

9. L'univers des métadonnées

Omar Boucelma

« Par ma foi ! Il y a plus de quarante ans que j'utilise des métadonnées sans que je n'en suse rien... ». Ainsi aurait pu s'exprimer de nos jours M. Jourdain, s'il avait consulté le catalogue d'une bibliothèque. Il aurait pu aussi, en navigant sur Internet, afficher le code HTML d'une page Web, et y découvrir les données « meta » (figure 1) qu'utilisent les moteurs de recherche pour un référencement optimisé d'un site Web. Ces données (URL, nom, contenu, format, liens entrants et sortants, etc.), qui décrivent une autre donnée (ici, une page Web), sont appelées « métadonnées ». Autres exemples de métadonnées : celles associées à un livre (titre, auteur, éditeur, copyright, etc.), à une image (type, résolution, etc.), à un morceau de musique (titre, artiste, etc.), ou à une base de données (structure, droits d'accès, etc.). Ces métadonnées sont donc des descripteurs d'objets, servant à les identifier, les indexer, les classer, les décoder, les utiliser ou plus généralement « à en parler » sans détailler leur contenu.

Les métadonnées sont codées avec des langages standardisés. Elles sont spécifiques à des domaines et

favorisent l'interopérabilité des systèmes qui manipulent les données qu'elles décrivent (figure 2).

Les métadonnées pour les ressources bibliographiques

Les bibliothèques ont une pratique ancienne, mais en constante évolution, des métadonnées. Par exemple, les notices bibliographiques peuvent être décrites en FRBR (*Functional Requirements for Bibliographic Records*), un modèle conceptuel indépendant de toute implémentation. La description en FRBR ou « FRBRisation » permet à des bibliothèques, comme la BNF par exemple, de diffuser sur le Web des notices au format RDF. Par ailleurs, pour exposer sur le Web sémantique (cf. V.4) une plus large catégorie d'objets (journaux, livres, etc.), on peut utiliser l'ontologie FaBIO, dérivée de FRBR, qui permet de référencer des entités bibliographiques encodées avec une autre ontologie.

Un autre exemple concerne les citations des ressources numériques (un article, une conférence, etc.) par des chercheurs. Ces derniers ont besoin de citer les travaux d'autres chercheurs et de s'assurer que les leurs peuvent être cités. Les citations sont décrites avec le vocabulaire *Datacite Schema* promu par l'ONG DataCite, qui fournit aussi un identifiant d'objet numérique unique (DOI) à chaque ressource.

Enfin, une œuvre (culturelle, scientifique, etc.) peut être sujette à des droits (copyright), et toute personne détentrice de ces droits peut les céder en totalité ou en partie via une licence *Creative Commons* (CC). Les métadonnées associées à cette licence sont générées automatiquement en HTML, en vue de leur publication sur le site Web de l'œuvre.

Les métadonnées associées aux ressources multimédias

Pour accommoder la diversité des ressources multimédia (texte, son, image, vidéo) et de leurs supports (télévision numérique, Web), il existe plusieurs normes et métadonnées. Par exemple, le standard MPEG-7

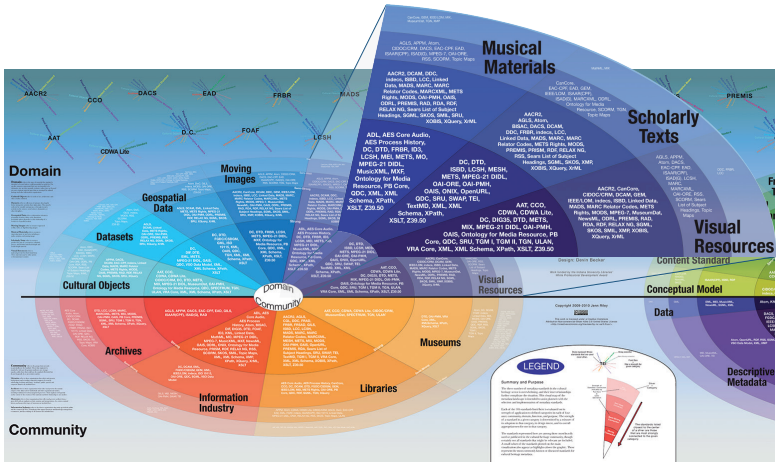


Fig. 2 – Zoom sur l'« univers des métadonnées », formalisé par Jenn Riley, et chercheuse bibliothécaire numérique, dans lequel elle répertorie 105 standards du patrimoine culturel. Source Riley and Becker 2009-2010, CC by NC SA 3.0. ■

permet l'indexation et la recherche de documents multimédia à l'aide de descripteurs, comme le titre d'un document, l'histogramme des couleurs, ou le timbre d'un son. Dans le domaine du e-Learning et des MOOC, qui fait un usage intensif de ressources diverses, un cours est décrit en LOM (*Learning Object Metadata*), et les systèmes de e-learning (LMS) sont construits selon les directives du modèle SCORM pour structurer le contenu d'une formation e-learning.

Les métadonnées géographiques

Avec des éditeurs de carte collaboratifs en ligne comme OpenStreetMap ou Wikimapia, la production de données et de métadonnées géographiques n'est plus l'apanage du seul spécialiste. Un utilisateur peut, par exemple, éditer une carte pour baliser de nouveaux chemins de randonnées en récupérant des traces GPS : on parle alors d'information géographique participative (VGI). Grâce aux métadonnées standard ISO 19115 ou bien celles du FGDC (USA), l'utilisateur peut

décrire aussi bien des attributs géographiques (latitude, longitude, etc.) que des attributs de qualité, comme l'exactitude (l'objet est représenté au bon endroit sur une carte) ou la précision (la taille de l'objet est très proche de sa taille réelle). La qualité des données est très importante pour la prise de décision et le principe d'« aptitude à l'emploi » (*fitness for use*) des données.

Les métadonnées pour la préservation des données numériques

L'archivage et la préservation des données sont cruciaux pour certaines disciplines. Pouvoir lire et traiter des données vieilles de plusieurs décennies est un enjeu

important, comme l'illustre l'histoire du « projet valise », rapportée par Guillaume Garvanèse dans le Journal du CNRS : une équipe de chercheurs a « exhumé » des données vieilles de vingt ans, stockées sur des bandes magnétiques rangées dans une valise. Les chercheurs ont pu décrypter ces données, suite à un laborieux travail de récupération et de lecture des bandes. Avec un standard comme PREMIS, qui décrit les formats de données, les supports de stockage, les traces de modification entre deux dates, la technologie du support utilisé, etc., le décryptage aurait été sans doute plus facile. La préservation à long terme des données nécessite des métadonnées riches décrivant le contexte dans lequel les données ont été produites, les technologies de stockage et de lecture utilisées, leur degré de fiabilité en fonction de la durée du stockage, les risques d'altération physiques ou chimiques, et toute information restituant le contexte historique permettant l'interprétation des données.

L'univers des métadonnées ne cesse d'évoluer grâce, notamment, aux avancées du Web sémantique (cf. II.5) et à l'émergence du Web des données ouvertes (Linked Open Data). Pour une gestion durable de cet univers, il est nécessaire de concilier les standards de droit (normes ISO) et de fait (communautés d'utilisateurs) en assurant la publication et la maintenance des différents vocabulaires, comme le propose le projet LOV.

Références bibliographiques

- J. RILEY et D. BECKER – *Seeing Standards: A Visualization of the Metadata Universe*, <http://jennriley.com/metadatamap>, 2009-2010.
- G. CARVANESE – *Préserver les données de la recherche à l'ère du Big Data*, Le Journal du CNRS, 09/09/2016.
- *Understanding Metadata*, NISO Press, www.niso.org/publications/press/UnderstandingMetadata.pdf 2004.

```
<!-- This site is optimized with the Yoast SEO plugin v3.4.2 -->
<meta name="description" content="Le bourgeois gentilhomme au Théâtre Michel. Mr Jourdain est un riche bourgeois ..." />
<meta name="twitter:title" content="Le bourgeois gentilhomme au Théâtre Michel" />
<meta property="article:tag" content="Le bourgeois gentilhomme" />
```

Fig. 1 – Exemple de métadonnées. Source : www.theatre-michel.fr. ■