http://catalogueservice.bnf.fr/SRU?version=1.2&operation=searchRetrieve&query=ISBN%20all%20%229782953665765%22&recordSchema=arkID	9		9782953665765	978	ELC3351654	Electre
8.	1000006	978-2-36857-062-3	9782368570623		9782368570623	http://catalogueservice.bnf.fr/SRU?version=1.2&operation=searchRetrieve&query=ISBN%20all%20%229782368570623%22&recordSchema=arkID	9		9782368570623	978	ELC3338509	Electre
9.	1000007	978-2-88944-007-8	9782889440078		9782889440078	http://catalogueservice.bnf.fr/SRU?version=1.2&operation=searchRetrieve&query=ISBN%20all%20%229782889440078%22&recordSchema=arkID	9		9782889440078	978	ELC3182858	Electre
10.	1000008	978-2-7619-4322-2	9782761943222		9782761943222	http://catalogueservice.bnf.fr/SRU?version=1.2&operation=searchRetrieve&query=ISBN%20all%20%229782761943222%22&recordSchema=arkID	9		9782761943222	978	BTL0000324178	BTLF

**Using facets and filters**

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

- open source
- multiplateforme
- traitement de données
- nettoyage de données
- exploitation API



# Quel outil de traitement automatique ?

## Expérimentations avec Open Refine

The screenshot shows the Open Refine interface. The 'Edit column' menu is open, with 'Add column based on this column...' selected. Below it, the 'Add column based on column ISBN' dialog box is shown. The 'New column name' field contains 'longueurISBN'. The 'Expression' field contains 'length(value)'. The 'Language' dropdown is set to 'Google Refine Expression Language (GREL)'. The 'Preview' tab is active, showing a table with two rows of data.

row	value	length(value)
1	97826566531226	13
2	9783727745437	13

### S'aligner avec les données de la BnF : méthodologie

A travers l'expérience de la BM de Montpellier

Etienne Cavalé (BnF / MET / SM)  
Sylvain Franceschi (BM Montpellier)

### Argumentaire

La Transition bibliographique, pour la France, consiste à basculer progressivement dans un nouveau modèle de description des collections des bibliothèques : passer du modèle ISBD, où des « notices bibliographiques » décrivent des « éditions » d'un document, au modèle LRM (ex-FRBR) où les informations décrivant un ouvrage ou une ressource se rattachent aux différentes entités intellectuelles concrétisées par le document :

- La **manifestation**, qui porte les informations spécifiques à l'édition de l'ouvrage (date de parution, nombre de pages, éditeur et lieu d'édition)
- L'**expression**, qui décrit la version du texte publié
- L'**œuvre**, création intellectuelle qui se concrétise dans ses diverses expressions (et notamment les traductions de l'œuvre dans différentes langues)

Il y a donc un énorme chantier à conduire, pour extraire des notices bibliographiques actuelles, les différents éléments d'information qui vont se répartir entre ces 3 entités LRM. Ce chantier nécessite :

- d'**exploiter d'importantes bases de connaissances** (pour savoir, autant que possible, quelles sont les différentes éditions successives d'un même ouvrage, et notamment quelle est la date de sa première parution, qui se confond avec la date de création de l'œuvre)
- de disposer d'**algorithmes de rapprochement** permettant de reconnaître que *Le tiers livre des faits et dits du bon Pantagruel* est bien la même œuvre que *Le tiers livre des faits et dicitz heroïques du noble Pantagruel* et que *Le tiers livre*, tout simplement

C'est pourquoi les deux agences nationales, l'Abes et la BnF, se sont lancées dans des chantiers de FRBRisation incluant de lourds processus automatiques (et des équipes pour les exécuter et les contrôler). Les autres bibliothèques ont plus intérêt à bénéficier de leurs travaux qu'à réaliser de la FRBRisation chacune de son côté.

Cela dit, les résultats de travaux de ces agences ne sont pas attendus avant quelques mois ou années.

# Quel outil de traitement automatique ?

Piste à creuser : vers un logiciel spécifique

Programme d'alignement de données bibliographiques avec la BnF

**En entrée :**

Fichier contenant les notices

Séparateur TAB, Encodage UTF-8

Type de documents

Documents imprimés (monographies)  
Format : Num Not | FRBNF | ARK | ISBN | Titre | Auteur | Date

Audiovisuel (CD / DVD)  
Format : Num Not | FRBNF | ARK | EAN | N° commercial | Titre | Auteur | Date

**En sortie :**

1 fichier

Plusieurs fichiers  
(Pb / 0 / 1 / plusieurs ARK trouvés)

ID traitement (optionnel)

OK

Besoin d'aide ?

Annuler

Version 0.3

# Extraction / injection

- Mise à jour du format Unimarc pour la zone 003
- Fourniture d'un simple fichier tabulé (id > ARK)
- Test d'injection en environnement d'intégration : validé
- Quels traitement pour les dérivations courantes ?

# Et pour la suite

Publication et partage de la méthodologie et des outils employés

Atelier le lundi 19 mars :

- venez avec vos données
- venez avec votre ordinateur
- venez avec Open Refine

Repartez avec des alignements

*And we will keep in touch*