Documenter ses données

La documentation des données renvoie à un ensemble de pratiques d'organisation, de structuration, de description et de communication de ses matériaux de recherche et de leur contexte de production. Documenter ses données constitue une étape essentielle d'un projet de recherche, permettant de les préserver, les réutiliser et les partager. Afin de rendre ses données faciles à trouver, accessibles, interopérables et réutilisables, selon les principes FAIR, il est important de leur associer une bonne documentation.

La documentation des données est formalisée par la saisie de **métadonnées**, soit des données qui qualifient une donnée. Il existe différents standards de métadonnées selon les types de données traitées. L'objectif principal de l'utilisation de métadonnées standardisées est de contextualiser les données produites à la fois pour les hommes et pour les machines.

Le choix du standard à utiliser dépend de quatre facteurs

- le type de ressources décrites
- le domaine scientifique dont sont issues les données
- la communauté concernée par la production des données
- l'entrepôt choisi pour stocker les données

Parmi les standards de métadonnées les plus usités en Sciences Humaines et Sociales, on trouve :

- Le **Dublin Core**, standard de métadonnées qui fournit un socle commun d'éléments descriptifs pouvant s'adapter à différents types de données. Il est composé de 15 propriétés de base (titre, sujet, description, source, langue, relation, couverture, créateur, contributeur, éditeur, gestion des droits, date, type, format, identifiant de la ressource) qui peuvent être enrichies. On parle alors de Dublin Core qualifié avec trois éléments supplémentaires (*audience*, *provenance*, *rightsholder*). C'est notamment le Dublin Core qui est utilisé dans l'entrepôt de données Nakala.
- L'**EAD** (*Encoded Archival Description*) est un standard essentiellement utilisé dans le monde des archives. Exprimé dans un langage XML, il permet notamment de respecter l'organisation hiérarchisée des instruments de recherche tout en conservant le principe d'héritage des informations entre les différents niveaux.
- La **DDI** (*Data Documentation Initiative*) est un standard adapté aux données issues des sciences sociales, comportementales et économiques. Il est plus particulièrement adapté aux données d'enquête et informations statistiques.
- La **TEI** (*Text Encoding Initiative*) est un consortium dédié à la maintenance et au développement d'un langage de description des textes en milieu numérique. S'il est un langage de structuration de données textuelles exprimé en XML, il permet aussi de saisir des métadonnées associées aux textes.

Toutefois, la documentation ne doit pas concerner que les données mais aussi les méthodes employées, afin de rendre la méthodologie du projet explicite, favoriser la confiance dans les résultats obtenus et permettre la réutilisation des données. Cette documentation peut prendre la forme d'un simple texte explicatif (fichiers readme au format txt ou Markdown tout comme s'insérer dans la pratique des écritures exécutables (jupyter notebooks).

Contact
Pour toute question sur les données de recherche

guichet-ardoise@groupes.renater.fr
Fiona Edmond

Chargée des données de recherche - SCD

fiona.edmond@univ-rennes2.fr
Oanez Helary

Animatrice Datalab MSHB Rennes

oanez.helary@mshb.fr

Aller plus loin

- Les standards de métadonnées : pourquoi et lequel ? - DoRANum
- Métadonnées, standards, formats Comment décrire les données ? - DoRANum
- Cours introductif sur les métadonnées
- Fiche synthétique - Métadonnées, standards, formats
- Guide de rédaction de Cornell d'une documentation Readme
- Modèle de README
- Faciliter la documentation des données grâce aux métadonnées dans un projet de …