

Data.bnf.fr : Entre outil de navigation et matériau brut ouvert à l'exploration

"Depuis novembre 2015, data.bnf.fr, site web exposant les données des catalogues de la BnF en les fédérant autour d'entités, contient près de la totalité des notices validées du Catalogue Général. Cette exhaustivité d'information constitue un terrain privilégié à l'exploration scientifique dans le cadre de recherches, notamment autour de l'histoire du livre et de l'histoire de la littérature, comme en témoignent les récents travaux de Frédéric Glorieux (Observatoire de la Vie Littéraire) effectués à partir des données de data.bnf.fr et publiés sur hypotheses.org. Cependant, la modélisation autour d'entités, telle que le modèle FRBR (Functional Requirements for Bibliographic Records) pour les bibliothèques, ainsi que la formalisation en RDF n'est pas sans poser problème à celui qui voudrait se lancer dans ce type de recherche. En effet, le travail de statistique et d'interprétation du chercheur s'appuie sur une donnée qui se voudrait la plus complète possible, notamment lorsqu'il s'agit de décrire les exemplaires physiques d'une œuvre : il devient alors crucial que des variables comme le titre, le type, la date de publication, le lieu de publication, la langue de l'ouvrage, soient renseignées de manière rigoureuses. Or, le modèle FRBR, utilisé pour simplifier la présentation des notices bibliographiques sur des interfaces destinées au grand public, rend de ce fait plus complexe l'accès à ce type d'information. D'une manière générale, s'expriment dans le RDF des logiques analytiques et représentationnelles propres au web et qui sont à l'opposées de la philosophie de mise en série et d'interprétation qui caractérisent à la fois les sciences humaines et les données massives.

Dès lors, l'objectif d'un projet comme data.bnf.fr pourrait être de se concentrer sur la création de synergies et de collaborations afin de répondre à ces deux logiques opposées mais incontournables du mouvement d'ouverture des données : comment en effet proposer une modélisation des données qui puissent répondre à la fois à des contraintes qui sont celles de l'ergonomie du web aujourd'hui tout en satisfaisant aux obligations de rigueur propres à la recherche scientifique, notamment en sciences humaines ? "