

Bibliostratus

STRatégie d'Alignement d'URIs pour la Transition bibliographique

Etienne Cavalié – Atelier Corpus « Données liées et données à lier »
10/07/2018

Plan

- Un contexte
- Le logiciel
- Des cas d'utilisation
- Quel calendrier ?

Contexte

Une demande récurrente
des bibliothèques



Le contexte de la
Transition bibliographique

Contexte

- Outil d'expérimentation : Open Refine
- Extraire, nettoyer, normaliser les données
- Tests d'alignements réussis

Refine OPEN Ensemble1_MONISBN Permalink

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 > »

Using facets and filters

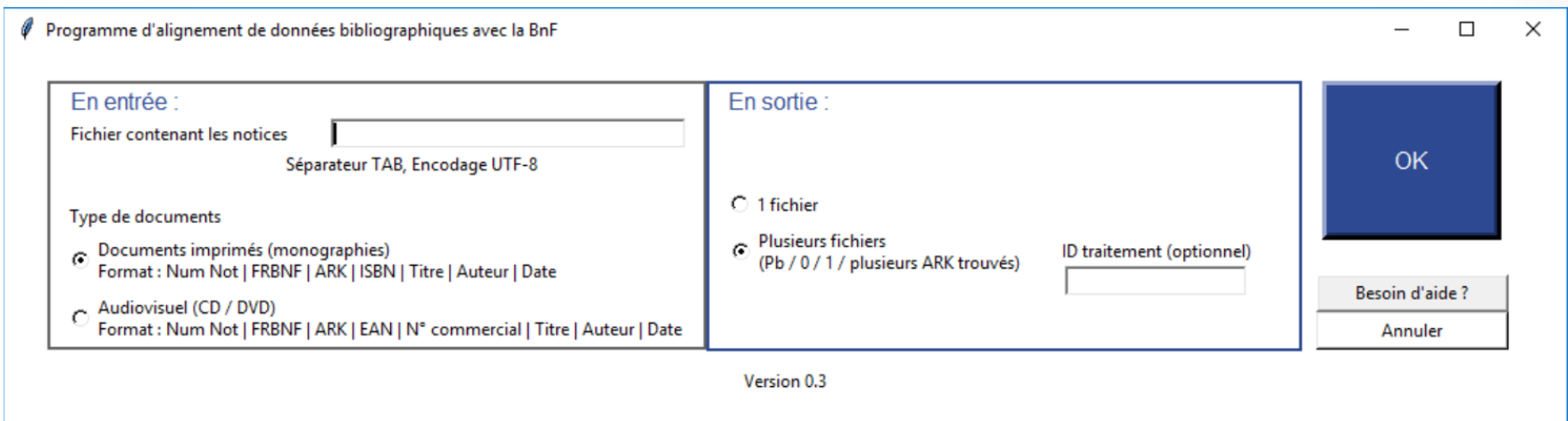
Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

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

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	M0500310090701	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	BT10000324178	BTIF

Contexte

- Décembre 2017 : nouvelle piste – Un logiciel spécifique



Le logiciel : principes

- Écrit en Python (par un non-développeur)
- Open source
- Porté par le groupe Systèmes & Données de la Transition bibliographique
- Applique les règles métier de manière systématique
- Conçu dans un processus de travail (destiné aux bibliothèques qui disposent d'un catalogue au format Unimarc)
- Code source et version précompilée sur Github
- Documentation fonctionnelle et technique sur Github
- Retours d'expériences sur transition-bibliographique.fr

Le logiciel

Bibliostratus : Stratégie d'alignement d'URIs pour la Transition bibliographique

Aligner ses données avec la BnF

Aligner ses données BIB
avec le catalogue BnF
à partir de fichiers tableaux

Aligner ses données AUT

Outils d'accompagnement

Avant alignement

Préparer ses données
(constitution de tableaux
à partir d'un export catalogue)

Convertir un fichier Unimarc
en tableaux

Après alignement

Exporter les données BnF après alignement

Exporter une liste d'ARK BnF
en notices

Documentation en ligne

Forum utilisateurs

Annuler

Bibliostratus - Version 1.2 - 21/06/2018

Le logiciel : module d'alignement

Programme d'alignement de données bibliographiques avec la BnF

En entrée :

Fichier contenant les notices : Séparateur TAB, Encodage UTF-8

Type de documents

- [TEX] Monographies texte [exemple](#)
(Colonnes : Num Not | FRBNF | ARK | ISBN | EAN | Titre | Auteur | Date | Volume-Tome | Editeur)
- [VID] Audiovisuel (DVD)
(Num Not | FRBNF | ARK | EAN | N° commercial | Titre | Auteur | Date | Editeur)
- [AUD] Enregistrements sonores [exemple](#)
(Num Not | FRBNF | ARK | EAN | N° commercial | Titre | Auteur | Date | Editeur)
- [PER] Périodiques [exemple](#)
(Num Not | FRBNF | ARK | ISSN | Titre | Auteur | Date | Lieu de publication)

Aligner de préférence :

- Avec la BnF (et à défaut avec le Sudoc)
- Avec le Sudoc (et à défaut avec la BnF)

En sortie :

Nombre de fichiers

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

Récupérer les métadonnées bibliographiques BnF [Dublin Core]

Préfixe fichiers en sortie

```
C:\Windows\system32\cmd.exe
86574. FRBNF346239610000002 : ark:/12148/cb34623961a
86575. FRBNF346243910000006 : ark:/12148/cb34624391b
86576. FRBNF346243910000003 : ark:/12148/cb34624391c
86577. FRBNF346331510000006 : ark:/12148/cb34633151g
86578. FRBNF346331510000002 : ark:/12148/cb34633151e
86579. FRBNF346332410000005 : ark:/12148/cb34633241f
86580. FRBNF346344410000000 : ark:/12148/cb34634441?
86581. FRBNF346209810000001 : ark:/12148/cb34620981p
86582. FRBNF346219110000002 : ark:/12148/cb34621911b
86583. FRBNF346211940000009 : ark:/12148/cb346211941
86584. FRBNF346214410000003 : ark:/12148/cb34621441f
86585. FRBNF346214610000006 : ark:/12148/cb34621461x
86586. FRBNF346217740000009 : ark:/12148/cb346217721
86587. FRBNF346218510000001 : ark:/12148/cb34621851q
86588. FRBNF346047810000001 : ark:/12148/cb34604781p
86589. FRBNF346052910000005 : ark:/12148/cb34605291o
86590. FRBNF346054910000006 : ark:/12148/cb34605491d
86591. FRBNF346054100000003 : ark:/12148/cb34605411k,ark:/12148/cb345931161
86592. FRBNF345485110000006 : ark:/12148/cb34548511e
86593. FRBNF345489110000008 : ark:/12148/cb34548911p
86594. FRBNF345489510000003 : ark:/12148/cb34548951e
86595. FRBNF345495910000001 : ark:/12148/cb34549591q
86596. FRBNF345500010000008 : ark:/12148/cb34550001s
86597. FRBNF361956110000003 : ark:/12148/cb36195611b
86598. FRBNF361957210000000 : ark:/12148/cb36195721b
86599. FRBNF361957310000007 : ark:/12148/cb36195731n
86600. FRBNF361957510000003 : ark:/12148/cb36195751b
86601. FRBNF361962410000009 : ark:/12148/cb36196241v
86602. FRBNF361962510000007 : ark:/12148/cb361962514
86603. FRBNF361962910000002 : ark:/12148/cb36196291c
86604. FRBNF361963210000003 : ark:/12148/cb36196321g
86605. FRBNF361963740000005 : ark:/12148/cb361963741
86606. FRBNF361965710000006 : ark:/12148/cb36196571r
86607. FRBNF361973010000005 : ark:/12148/cb36197301z
86608. FRBNF361974910000008 : ark:/12148/cb361974915
86609. FRBNF361977010000007 : ark:/12148/cb361977016
86610. FRBNF36198210000006 : ark:/12148/cb361982114
86611. FRBNF361982740000009 : ark:/12148/cb36198271?
86612. FRBNF361982810000005 : ark:/12148/cb36198281j
86613. FRBNF361985110000001 : ark:/12148/cb36198511e
86614. FRBNF361986010000006 : ark:/12148/cb361986015
86615. FRBNF361986740000000 : ark:/12148/cb36198674b
86616. FRBNF361986810000007 : ark:/12148/cb36198681n
86617. FRBNF361988110000003 : ark:/12148/cb361988111
86618. FRBNF361988410000002 : ark:/12148/cb36198841z
86619. FRBNF36597510000007 : ark:/12148/cb36597511d
86620. FRBNF361991110000008 : ark:/12148/cb36199111b
86621. FRBNF361992010000009 : ark:/12148/cb361992019
86622. FRBNF361995510000007 : ark:/12148/cb36199551p
86623. FRBNF361998110000007 : ark:/12148/cb36199811p
86624. FRBNF362000740000001 : ark:/12148/cb362000712
86625. FRBNF362004610000007 : ark:/12148/cb36200461o
86626. FRBNF362005010000004 : ark:/12148/cb36200501b
86627. FRBNF362006110000006 : ark:/12148/cb36200611v
86628. FRBNF362007610000003 : ark:/12148/cb36200761p
```

BiblioStratus - Version 1.2 -

Le logiciel : résultats

1	NumNot	nbARK	ark trouvé	Méthode	ark initial	FRBNF	ISBN	EAN	Titre	auteur	date	Tome/Volume	editeur
2	FRBNF43476742000001	2	ark:/12148/cb43476742b,ark:/12148/cb434768580	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			978-2-7532-0274-0	9782753202740	A la ferme		2010		Elcy
3	FRBNF434872150000005	2	ark:/12148/cb43487215h,ark:/12148/cb434330912	ISBN + contrôle Titre 2005a			2-841-29811-6		La Planète Si christine ma		2001		Edicef Nouvelles Ed
4	FRBNF43487199000000X	2	ark:/12148/cb434871993,ark:/12148/cb43441575x	ISBN + contrôle Titre 2005a			2-85049-682-0		Mémoires d salif Iherme		1996 T. 2		Les classiques africa
5	FRBNF434618110000004	2	ark:/12148/cb43461811x,ark:/12148/cb43462833g	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-01-224950-7	9782012249509	La galerie des bandits		2005		Hachette Jeunesse
6	FRBNF434009240000001	2	ark:/12148/cb434866966,ark:/12148/cb434009242	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			978-2-7507-0089-8	9782750701055	Kebebi		2004	9	Tokebi
7	FRBNF434866960000000	2	ark:/12148/cb434866966,ark:/12148/cb434009242	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-7507-0105-8	9782750701055	Kebebi		2004	10	Tokebi
8	FRBNF434872030000004	4	ark:/12148/cb43487203h,ark:/12148/cb43445835g,ark:/12148/cb43445835g,ark:/12148/cb43445835g	ISBN + contrôle Titre 2005a			2-850-69448-7		Halimatou kanta sylvie		1987		Les Nouvelles Editi
9	FRBNF434768580000007	2	ark:/12148/cb43476742b,ark:/12148/cb434768580	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			978-2-7532-0274-0	9782753202740	A la ferme		2010		Elcy
10	FRBNF434900270000007	2	ark:/12148/cb43490027g,ark:/12148/cb442904230	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre BNF contenu dans titre initial,i	978-2-8096-4993-2		9782809649932	So fashion so chic bijoux			2013		Play Bac
11	FRBNF43476857000000X	2	ark:/12148/cb434767410,ark:/12148/cb43476857n	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			978-2-7532-0268-9	9782753202689	A la ferme		2010		Elcy
12	FRBNF434722670000005	2	ark:/12148/cb43472267b,ark:/12148/cb43469717n	EAN + contrôle Titre BNF contenu dans titre initial,EAN > ARK,EAN + contrôle Titre BNF co	978-2-8094-0306-0;2-		9782809403060	City Hunter X illustrations			2008	1	Panini manga
13	FRBNF434705770000009	4	ark:/12148/cb43470579p,ark:/12148/cb43470580w,ark:/12148/cb43470580w,ark:/12148/cb43470580w	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK,EAN + contr	978-2-01-391393-5;2-		9782013913935	Je colorie			2008	1	Deux coqs d'or
14	FRBNF434624670000002	2	ark:/12148/cb43462467p,ark:/12148/cb43414779z	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-89021-284-X	9782890212848	Les fantaisie raymonde munch michael robert longval m				COURTE ECHELLE
15	FRBNF434642270000005	2	ark:/12148/cb43464227c,ark:/12148/cb41183723r	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-35100-142-7	9782351001424	Au nom du p christopher		2006	5	Carabas
16	FRBNF434645770000003	2	ark:/12148/cb43464577r,ark:/12148/cb43463519b	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-88890-074-2	9782888900740	Les fugitifs d roke aja		2006	2	Paquet
17	FRBNF434869820000001	2	ark:/12148/cb43486982b,ark:/12148/cb434623949	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-7338-0791-9	9782733807910	Milly Molly gill fabienne		2005	2005	Auzou
18	FRBNF434951020000006	2	ark:/12148/cb43495102m,ark:/12148/cb442595913	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2255a,EAN > ARK			978-2-01-204546-0	9782012045460	Hello Kitty		2014	2	Hachette Jeunesse
19	FRBNF434640620000003	2	ark:/12148/cb43464062b,ark:/12148/cb41075880z	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-84946-628-X	9782849466285	Gothic sport:hage anike		2006	2	Soleil
20	FRBNF43465812000000X	2	ark:/12148/cb434658127,ark:/12148/cb40997817h	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-915492-12-3	9782915492125	La vénéneus stephane bli		2007		Cornélius
21	FRBNF434756020000006	2	ark:/12148/cb43475602k,ark:/12148/cb21116977	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,[demi-titre-8caractères]	978-2-302-00925-7		9782302009257	Couleur de f jing			2009		Quadrants
22	FRBNF434764820000000	2	ark:/12148/cb43476482r,ark:/12148/cb42191350d	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			978-2-211-20153-7	9782211201537	Anticondes d gregoire solc		2010		Ecole des loisirs
23	FRBNF434893950000000	2	ark:/12148/cb43489395r,ark:/12148/cb44290452k	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			978-2-8096-4996-3	9782809649963	Dance		2013		Play Bac
24	FRBNF434623950000001	2	ark:/12148/cb43462395n,ark:/12148/cb434869810	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-7338-0789-7	9782733807897	Milly Molly gill fabienne		2005	2005	Auzou
25	FRBNF434853250000008	2	ark:/12148/cb43485325m,ark:/12148/cb43847965v	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre BNF contenu dans titre initial,i	978-2-01-394219-5		9782013942195	Les contes de fées			2013		Deux coqs d'or
26	FRBNF434853350000004	2	ark:/12148/cb43485335x,ark:/12148/cb43847976h	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre BNF contenu dans titre initial,i	978-2-01-394220-1		9782013942201	Les dinosaur frederic kuci			2013		Deux coqs d'or
27	FRBNF434870350000007	2	ark:/12148/cb43487035k,ark:/12148/cb43391386w	ISBN + contrôle Titre 2005a			2-226-17013-8		Peau de pécl jodi rchetin		2006		Albin Michel-Jeune
28	FRBNF43494259000000X	2	ark:/12148/cb43494259m,ark:/12148/cb45257062z	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			978-2-87833-698-6	9782878336986	Le monde de samoyault ti		2014		Circonflexe
29	FRBNF434632900000003	2	ark:/12148/cb43463290r,ark:/12148/cb40153918z	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre BNF contenu dans titre initial,i	5371126996		3505371126996	Alexandra Le emme			2005		Emme
30	FRBNF434623890000004	2	ark:/12148/cb43462389q,ark:/12148/cb43486980n	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-7338-0792-7	9782733807927	Milly Molly gill fabienne		2005	2005	Auzou
31	FRBNF434628090000000	2	ark:/12148/cb43462809m,ark:/12148/cb40960266r	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2255a,EAN > ARK			2-7316-6283-2	9782731662832	Songes terry dodsor		2005	1	Humanoïdes associ
32	FRBNF434705790000003	4	ark:/12148/cb43470579p,ark:/12148/cb43470580w,ark:/12148/cb43470580w,ark:/12148/cb43470580w	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK,EAN + contr	978-2-01-391393-1		9782013913935	Je colorie			2008	3	Deux coqs d'or
33	FRBNF434900290000001	2	ark:/12148/cb43490029s,ark:/12148/cb44290416q	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre BNF contenu dans titre initial,i	978-2-8096-4992-5		9782809649925	So fashion so chic paillet			2013		Play Bac
34	FRBNF434868700000005	2	ark:/12148/cb43486870z,ark:/12148/cb43486964d	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-7373-3618-X	9782737336188	Je colorie la ehrhard dorr		2005		Ouest-France
35	FRBNF434869800000007	2	ark:/12148/cb43486980q,ark:/12148/cb43486980n	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			2-7338-0792-7	9782733807927	Milly Molly gill fabienne		2005	2005	Auzou
36	FRBNF434703400000008	2	ark:/12148/cb43470340s,ark:/12148/cb41206309b	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK			978-2-7234-6249-5;2-	9782723462495	Cendrillon jobert marle		2008		Atlas
37	FRBNF434705800000004	4	ark:/12148/cb43470579p,ark:/12148/cb43470580w,ark:/12148/cb43470580w,ark:/12148/cb43470580w	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre 2005a,EAN > ARK,EAN + contr	978-2-01-391393-1		9782013913935	Je colorie			2008	4	Deux coqs d'or
38	FRBNF434751100000000	2	ark:/12148/cb43475110t,ark:/12148/cb42095257d	EAN + contrôle Titre 2005a,EAN > ARK,EAN + contrôle Titre BNF contenu dans titre initial,i	978-2-508-00411-7		9782508004117	Disney princ developper			2009		Hemma

Cas d'utilisation

- Bibliothèque de lecture publique
 - Pour récupérer les notices BnF sur leur fond « ancien »
 - Pour intégrer un fonds particulier (dons volumineux, etc.)
 - Pour récupérer les notices d'autorité BnF
- Bibliothèques qui entrent dans le réseau Sudoc
- Baromètre de la lecture publique :
à partir d'une liste d'ISBN, récupération des notices BnF pour avoir des métadonnées homogènes
- Bibliothèque nationale d'Espagne : récupérer une liste d'ISSN
- BnF : opérations de dédoublonnage
- Chercheurs... ?

Calendrier à venir

- Projet encore tout récent
- Ateliers déjà tenus ou programmés pour 2018 à destination des bibliothèques
- Encore beaucoup de marge de progression :
 - Réécriture du code selon les règles de rédaction Python (PEP8)
 - Modules d'alignement et d'extraction : équilibre BnF / Sudoc
 - Module rouge : Extraction d'une sélection de métadonnées
 - Vers un web service / une version cloud ?
- Envisager d'autres cas d'usages
Intégrer et faciliter d'autres workflows