

Auteurs et autorités dans les systèmes de bibliothèque après l'adoption de RDA : étude de cas

French translation of the original paper : « Authors and Authorities in Post-RDA Library Systems : A Case Study »

Translated by: Sylvie Sollier, Bibliothèque nationale de France, Paris, France.

Le texte de ce document a été traduit en français et peut contenir des différences par rapport au texte original. Cette traduction est uniquement fournie à titre indicatif.

Heather Lea Moulaison

The iSchool at the University of Missouri, Columbia, USA.



This is a French translation of "*Authors and Authorities in Post-RDA Library Systems : A Case Study* " copyright © 2014 by Sylvie Sollier. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

Résumé :

Les personnes sont complexes et leur représentation dans les notices d'autorité des bibliothèques est rendue de plus en plus complexe par l'ajout d'attributs conformément aux règles de RDA. Cette étude de cas examine, en utilisant une approche longitudinale, les attributs des notices d'autorité des bibliothèques universitaires du Merlin Cluster au bout de six mois, puis d'un an, d'adoption officielle de RDA, pour évaluer les tendances dans l'utilisation des champs d'attributs.

Cette étude a révélé qu'au bout d'un an de catalogage en RDA, 87,58 % des notices n'avaient aucun attribut, mais que ce pourcentage était en baisse de -2,43% par rapport à ce qu'il était au bout de 6 mois. Presque 8% des notices d'autorité avaient au moins un attribut au bout d'un an ; et presque 5% en avaient au moins deux. En ce qui concerne les attributs renseignés, les dates enregistrées dans le 046 des notices d'autorité étaient les plus répandues avec 7,1% de notices en contenant au bout de six mois de catalogage en RDA, et presque 10% de notices en contenant au bout d'un an de catalogage en RDA. Venaient ensuite comme attributs les plus renseignés le 678 (données biographiques ou historiques) et le 375 (sexe).

Globalement les attributs renseignés sont rares et concentrés sur les personnes anglophones, de sexe masculin et celles ayant un lien quel qu'il soit, avec les universités. Bien que le titre de cette enquête soit l'analyse de la qualité des métadonnées à travers le degré de complétude des attributs dans les notices d'autorité après l'adoption de RDA, on a pu faire quelques observations informelles au cours de l'enquête portant sur des points relatifs à la démographie des personnes du fichier d'autorité.

Mots-clés : notice d'autorité Nom de personne, qualité des métadonnées, RDA, études de cas

Introduction :

Les notices catalographiques basées sur RDA (*Ressources : description and accès*), y compris les notices d'autorité, ont pour but de rassembler de l'information sur les entités de l'univers bibliographique comme jamais on ne l'a fait auparavant. D'un point de vue historique, les catalogues de bibliothèque fournissaient des informations sur les documents de leurs collections afin de permettre ensuite de les retrouver. Toutefois, ces documents n'existent pas indépendamment des gens qui les ont créés ou qu'ils concernent. Pour les *Spécifications fonctionnelles des notices bibliographiques* (FRBR : *Functional requirements for bibliographic records*) la première opération effectuée par l'utilisateur dans le modèle conceptuel FRBR est *trouver*, suivie d'*identifier*, *sélectionner* et *obtenir* (FRBR 2009). Une fonctionnalité *trouver* réellement robuste devrait permettre à l'utilisateur d'un catalogue de bibliothèque d'effectuer une recherche non seulement sur les documents et leurs attributs, mais aussi sur les attributs des personnes qui leur sont associées et sur les relations entre ces personnes. Ceci pour permettre ensuite une recherche sur des critères jusqu'ici inconcevables. Être capable de trouver un ensemble de documents écrits en chinois par des femmes dentistes qui sont aussi poètes n'est qu'un rêve avec les systèmes de bibliothèques actuels, mais ce pourrait être une réalité dans un futur très proche.

Les personnes sont complexes et c'est peu dire que d'affirmer que les gens sont beaucoup plus complexes que la description qu'en donnent les métadonnées de bibliothèques. Les personnes sont représentées de deux façons dans les systèmes des bibliothèques : par la chaîne de caractères Nom de personne utilisée comme point d'accès ou par un *identifiant* dans les notices bibliographiques, et par les informations conservées dans la notice d'autorité. RDA définit l'identifiant d'une personne comme «une chaîne de caractères associée de manière unique à une personne ou au substitut d'une personne (par exemple, une notice d'autorité). L'identifiant sert à distinguer cette personne d'autres personnes » (RDA Toolkit, 2014) et est une aide à la fonction *trouver* des catalogues. Les attributs des personnes, avec l'identifiant et la justification de l'identifiant, sont enregistrés dans les notices d'autorité RDA, notices qui traditionnellement n'ont pas contribué de façon robuste à la recherche documentaire (Yee, 2005). Les attributs dans les notices d'autorité aident à identifier les personnes et peuvent également aider à les distinguer. Les notices d'autorité Nom de personne basées sur RDA permettent maintenant d'enregistrer en tant qu'attributs des informations supplémentaires telles que : sexe, occupation, lieu associé, langue et autres (RDA Toolkit, 2014) et ces données sont conçues pour être utilisables dans le futur.

Dans le système de bibliothèque du futur il y a une interrelation entre la chaîne de caractères représentant les personnes, leurs attributs et le rôle des autorités. Comme la nature des identifiants Nom de personne évolue, les données doivent être présentes sous forme d'attributs pour aider l'utilisateur lors de l'étape *identifier*. Bien qu'avec RDA on puisse à présent ajouter plus d'informations sur les individus qu'en utilisant les *Règles de catalogage anglo-américaines, 2^e édition* (RCAA2), on ne fournit pas nécessairement de contenu supplémentaire dans les notices d'autorité établies en suivant RDA. Quels sont les attributs réellement inclus dans les notices d'autorité et quelles conclusions peut-on en tirer ? Les perspectives potentielles apportées à l'utilisateur final par l'ajout de données sur les personnes sont grandes, en particulier si on peut simplifier et différencier les chaînes de caractères dans les identifiants noms de personne.

FRAD et RDA : de nouveaux modèles pour les accès

Les *Spécifications fonctionnelles des données d'autorité* (FRAD)(2013) servent d'appui aux FRBR (2009) en identifiant les catégories d'attributs des noms de personne qui aident à l'organisation et à la recherche au sein de l'univers bibliographique. FRAD identifie quatorze attributs pour les personnes, dont onze s'ajoutent aux attributs déjà identifiés dans les FRBR. Les trois attributs des personnes identifiés dans les FRBR sont dates, titre et autre information associée à la personne. Les attributs supplémentaires identifiés dans FRAD sont sexe, lieu de naissance, lieu de mort, pays, lieu de résidence, affiliation, adresse, langue, champ d'activité, profession et biographie/histoire. Certains de ces attributs ont une relation à l'entité personne qu'elles décrivent (FRAD, 2013). Ces attributs aident

aux opérations effectuées par les utilisateurs définies dans FRAD : *trouver, identifier, contextualiser et justifier*. L'utilisateur final des systèmes de bibliothèques sera plus intéressé par les trois premières : *trouver, identifier et contextualiser*. Dans un système indexant le plein-texte, n'importe quel attribut pourra potentiellement être une aide à la fonction *trouver*. Les attributs des personnes aident également à l'identification. De même, on considère que les dates associées à la personne et la biographie permettent de contextualiser.

En outre, FRAD (2013) identifie une série de relations potentielles entre les personnes et les chaînes de caractères qui les représentent, dans le cadre des fonds documentaires des bibliothèques. Dans les points d'accès, la personne est d'abord connue sous le nom choisi par les bibliothécaires en conformité avec les instructions de RDA. La chaîne de caractères, comme toute métadonnée, est un outil efficace d'organisation de la connaissance, ce n'en est pas moins un produit artificiel de la création de métadonnées de bibliothèque.

Ressources : description et accès (RDA) est la norme de contenu couramment utilisée aux Etats-Unis. Le PCC (Program for cooperative cataloging, Programme de catalogage en coopération) a annoncé que le Jour Un du PCC pour les notices d'autorité en RDA (*PCC Day One for RDA Authority Records*) serait celui de l'adoption officielle de RDA par la Bibliothèque du Congrès le 31 mars 2013 (« PCC Day One », 2011). On a ajouté aux notices d'autorité des champs complémentaires en MARC pour les nouvelles données qui pouvaient être enregistrées. Le tableau 1 montre les champs d'autorité en MARC pour l'encodage des attributs ; tous sauf le 678 (Données biographiques ou historiques) ont été créés pour RDA (« Marc 21 encoding », 2012 ; « RDA cataloger training », 2013).

Tableau 1 : champs en MARC relatifs aux personnes pour les autorités établies selon RDA
(R) = répétable (NR) = non répétable (« MARC 21 », 2013)

MARC autorité Code-champs	Description
046	Dates spéciales codées (R)
370	Lieu associé (R)
372	Domaine d'activité (R)
373	Groupe associé (R)
374	Profession (R)
375	Sexe (R)
376	Information sur la famille (R)
377	Langue associée (R)
378	Forme développée du Nom de personne (R)
678	Données biographiques ou historiques

Parce que RDA gère les notices bibliographiques mais aussi les points d'accès et les autorités, les points d'accès Nom de personne et les notices d'autorité sous-jacentes ont subi des changements massifs lors de l'adoption de RDA (cf « RDA cataloger training », 2013 ; « PCC Day One », 2012). Ces changements n'expliquent pas, cependant, la copie massive d'information des points d'accès vers les champs attributs des notices d'autorité. Tous les attributs ont bien sûr été saisis manuellement par les catalogueurs.

Contribution au contenu des notices

Actuellement, la Bibliothèque du Congrès et tous les membres du PCC (<http://www.loc.gov/aba/pcc/>) créent des notices bibliographiques et des notices d'autorité. Les membres du Name Authority

Cooperative Program (NACO) (<http://www.loc.gov/aba/pcc/naco/index.html>) ont été formés pour fournir des notices d'autorité conformes à RDA, et de la documentation sur les nouveaux champs RDA est à leur disposition sur le site web du NACO. De nombreux canalisateurs (funnels) de NACO, y compris ceux qui se consacrent au domaine des beaux-arts comme Art NACO funnel, NACO-AV, et NACO-Music sont très actifs. Bien que NACO soit un programme ouvert à tous, seules les bibliothèques disposant du budget, du personnel et du temps nécessaire peuvent envisager de former les meilleurs et les plus brillants de leurs bibliothécaires ; sinon les bibliothécaires doivent d'eux-mêmes entreprendre de suivre la formation NACO et rejoindre un canalisateur s'ils souhaitent participer au programme. Dans les deux cas, une fois formés, les catalogueurs doivent disposer du temps nécessaire pour créer les nouvelles notices et mettre à jour celles qui existent. S'ils créent de nouvelles notices d'autorité, ces catalogueurs sont intégrés dans un circuit de catalogage de documents que la Bibliothèque du Congrès ne possède pas ; les documents qu'ils cataloguent sont uniques, sinon on aurait recouru au catalogage par copie. Les catalogueurs formés à NACO constituent donc un groupe quelque peu fermé travaillant pour la plupart dans des bibliothèques universitaires qui les ont encouragés à participer à ce prestigieux programme national.

Qualité des métadonnées : complétude

La complétude est un des indicateurs permettant de mesurer la qualité des métadonnées. (Zeng & Qin, 2009). La complétude couvre l'« éventail de détails descriptifs fournis dans une notice donnée et aussi la répartition générale des différents niveaux de notices dans une base ou un entrepôt de métadonnées » (Zeng & Qin, 2009, p. 254). Cependant, tous les champs d'une notice donnée sont-ils remplis ? Une seule notice ayant des données complètes est de peu d'utilité en terme de retour si les autres notices de la collection ont des données incomplètes. Tous les champs de toutes les notices sont-ils complètement remplis pour les collections d'une bibliothèque ou d'un consortium de bibliothèques ?

La complétude des notices d'une collection est un but louable. On peut tout à fait considérer que les notices d'autorité complètes et de grande qualité d'une seule institution, une fois combinées à des données de qualité inférieure, influencent de façon négative, au moins temporairement la qualité de l'ensemble des notices. On a considéré l'adage « l'avenir est plus long que le passé » comme une valeur de base du catalogage (Miller, 2007) et il a permis aux agences de catalogage de s'engager dans RDA, en particulier au début. Dans l'idéal, toutes les notices d'autorité, grâce à l'action commune des bibliothèques du PCC, seront à terme mises à jour et complétées. Actuellement, cependant, on ne sait pas exactement combien de notices ont été mises à jour et ce qu'on peut en déduire sur le niveau général de qualité.

Les notices d'autorité Nom de personne

On a toujours préféré que les notices d'autorité ne concernent qu'une seule personne (principe des points d'accès distincts). Parfois, cependant, les caractéristiques d'une personne ne sont pas suffisamment claires pour permettre au catalogueur d'enregistrer les informations la concernant sous une notice distincte à cause des similitudes dans les chaînes de caractères des noms. Quand plusieurs individus sont désignés par la même chaîne de caractères, on crée une seule notice d'autorité dite *indifférenciée*. RDA définit les indicateurs de nom indifférencié comme « une catégorisation indiquant que les éléments fondamentaux enregistrés sont insuffisants pour distinguer deux personnes ou plus portant le même nom » (RDA Toolkit, 2014, sect. Indicateur de nom indifférencié). En 2013, un livre blanc signé Reimer et Schreur (2013) recommandait de dissocier les notices d'autorité des noms indifférenciés. Les technologies du WEB sémantiques, y compris le Fichier d'autorité international virtuel (VIAF), nécessitent des notices d'autorité distinctes et adressables. Même si les chaînes de caractères sont identiques, les notices d'autorité correspondantes doivent être uniques pour chaque personne.

MARC comme système d'encodage

Les systèmes intégrés de gestion de bibliothèque (SIGB), contexte dans lequel on trouve les données des bibliothèques, sont ordinairement adaptés pour utiliser le système d'encodage MARC. Étant donné l'intérêt des bibliothèques pour l'environnement du web sémantique et du web des données, la Bibliothèque du Congrès recherche les moyens de faire progresser l'encodage utilisé par les bibliothèques de façon à mettre en évidence les relations. On présente généralement l'initiative BIBFRAME (<http://www.loc.gov/marc/transition/pdf/marclid-report-11-21-2012.pdf>), (Kroeger, 2013) comme le principal prochain système d'encodage des notices catalographiques. Quoi qu'il en soit, tant que ce système d'encodage ne sera pas parvenu à maturité, il faudra continuer à créer des notices d'autorité en MARC.

Les identifiants Nom de personne

Les identifiants Nom de personne dans les notices des bibliothèques contiennent des chaînes de caractères représentant le nom de famille de la personne, le prénom, et d'autres informations servant à distinguer cette chaîne de caractères des autres, à présent ou plus tard. Le tableau 2 montre les sous-champs les plus répandus dans les points d'accès se rapportant aux personnes. Les éléments montrés à l'utilisateur final comprennent des titres, des dates, des termes ou des codes de relation, des qualificatifs contenant la forme complète d'un nom. Les numéros de contrôle et les liens sont potentiellement utiles aux systèmes mais ne sont généralement pas affichés.

Tableau 2 : Sous-champs des points d'accès noms de personne des notices bibliographiques
(100 et 700)
(R) = répétable (NR) = non répétable

Code de sous-champ	Description
‡a	Nom de personne (NR)*
‡c	Titres et autres termes associés à un nom (R)**
‡d	Dates associées à un nom (NR)**
‡e	Terme de relation (R)***
‡q	Forme complète du nom (NR)**
‡0	Numéro de contrôle de la notice d'autorité (R)***
‡4	Code de relation (R)***
‡6	Lien (NR)***

Légende : Obligatoire*, Obligatoire si applicable**, Optionnel***

Les points d'accès utilisés dans les bibliothèques ont fait l'objet de critiques à cause des données supplémentaires qu'ils contiennent. Dans quelle mesure est-il pertinent, pour les utilisateurs, de voir les dates de naissance et de mort qui sont des moyens de distinguer les auteurs ? Des libraires comme Amazon (<http://amazon.com>) n'incluent pas, dans le point d'accès du nom, de termes servant à distinguer des auteurs, par exemple. « Le contrôle d'autorité traditionnel s'attache à exprimer l'identité au moyen de points d'accès distincts [...]. Le point d'accès unique au nom d'une personne garantissait que les notices bibliographiques où cette personne figurait comme créateur, contributeur ou sujet seraient rassemblées sous un seul point d'accès » (« Report for PCC, 2013, p. 13).

Avec ces données superflues dans la chaîne de caractères du point d'accès, la mise en correspondance (mapping) des vocabulaires contrôlés de bibliothèques avec d'autres systèmes est plus difficile. BDpedia (<http://dbpedia.org>About>), la version en données liées de Wikipédia, n'utilise pas les dates pour distinguer les personnes. Des systèmes comme l'Internet Archives Open Library

(<https://openlibrary.org>), qui récupère des données venant de bibliothèques tout comme des jeux de données liées, peut éprouver des difficultés pour rapprocher automatiquement des notices. Ceci n'est pas étonnant : au départ le monde des bibliothèques a rencontré une série de difficultés en rapprochant dans VIAF, le Fichier d'autorité international virtuel, des notices d'autorité provenant de différents systèmes internationaux (Bennett, Hengel-Dittrich, O'Neill, & Tillett, 2006).

Les identifiants Nom de personne dans WorldCat : analyse

Les personnes qui sont des créateurs (auteurs, acteurs etc.) sont représentées par des points d'accès dans les notices bibliographiques en MARC dans les champs 100 et 700. Les points d'accès dans les notices bibliographiques comptent un ou plusieurs sous-champs. On peut analyser les tendances de la complexité des points d'accès en se basant sur les champs utilisés dans les notices bibliographiques. Le projet *Utilisation de MARC dans WorldCat (Marc usage in WorldCat)* (<http://oclc.org/research/activities/marcusage.html>), publié par OCLC Research, met à disposition les données sur le contenu de WorldCat. Grâce aux données sur l'utilisation des champs 100 et 700, on peut évaluer l'utilisation des sous-champs dans les notices Nom de personne des créateurs dans WorldCat. Tout au long de la période de 13 mois retenue pour l'étude, on a continué d'ajouter des dates, des formes complètes du nom (i.e. qualificatifs) et des titres, mais pas plus rapidement. On a ajouté moins souvent des qualificatifs aux points d'accès contrôlés Nom de personne, car leur pourcentage d'utilisation dans les points d'accès diminue au long de l'année. Voir le tableau 3 pour la présentation sous forme de tableau des données d'OCLC Research avant et après l'adoption de RDA. Les notices d'autorité évoluent, comme nous l'avons montré au paragraphe précédent, mais également les formes utilisées pour créer les points d'accès.

Tableau 3 : complexité des identifiants Auteur dans WorldCat d'OCLC
(i.e. 100 et 700 du MARC bibliographique)

‡	Description ¹	janvier 2013 occurrences	janvier 2013 % ²	janvier 2014 occurrences	janvier 2014 % ²	Evolution du % janvier 2013- janvier 2014
a	Nom de personne (NR)	291.478.411	100,00%	311.607.120	100,00%	0,00%
d	Dates associées à un nom (NR)	85.309.924	29,27%	91.269.104	29,29%	+0,02%
4	Code de relation (R)	43.604.446	14,96%	48.798.542	15,66%	+0,70%
0	Numéro de contrôle de la notice d'autorité (R)	25.855.327	8,87%	29.363.041	9,42%	+0,55%
e	Terme de relation (R)	19.551.901	6,71%	23.745.659	7,62%	+0,91%
6	Lien (NR)	15.406.166	5,29%	18.011.782	5,78%	+0,49%
q	Forme complète du nom (NR)	10.308.492	3,54%	10.837.508	3,48%	-0,06%
c	Titres et autres termes associés à un nom (R)	8.785.193	3,01%	9.509.560	3,05%	+0,04%

¹ Description provenant de <http://www.loc.gov/marc/bibliographic/bd100.html>

² A chaque fois, on a considéré que 100% des identifiants Nom de personne contenaient le sous-champ obligatoire non répétable a.

L'étude de cas

On ne trouve pas, dans la littérature, d'étude mesurant le volume des attributs fournis dans les notices d'autorité pour les personnes ni le degré de complétude des ajouts quels qu'ils soient. Sans métadonnées complètes, on peut douter de notre capacité à constituer un jour des jeux complets de données basés sur les attributs.

La présente étude, fondée sur une étude de cas longitudinale des notices d'autorité d'un groupe de bibliothèques universitaires américaines, explore et analyse la complétude de l'information contenue dans les attributs des notices d'autorité Nom de personne. De plus, elle examine les possibilités futures étant donné la situation actuelle.

Méthode

La présente étude de cas examine les attributs d'un jeu de notices d'autorité d'un consortium de bibliothèques universitaires situé dans l'État du Missouri aux États-Unis (voir annexe 1). Cette étude examine, en utilisant une approche longitudinale, les attributs des notices d'autorité six mois (i.e. septembre 2013) puis un an (i.e. avril 2014) après l'adoption officielle de RDA, pour évaluer les tendances dans l'utilisation des champs d'attributs.

Le 31 septembre 2013, six mois après l'adoption officielle de RDA par la Bibliothèque du Congrès, les champs des attributs des notices d'autorité Nom de personne (1 156 315 notices) étaient extraits avec des champs délimités en quatre fichiers texte. On a pu analyser plus finement les trois premiers fichiers (contenant respectivement 300 000, 300 000 et 299 999 notices) en utilisant Microsoft Excel. Le quatrième fichier, qui contenait 256 317 notices, était lisible mais corrompu et ne pouvait être exploité sous Excel. Pour s'assurer que le quatrième fichier ne différait pas statistiquement des trois premiers, 21 058 notices (8,2%) en ont été extraites de manière aléatoire et examinées manuellement afin de vérifier que ces données étaient représentatives de celles des trois autres fichiers. Comme il n'y avait pas de différences statistiquement significatives, on a considéré que l'analyse des fichiers de données utilisables était représentative.

Le 7 avril 2014, au bout d'un an de mise en œuvre de RDA, on a à nouveau extrait le même jeu de champs d'attributs de toutes les notices d'autorité Nom de personne du Merlin Cluster (1 274 618 notices). Les notices, extraites sous la forme de 13 fichiers texte, ont toutes été analysées sous Excel. Les résultats ci-dessous sont indiqués, selon les cas, pour un champ ou pour une notice, les champs répétés dans la même notice ou les contenus répétés dans un champ donné n'ont été comptés qu'une fois pour cette notice ou ce champ. De plus, cette étude a appliqué la méthode d'analyse « prenez ce que vous voyez » (« take what you see ») : tout contenu erroné ou saisi dans un mauvais champ est néanmoins compté comme saisi. Étant donné la taille du corpus, si 30 des 1 274 618 notices ont un champ 375 Sexe erroné, le contenu a été néanmoins compté comme une saisie dans le champ 375.

Résultats

Complétude des attributs au niveau du Consortium

Les dates enregistrées dans le 046 des notices d'autorité sont l'attribut le plus fréquent contenu dans les notices d'autorité, avec 7,1% des notices en contenant après six mois de catalogage en RDA et presque 10% des notices d'autorité en contenant après un an de catalogage en RDA. Après le 046, les données apparaissant dans le 678 sont les plus fréquentes, le 678 est un champ qui était utilisé dans le passé et qui est à nouveau utilisé. Bien que la tendance générale montre pour cette donnée une légère réduction de la quantité de 678, cette tendance peut être due à des incohérences dans le jeu de données étudié en septembre 2013. Il faudrait une analyse plus approfondie pour évaluer ce champ, à cause de son usage antérieur dans le catalogue. Tous les nouveaux champs RDA 3XX ont vu leur usage augmenter au cours de la période étudiée. Voir les résultats tableau 4.

Tableau 4 : utilisation des champs dans les notices d'autorité nom de personne de MERLIN

Champ MARC	Description de l'élément	6 mois : % de l'échantillon (N=899,999)	1 an : % du total (N=1,274,618)	Évolution
046	Dates spéciales codées (R)	7.10%	9.62%	+2.52%
678	Données biographiques ou historiques	2.84%	2.31%	-0.53%
375	Sexe (R)	1.86%	3.39%	+1.51%
374	Profession (R)	1.78%	3.12%	+1.34%
370	Lieu associé (R)	1.49%	2.74%	+1.25%
377	Langue associée (R)	1.33%	2.44%	+1.11%
378	Forme complète du nom (R)	1.07%	1.45%	+0.38%
372	Domaine d'activité (R)	0.90%	1.69%	+0.79%
373	Groupe associé (R)	0.89%	1.70%	+0.81%

Complétude des attributs au niveau de la notice

Dans les notices d'autorité, les attributs sont indiqués dans le champ MARC correspondant. Pour mesurer la complétude des notices, chaque fois qu'une donnée était saisie dans un champ, ce champ était décompté, on n'a pas tenu compte des champs répétés. Au bout de six mois, la grande majorité (90%) des notices n'avait pas d'attributs, 7% seulement avaient un attribut de plus dans un champ MARC, et 3% seulement avaient plus de deux attributs. Au bout d'un an, il y avait en général plus d'attributs dans les notices d'autorité. Le pourcentage de notices sans attributs avait diminué de 2,43 pour cent ; seulement 88 pour cent (1116259 notices) des notices n'avaient pas d'attributs. Presque 8 pour cent des notices d'autorité avaient au moins un attribut au bout d'un an, et presque 5 pour cent en avaient deux ou plus. Voir tableau 5 la ventilation complète du nombre des attributs dans les notices.

Tableau 5: Nombre de champs dans les notices d'autorité (par attribut)

Nombre d'attributs par champ MARC	6 mois : % du total	1 an : % du total	Évolution en pourcentage
0	90,01%	87,58%	-2,43%
1	7,05%	7,63%	+0,58%
2	0,89%	1,15%	+0,26%
3	0,28%	0,53%	+0,25%
4	0,49%	0,82%	+0,33%
5	0,50%	0,88%	+0,38%
6	0,45%	0,80%	+0,35%
7	0,26%	0,49%	+0,23%
8	0,08%	0,13%	+0,05%
9	0,01%	0,01%	+0%

Étant donné la fréquence du 046 (9,62% = 122676 notices) dans les notices d'autorité examinées dans cette étude, et étant donné que 9,31% (118696 notices) des notices étudiées contiennent entre un et trois attributs, il est raisonnable penser que si une notice contient un attribut, cet attribut est dans la grande majorité des cas un champ 046. Ce qui n'est pas surprenant, car recopier le contenu du \$d de l'identifiant dans le champ 046 de l'autorité doit être une procédure simple pour les catalogueurs si

cette donnée est déjà présente, et c'était le cas, comme nous l'avons indiqué, dans presque 30% des points d'accès de WorldCat.

Attributs démographiques des personnes

Bien que l'analyse complète du contenu des attributs dans les notices MARC ne soit pas le but de cette étude, un petit nombre de statistiques descriptives fournissent un point de départ pour une réflexion future. Dans le second, et par conséquent plus complet, jeu de notices extrait en avril 2014, on a comptabilisé 375 attributs « Sexe » et 377 attributs « Langue ». Les hommes (80% = 34515 notices) étaient massivement représentés dans les métadonnées fournies. Selon RDA 9.14.1.1, la langue « de la personne est une langue qu'une personne utilise lorsqu'elle écrit en vue d'une publication, d'une diffusion, etc. » (RDA Toolkit, 2014). Les auteurs écrivant en anglais (73% = 22666 notices), et éventuellement dans d'autres langues, étaient aussi très représentés. Le tableau 7 montre un échantillon des langues indiquées dans les champs 377. Puisque ce champ est répétable, on a pu indiquer plus d'une langue dans une seule notice d'autorité.

Tableau 6: attributs "sexe" enregistrés dans les notices d'autorité de MERLIN

375 entrées Sexe	Nombre de notices contenant le texte (N=43077)	Dans les notices avec 375 % contenant le texte
Masculin	34515	80,12%
Féminin	8544	19,83%
Donnant à la fois masculin et féminin	(-12)	(-0,03%)
Erreurs présumées de catalogage	30	0,07%
Total	43077	100,00%

Tableau 7 : attributs « langue » enregistrés dans les notices d'autorité de MERLIN

377 entrées Langue	Nombre de notices contenant le texte (N=31149)	Dans les notices avec 377 % contenant le texte
eng	22666	72,77%
fre	2066	6,63%
ger	1961	6,30%
spa	1505	4,83%
rus	1032	3,31%
lat	610	1,96%
chi	420	1,35%
jpn	269	0,86%
hin	117	0,38%
swe	92	0,30%
ice	11	0,04%

Il existe des limites à ce type d'analyse des attributs dans les notices : le fait qu'on ne puisse pas connaître le degré de représentativité de cet échantillon de personnes appartenant au fichier d'autorité

de ce consortium de bibliothèques universitaires n'est pas l'une des moindres. Les hommes écrivant en anglais sont-ils tellement favorisés dans les collections universitaires aux États-Unis, à la fois en tant qu'auteurs et sujets ? Si oui, y-a-t-il lieu d'ouvrir le débat ? Si on laisse de côté la question de la représentativité, les métadonnées fournies intéressent un groupe de personnes remarquablement peu représentatives de la population universitaire et savante au fil du temps.

Au cours de l'enquête sur les données démographiques relatives aux personnes, on a aussi examiné les informations restreintes concernant les occupations. Les occupations se présentent généralement sous forme de chaînes de caractères basées sur les vedettes-matière de la Bibliothèque du Congrès et sont sujettes à toutes les incohérences des champs en texte libre, telles que des erreurs de saisie et des fautes de frappe. On a constitué un échantillon de 8 occupations assez diverses à partir du champ 374 Occupation (=39772) pour en dégager des constantes. Environ 15 pour cent (= 6517) des entrées 374 Occupation comprennent le terme *auteur* ou *auteurs*. L'activité des catalogueurs de documents non-textuels parmi les membres de NACO est mise en évidence par la création de champs 374 relatifs aux arts du spectacle tels que *Acteur*, *Pianiste*, et *Musicien*. Voir tableau 8 pour plus d'information.

Tableau 8: Échantillon de « Professions » enregistrée dans les notices d'autorité de MERLIN

374 entrées Profession	Nombre de notices contenant le texte (N=39772)	Dans les notices avec champ(s) 374 % contenant ce texte
Auteur*	6517	16,39%
Professeur*	2885	7,25%
Acteur*	1962	4,93%
Poète*	1826	4,59%
Historien*	1724	4,33%
Juriste*	1320	3,32%
Pianiste*	1130	2,84%
Musicien*	1109	2,79%

On a aussi examiné avec succès un nombre limité de Lieux associés (MARC champ 373). Les Lieux associés contiennent des informations en texte libre concernant des institutions qui sont ou qui ont pour base des points d'accès Collectivité. Parmi les champs 373 contenant des données (21 620 notices), plus de la moitié contiennent le terme « University » (11 040 notices) et plus de 10% contiennent le terme « College » (25 17 notices) ce qui implique que l'information fournie dans le 373 concerne principalement des personnes instruites ; si le 373 est significatif, ces personnes ne sont pas toutes des enseignants mais peuvent avoir été étudiants. Voir tableau 9 pour la représentation de ces données.

Tableau 9: Échantillon de « Lieux associés » enregistré dans les notices d'autorité de MERLIN

373 entrées Lieux associés	Nombre de notices contenant du texte (N=21620)	Dans les notices avec champ 373 % contenant ce texte
University	11040	51,06%
College	2517	11,64%

Synthèse

L'ajout des données personnelles offre un grand nombre de possibilités à l'utilisateur final. Cependant, pour que les systèmes répondent aux besoins des utilisateurs, il faudrait que d'autres personnes que les auteurs masculins anglophones et instruits soient prises en considération. Si l'on considère la

complétude comme un indicateur de qualité des données, alors les données concernant les personnes ne sont pas de grande qualité, ni par la complétude des données dans les notices individuelles que l'on a étudiées, ni par la complétude des données des champs sélectionnés dans cet ensemble de notices.

Parmi les contraintes pouvant empêcher les données personnelles d'entrer dans les notices d'autorité partagées, on trouve des problèmes liés aux collections et aux travaux menés par des bibliothèques particulières. Les catalogueurs autorisés par NACO à créer des notices d'autorité pour les personnes ne parviennent pas à suivre le processus de création de données manuelle demandé par RDA. En se basant sur les résultats présentés ci-dessus, on peut en déduire que les catalogueurs ajoutent des informations codées pour des hommes ayant des relations avec l'éducation supérieure, éventuellement avec leur propre université, là où ils travaillent. Les catalogueurs peuvent aussi fournir du contenu additionnel pour des personnes faisant partie du programme de travail dans le domaine des beaux-arts. Cet apport sporadique de métadonnées ne fournit aux systèmes de bibliothèques que des fragments d'information sur une toute petite partie des personnes de l'univers bibliographique. Dans les systèmes de bibliothèques du futur, les notices d'autorité ne serviront pas seulement à distinguer des individus, mais permettront aussi des regroupements par attributs, si et seulement si ces attributs sont renseignés. La richesse et la complexité de l'information sur les personnes pourrait être une aide à l'établissement des relations et des liens dans les futurs systèmes, tels que ceux basés sur BIBFRAME, et cette information doit être complète pour donner tout son potentiel.

L'ajout d'attributs dans les notices d'autorité rendra moins nécessaire d'ajouter du texte au nom de personne dans l'identifiant Nom de personne, en accord avec les suggestions du Groupe sur la création et la fonction des autorités Nom dans un contexte non-MARC du PCC (Report for PCC, 2013). Les identifiants complexes remarqués dans WorldCat pourraient être restructurés grâce à l'ajout systématique d'attributs à la notice d'autorité. Presque 30% (29,29% = 91 269 104 notices) des identifiants Nom de personne dans WorldCat, comportent des dates, on pourrait toutes les transférer ou les copier dans la zone attributs de la notice d'autorité. Bien que l'étude dont il est question ici n'examine pas les notices d'autorité stockées dans WorldCat, cette recommandation pourrait être d'abord réalisée, en tant que projet pilote, sur un plus petit corpus. La même méthode de fourniture d'attributs se vérifie pour les formes complètes du nom, généralement enregistrées comme qualificatifs (dans 3,51% des identifiants = 10 837 508 notices) et pour les titres et autres mots associés à un nom (dans 3,06% des identifiants = 9 509 560 notices)

Une des limites de cette l'étude est que les entrées Nom indifférencié ne peuvent pas contenir de champs spécifiques aux personnes puisque, par définition, les personnes représentées par l'identifiant ne sont pas connues. Il n'a pas été possible de déterminer de façon certaine le nombre d'entrées Nom indifférencié dans les données extraites pour cette étude. Un jour viendra où il sera nécessaire de distinguer ces entrées, en scindant les notices indifférenciées pour créer autant de notices d'autorité que nécessaire. Cependant, il n'y aura pas forcément enregistrement des données relatives aux personnes dans les notices d'autorité correspondant à ces identifiants. On pourrait peut-être montrer, dans une prochaine étude, quand et comment les notices d'autorité sont alimentées en données d'attributs de personnes, en particulier s'il s'agit d'un Nom indifférencié. Si l'auteur n'est ni prolifique ni célèbre, il ne semble guère concevable qu'on puisse ajouter beaucoup de contenu à sa notice d'autorité, excluant de fait ses œuvres des recherches effectuées à partir des attributs des personnes, à moins de mettre en place un plan d'action.

Prochains travaux

Pour que le catalogue puisse fournir un accès impartial basé sur les attributs, il faut que les attributs soient répartis de manière égale dans toutes les notices. D'autres personnes que les auteurs masculins anglophones et instruits devront être pris en considération. A l'avenir, le système de bibliothèque idéal possédera des jeux complets de métadonnées, afin d'apporter des réponses homogènes à toutes les opérations effectuées par l'utilisateur telles qu'identifiées dans FRAD (2013) : trouver, identifier, contextualiser et justifier. A cet effet, il faut que les bibliothèques trouvent le moyen d'exploiter les données librement disponibles sur le WEB sémantique. Qu'un groupe de bibliothécaires hautement

qualifiés fasse de la saisie ponctuelle de contenu ne peut pas être et ne sera pas suffisant. Comme le fait remarquer le PCC : « intégrer des données externes à ses propres données, plutôt qu'exposer ses propres données, est ce qui permettra, si toutes les parties intéressées s'y engagent, une navigation plus fluide et une exploitation de la richesse de l'environnement du web des données pour améliorer la recherche » (PCC Day One, 2013, p. 10). Les données d'autorité sont uniques et précieuses, mais face aux exigences actuelles, les bibliothécaires doivent se résoudre à abandonner le contrôle complet et entier du contenu des attributs des notices d'autorité et à travailler sur du contenu provenant de ressources structurées telles que Wikipédia via DBPédia.

Le point de vue présenté ici n'est pas incompatible avec celui de BIBFRAME, où, selon le PCC « les autorités ne sont pas destinées à rivaliser avec les projets en cours ou à les remplacer, mais plutôt à fournir une simple couche d'abstraction, légère, qui couvrirait divers projets relatifs aux autorités sur le web pour les rendre encore plus efficaces (« PCC Day One », 2013, p. 13). Le projet ISNI a des buts similaires (MacEwan, Angjeli, & Gatenby, 2014). Les pages Auteurs de Wikipédia montrent un exemple similaire. Sous l'entrée de l'auteur Chris Anderson http://en.wikipedia.org/wiki/Chris_Anderson_%28writer%29), une section comprend des notices d'autorité identifiant cet auteur provenant de systèmes de bibliothèque. On y voit des liens, lisibles par les humains (et les machines), à sa notice d'autorité dans WorldCat, VIAF, LC (NAF), ISNI, GND, SUDOC, BNF et NDL.

Il est recevable que toutes les personnes associées au contenu des collections des bibliothèques ne soient pas représentées sur le web sémantique ; les catalogueurs devront accorder une attention particulière à celles qui n'y sont pas, mais ce travail devra être mené dans le cadre d'un plan d'action. Dans l'idéal, la somme de travail effectuée par les catalogueurs devrait être inversement proportionnelle à la somme de données qui existe déjà sur le web sémantique.

En outre, les attributs contenus dans les notices d'autorité devront être fournis sous une forme lisible par les machines. Hétérogènes par nature, les informations saisies en texte libre sont un obstacle à la recherche et la récupération systématiques. Les identifiants que le web sémantique autorise et les possibilités qu'ils offrent devraient être étudiés en tant que solution possible aux problèmes et aux aberrations des informations en texte libre, qui sont de simples chaînes de caractères, et non des *données*.

Conclusion

Cette étude a révélé qu'au bout d'un an de catalogage selon RDA, les attributs les plus communément saisis dans les notices d'autorité sont ceux qui sont le plus facilement accessibles, tels que les dates. Les autres attributs renseignés sont rares et concentrés sur les personnes anglophones, de sexe masculin et celles ayant un lien, quel qu'il soit, avec les universités. Bien que le but de cette enquête soit l'analyse de la qualité des métadonnées à travers le degré de complétude des attributs dans les notices d'autorité après l'adoption de RDA, on a pu faire quelques observations informelles au cours de l'enquête portant sur des points relatifs à la démographie des personnes du fichier d'autorité. Il semble que les bibliothécaires ne créent pas nécessairement de notices pour constituer un corpus de personnes sur lequel on pourrait faire des recherches. Il semble aussi, en toute logique, que les bibliothécaires des grands établissements d'enseignement disposant de ressources financières, créent la majorité des notices, ce qui peut biaiser les catégories de personnes décrites. Ces observations mériteraient une étude plus approfondie au cours de recherches ultérieures.

RDA ouvre la voie vers des systèmes de bibliothèque plus riches et plus performants, car on y trouvera, en plus des ressources, les personnes. Les évolutions nouvelles nécessaires à l'intégration des attributs concernant les auteurs dans les métadonnées ne seront pas sans conséquences. Parce que RDA fait évoluer les choses dans le bon sens, les bibliothécaires devront inventer les moyens de mener à bien ces nouvelles tâches en conciliant précision et volume de données. Il faut explorer les possibilités que laisse entrevoir le web sémantique, cependant que sont repensés systèmes et pratiques.

Annexe I

Consortium de bibliothèques MERLIN

- Missouri University of Science and Technology
- University of Missouri (MU)
- MU Law
- University of Missouri - Kansas City
- UMKC Law
- University of Missouri - St. Louis
- University of Missouri Archives and WHMC

Remerciements

Une profonde reconnaissance à Felicity Dykas, Head of Digital Services at MU's Ellis Library. L'auteur aimerait aussi remercier Roy Tennant et l'équipe d'OCLC Research pour avoir rendu possible la réutilisation de leurs données MARC.

Références bibliographiques

Bennett, R., Hengel-Dittrich, C., O'Neill, E., & Tillett, B. (2006). VIAF (Virtual International Authority File): Linking Die Deutsche Bibliothek and Library of Congress name authority files. In *World Library and Information Congress: 72nd IFLA General Conference and Council*. Retrieved from:

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.5328&rep=rep1&type=pdf>

En français : http://archive.ifla.org/IV/ifla72/papers/123-Bennett_trans-fr.pdf

Functional Requirements for Bibliographic Data (FRAD). (2008, 2013 rev.). Retrieved from

http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf

En français : http://www.bnf.fr/documents/frad_rapport_final.pdf

Functional Requirements for Bibliographic Records (FRBR): Final report. (1997, 2009 rev.). International Federation of Library Associations and Institutions. Retrieved from

http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf

En français : http://www.bnf.fr/documents/frbr_rapport_final.pdf

Kroeger, A. (2013). The road to BIBFRAME. *Cataloging & Classification Quarterly*, 51(8), 873-890.

MacEwan, A. Angjeli, A., & Gatenby, J. (2014). The International Standard Name Identifier (ISNI): The evolving future of name authority control. *Cataloging & Classification Quarterly*, 51(1-3): 55-71.

MARC 21 encoding to accommodate new RDA elements 046 and 3XX in NARs and SARs. (2012, October).

Retrieved from http://loc.gov/aba/pcc/rda/PCC%20RDA%20guidelines/RDA%20in%20NARs-SARs_PCC.pdf

MARC 21 format for authority data. (1999, rev. September 2013). Library of Congress Network Development and MARC Standards Office. Update No. 1 (October 2001) through Update No. 17 (September 2013). Retrieved from <http://www.loc.gov/marc/authority/ecadhome.html>

- MARC Usage in WorldCat*. (2014). OCLC Research. Retrieved October 2013 and April 2014 from <http://experimental.worldcat.org/marcusage/>
- Miller, D. (2007). Basic values and the future of cataloging. *ALCTS*. Retrieved from <http://www.ala.org/alcts/ianda/bibcontrol/07futcat>
- PCC Day One for RDA Authority Records. (2011, Dec. 20; updated 2012, Mar. 23). Retrieved from <http://www.loc.gov/aba/pcc/rda/PCC%20Day%20One/PCC%20Day%20One%20for%20RDA%20Authority%20Records.docx>.
- RDA Cataloger Training (2013, Feb. 18). *Indiana University Libraries*. Retrieved from <http://www.indiana.edu/~libslav/RemoteRDATraining/IULtraining2-authorityrecords.doc>
- RDA Toolkit: Resource description and access. (2014). American Library Association, Canadian Library Association, and CILIP: Chartered Institute of Library and Information Professionals. Retrieved from <http://www.rdatoolkit.org/>
- Report for PCC task group on the creation and function of name authorities in a non-MARC environment. (April 5, 2013). Retrieved from http://www.loc.gov/aba/pcc/rda/RDA%20Task%20groups%20and%20charges/ReportPCCTGonNameAuthInA-NonMARC_Environ_FinalReport.pdf
- Riemer, J. J. & Schreur, P. E. (2012, February 14). The future of undifferentiated personal name authority records and other implications for PCC authority work. End notes updated May 8, 2012. Retrieved from <http://www.loc.gov/aba/pcc/Undiff%20Personal%20NARs%20Discussion%20Paper%20March%202012.doc>
- Yee, M. M. (2005). FRBRization: A method for turning online public finding lists into online public catalogs. *Information Technology & Libraries*, 24(2), 77-95.
- Zeng, M. L., & Qin, J. (2009). *Metadata*. New York: Neal-Schuman.