

### Diapo 35 : le travail des réseaux

L'ABES tient à saluer le travail de fond réalisé quotidiennement par les réseaux... Un exemple, parmi d'autres, tiré des derniers messages parus sur les listes de diffusion. Il s'agit d'un gros travail de dédoublonnage dans le catalogue Sudoc réalisé par la Bibliothèque interuniversitaire de la Sorbonne

### Diapo 36 : SRU sudoc / Worldcat

L'année 2013 s'est achevée avec la mise en production -attendue de longue date- du processus de synchronisation automatique via le protocole SRU (*Search/Retrieval via URL*) entre le CBS et WorldCat, catalogue en ligne d'OCLC.

Afin d'obtenir des données de localisation cohérentes, la mise en place d'une synchronisation du fichier des bibliothèques a constitué le préalable à cette mise à jour.

La mise en œuvre de la synchronisation sur les notices bibliographiques a été activée début novembre 2013.

Il est à noter que la synchronisation pour les notices ISSN est entrée en production en janvier 2014. L'ABES n'ayant pas l'autorisation du Centre International d'exporter les notices venant d'ISSN, seule une notice minimale est envoyée : si elle existe déjà dans WorldCat, cela permet d'y rattacher les localisations. L'ABES ne peut donc pas assurer l'exposition exhaustive de ces données dans WorldCat.

Une nouvelle campagne d'adhésion à WorldCat pour l'année 2015 sera lancée auprès des établissements n'ayant pas souscrit en 2013 dès que les conditions d'entrée tarifaires auront été précisées par OCLC (la négociation du contrat est en cours).

### Diapo 37 : des évolutions du format Unimarc

Depuis 2011, le pôle Métadonnées a œuvré pour la prise en compte des principales évolutions du format Unimarc à intégrer dans les formats de catalogage et d'export du Sudoc.

Deux vagues successives (en juillet 2012 et novembre 2013) ont permis de mettre à niveau les formats des notices bibliographiques et d'autorités : ajout ou répétabilité de sous-zones ou de zones comme par exemples la zone 003 *identifiant pérenne de la notice* ou la 033 *identifiant pérenne de la notice dans une autre base*

La prochaine vague, essentiellement centrée sur les autorités, est quant à elle programmée pour l'automne 2014.

Ce travail de longue haleine vise aussi à s'adapter aux évolutions de formats et à la « FRBRisation » du catalogue Sudoc.

### Diapo 38 : SUDOC FRBR

Depuis la mise en ligne, en février 2013, des premières consignes de catalogage (en vue de FRBRiser le Sudoc), l'équipe en charge de ce projet à l'ABES réfléchit à la mise en œuvre d'un catalogage FRBRisé, doublée d'une FRBRisation rétrospective des données du catalogue.

Cette vaste entreprise s'avère complexe et s'attache dans un premier temps à produire de nouvelles règles visant :

- à isoler la date de copyright (qui exprime potentiellement la date de création de l'expression) dans une sous-zone dédiée de la zone 210

- à implémenter la zone 0 de l'ISBD (qui exprime les **Caractéristiques de contenu de l'Expression** (zone 181) et les **Caractéristiques du type de médiation de la Manifestation** (zone 182)

- à remplacer partiellement l'usage de la zone 423 pour lui substituer une zone 464 « contient » (quand cette 423 n'exprime pas la publication conjointe de 2 ressources distinctes)

Ceci en attendant de proposer des enrichissements des notices d'autorité *Titre uniformes* pour les faire évoluer vers des Notices d'Œuvre.

--> Ce début 2014 a été l'occasion d'un bilan et d'un recadrage des réflexions et orientations FRBR esquissées par l'ABES pour le Sudoc depuis un peu plus d'un an. Il en est sorti une demande de prestation à OCLC, validée en avril 2014 par le comité de direction de l'ABES. Il s'agirait d'initier (d'amorcer) la mise en œuvre d'une FRBRisation au sein du Sudoc à l'aide des algorithmes développés par OCLC dans le CBS (dans le cadre d'un projet similaire mené avec un réseau de bibliothèques de lecture publique néerlandaises.

La demande ABES est en attente d'une première réponse d'OCLC (faisabilité, les délais, les coûts ...)

#### Diapos 40 et 41 : SUDOC RDF

Concernant la sortie RDF du SUDOC, des corrections et des enrichissements sur les différents types de documents ont été effectués à partir d'octobre 2013. Des enrichissements des zones de liens bibliographiques sont en cours, notamment pour les périodiques (zones 412, 413, etc.) ainsi que des spécifications sur les zones des responsabilités et les données codées.

Une documentation a été mise en ligne l'été dernier à l'adresse qui apparaît en bas de l'écran : [http://documentation.abes.fr/sudoc/manuels/administration/sudoc\\_rdf/index.html](http://documentation.abes.fr/sudoc/manuels/administration/sudoc_rdf/index.html)

Cette documentation comporte :

- une introduction succincte sur le web de données et les standards utilisés ;
- un aperçu illustré d'exemples concrets;
- le modèle de données avec les espaces de noms, les vocabulaires utilisés et le tableau de correspondance Unimarc-RDF.

Notice exemple

#### Diapo 42 : Qualinca 1/5

Qualinca = Qualité et interopérabilité de grands catalogues documentaires

Projet multipartenaires développant chacun des méthodes différentes de traitement sur les métadonnées et financé par l'Agence Nationale de la recherche

Deux institutions nationales mettent à disposition leurs données :

- l'ABES → des métadonnées bibliographique de l'enseignement supérieur → Sudoc
- l'INA → des métadonnées audiovisuelles de l'archive nationale de l'audiovisuel

Trois laboratoires en informatique analysent ces métadonnées :

- Le LIRMM (laboratoire d'informatique de robotique et de microélectronique de Montpellier)
- Le LRI (laboratoire de recherche en informatique de Paris)
- Le LIG (laboratoire d'informatique de Grenoble)

Ils travaillent avec les technologies sémantiques sur la gestion de grandes bases de connaissances, comme par exemple Sudoc : le Sudoc n'est pas une base parfaite. Qualinca doit permettre de développer en priorité des mécanismes capables de vérifier la cohérence des liens au sein du Sudoc et de valider ou invalider ces liens. Il sera alors possible d'évaluer et de décrire le niveau de qualité de la base et ensuite d'améliorer et de maintenir le niveau de qualité (trouver et corriger les erreurs de liage, dédoublonner, enrichir et/ou créer des notices autorités)

A mi-parcours le projet dans son ensemble a été évalué positivement par l'ANR le 7 avril dernier. Ont été salués :

- la cohérence scientifique du projet et la bonne coordination entre les partenaires
- ainsi que l'effort de médiation réciproque entre scientifiques et experts métier

Chaque laboratoire essaie des techniques différentes, sur les données de l'INA d'un côté, et, sur celles de l'ABES de l'autre.

#### **Diapo 43 : Qualinca suite 2/5**

Côté ABES, Qualinca fait suite au projet SudocAD en partenariat avec le LIRMM qui a abouti à la construction d'un prototype de liage de notices d'autorités à des notices bibliographiques.

Actuellement, SudocAD est utilisé de façon ponctuelle dans le cadre du Hub de métadonnées. Des tests sont également effectués en interne sur d'autres corpus pour IDRef. Cependant, cet outil a ses limites car il est trop générique, il ne peut s'adapter aux spécificités de certaines données et il n'est pas prévu de le faire évoluer, à part quelques ajustements mineurs.

Une autre solution a été retenue pour poursuivre ce travail. Il s'agit de rendre plus modulable un service de liage aux autorités. Ainsi, il sera possible d'exploiter les informations les plus pertinentes de chaque corpus en fonction de la nature et de l'origine des données à analyser.

Aujourd'hui, dans le cadre du projet, l'ABES travaille avec le LIRMM sur le développement d'un nouveau prototype, « Sudoqual » dont l'objectif est d'évaluer la qualité des liens entre notices aut/bibliogr ; corriger les faux liens et générer de bons liens. A terme, l'ensemble de ces outils devraient nous permettre d'améliorer la qualité globale de nos bases en proposant du liage automatique

#### **Diapos 44 : Exemples Qualinca 1**

Exemple 1 : Qualinca intégrée à IDRef pour une aide à la décision de liage quand il y a trop d'homonymes (ici choisir entre 53 Alain Bernard)

#### **Diapos 45 : Exemples Qualinca 2**

Ou bien comme un outil de correction interne des liens entre les notices bibliographiques et les notices autorités liées.

#### **Diapo 46 : Qualinca 5 /5 Conclusion**

Demain : il sera possible de partager les résultats et, pourquoi pas, de mixer et d'enrichir les méthodes.

Le LIRMM et l'ABES fondent leur démarche sur les informations uniquement présentes dans le Sudoc alors que le LIG et l'INA ont adopté une autre approche en récupérant aussi des informations dans des bases externes du *linked open data* (dont DBPedia).

Le LRI quant à lui travaille sur la fusion de données en testant des échantillons des deux bases (Sudoc et archives INA) afin de créer des autorités enrichies et d'en dédoublonner d'autres. Voire créer des notices d'œuvre...

A ce stade du projet, il est nécessaire de commencer à envisager la question de l'évaluation, qui comporte deux volets : évaluer les méthodes scientifiques et évaluer les prototypes réalisés. [Pour le second volet, l'équipe envisage, notamment, de faire appel au réseau en s'appuyant sur l'expertise métier des catalogueurs].

#### Diapo 48 : HUB 1/3

Le hub de métadonnées, c'est : **un projet mais aussi des services actifs** qui forment un ensemble d'outils et de méthodes se construisant et s'affinant au fur et à mesure des traitements.

Ces services sont réalisés par une équipe de 8 personnes mise en place durant l'été 2013 (= 4.3 ETP) pour qui, il s'agit de:

- traiter des données de toutes sortes
- les convertir et les stocker dans une base RDF sans perte d'information
- les corriger et les enrichir de façon automatique ou non, certaines tâches ne pouvant se passer de l'intervention humaine
- fournir ces données traitées dans divers formats

Les actions menées en priorité visent :

- le traitement des Licences Nationales ISTEK ;
- le signalement dans le catalogue Sudoc
- la mise à disposition des fichiers de métadonnées au format Kbart pour les bases de connaissances sur lesquelles s'appuient les outils de découverte et les résolveurs de lien ...
- demain, le web de données (LOD - Linked Open Data).

#### Diapo 49 : HUB 2/3 : chronologie

Avant la signature d'un contrat Licence Nationale :

- le département ADELE négocie avec les fournisseurs et communique les offres à l'équipe HUB qui peut démarrer l'analyse et la correction des listes de documents pour obtenir une liste contractuelle. Cette tâche est facilitée par « Métarevues », outil développé pour les besoins du hub et qui permet de « déplier l'historique d'un titre » en vue de l'établissement des listes contractuelles.
- dans le même temps, l'équipe Hub analyse et réalise l'évaluation des échantillons (s'appuie sur une *Grille d'analyse et d'évaluation* ; les notices XML sont chargées telles quelles dans une base Oracle ; des requêtes prédéfinies ou non sont lancées pour « ausculter » les données).
- puis une mesure de la présence dans le Sudoc de ce corpus est effectuée (notices papier et électroniques ; nb d'exemplaires ; nb bibliothèques localisées ; années cumulées) afin de suivre l'évolution de la gestion des collections par le réseau :

Après la signature du contrat, il faut réaliser :

- une comparaison des échantillons et de la liste éditeur avec les données réellement livrées

- la modélisation RDF et les conversions :
- la création et/ou la mise à jour de notices du Sudoc puis l'exemplarisation et une sortie au format KBart
- les enrichissements avec priorité donnée aux autorités (auteur et vedette RAMEAU) dans la mesure du possible. Nécessite la mise en place d'un suivi de chantiers d'enrichissement via le réseau pour les améliorations restant à faire pour la zone 309 du Sudoc.
- une liste des collections imprimées est alors fournie pour faciliter le désherbage
- une communication des informations nécessaires est faite auprès du réseau

A noter : le programme Métarevues a vu son usage étendu à Périscope comme aide à l'amélioration de la qualité des données du Sudoc. Par ailleurs, une nouvelle version de Périscope est actuellement en phase de test. Elle intègre une technique utilisée pour affiner le signalement les états de collections et donc le repérage des lacunes. Cette même technique permet de réaliser les « photographies » des corpus de collections d'imprimés à un instant donné.

#### **Diapo 50 : HUB 3/3 tableau récapitulatif des traitements**

Tableau récapitulatif des traitements et de leur état d'avancement pour chaque corpus (mai 2014) : il reprend les principales étapes décrites dans la diapo précédente (en vert : ce qui a été fait / en jaune-orangé : le travail en cours / en blanc : ce qu'il reste à faire).

Les négociations en cours sont visibles sur le site Licence Nationale.

#### **Diapo 51 : Convention ABES / CTLES**

Le dernier CA de l'ABES a validé une convention-cadre définissant les grands principes de la coopération entre l'ABES et le CTLes pour les actions visant à soutenir, développer et mettre en œuvre des plans de conservation partagée des périodiques, en Ile-de-France et en région.

La conservation partagée figure à la fois dans le projet d'établissement de l'ABES et dans les missions du CTLes : il s'agit pour chaque partenaire de coopérer avec les moyens humains et techniques à disposition pour mettre en place un environnement de travail collaboratif où chacun reste sur son domaine d'expertise : le signalement et les outils qui lui sont dédiés pour l'ABES ; la gestion du PCP pour le CTLes

Cette convention est le pendant « enseignement supérieur » de celle élaborée avec la FILL (Fédération interrégionale du livre et de la lecture) côté lecture publique

#### **Diapo 52 : BACON**

Il s'agit du premier véritable projet conjoint ABES-Couperin + 8 établissements associés. Il portait à la fois sur les données et le test d'outils. Un rapport a dressé les constats communs :

- les outils actuels de signalement des ressources électroniques ont de fortes lacunes en particulier pour le contenu français/francophone.
- les établissements ont beaucoup plus besoin de données de qualité à intégrer à leurs outils (listes alphabétiques, bases de connaissances de résolveurs de liens) que d'outil à proprement parler

#### **Les orientations à très court terme :**

- obtenir des éditeurs français/francophones des fichiers suivant la recommandation KBART

- mettre en place un service de données de référence pour les fichiers KBART concernant le secteur académique français et remis à disposition sous une licence libre et enfin
- participer à la dynamique internationale de signalements tels que KB+ et GOKb

#### L'orientation à moyen terme :

- obtenir des éditeurs français toutes les métadonnées d'articles, les enrichir et les exposer sous une licence libre

#### Diapo 54 : Convention ABES / BnF 1/2

La convention triennale de transfert de données entre l'ABES et la BnF arrivait à son terme cette année. Une nouvelle convention a été présentée et approuvée au CA de l'ABES le 12 mai dernier. Elle s'appelle désormais convention de coopération

Cette nouvelle convention a pour premier objectif de **sécuriser et consolider les flux de données existants**, tant pour les chargements de données bibliographiques et d'autorité venant de la BnF vers le Sudoc que pour les informations concernant les bibliothèques (les notices RCR) que l'ABES fournit à la BnF pour les besoins du Répertoire national des bibliothèques et fonds documentaires (RNBFD).

Cette convention se doit aussi **d'ouvrir la voie à de nouvelles coopérations entre l'ABES et la BnF** pour faciliter notamment l'articulation des stratégies de chacune, face à l'expansion des collections numériques et la volonté partagée de parier/s'appuyer sur les technologies du Web sémantique pour repenser les catalogues de bibliothèque, la diffusion et le partage des données.

#### Diapo 55 : Convention ABES / BnF 2/2 axes de travail

Voici les 5 axes prioritaires:

**1 – la BnF propose à l'ABES d'œuvrer à la mutualisation des données d'autorité** (œuvres et titres, personnes, collectivités, sujets) produites par les deux établissements et leurs réseaux.

L'ABES et BnF, avec RAMEAU, ont déjà fait l'expérience de la construction d'un fichier d'autorité commun, qui doit, aujourd'hui, être modernisé.

Les nouveaux axes de collaboration envisagés privilégient l'utilisation des technologies du Web sémantique, c'est-à-dire des techniques d'alignement entre référentiels, indépendantes des systèmes de production respectifs.

L'objectif de cette mutualisation des autorités serait de **disposer d'une base de données à la fois ouverte juridiquement** (réutilisable par des tiers sous licence ouverte de l'Etat) **et maîtrisée** (les opérateurs français conserveraient une capacité de retraitement indépendamment des systèmes utilisés).

Ce référentiel autorité mutualisé, véritable centre névralgique pour l'ensemble des données et des bases utilisatrices (gérées ou non par l'ABES ou la BnF), renforcerait la stratégie adoptée par les 2 établissements pour l'évolution de leurs catalogues et leurs services bibliographiques.

Par ailleurs, les réalisations conjointes de l'ABES et de la BnF en termes d'alignements pourraient opportunément **orienter la réflexion stratégique en cours sur les systèmes d'information et les**

**référentiels de la recherche auprès d'autres acteurs de l'Enseignement supérieur et de la recherche comme de la sphère Culture.**

A court terme un axe prioritaire de collaboration concerne le rapprochement des **fichiers d'autorité noms de personnes et collectivités via un alignement des identifiants ISNI** (utilisés par la BnF) et **IdRef** (utilisés par l'ABES).

## **2 –l'INTEROPERABILITE DES FORMATS :**

L'ABES et la BnF **poursuivent leur collaboration en matière de modélisation dans le but de définir un modèle de données commun permettant d'exprimer en RDF et selon le modèle FRBRoo toute la richesse et la granularité de leurs formats MARC (Unimarc et InterMarc).** Ce travail, qui vise des usages professionnels nécessitant la récupération des données « complètes » des catalogues, sera conduit de façon pragmatique et phasée. Des confrontations et des analyses effectives de données seront effectuées.

Ce travail devra aussi s'accompagner d'études conjointes relatives à l'utilisation de **BIBFRAME** ou d'autres solutions d'échange plus simples et plus génériques, à des fins d'interopérabilité et de réutilisation par des acteurs/utilisateurs (*autres que les bibliothèques*) qui n'attendent pas un niveau d'information aussi fin ni une structuration spécifique aux données de bibliothèque.

L'amélioration des données exposées en RDF dans une version que l'on peut qualifier de « grand public » sera par ailleurs menée via les entrepôts et requêteurs SPARQL.

## **3 – pour le troisième axe, il est question de DEMATERIALISATION DU CIRCUIT DES DEMANDES ABES AU CENTRE ISSN FRANCE**

Comme vu précédemment dans le diaporama, la BnF (qui héberge le Centre ISSN France) et l'ABES viennent de lancer un projet de développement applicatif visant la **modernisation du circuit de gestion des demandes de correction et de numérotation ISSN de ressources continues** issues des établissements coordonnés par l'ABES (**c'est-à-dire permettre et gérer la dématérialisation des justificatifs**).

## **4 –la CONCERTATION NATIONALE AUTOUR D'UN OUTIL DE CATALOGAGE EN EAD POUR LES FONDS D'ARCHIVES ET DE MANUSCRITS**

La BnF a proposé à l'ABES de s'associer à une **concertation nationale** (étendue aux partenaires du CCFr et de l'Equipex Biblissima) **autour des outils et des pratiques de catalogage au format EAD pour les collections d'archives et de manuscrits**. Cette collaboration pourra s'appuyer sur les liens étroits déjà tissés entre les communautés d'experts travaillant dans BAM (l'outil de la BnF) et dans Calames (l'outil ABES). Cette proposition intervient dans le contexte où, d'une part, l'usage de l'EAD s'est largement répandu et normalisé, en France comme à l'étranger, et où, d'autre part, les outils de production des deux établissements atteignent certaines limites et présentent des dysfonctionnements et des usages comparables.

**Cette collaboration ne vise que l'outil de production les interfaces de consultation en sont exclue.**

## **5 –la nécessaire continuité de l'Alimentation et des évolutions du Catalogue collectif de France**