

# Des référentiels et de leur usage aujourd'hui

**Le catalogage se transfigure : d'un monde où l'on normalisait des notices pour les échanger, on passe à un monde où ce que l'on expose, ce sont des données structurées pour les partager. Le classement hiérarchique cède alors le pas au contrôle des éléments de description d'une ressource.**

**E**n 1627, Gabriel Naudé, dans son *Advis pour dresser une bibliothèque*, comparait le travail de classement d'un bibliothécaire au rôle d'un médecin. Afin de pouvoir maîtriser la « quantité infinie » des publications, il appelle à établir un ordonnancement, à des fins de sélection.

Le classement, perçu comme un instrument de domination et de sélection dans un univers fini, sert à dire ce qui est digne ou non de figurer dans une bibliothèque : « Toutefois pour ne laisser cette quantité infinie ne la définissant point, et aussi pour ne jeter les curieux hors d'esperance de pouvoir accomplir et venir à bout de cette belle entreprise, il me semble qu'il est

avec l'essor du web, qui implique non seulement une croissance exponentielle du nombre de ressources, mais renforce encore le caractère composite et évolutif d'une ressource, qui n'est plus que la décomposition et la recombinaison à l'infini de données.

La notion de web de données (*linked data*) consacre l'explosion de l'idée de « ressource ». Il s'agit de faciliter l'exploitation de données structurées, c'est-à-dire de données qualifiées et encadrées, de sorte que des machines puissent les interpréter facilement. Le web de données consiste à exposer ces données structurées sur le web et les relier entre elles, ce qui permet d'accroître leur visibilité et leur réutilisation. En outre, le boom de la production de données, de l'ordre de 2,5 trillions d'octets chaque jour, provenant souvent, par le retraitement automatisé, de données d'origines diverses, implique l'essor de nouvelles technologies de stockage et de traitement, appelées Big Data (ou mégadonnées). La manière dont les données utilisées ou produites par les bibliothèques s'insèrent dans le Big Data est un enjeu désormais identifié par l'IFLA<sup>2</sup>.

## Peut-on encore « ordonner la quantité de drogues suivant la qualité d'icelles »?

*à propos de faire comme les Médecins, qui ordonnent la quantité de drogues suivant la qualité d'icelles, et de dire que l'on ne peut manquer de recueillir tous ceux qui auront les qualités et conditions requises pour être mis dans une Bibliothèque*<sup>1</sup>.

Dans cette logique, la notion de description et de classement est vécue de manière hiérarchique et verticale.

Gabriel Naudé, même s'il n'emploie pas le terme, conçoit le rôle du bibliothécaire comme un médiateur : un patient a besoin du médecin pour se soigner ; un utilisateur a besoin d'un bibliothécaire pour accéder à des ressources. Un certain nombre d'éléments structurants amènent à repenser de telles conceptions.

### EXPLOSION DE L'IDÉE DE « RESSOURCE »

Les pratiques et les usages ont d'ores et déjà évolué en remettant en cause l'idée d'un classement hiérarchique. L'utilisateur prend des concepts ou des notions et navigue entre elles pour accéder à des ressources. Les modes de recherche sur le web incitent à une telle évolution. La logique de graphe, composé d'entités reliées par des relations qualifiées, prend donc le pas sur celle de classification hiérarchisée.

La question de la « quantité infinie » évoquée par Gabriel Naudé est devenue encore plus critique

### DE LA NOTICE NORMALISÉE À LA DONNÉE STRUCTURÉE

Nous sommes donc en présence d'une contradiction apparente : comment concilier la nécessité de données structurées, soulignée par les réalisations du web de données, et l'essor exponentiel des données produites, impulsé par le Big Data ?

Un peu de dialectique permet de la dépasser : c'est grâce à la structuration renforcée des données que celles-ci pourront être traitées plus rapidement et plus efficacement. Si l'on applique ce principe aux données produites ou utilisées par les bibliothèques, cela conduit à passer d'un monde où l'on normalisait des notices pour pouvoir les échanger à un monde où l'on structure des données pour pouvoir les partager. Le concept de contrôle bibliographique universel, au cœur de l'écosystème international des données de bibliothèque, s'en trouve redéfini. Afin de prendre en compte ces différents éléments structurants, la France s'est lancée dans le programme de la « transition bibliographique »<sup>3</sup>, dont l'objectif est de se doter de règles et d'outils permettant de mettre en pratique ce changement de

[1] Sur Gallica, [gallica.bnf.fr/ark:/12148/bpt6k1041429f](http://gallica.bnf.fr/ark:/12148/bpt6k1041429f), p. 41-42.

[2] Le rapport de développement stratégique de l'IFLA de 2013 (<https://trends.ifla.org/literature-review/big-data>) a amené en 2016 la constitution d'un groupe d'intérêt spécial consacré au Big Data.

[3] [www.transition-bibliographique.fr](http://www.transition-bibliographique.fr)

[4] La version définitive de ce modèle conceptuel fusionnant FRBR-ER (FRBR OO n'est pas concerné), FRAD et FRSAD est attendue par l'automne 2017. Voir le projet diffusé pour enquête internationale ([www.ifla.org/files/assets/cataloguing/frbr-irm/frbr-irm\\_20160225.pdf](http://www.ifla.org/files/assets/cataloguing/frbr-irm/frbr-irm_20160225.pdf)) et une présentation du modèle par Pat Riva (congrès IFLA 2016) (<http://library.ifla.org/1084/7/207-riva-fr.pdf>)

[5] Gabriel Naudé, *op. cit.*, chapitre 9, « quel doit être le but principal de cette bibliothèque ».

paradigme. Concrètement, cela suppose d'appliquer le modèle de données FRBR, désormais sous sa forme dite « IFLA-LRM »<sup>4</sup>, d'en tirer un code et des règles de catalogage en s'inspirant de RDA, de former la communauté professionnelle aux enjeux et aux normes produites dans ce cadre, et de pousser les systèmes et les outils à s'adapter.

## PÉRIMÈTRE ÉLARGI

Pleinement concernés par la transition bibliographique, les référentiels recouvrent le périmètre des données d'autorité. Cependant, la notion de référentiel va plus loin, car elle annexe l'ensemble des éléments contrôlés d'une description de ressource (langues, types de document comme la nouvelle zone 0 de l'ISBD...).

Ainsi, la notion de données d'autorité, opposée à celle de données bibliographiques, n'est plus pertinente aujourd'hui. Nous avons désormais affaire à des entités dont les éléments de description, les données, sont contrôlées par des référentiels. Par exemple, pour décrire une entité « manifestation », nous avons besoin des entités « expression », « œuvres », « personne » accompagnées de liens qualifiés (« auteur », « sujet »). Il nous faut également une entité « date », une entité « langue »... qui ne font pas partie du périmètre traditionnel des données d'autorité.

Cela implique d'élargir et d'affiner le périmètre des données structurées. Les bibliothèques en possèdent, de qualité, bien structurées. Le format Marc étant un peu « le web de données de Monsieur Jourdain », il ne s'agit donc pas de crier à la « fin du catalogage », mais d'accompagner sa transfiguration. D'autres organismes publics en possèdent également : les institutions patrimoniales (archives, musées...), les éditeurs publics (DILA, éditeurs universitaires...). La mutualisation de ces données – qui permet tout à la fois de les confronter, donc de renforcer la confiance qu'elles inspirent à des tiers susceptibles de les récupérer, et de faire masse – est devenue une nécessité... et un atout majeur. En effet, la production mutualisée et la réutilisation des données publiques constituent les fondements d'un véritable service public de la donnée.

Différentes initiatives sont prises dans ce sens, à l'image de la perspective d'un fichier national d'entités (voir l'article « FNE » page 16), fusionnant dans un premier temps les données d'autorité de la BnF et de l'Abes et offrant à ses utilisateurs une base de coproduction mutualisée.

Gabriel Naudé disait également : « *car de s'imaginer qu'il faille après tant de peine et de despense cacher toutes ces lumières sous le boisseau, et condamner tant de braves esprits à un perpétuel silence et solitude, c'est mal reconnoître le but d'une bibliothèque* »<sup>5</sup>. Les « lumières » sont les données ; les « braves esprits » les catalogueurs qui les produisent. Dans le nouvel écosystème international des données,



© Numelyo, Bibliothèque municipale de Lyon, Domaine public, licence ouverte.



**Robert Marcaire Médecin, lithographie coloriée d'Honoré Daumier.**  
Du temps de Naudé et jusqu'à récemment, le lecteur avait besoin du bibliothécaire comme le malade du médecin. Changement de paradigme !

la généralisation et la mutualisation des référentiels, par leur inscription dans le web de données et dans le Big Data, est un outil pour remplir cet objectif.

**VINCENT BOULET**

*Chef du service des référentiels, BnF  
vincent.boulet@bnf.fr*



### POUR EN SAVOIR PLUS

- Définition et estimations quantitatives du Big Data :

[lebigdata.fr/definition-big-data](http://lebigdata.fr/definition-big-data)

- Sur le site de la Fédération internationale des associations et institutions de bibliothèques (IFLA) se trouvent de nombreux documents utiles pour comprendre le changement de paradigme.

La définition du web de données, avec des exemples concrets d'application dans le monde des bibliothèques, a été étudié au cours de la conférence « *Linked Data in Libraries: Let's make it happen!* » :

<http://ifla2014-satdata.bnf.fr/program.html>

Sur le concept de contrôle bibliographique universel, voir l'article de Françoise Bourdon et Gildas Illien

« *À la recherche du temps perdu, retour vers le futur : CBU 2.0* »  
<http://library.ifla.org/956/1/086-illien-fr.pdf>