

Vers des référentiels globalisés

Comment identifier de manière précise et durable Raymonde Moulin, le concept de « moulin » ou la ville de Moulins ? Dans notre monde à nous, celui de l'information bibliographique, on s'appuie sur nos propres référentiels, historiquement liés au Sudoc : fichier d'autorité de noms de personne, vocabulaire contrôlé Rameau ou MeSH, fichier d'autorité de noms géographiques, etc. Dans le périmètre de l'enseignement supérieur et de la recherche en France, on a souhaité donner à ces référentiels une autonomie fonctionnelle et technique, afin de permettre à d'autres applications que le Sudoc d'utiliser et d'alimenter IdRef, comme il faut désormais appeler ces « autorités Sudoc ».

Des référentiels plus ouverts

Le mouvement d'ouverture d'IdRef vers d'autres applications illustre à l'échelle nationale et universitaire un mouvement global d'ouverture des référentiels. Dans le monde des bibliothèques, il faut citer VIAF, fichier d'autorité international virtuel – virtuel au sens où il est l'agrégation et surtout l'interconnexion de plusieurs fichiers d'autorité

nationaux, dont celui de la Bibliothèque nationale de France et IdRef. Mais VIAF est lui-même agrégé et interconnecté dans une base plus large et moins spécifiquement bibliographique : ISNI (norme ISO ; *International Standard Name Identifier*). Cette tendance accompagne et enrichit la croissance du web et notamment du web de données, avide d'identifiants globaux pour référencer et décrire de manière univoque toutes sortes d'entités : personnes, organismes, concepts, lieux, époques, microbes, émotions, événements, langues, etc. Tout ce dont on peut parler, il faut l'identifier, au moyen d'un code (un numéro PPN dans IdRef ; une URL sur le web). Le nouveau contexte est global en deux sens : il dépasse les frontières locales ou nationales ; il dépasse les frontières professionnelles (le bibliothécaire a besoin de référencer sans ambiguïté la ville de Moulins, mais c'est aussi le cas de l'administra-

tion, des offices du tourisme, des épidémiologues, des biographes de Georges Bidault et même des indépendantistes bourbonnais). Très récemment, Google et d'autres moteurs de recherche majeurs ont investi la question des référentiels, en mettant sur la table schema.org¹. Jusqu'à présent, le web de données se construisait de manière relativement décentralisée, grâce à la rencontre de diverses initiatives, de vocabulaires de description multiples et de référentiels variés. Avec schema.org, Google et ses alliés entendent remettre de l'ordre dans cette auberge espagnole, en imposant un vocabulaire unique et en adoubant, voire en absorbant, certains référentiels – et pas d'autres².

Faire cohabiter des référentiels locaux et globaux

Avec schema.org, la mondialisation des métadonnées risque de prendre une tournure unilatérale, appauvrissant de ce fait la richesse des métadonnées et donc les métadonnées elles-mêmes... à moins qu'on en fasse un bon usage. Car il faut plutôt voir dans schema.org (et même VIAF) une *lingua franca*, outil pratique de communication qui ne doit pas se substituer aux langues vernaculaires plus expressives et enracinées dans des pratiques professionnelles voire culturelles – et non décrétées depuis Mountain View, CA. Par exemple, nous devons conserver la maîtrise d'un référentiel de personnes physiques et de collectivités constitué et entretenu à des fins bibliographiques et universitaires, même s'il faut se brancher sur le niveau global en liant IdRef à VIAF ou DBpedia (Wikipedia version web de données). Au contraire, nous avons pris l'habitude d'utiliser directement les normes ISO qui référencent les pays ou les langues – en tolérant au passage certains sacrifices délicats.

Il n'est pas toujours facile de choisir : s'il n'est pas question de gérer dans un référentiel propre à notre métier, comme Rameau ou même MeSH, la diversité et le volume d'information sur la drosophile que contient Flybase, on peut se poser la question pour identifier des lieux (noms géographiques), des marques commerciales, des genres littéraires, des concepts mathématiques, etc. Il s'agit là de faire des choix stratégiques de long terme, en prenant en considération des facteurs fonctionnels (structure et finesse de l'information, interopérabilité) et économiques (capacité à maintenir collectivement un référentiel, viabilité des organisations extérieures chargées de maintenir un référentiel).

Yann Nicolas

Chef du département Études et Projets de l'ABES

nicolas@abes.fr

¹ <http://schema.org>

² <http://www.w3.org/wiki/WebSchemas/ExternalEnumerations>



Le Vieux Moulin(s) à Moulins (03).
Phot. Thomas Durand sur Wikimedia
Commons (CC BY-SA 3.0)