

Quand **data.bnf.fr**  
calcule des œuvres

*automatiquement*

Etienne Cavalié

BnF – Département des métadonnées

Chef du service Ingénierie des métadonnées

**Journée sur la diffusion des données de la BnF**

28 mai 2021

Quand **data.bnf.fr**  
calcule des œuvres

*automatiquement*

Etienne Cavalié

BnF – Département des métadonnées

Chef du service Ingénierie des métadonnées

**Journée sur la diffusion des données de la BnF**

28 mai 2021

# Passé, présent, avenir

- **Rappel des épisodes précédents**
  - **10 ans (2011-2021) :**  
[data.bnf.fr](http://data.bnf.fr) et le modèle FRBR
  - **5 ans (2017-2021) :**  
RobotDonnées et les oeuvres calculées
- **Aujourd'hui et demain**
  - Processus de travail
  - Résultats dans [data.bnf.fr](http://data.bnf.fr)
  - Et dans le catalogue

# 10 ans avec data.bnf.fr

## • 2011 :

- Implémentation du modèle FRBR à partir des données du catalogue
- 14 millions de notices bibliographiques
- 300.000 notices d'autorité Titre

8	1963
<b>Christie, Agatha (1890-1976)</b> Ten little niggers (français) Dix petits nègres / Agatha Christie ; traduit de l'anglais par Louis Postif Livres de poche Livres	
9	1974
<b>Christie, Agatha (1890-1976)</b> Ten little niggers (français) Dix petits nègres... / Agatha Christie ; traduit de l'anglais par Louis Postif J. Tallandier Livres	
10	1975
<b>Christie, Agatha (1890-1976)</b> Ten little niggers (français) Dix petits nègres / Agatha Christie ; [traduit par Louis Postif] Éditions de Saint-Clair Livres	
11	1976
<b>Christie, Agatha (1890-1976)</b> Ten little niggers (français) Dix petits nègres / Agatha Christie ; traduit de l'anglais par Louis Postif Gallimard Livres	

< 10% notices BIB  
contiennent un lien  
à une oeuvre

6% pour le texte

Notice de titre conventionnel

Notice Au format public

[Christie, Agatha \(1890-1976\)](#)  
**Ten little niggers** *forme internationale anglais*

Langue(s) : anglais  
Date de l'oeuvre : 1939  
Première édition : 1939  
Forme de l'oeuvre : Roman policier

Forme(s) rejetée(s) :  
And then there were none *anglais*  
Dix petits nègres *français*  
Ils étaient dix *français*

Forme(s) associée(s) :  
>> << A pour adaptation : [Ten little Indians \(film\)](#)  
>> << A pour adaptation : [And then there were none \(film\)](#)

Source(s) :

# 10 ans avec **data.bnf.fr**

Exposition des données sur le web

## • 2011 :

- Implémentation du modèle FRBR à partir des données du catalogue
- 14 millions de notices bibliographiques
- 300.000 notices d'autorité Titre

8	1963
<b>Christie, Agatha (1890-1976)</b> Ten little niggers (français) Dix petits nègres / Agatha Christie ; traduit de l'anglais par Louis Postif Livres de poche Livres	
9	1974
<b>Christie, Agatha (1890-1976)</b> Ten little niggers (français) Dix petits nègres... / Agatha Christie ; traduit de l'anglais par Louis Postif J. Tallandier Livres	
10	1975
<b>Christie, Agatha (1890-1976)</b> Ten little niggers (français) Dix petits nègres / Agatha Christie ; [traduit par Louis Postif] Éditions de Saint-Clair Livres	
11	1976
<b>Christie, Agatha (1890-1976)</b> Ten little niggers (français) Dix petits nègres / Agatha Christie ; traduit de l'anglais par Louis Postif Gallimard Livres	

< 10% notices BIB  
contiennent un lien  
à une oeuvre

6% pour le texte

Notice de titre conventionnel

Notice Au format public

[Christie, Agatha \(1890-1976\)](#)  
**Ten little niggers** *forme internationale anglais*

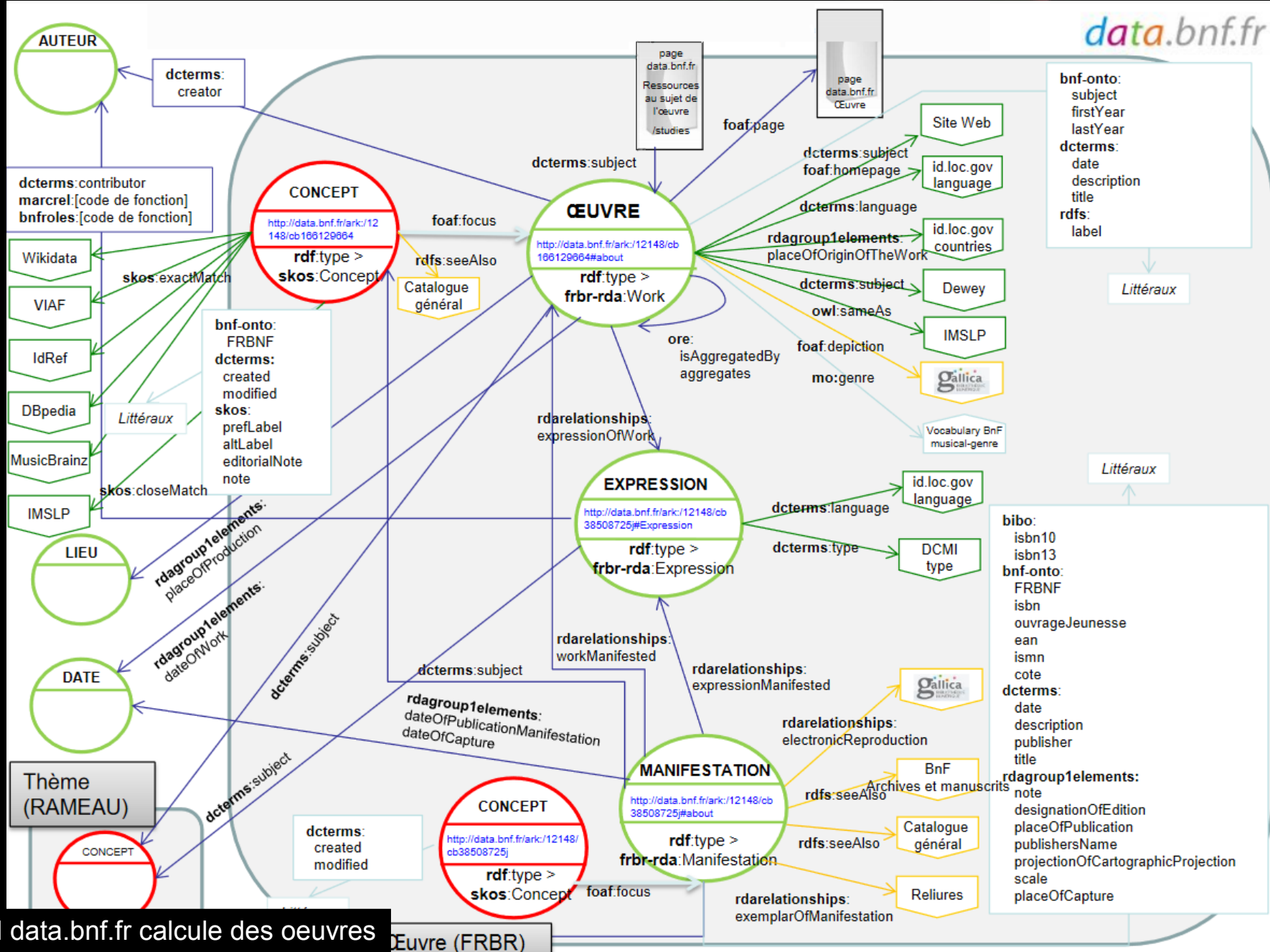
Langue(s) : anglais  
Date de l'oeuvre : 1939  
Première édition : 1939  
Forme de l'oeuvre : Roman policier

Forme(s) rejetée(s) :  
And then there were none *anglais*  
Dix petits nègres *français*  
Ils étaient dix *français*

Forme(s) associée(s) :  
>> << A pour adaptation : [Ten little Indians \(film\)](#)  
>> << A pour adaptation : [And then there were none \(film\)](#)

Source(s) :

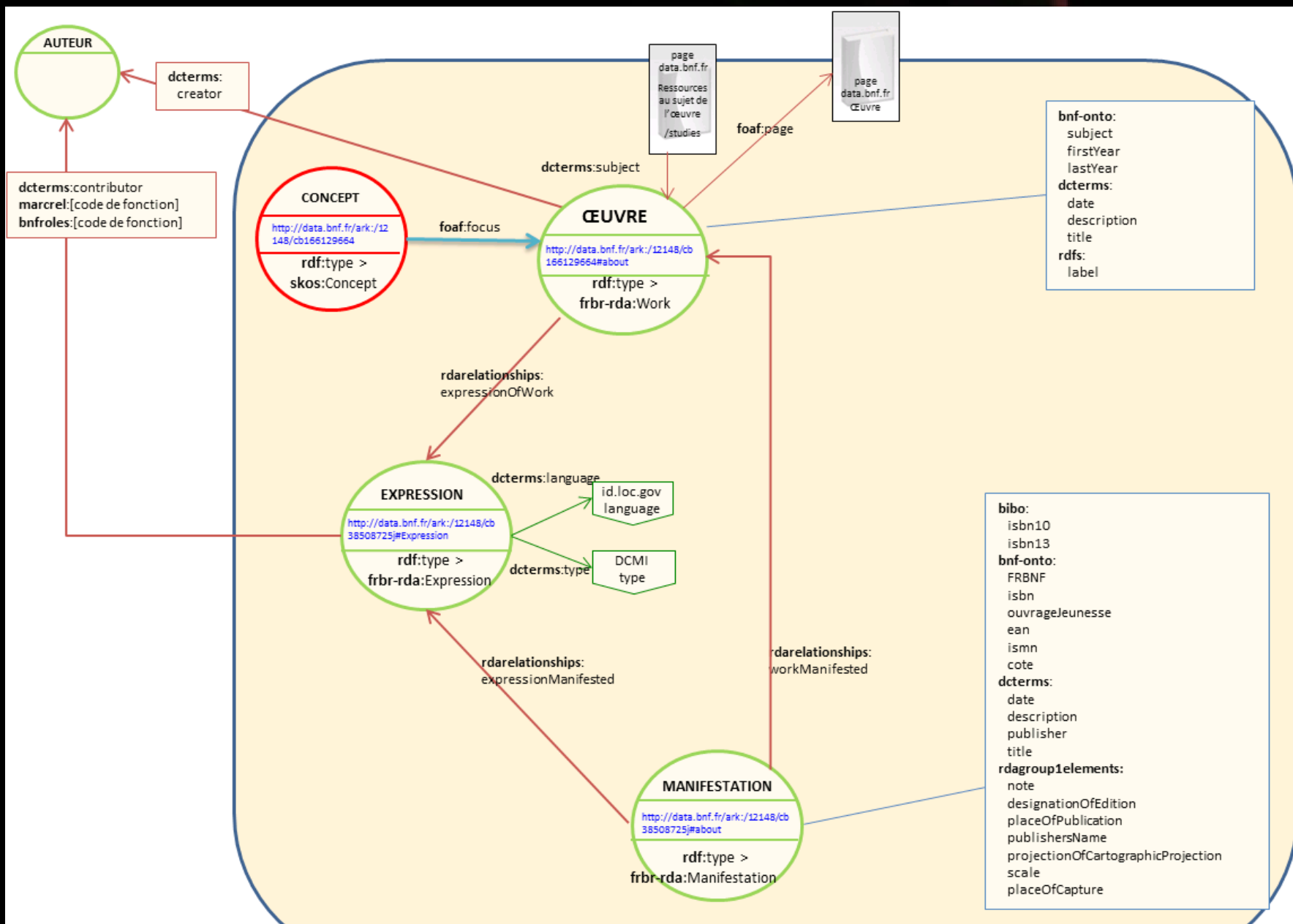
# Data.bnf.fr : tisser des liens



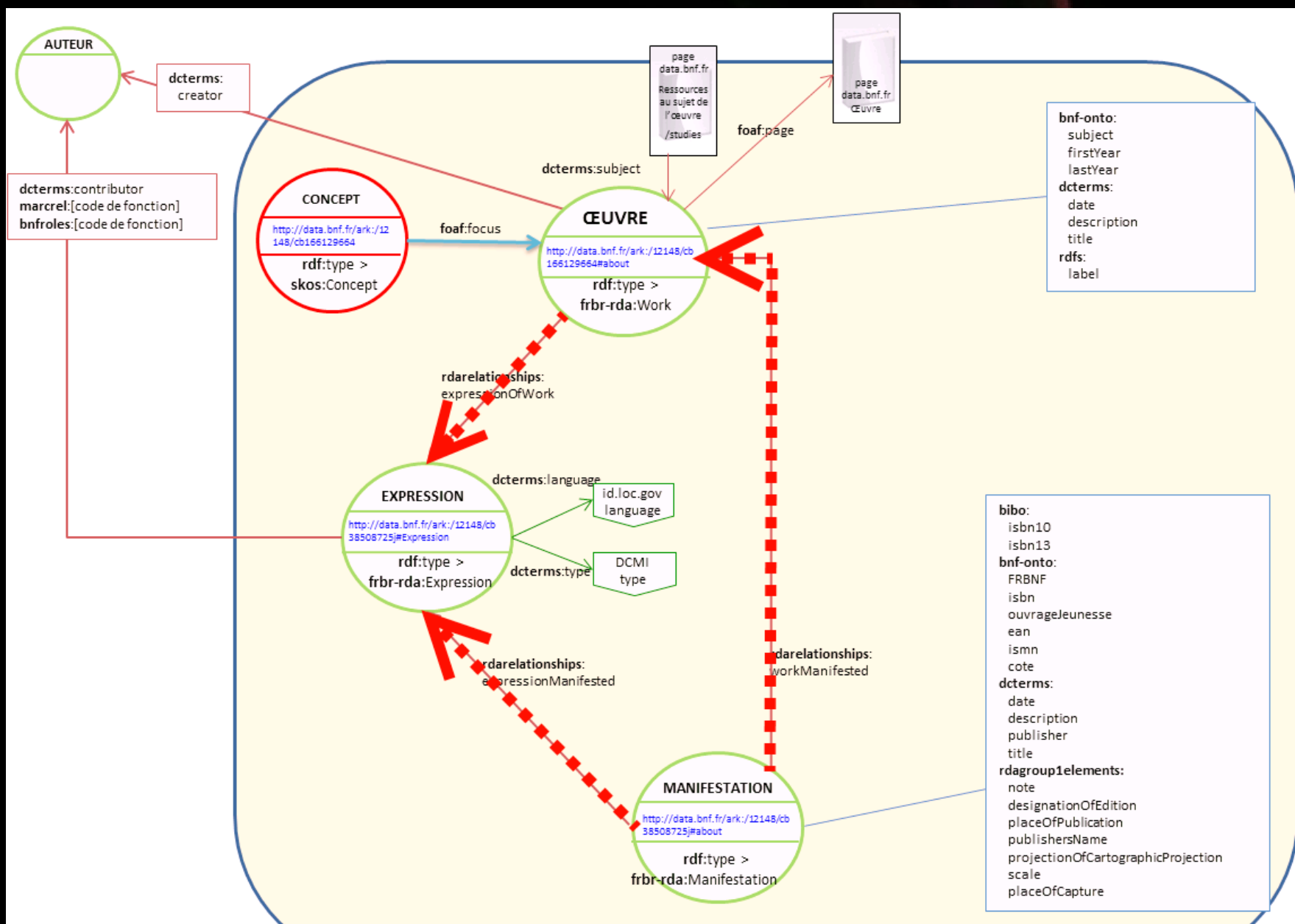
E. Cavalié – Quand data.bnf.fr calcule des oeuvres

Œuvre (FRBR)

# Data.bnf.fr : tisser des liens



# Data.bnf.fr : tisser des liens





# PRR (2016) :

## 165.000 liens dans le catalogue

- Algorithme rejoué tous les 1 ou 2 ans + 5000 à 10.000 liens par an

**Notice de titre conventionnel**

Notice Au format public

[Christie, Agatha \(1890-1976\)](#)  
*Ten little niggers* forme internationale anglais

Langue(s) : anglais  
Date de l'oeuvre : 1939  
Première édition : 1939  
Forme de l'oeuvre : Roman policier

Forme(s) rejetée(s) :  
*And then there were none* anglais  
*Dix petits nègres* français  
*Ils étaient dix* français

Forme(s) associée(s) :  
>> << A pour adaptation : [Ten little Indians \(film\)](#)  
>> << A pour adaptation : [And then there were none \(film\)](#)

Source(s) :

Outils

Citer la notice : <https://catalogue.bnf.fr/ark:/12148/cl>

Télécharger/Imprimer

Envoyer par courriel

Ajouter à mes références

Signaler une erreur sur cette notice

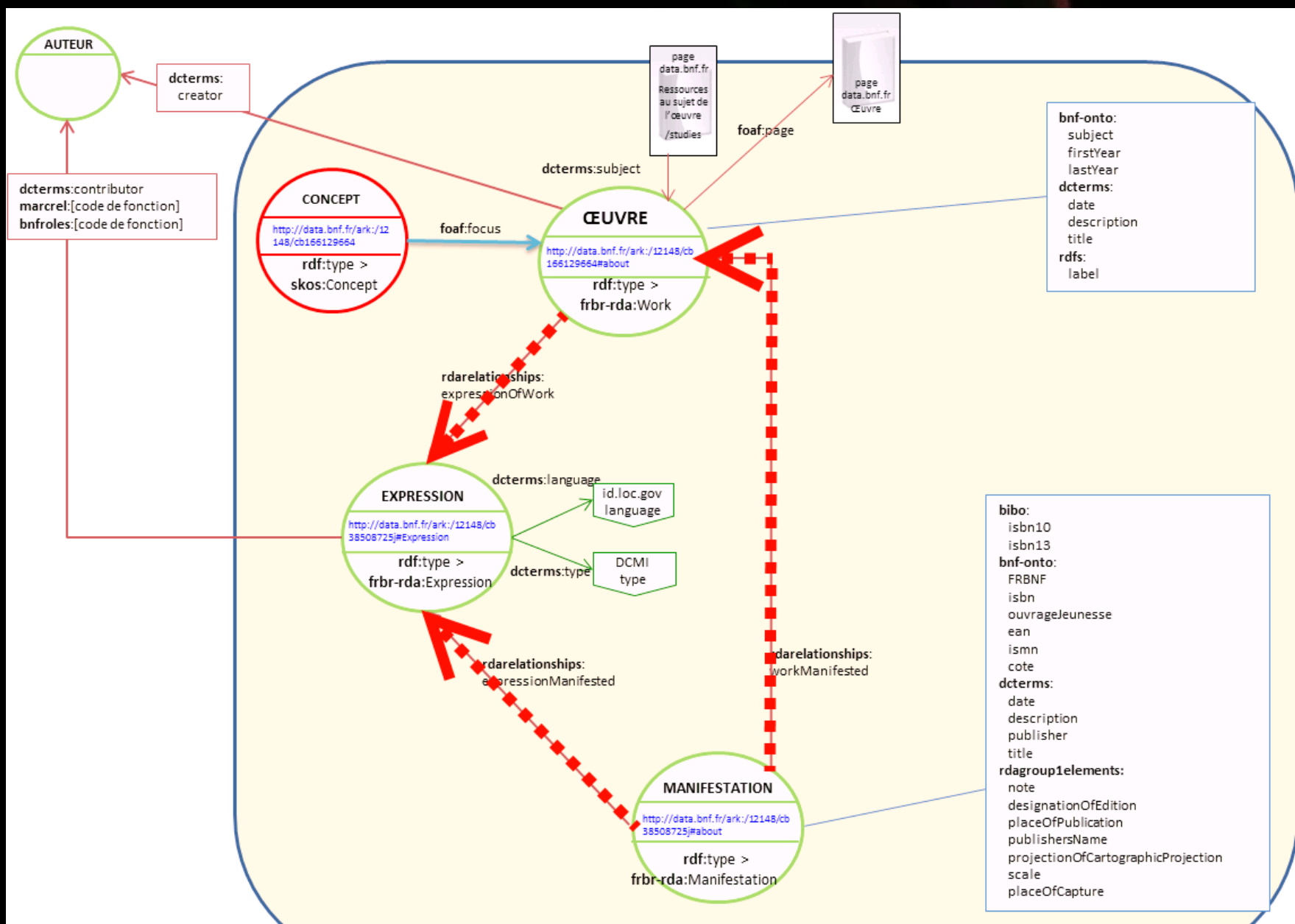
**NOTICES BIBLIOGRAPHIQUES LIÉES**

Voir les notices liées en tant que :

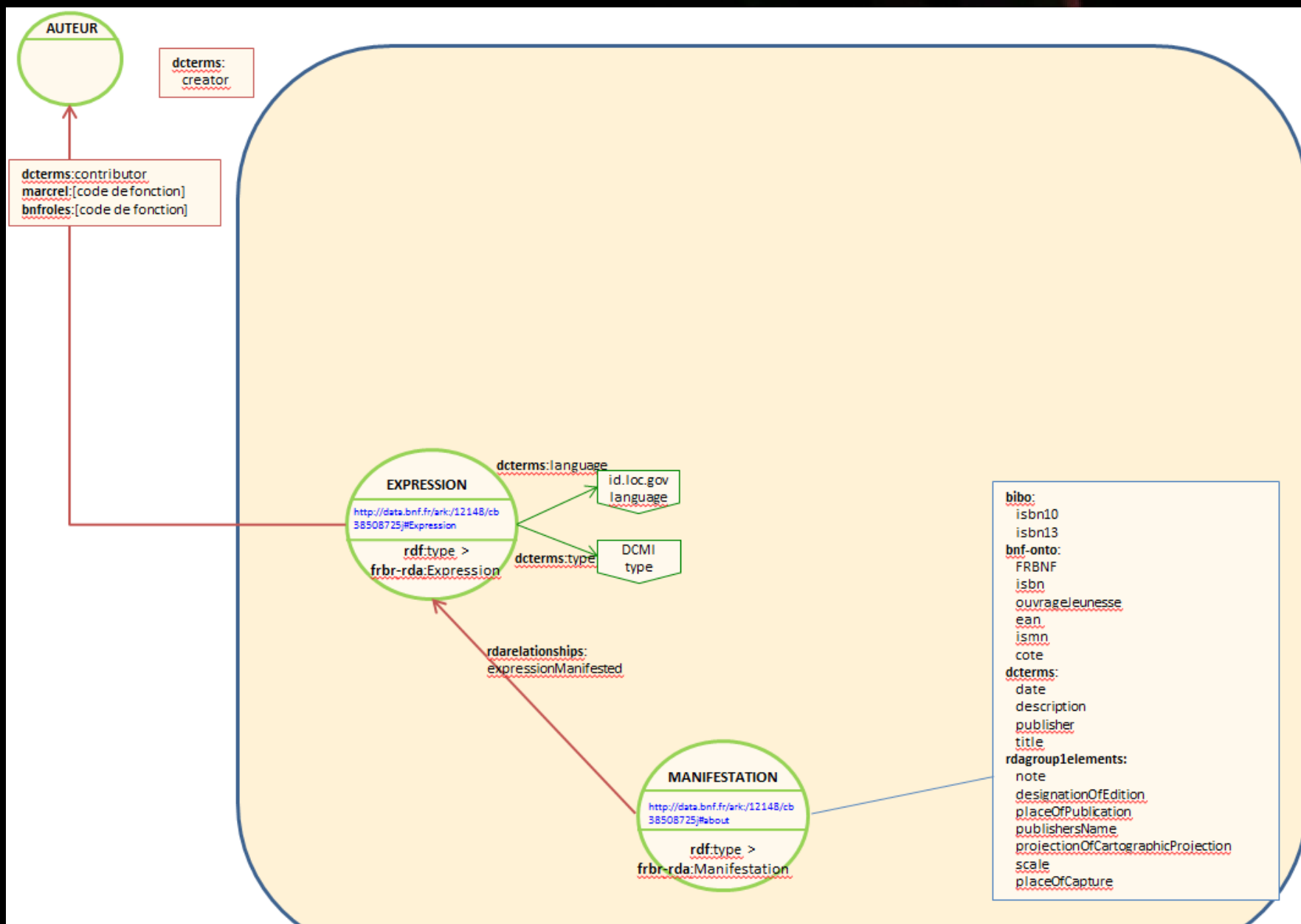
- sujet (1)

Voir toutes les notices liées (1)

# Data.bnf.fr : tisser des liens




# Data.bnf.fr : tisser des liens



# RobotDonnées



- Plate-forme
- Règles algorithmiques
- Processus de travail

 RobotDonnées

---

### WsCatalog

**Authors**

**Notice types**

**Doc types**

**Title zones**

zones de titre, e.g 245\$a+245\$i; 245\$a; 245\$a+\$r; 290\$a; 142\$a; 247\$a; 750\$a.

Il est également possible d'utiliser des conditions, en utilisant les notations suivantes :

- '245\$e[\$e%histoire]' pour récupérer les 245\$e si et seulement si elles contiennent 'histoire'.
- '750\$a[ind2=3]' pour récupérer les 750\$a si et seulement, l'indicateur 2 est égale à 3
- '245\$e[\$e~.\*(dictionnaire|actes|note).\*]' pour récupérer les 245\$e si et seulement si elles contiennent 'dictionnaire'.

**Kinds**

liste de genres séparés par des ";". Si un champ 245\$e contient un de ces genres, il ne sera pas récupéré.

add some files

OR

# RobotDonnées : une plate-forme



- Travail par corpus
- semi-manuel ou semi-automatique
- Méthodologie récursive

Run 5700773 of minhashing.py

wscatalog.py #5700748 > minhashing.py #5700773

Parameters Log **Result** Statistics

+ create dedupe run from these clusters + create merge run from these clusters + create works builder run from these clusters

▼ Clusters (raw csv) :

Trier par : cluster id nombre de manifs nombre de titres distincts ▼ titre de l'œuvre

Group 11024543000005 (1 manifestations - 2 formes de titre) - *possibilites actuelles me*

Group 16693692000010 (1 manifestations - 2 formes de titre) - *tunisien ne s lorraine*

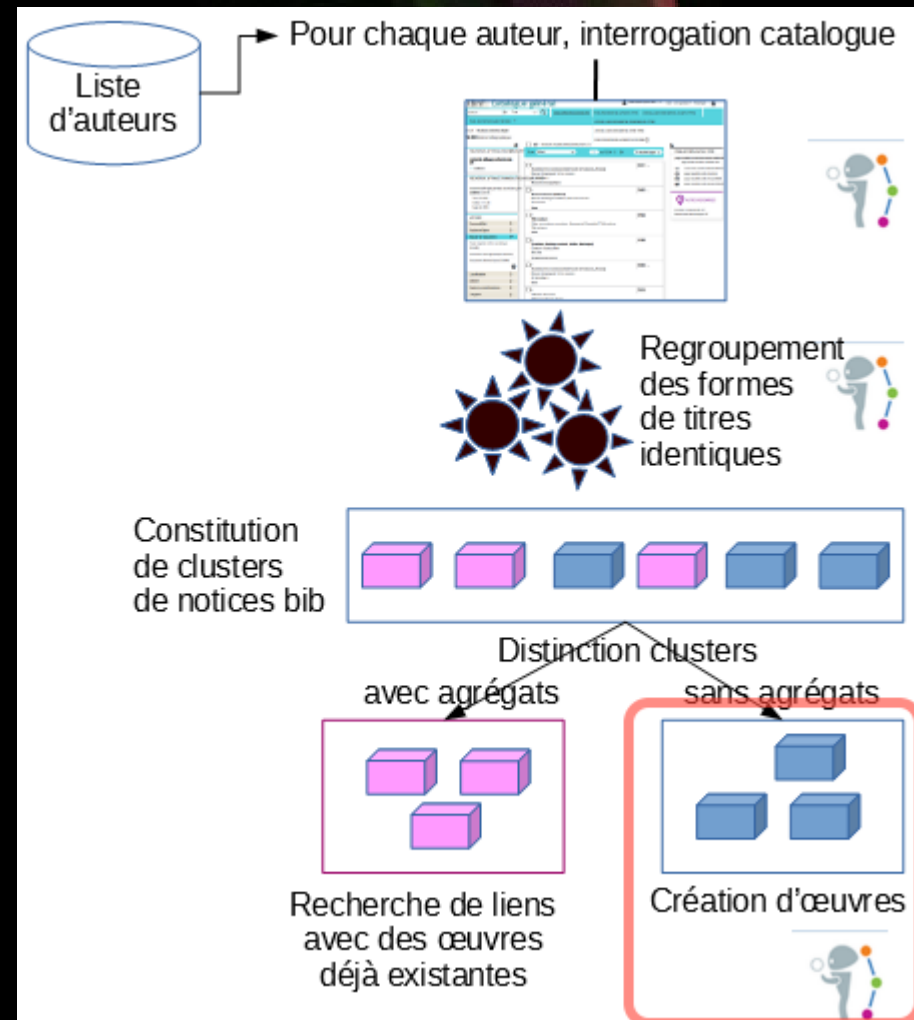
Group 10632235000038 (2 manifestations - 2 formes de titre) - *eclaircissemens litterair*  
*alphabetique sur l histoire litteraire cave sur quelques autres ouvrages semblables pour bie*

delete	Titre Manifestation	ARK de la manifestation	Cluster Id
<input checked="" type="checkbox"/>	eclaircissemens litteraires sur projet bibliotheque alphabetique sur l histoire litteraire cave sur quelques autres ouvrages semblables pour bien ecrire	<a href="http://ark:/12148/cb30895441h">ark:/12148/cb30895441h</a>	10632235
<input checked="" type="checkbox"/>	eclaircissemens litteraires sur projet bibliotheque alphabetique sur l histoire litteraire cave sur quelques autres ouvrages semblables pour bien ecrire	<a href="http://ark:/12148/cb33358726w">ark:/12148/cb33358726w</a>	10632235
<input checked="" type="checkbox"/>	eclaircissemens litteraires sur projet bibliotheque alphabetique sur l histoire litteraire cave sur quelques autres ouvrages semblables pour bien ecrire ouvrage periodique	<a href="http://ark:/12148/cb33358726w">ark:/12148/cb33358726w</a>	10632235
<input checked="" type="checkbox"/>	eclaircissemens litteraires sur projet bibliotheque alphabetique sur l histoire litteraire cave sur quelques autres ouvrages semblables pour bien ecrire	<a href="http://ark:/12148/cb30895441h">ark:/12148/cb30895441h</a>	10632235
<input checked="" type="checkbox"/>	eclaircissemens litteraires sur projet bibliotheque alphabetique sur l histoire litteraire cave sur quelques autres ouvrages semblables pour bien ecrire	<a href="http://ark:/12148/cb33358726w">ark:/12148/cb33358726w</a>	10632235

# RobotDonnées : des algorithmes



- Conditions d'extractions
- Conditions de regroupement
- Métadonnées générées :
  - Titre, Auteur(s), Date, Langue
  - Autres titres
  - Manifestations liées



# RobotDonnées : des algorithmes



- Règles pour les métadonnées des oeuvres

- **Les titres**

- Titre original s'il en existe un dans le cluster
- Sinon, titre de la notice bibliographique la plus ancienne
- Récupération de toutes les formes de titre disponibles avec leurs codes de langue

- Les auteurs

- Les identifiants d'oeuvres

## Feuilles de la vie et du rêve (Run #5394075) xml

Langue	fre			
Date	1911			
Notices bib.	<a href="ark:/12148/cb31779915v">ark:/12148/cb31779915v</a>			
Formes de titre				
	<b>Titre</b>	<b>Langue</b>	<b>Zone</b>	<b>Notice d'origine</b>
	Feuilles de la vie et du rêve	fre	245\$a	<a href="ark:/12148/cb31779915v">ark:/12148/cb31779915v</a>
Auteurs				
	<b>Nom</b>	<b>Ark</b>	<b>Rôle</b>	<b>Zone</b>
	Albert Bazouin	10990692	70	100
Id. robotdonnées	5394107			
Id. pérenne	5e38779b15556b9281240771f804ca7a			
Date de création	2021/03/22			

## Structure de données brute

```
{
  "language": [
    "fre"
  ],
  "preferred_title": {
    "title": "Feuilles de la vie et du rêve",
    "lang": "fre",
    "zone": "245$a",
    "ark": "ark:/12148/cb31779915v",
    "pipeindex": -1
  },
  "alternative_titles": [],
  "authors": [
    [
      "Albert Bazouin",
      70,
      100,

```

# RobotDonnées : des algorithmes



- Règles
  - Les titres
  - **Les auteurs**
    - Codes fonctions pertinents
    - Essentiellement :
      - Auteur
      - Adaptateur
      - Editeur scientifique
      - Directeur de publication
  - Les identifiants d'oeuvres

Run 5700996 of *worksbuilder.py*

[worksbuilder.py #5700996](#)

Parameters

Edit

## Clusters

keep file from copied run

## Kinds

liste de genres séparés par des ";" . Si un champ 245\$e con

keep file from copied run ( [↓](#) )

## Skip keywords

liste de motifs-clés (e.g. ", roman; , comédie") identifiant les

keep file from copied run ( [↓](#) )

## Author roles

70;10;11;13;330;360

Liste de rôles séparés par des ";" correspondant à une fonc

## Oldest publication year

1870

date de publication en dessous de laquelle la date d'une œ



# RobotDonnées : des algorithmes

- Règles
  - Les titres
  - Les auteurs
  - **Les identifiants d'oeuvres**
    - Un *hash* concaténation des ARK biblio + titre → hachage<sup>1</sup>
    - Une URI data.bnf.fr
    - ~~Un numéro de notice~~
    - ~~Un ARK~~



```
{
  "language": [
    "fre"
  ],
  "preferred_title": {
    "title": "Feuilles de la vie et du rêve",
    "lang": "fre",
    "zone": "245$a",
    "ark": "ark:/12148/cb31779915v",
    "pipeindex": -1
  },
  "alternative_titles": [],
  "authors": [
    [
      "Albert Bazouin",
      70,
      100,
      10990692
    ]
  ],
  "manifs": [
    "ark:/12148/cb31779915v"
  ],
  "creation_date": "2021/03/22",
  "notice_type": "TIC",
  "custom_fields": {},
  "043": [
    [
      "o",
      "te"
    ]
  ],
  "065": [],
  "id": "5e38779b15556b9281240771f804ca7a",
  "genre": {
    "code": "te",
    "label": "Oeuvres textuelles"
  },
  "subgenre": {
    "code": null,
    "label": null
  },
  "date": 1911,
  "date_str": "1911",
  "runid": 5394075,
  "internal_RD_id": "5394075/5394107"
}
```

<sup>1</sup> [https://fr.wikipedia.org/wiki/Fonction\\_de\\_hachage](https://fr.wikipedia.org/wiki/Fonction_de_hachage)

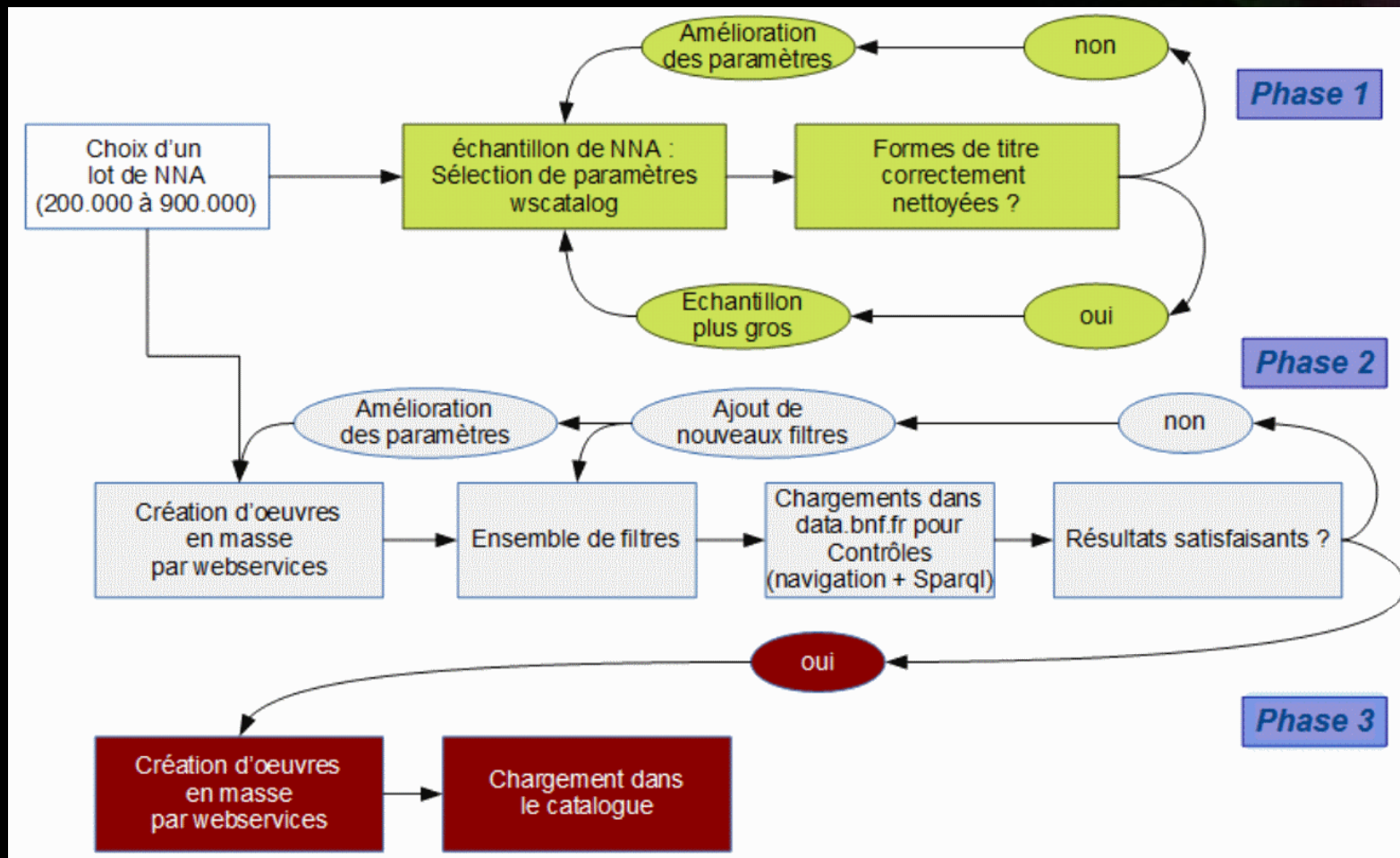
# RobotDonnées : des algorithmes



- Contrôles des résultats :
  - Relecture de fichiers
  - Contrôles `data.bnf.fr`
    - Interface
    - Sparql Endpoint
  - Filtres

# De data.bnf.fr au catalogue

- RobotDonnées -> data.bnf.fr



Phase 1 :  
Exécutions  
manuelles

Phase 2 :  
Exécutions  
programmées

Phase 3 :  
Chargements

- RobotDonnées -> catalogue.bnf.fr

# Premières notices (Fred Vargas)

## Petit traité de toutes vérités sur l'existence Fred Vargas

Page générée automatiquement



Langue : Français  
Genre ou forme de l'œuvre : Œuvres textuelles  
Date : 2001

## Éditions de Petit traité de toutes vérités sur l'existence (4 ressources dans data.bnf.fr)

### Livres (4)

Affichage data.bnf.fr

Trier par

- Petit traité de toutes vérités sur l'existence (2013), Fred Vargas, Paris : Librio , impr. 2013
- Petit traité de toutes vérités sur l'existence (2008), Fred Vargas, Paris : Librio , impr. 2008
- Petit traité de toutes vérités sur l'existence (2003), Fred Vargas, Paris : E.J.L , 2003
- Petit traité de toutes vérités sur l'existence (2001), Fred Vargas, [Paris] : V. Hamy , 2001

# Premières notices (Fred Vargas)

Petit traité de toutes vérités sur l'existence Fred Vargas

Page générée automatiquement



## Notice de titre conventionnel

Notice Au format public

[Vargas, Fred \(1957-....\)](#)

**Petit traité de toutes vérités sur l'existence** *forme internationale*

**Langue(s)** : français

**Date de l'oeuvre** : 2001

Première édition : 2001

**Source(s)** :

- Petit traité de toutes vérités sur l'existence / Fred Vargas

[Affichage catalogue public](#)

**Identifiant de la notice** : ark:/12148/cb17105398w

**Notice n°** : FRBNF17105398

**Création** : 21/04/27

Éditions de Petit traité

**Livres (4)**

Trier par

Petit traité de toutes vérités

Petit traité de toutes vérités

Petit traité de toutes vérités

Petit traité de toutes vérités

# Premières notices (Fred Vargas)

Petit traité de toutes vérités sur l'existence Fred Vargas

Page générée automatiquement

## Notice de titre conventionnel



Noti

### Notice de titre conventionnel

Vargas, Fr

Petit trait

Notice Au format Intermarc ▼

Langue(s)

Date de l'

Première

## Éditions de Petit trait

### Livres (4)

Trier par

Petit traité de toutes véri

Petit traité de toutes véri

Petit traité de toutes véri

Petit traité de toutes véri

000 00444c5 as2200027 45

001 FRBNF171053981

008 210427 yyfre 2001 1

043 .. \$o te

100 .. \$3 [12058359](#) \$w 0 2b..... \$a Vargas \$m Fred \$d 1957-....

145 16 \$w .0..b..... \$a Petit traité de toutes vérités sur l'existence

Source(s)

609 .. \$r edi1 \$d 2001

- Petit tra

610 .. \$3 [37647753](#) \$a Petit traité de toutes vérités sur l'existence / Fred Vargas

690 .. \$a RDWorksB

919 .. \$a RDW \$o 970e65dcab3bba72b4f6d4696d2177ea

991 .. \$a Création\_oeuvres \$b 5700258 \$d 20210416

Identifiant

Notice n° : FRBNF17105398

Création : 21/04/27

Affichage intermarc



# Premières notices (Fred Vargas)

Petit traité de toutes vérités sur l'existence Fred Vargas

Page générée automatiquement

## Notice de titre conventionnel



Notice

## Notice de titre conventionnel

Vargas, Fre

Petit traité

Notice

## Notice de titre conventionnel

Notice Au format Unimarc ▼

Affichage Unimarc

Langue(s)

000 00444c5 a

Date de l'o

001 FRBNF17

Première é

008 210427 yy

100 .. \$3 1205

000 cx h22 45

Source(s) :

043 .. \$o te

- Petit tra

001 FRBNF171053981

610 .. \$3 3764

003 <http://catalogue.bnf.fr/ark:/12148/cb17105398w>

690 .. \$a RDW

100 .. \$a 20210427|frey50 ba0

919 .. \$a RDW

101 0. \$a fre

991 .. \$a Créat

102 .. \$a XX

Identifiant de la notice

103 .. \$a 2001

Notice n° :

106 .. \$a 1

Création :

154 .. \$a |a

240 .. \$a Vargas \$b Fred \$f 1957-.... \$9 0 \$7 ba0yba0y \$8 fre \$t Petit traité de toutes vérités sur l'existence

300 0. \$a Catégorie de l'oeuvre : Oeuvres textuelles

349 .. \$r Première édition \$d 2001

801 .. \$a FR \$b FR-751131015 \$c 20210427

810 .. \$a Petit traité de toutes vérités sur l'existence / Fred Vargas

Éditions de Petit tra

Livres (4)

Trier par

Petit traité de toutes vér

Petit traité de toutes vér

Petit traité de toutes vér

Petit traité de toutes vér

# Quelle diffusion aujourd'hui et demain

- **data.bnf.fr** :

- dump des œuvres JSON sur [api.bnf.fr](https://api.bnf.fr)

- Sparql Endpoint :

- `?w rdagrouplelements:statusOfIdentification "provisional"@en`

<https://c.bnf.fr/NIF>

- Corpus :

- Auteurs français 1830-2021  
900.000 auteurs, 1,5 millions d'œuvres
- Puis :
  - Auteurs anglais 1830-2021
  - Films



# Quelle diffusion aujourd'hui et demain

- **dans le catalogue :**  
**auteurs français 1830-2021**
  - Calendrier : été, automne, hiver 2021
  - une fois chargés :
    - Des ARK et du Marc
    - Mot-clé de requête pour les œuvres : RDWorksB
    - Mot-clé de requête pour les notices BIB : RDW\*
    - Produit à façon
    - SRU, Z39.50

# Conclusion

- Chargement imminent de lots d'oeuvres dans le catalogue
- Articulation avec le projet Noemi
- Articulation avec les clusters Sudoc (AlgOCLC)
  
- Quels services associés ?
  - Jeux test Intermarc ?
  - listes d'ARK biblio ?
  - listes d'auteurs ?

# Temps de questions et d'échange

