

En réponse à trois questions centrales du dossier, des experts de l'Abes livrent ici leurs réflexions, issues de leur expérience quotidienne. François Mistral est responsable d'IdRef, Olivier Rousseaux est chef du service Métadonnées, Yann Nicolas, expert métadonnées, et Philippe Le Pape, chargé de mission Normalisation.

3 fois sur le métier remettons les notions

1. Quelle distinction faites-vous entre « référentiel » et « fichier d'autorité » ?

OLIVIER ROUSSEAU : De « référentiel », le Larousse donne la définition suivante : « ensemble auquel doivent appartenir les éléments, les solutions d'un problème posé ». Du fait qu'ils organisent et contraignent les métadonnées produites par les réseaux de signalement animés par l'Abes, les fichiers d'autorité appartiennent bien à un ensemble : fonctionnant comme un tout, avec des règles d'interdépendance et d'interopérabilité de ses constituants, ils permettent le contrôle de la cohérence des métadonnées bibliographiques.

PHILIPPE LE PAPE : Les deux termes s'appliquent à des données « de référence », statut consacré soit par une labellisation, soit par l'usage. On pourrait nommer « référentiel » un ensemble de « données outils » (modèles, règles de catalogage, formats, vocabulaires normalisés...) auquel on recourt pour garantir qualité et interopérabilité des métadonnées qu'on produit et/ou administre. Et on nommerait « données d'autorité » celles des métadonnées qu'on produit ou/et administre qui ont vocation à faire référence à leur tour. Dans le Sudoc, et plus encore

dans theses.fr, cette définition s'applique aussi (à mon sens) aux données « bibliographiques ».

FRANCOIS MISTRAL : De « référentiel », je retiens l'idée d'un jeu de données suffisamment fiables pour être utilisées en confiance et en produire ou en agréger d'autres. De « fichier d'autorité », je retiens la double dimension de contrôle et de légitimité à assurer ce contrôle, par une pratique fondamentale : la mention explicite des sources. Ce point établi, je voudrais surtout souligner leur horizon de convergence : la qualité. Les bibliothécaires produisent ainsi des données d'autorité qui ont toute légitimité à devenir des référentiels.

YANN NICOLAS : Les référentiels des uns sont les données centrales des autres : si, à l'Abes, par « référentiel », on entend « données permettant de décrire les documents en minimisant le recours à du texte libre », pour un autre système d'information, nos documents correspondent à des entités secondaires... On observe également la tendance à transformer les référentiels « listes fermées » en référentiels « entités identifiées » : on passe de « littéraux » (chaînes de caractères simples ou typées) à des « ressources » pouvant elles-mêmes comporter des attributs, être décrites. Connecter nos données à ces référentiels les enrichit, par transitivité, de beaucoup d'autres informations. Autre tendance observée : nos autorités traditionnelles se normalisent, rejoignent d'autres référentiels. Cette normalisation est à la fois intellectuelle et technique. Le paradigme « web sémantique » est le vecteur principal de cette normalisation : tout devient ressource, identifiée de manière univoque à l'échelle universelle grâce à sa (ou ses !) URI, et ce sont les triplets RDF qui en parlent le mieux...

2. Comment la montée en puissance du rôle des référentiels impacte-t-elle votre travail et les pratiques de catalogage des professionnels des réseaux ?

O. R. : L'Abes est confrontée à ces questions pour avoir posé comme principe, dès la mise en place du Sudoc, la réutilisation de référentiels existants : Rameau ou FMeSH pour les accès matière, fichiers d'autorité de la BnF pour les accès auteur ou titre uniforme ; codes ISO ou Unimarc. Aujourd'hui, la



réflexion porte sur la manière d'appréhender les métadonnées de référentiels tiers (Sirene, RNSR... au niveau national; VIAF, au niveau international). Pour le catalogage au quotidien, l'expansion des référentiels offre des opportunités d'améliorer les méthodes de liage automatique, ou de développer des outils d'aide à la décision.

P. LP : Dans le Sudoc, la normalisation (et l'emploi de référentiels) a toujours été cruciale, et avec les règles RDA, le recours à des vocabulaires contrôlés (type de contenu, de médiation, de support...) s'accroît encore.

F. M. : Ma démarche au poste de responsable IdRef se déroule selon trois axes : le réseau des catalogueurs (dont les Correspondants autorité); l'outillage visant à accroître leur maîtrise de la production des données d'autorité; et la dissémination multicanaux, multiformats, dans et hors les murs, de ces données. Il s'agit de rendre plus attractifs les données mais aussi les services associés, notamment les alignements¹ afin d'en faire un levier pour démarcher des partenaires et intégrer de nouveaux projets, l'un des apports « spécifiquement Abes » étant la mutualisation.

Y. N. : Je ne pense qu'à ça ! Ces dernières années, nous avons travaillé sur les projets Qualinca et Hub de métadonnées. Qualinca produit des algorithmes qui auscultent et améliorent la qualité des liens entre notices bibliographiques et notices d'autorité. Entre Sudoc et IdRef, par exemple. Côté Hub de métadonnées, nous récupérons des données en provenance des éditeurs, nous les homogénéisons et enrichissons grâce aux référentiels de toutes sortes (langues, auteurs, sujets, types de document...). Nous remplaçons (ou complétons) des mots par des identifiants : un nom d'auteur par un lien vers une URI IdRef, ISNI, Orcid, etc. Dans les deux cas, nous connectons l'information bibliographique à des référentiels, pour mieux la structurer et la connecter à l'échelle du web et pas seulement à un catalogue, fût-il national !

3. Comment imaginez-vous le rôle des référentiels dans le paysage de l'IST voire au-delà ?

O. R. : Je ne le vois pas évoluer radicalement dans l'immédiat. La tendance est aux rapprochements de référentiels existants. Mais un référentiel tiers est à aborder avec prudence, et chaque fois, les mêmes questions devront être examinées : à quels objectifs, quelles contraintes répond-il ? Sur quel modèle de données est-il fondé ? Quelles en sont les règles d'alimentation ? Couvre-t-il nos besoins en granularité des données et traçabilité des futures évolutions ? Quelle gouvernance technique et scientifique en résultera-t-il, et sera-t-elle adaptée à notre contexte ? Un exemple : le projet de FNE amorcé en mars 2017 entre la BnF et l'Abes répond à ces questions en recherchant une solution de fusion des

fichiers d'autorité existant de part et d'autre au profit d'un fichier national unique géré en coproduction.

P. LP : On va vers une importance grandissante des identifiants de confiance dans lesquels le « nom », la « forme d'autorité », les données elles-mêmes se trouvent de plus en plus ramassés : le passage des métadonnées bibliographiques de systèmes fermés au Web renforce la nécessité de les normaliser et de les étiqueter selon des systèmes d'identification qui, pour être efficaces, doivent jouir d'une large reconnaissance.

F. M. : Les données produites par les bibliothécaires sont promises à un grand avenir, l'enjeu de l'information se situant dans sa structuration. Avec un point délicat : les catalogueurs sont à la fois acteurs et spectateurs aux premières loges de ce phénomène. Un accompagnement est à inventer... On imaginerait volontiers les référentiels jouer un rôle structurant pour l'IST. Ils pourraient amener une reconfiguration des missions des opérateurs nationaux, façonnée à leur image : tout entière de spécialisation et de coopération, pour un service rendu à la communauté ESR (et au-delà) de haut niveau.

Y. N. : Vu de l'Abes, en caressant du regard le paysage un peu cacophonique de l'IST en France, j'espère une politique publique des référentiels claire. Que chacun joue sa partition, maintienne et mette à disposition les référentiels de son ressort. Qu'on évite les doublons ! Par exemple, que le RNSR soit, de droit et de fait, reconnu comme LE service public national qui fournit identifiants et attributs de référence pour les laboratoires français. Ce qui n'empêche pas des clients comme STAR et theses.fr de gérer leurs propres attributs complémentaires, en sus des attributs RNSR, à des fins propres, bibliographiques ou pas. En revanche, accrochons nos attributs aux identifiants RNSR ! Même chose pour les autorités de type Entreprise (avec Sirene, de l'INSEE). L'Abes doit être un bon client des référentiels des autres, et un bon fournisseur de référentiels : son positionnement, son organisation et son capital de données la rendent légitime. C'est déjà le cas pour le référentiel des thèses françaises ou celui des chercheurs français.

[1] L'alignement, auquel nous nous référons à de multiples reprises et qui est illustré dans les exemples du dossier, consiste à établir un lien, terme à terme, entre les entités de deux (ou plusieurs) bases, au moyen d'un identifiant unique et pérenne, de façon manuelle ou automatique.



POUR EN SAVOIR PLUS

Les réponses publiées ici ne sont qu'un bref résumé des réflexions de leurs auteurs. Retrouvez-les *in extenso* et accompagnées de nombreux liens sur <https://punktokomo.abes.fr/2017/04/20/autorites-vs-referentiels-3-questions-aux-experts-de-labes/>