



Quelques mots sur la FRBRisation des données à l'Abes

Description

Modèle de données, FRBR, LRM : quelques rappels théoriques et une expérimentation de leur application dans le Sudoc

Public

Coordinateur
Correspondant catalogage
Correspondant autorité



Intervenants

Experts : Olivier Rousseaux ; Philippe Le Pape
Modérateur : Laurent Piquemal

PLAN

- FRBRiser les données d'un catalogue
- Les données FRBRisées au 23 octobre 2017
- Et maintenant ?

Partie 1

FRBRiser les données d'un catalogue

FRBriser les données d'un catalogue

« Ce que nous appelons « FRBRisation » est le processus automatique qui vise à rapprocher du modèle FRBR les informations contenues dans les notices bibliographiques existantes. »

Olivier Rousseaux. *FRBRisation du Sudoc : une esquisse*, dans : *Arabesques*, no. 87, octobre novembre décembre 2017.

→ répartition des données selon les quatre entités
Œuvre, Expression, Manifestation, Item

FRBriser les données d'un catalogue

- Préalable indispensable : avoir défini quelles données ressortissaient
 - à l'entité Œuvre (titre, date, créateur...)
 - à l'entité Expression (langue, droits...)
 - à l'entité Manifestation (titre, adresse, collation, contributeurs...)
 - à l'entité Item (cote, provenance, fonds, reliure...)

Ce travail est actuellement en cours de réalisation en France à travers la rédaction du code RDA-FR, adaptation française du nouveau code de catalogage international RDA.

Adaptation réalisée par le programme Transition bibliographique : <https://www.transition-bibliographique.fr/>

Actuellement en cours également par les groupes de travail du programme Transition bibliographique : la rédaction d'un manuel de catalogage des entités Œuvre, Expression, Personne. Calendrier prévu : fin 2018

FRBriser les données du Sudoc

- Que veut-on ? → Faire en sorte que les données des notices bibliographiques actuelles permettent la **création de métadonnées d'entités** respectant le modèle FRBR
- Par quels moyens ? → Exploitation d'**algorithmes** conçus par OCLC et adaptés aux spécifications définies par l'Abes
- Où ? → Dans **CBS (le socle technique du Sudoc)**.
Dans un premier temps, résultats uniquement visibles sur Winibw
→ aucun impact sur le Sudoc public, ni en termes d'affichage ni en termes de recherche
- Que s'est-il passé le 23 octobre 2017 ? → Mise en place d'algorithmes spécifiques, qui regroupent quotidiennement les notices bibliographiques signalant une même œuvre et créent une notice d'autorité d'un nouveau type : une **notice de regroupement**, préfiguration de notices d'œuvre. Chaque notice bibliographique ayant servi à ce regroupement lui est liée.

Les algorithmes : une mécanique de regroupements

Les algorithmes s'appuient sur un nouvel index CLM, destinés à eux seuls.
Cet index permet d'établir les regroupements de notices bibliographiques

Clé d'index titre auteur 9 4 2 2 1 4

9 premiers caractères du premier mot du titre

4 premiers caractères du deuxième mot du titre

2 premiers caractères du troisième mot du titre

2 premiers caractères du quatrième mot du titre

1 (premier) caractère du cinquième mot du titre

4 premiers caractères du premier mot du nom de l'auteur

Exemple

200 1# \$a La @grande encyclopédie des lutins

700 #1 \$3 026839806 *Dubois, Pierre (1945-....)* \$4 070

702 #1_\$3 028597907 *Sabatier, Claudine (1946-....)* \$4 440

702 #1_\$3 027116344 *Sabatier, Roland (1942-.... ; illustrateur pour la jeunesse)* \$4 440

CLM grande ency de lu dubo

CLM grande ency de lu saba

Clé visible dans WinIBW par la commande aff cle

PARTIE 2

La FRBRisation du Sudoc au 23 octobre 2017

Ce qu'on voit dans WinIBW depuis le 23 octobre 2017 (1/5)

- ✓ une nouvelle zone dans les notices bibliographiques (579)
(pour 4,2M de notices bibliographiques)
- ✓ 1,4 million de notices d'autorité d'un nouveau type (Tr), dites *notices de regroupement*, créées automatiquement
- ✓ Un fonctionnement en base de production (sans modification possible par les catalogueurs)

Un processus dynamique

- ✓ Enrichissement quotidien sur toutes les mises à jour du catalogue (300 à 600 notices de regroupement mises à jour) → les regroupements ne sont pas figés

Ce qu'on voit dans WinIBW depuis le 23 octobre 2017 (1/5)

Côté notices bibliographiques

Création dans les notices regroupées par les algorithmes d'une nouvelle zone **579** qui fait le lien avec la ***notice de regroupement***.

Elle est construite ainsi :

- **##** : indicateurs non définis
- **\$1** : code du programme (0499)
- **\$3** ppn de la notice de regroupement Tr (+ expansion)

Exemple

579 **\$1** 0499 **\$3** 209509953 *La @grande encyclopédie des lutins / Dubois, Pierre (1945-....)*



Zone spécifiquement Sudoc, non définie dans le format Unimarc
La zone est verrouillée, il est impossible de la modifier / créer / supprimer

Ce qu'on voit dans WinIBW depuis le 23 octobre 2017 (2/5)

Côté autorités

- Un nouveau type de notice autorité : *notice de regroupement (008 \$a Tr)* pour lequel sont définies une vingtaine de zones du format Unimarc (A) :
 - . certaines déjà en usage pour d'autres types d'autorités (langue, dates, coordonnées, ...)
 - . certaines reprises du format bibliographique (résumé, note de thèse, accès sujet,...)
 - . certaines créées spécifiquement (titre de l'œuvre, type d'œuvre, ...)

A l'exception du type d'œuvre (189), les nouvelles zones sont décrites dans le format Unimarc, 3e édition (UNIMARC manual : authorities format / edited by Mirna Willer. - 3rd ed. - München : K.G. Saur, 2009).

Les mises à jour depuis 2012 sont disponibles en ligne :

<https://www.ifla.org/publications/unimarc-authorities--3rd-edition--updates>

Ce qu'on voit dans WinIBW depuis le 23 octobre 2017 (3/5)

notice de regroupement : zones du format

008 type d'entité et statut de la notice : \$a **Tr2**

→ statut 2 attribué par défaut

101 langue de l'expression représentative de l'œuvre

103 date associée à l'œuvre

→ date de la plus ancienne manifestation connue dans le Sudoc

104 Niveau intellectuel ou public destinataire

106 utilisation en accès sujet

→ valeurs par défaut

123 coordonnées géographiques [œuvre cartographique]

128 données codées musicales [œuvre musicale]

189 type d'œuvre

231 / 241 titre de l'œuvre anonyme/non anonyme

300 résumé

328 note de thèse

431 / 441 variante de titre de l'œuvre anonyme/non anonyme

500 / 510 / 520 créateur de l'œuvre (personne / collectivité / famille)

501 / 511 / 521 co-créateur de l'œuvre (personne / collectivité / famille)

60X indexation sujet contrôlée

610 indexation en vocabulaire libre

686 classification

→ \$2 TEF ou \$2 Code de classement de la Bibliographie nationale française

zone déjà existante dans le format autorités
zone inspirée du format bibliographique
Zone totalement nouvelle

Ce qu'on voit dans WinIBW depuis le 23 octobre 2017 (5/5)

La notice de regroupement est générée à partir des informations contenues dans sa grappe de notices bibliographiques

008 \$aTr2

101 ## \$a eng

103 ## \$a 1993

104 ## \$a m

106 ## \$a 0 \$b # \$c #

189 ## \$a Oeuvre Textuelle

241 ## \$t @Enchantment of the faerie realm

← langue

← date

public

← type d'œuvre

441 ## \$t @Enchantment of the faerie realm, communicate with nature spirits & elementals

titres

441 ## \$t Le @monde enchanteur des fées, communiquez avec les esprits de la nature et les élémentaux

500 #1 \$3 06045167X Andrews, Ted (1952-....) \$4 070

606 ## \$a Nature \$x Miscellanea \$2 lc

← concepteur

606 ## \$3 029868939 Elfes \$2 rameau

606 ## \$3 027257002 Nature \$3 028679326 Miscellanées \$2 rameau

606 ## \$3 02758206X Fées \$2 rameau

sujets

606 ## \$3 027365565 Esprits \$2 rameau

686 ## \$a 130 \$2 Cadre de classement de la Bibliographie nationale française

Ce qu'on voit dans WinIBW depuis le 23 octobre 2017 (4/5)

notice 1

100 0# \$a2012
 101 1# \$afre \$ceng
 [...] **notice 1**
 104 ## \$am \$by \$scy \$dba \$e0 \$ffre
 [...] **200** 1# \$a Le @monde enchanteur des fées \$e communiquez avec les esprits de la nature et les élémentaux \$f Ted Andrews \$g traduit de l'anglais par Janine Renaud
 [...] **454** ## \$t @Enchantment of the faerie realm
 579 \$10499 [\\$3212901281](#) @Enchantment of the faerie realm / Andrews, Ted (1952-....)
 606 ## [\\$302758206X](#) Fées \$2 rameau

notice 2

100 0# \$a1993
 101 0# \$aeng
 [...] **notice 2**
 104 ## \$am \$by \$scy \$dba \$e0 \$ffre
 [...] **200** 1# \$a @Enchantment of the faerie realm \$e communicate with nature spirits & elementals \$f Ted Andrews
 [...] **579** \$10499 [\\$3212901281](#) @Enchantment of the faerie realm / Andrews, Ted (1952-....)
 606 ## \$a Fairies \$2lc
 606 ## \$a Spirits \$2lc
 606 ## \$a Elves \$2lc
 606 ## \$a Nature \$x Miscellanea \$2lc
 606 ## [\\$302758206X](#) Fées \$2 rameau
 606 ## [\\$3029868939](#) Elfes \$2 rameau
 606 ## [\\$3027257002](#) Nature [\\$3028679326](#) Miscellanées \$2 rameau
 676 ## \$a 133.1/4 \$v20
 680 ## \$a BF1552 \$b.A53 1993
 [...] **700** #1 [\\$306045167X](#) Andrews, Ted (1952-....) \$4070

Annotations on the left:

- 008 \$a Tr2
- 101 ## \$a eng
- 103 ## \$a 1993
- 104 ## \$a m
- 106 ## \$a u \$b # \$c #
- 189 ## \$a Oeuvre Textuelle
- 241 ## \$t @Enchantment of the faerie realm
- 441 ## \$t @Enchantment of the faerie realm
- 441 ## \$t Le @monde enchanteur d'élémentaux
- 500 #1 [\\$3 06045167X](#) Andrews, Ted
- 606 ## \$a Nature \$x Miscellanea \$2
- 606 ## [\\$3 029868939](#) Elfes \$2 rameau
- 606 ## [\\$3 027257002](#) Nature [\\$3 028679326](#)
- 606 ## [\\$3 02758206X](#) Fées \$2 rameau
- 606 ## [\\$3 027365565](#) Esprits \$2 rameau
- 686 ## \$a 130 \$2 Cadre de classe

Index de recherche des notices de regroupement

Pour chercher des notices de regroupement Tr

- ONT [provisoire] :
 - . recherche les mots du titre uniquement dans les notices de regroupement
 - . affiche aussi les notices bibliographiques liées

A partir d'une notice d'agent (= autorités personne / collectivité / famille)

- REL : affiche toutes les notices autorités liées *dont les notices de regroupement*

A partir d'une notice de regroupement Tr

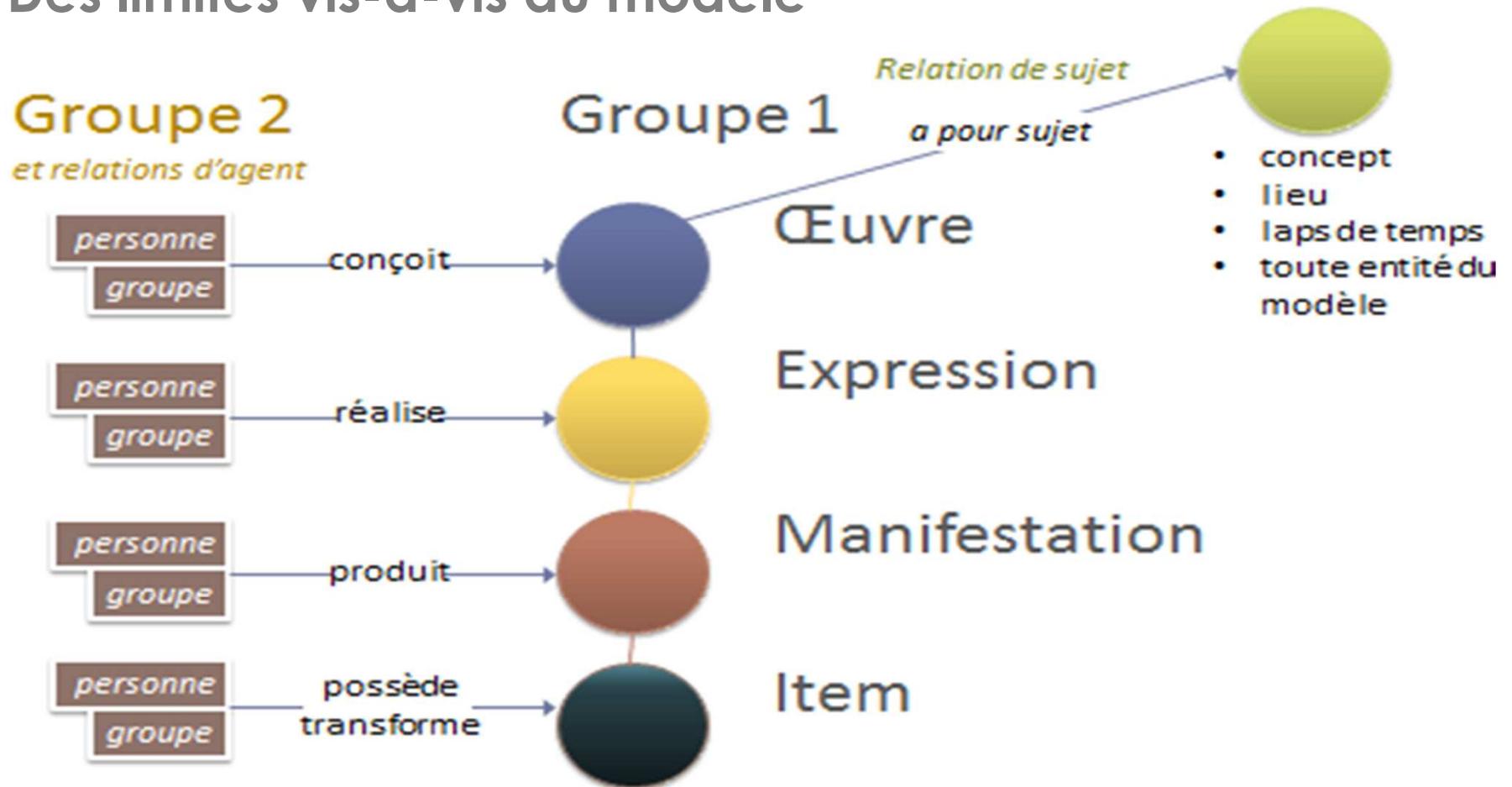
- REL TT : affiche toutes les notices bibliographiques liées (= regroupées)



Dans les notices bibliographiques, la zone 579 qui contient un lien vers la notice de regroupement est cliquable.

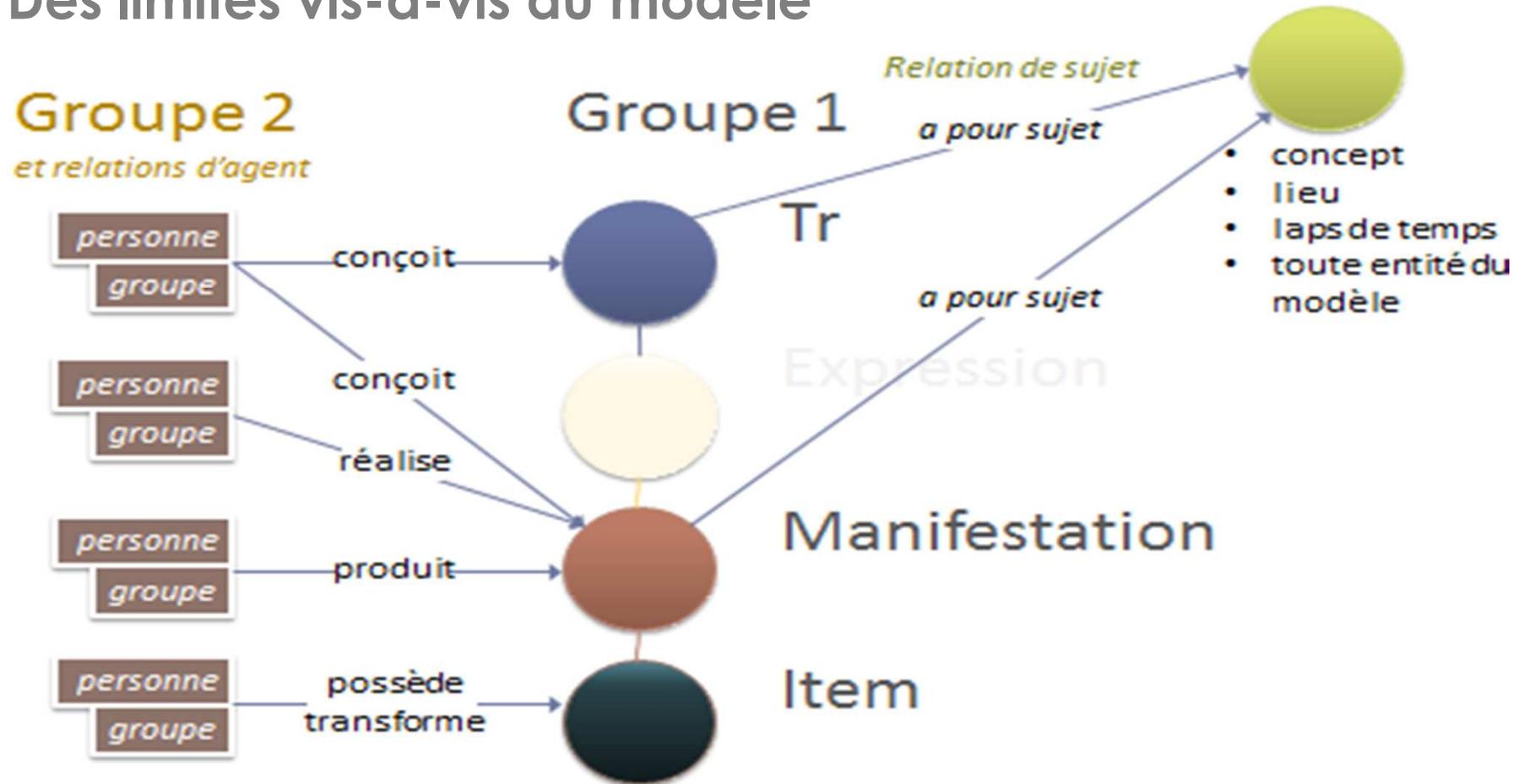
Résultats obtenus / attendus

Des limites vis-à-vis du modèle



Résultats obtenus / attendus

Des limites vis-à-vis du modèle



Dans cette FRBRisation des données du Sudoc, pas de notice d'*expression* générée par les algorithmes

→ l'articulation *Manifestation – Expression – Œuvre* demeure incomplète

Quelques remarques (1/4)

**Cette FRBRisation s'applique à tout le catalogue ...
mais ne peut traiter certains types de ressources**

- ✓ Sont exclues du traitement :
 - . les notices de ressources continues
 - . les notices d'agrégats (manifestations contenant + d'une œuvre) sans titre d'ensemble
 - . les notices de recueils factices

- ✓ Il faut au moins 2 notices bibliographiques représentant la même œuvre pour créer une notice de regroupement.
 - ➔ Pas de notice de regroupement générée pour une notice bibliographique seule à représenter une œuvre.

Quelques remarques (2/4)

Les notices de regroupement sont des pré-notices d'*œuvre*

- ✓ Les données actuelles sont uniquement générées par programme
- ✓ Ces données dépendent de la qualité des notices bibliographiques actuelles (et de la complexité à les interpréter automatiquement)
→ Qualité variable des résultats obtenus
- ✓ Les règles de catalogage pour ces notices restent à finaliser [dans le respect des règles de description des œuvres (RDA-FR)]

Ces notices ne sont donc pas (encore) des notices d'*œuvre* conformes au modèle

Quelques remarques (3/4)



auteur-titre / titre uniforme \leftrightarrow *œuvre*
quelles différences ?

Une succession inscrite dans le format Unimarc

- Le type d'entité *œuvre* a été créé en 2012 dans le format Unimarc(A), en association explicite avec l'application du modèle FRBR :
 - . nouveau type de notice autorité
 - . nouvelles étiquettes de liens bibliographiques
- Dans un catalogue s'appuyant sur ce modèle, les entités *œuvre* succéderont aux notices auteur-titre et titres uniforme

→ Les autorités Auteur-titre et Titre uniforme ont vocation à disparaître au profit des entités *œuvre*

Quelques remarques (4/4)



Autorité Auteur-titre / Autorité Titre uniforme /
Notice de regroupement : quelles différences ?

Une distinction temporaire

- La relation qui lie une notice bibliographique à une notice de regroupement Tr est pour l'instant une relation de type : Manifestation → Œuvre (zone 579)

Elle est assimilable à la relation qui lie les notices bibliographiques aux autorités Titre uniforme en accès titre 500.

- La relation de sujet d'une œuvre à une autre œuvre [assimilable aux actuels liens d'accès sujet (604/605) vers les autorités auteur-titre (Tq) / titre-uniforme (Tu)] n'existe pas encore dans le Sudoc.
Des zones dédiées existent dans le format Unimarc (B).
Leur mise en production dans le Sudoc reste à étudier

PARTIE 3

Et maintenant... ?

Et maintenant...

L'Abes vous propose dans un premier temps :

- d'observer ce qui se passe
- de transmettre vos observations si vous le souhaitez
→ Un [fichier partagé](#) est mis à disposition à cette fin, accessible à partir du Guide méthodologique :
<http://documentation.abes.fr/sudoc/regles/Catalogage/index.htm>
- de [poursuivre le travail sur la qualité des données](#) : en respectant les consignes de catalogage, vous contribuez à rendre efficaces les algorithmes de regroupement.

Un premier bilan sera fait à l'occasion des Journées Abes 2018

**Aucune de ces nouveautés n'est exportée vers les SIGB
Aucune n'est actuellement visible dans le Sudoc Public ni dans Idref**

Poursuivre le travail sur la qualité des données

Rester vigilant sur l'application des consignes de catalogage du GM
(rubrique « [FRBRisation du Sudoc : consignes de catalogage](#) »)

En particulier

- ✓ 1^{er} Indicateur de la zone 101 → langue originale de l'œuvre
- ✓ Zone 454 → est une traduction de

- ✓ Etiquette des zones 7XX
- ✓ Choix des rôles (codes de fonction \$4)

- ✓ Structuration des zones 200
- ✓ Utilisation des zones 463/464

Poursuivre le travail sur la qualité des données

- ✓ 1^{er} Indicateur de la zone 101 → langue originale de l'œuvre
- ✓ Zone 454 → est une traduction de

Notice 1 (langue originale):

101 0# \$a ger

200 1# \$a Die @Panne \$e ein Hörspiel und eine Komödie \$f Friedrich Dürrenmatt

Notice 2 (traduction):

101 1# \$a fre \$c ger

200 1# \$a La @panne \$e Une histoire encore possible \$f Friedrich Dürrenmatt \$g trad.
de l'allemand par Armel Guerne

454 ## \$t Die @Panne

L'impact éventuel d'une modification de notice bibliographique sur la FRBRisation est visible dès le lendemain :

- . modification de l'appartenance de la notice à un regroupement
- . modification de la notice de regroupement

Poursuivre le travail sur la qualité des données

- ✓ Etiquette des zones 7XX
- ✓ Choix des rôles (codes de fonction \$4)

Exemple :

200 1# \$a @Harry Potter à l'école des sorciers \$f J. K. Rowling \$g **traduit de l'anglais par Jean-François Ménard**

700 #1 [\\$3050222937](#)Rowling, Joanne Kathleen (1965-....) \$4 070

~~701 #1 [\\$3027022617](#)Ménard, Jean-François (1948-....) \$4 070~~

702 #1 [\\$3027022617](#)Ménard, Jean-François (1948-....) \$4 730

L'impact éventuel d'une modification de notice bibliographique sur la FRBRisation est visible dès le lendemain :

- . modification de l'appartenance de la notice à un regroupement
- . modification de la notice de regroupement

Poursuivre le travail sur la qualité des données

- ✓ Structuration des zones 200
- ✓ Utilisation des zones 463/464

Exemple :

200 1# \$a @Physiologie du gout ou, Méditations de gastronomie transcendante \$f par Brillat-Savarin ~~\$g suivi de La gastronomie par Bercoux~~ \$c suivi de La gastronomie \$f par Bercoux

464 ## \$t @Physiologie du gout ou, Méditations de gastronomie transcendante **\$f** par Brillat-Savarin

464 ## \$t La @gastronomie **\$f** par Bercoux

Même si les agrégats ne sont actuellement pas traités, la structuration de leurs données est la condition impérative d'une FRBRisation automatique ultérieure

Poursuivre le travail sur la qualité des données

Le catalogage selon RDA-FR et la FRBRisation concourent au même but : tendre vers une application du modèle FRBR (et désormais, du modèle LRM)

Quelles suites pour 2018 ?

Ces premiers résultats sont une expérimentation en cours

Evaluation qualitative à mener

- . sur les données existantes du catalogue
- . sur les données générées :
 - paramètres à ajuster
 - limites et insuffisances des algorithmes

Un bilan en 2018 permettra de décider des suites à mener : poursuite ou non de l'expérimentation

- Travaux conjoints avec la Bibliothèque nationale de France : confrontation des résultats obtenus dans le Sudoc avec ceux de data.bnf.fr
 - réflexions autour d'une base d'entités Œuvres commune
 - dans le cadre de l'étude de faisabilité (en cours) d'un Fichier national des Entités (FNE)

Des questions ?