

Métadonnées, standards et formats

Les pratiques d'utilisation de métadonnées sont anciennes, les bibliothèques les utilisaient sous forme de fiches cartonnées pour référencer les documents. Le numérique a permis la transformation des fiches cartonnées en un catalogue de métadonnées interrogeables par un moteur de recherche. Des spécificités propres à de nombreux métiers ont fait naître des standards variés. Un enjeu majeur de l'Open Data pour la réutilisation des données de la recherche est la description standardisée de celles-ci.

Métadonnées, des données sur les données



Principe

Dans un cadre scientifique, les métadonnées contribuent à décrire les ressources, les données de recherche et les productions réalisées (article, dépôt, photo, mesure, logiciel, page Web, etc.).

Note

Les métadonnées sont présentes :

- Soit de manière **embarquée**, exemple dans un fichier informatique (photo, logiciel, document, ...).
- Soit de manière **externe**, exemple dans un catalogue d'accompagnement d'un jeu de données ou dans un annuaire d'entrepôts.



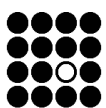
Usages



Gérer et partager

Mieux organiser et gérer l'ensemble des productions scientifiques, tracer leurs transformations, définir leurs usages, identifier l'auteur.

Préparer le partage et l'interopérabilité. Prévoir la pérennisation.



Trouver, comprendre et réutiliser

Faciliter la recherche des données de recherche et publications scientifiques.

Aider à comprendre les jeux de données (recueil, structure et contexte), pour leur validation (pairs) et/ou réutilisation.

Le saviez-vous ?

Les métadonnées sont embarquées dans un grand nombre d'objets numériques. Elles existent nativement et on peut les enrichir manuellement. La vérification intégrale de l'ensemble des métadonnées est importante pour partager uniquement les données souhaitées.

Structuration des métadonnées



Les standards de métadonnées

Le standard a pour objectif de fournir un ensemble d'éléments caractéristiques qui permettent de décrire les productions scientifiques. Ainsi la recherche peut être facilitée en portant sur les critères définis. La description des éléments peut elle-même être précisée par l'emploi de vocabulaires dédiés.

Le standard est choisi en fonction de la destination des données, dépôt, publication, archivage, etc. Il peut aussi être spécialisé par discipline, par type de données, etc., ainsi que son vocabulaire.



Les formats d'échange

Le format est une représentation sous forme numérique du standard et des métadonnées associées. Il est possible de parler de convention ou d'encodage. Cela permet l'utilisation de logiciels de représentation et de traitement des métadonnées, et autorise des protocoles d'échanges. Deux formats courants sont le XML et le CSV.

Exemples de standards

Dublin Core (interdisciplinaire), description des ressources numériques.

MARC (Machine-readable cataloging), description du contenu des bibliothèques.

EAD (Encoded Archival Description), description des archives.

DwC (Darwin Core), domaine de la biodiversité.

DDI (Data Documentation Initiative), domaine des sciences sociales, comportementales et économiques.

EXIF (Exchangeable image file format), description technique et automatique d'un cliché.

IPTC (International Press Telecommunications Council), description d'une image par l'auteur.

Exemple de standard interdisciplinaire usité : Dublin Core

➔ Présentation des 15 éléments de métadonnées ayant trait : au contenu - à la propriété intellectuelle - à la version

Le **Dublin Core simple** est utilisé par tous les gestionnaires de bibliothèques numériques et les plateformes généralistes de dépôt et publication de données. Ces 15 éléments restent cependant insuffisants pour représenter finement certaines données dans un univers dédié pour des utilisateurs et applications spécifiques. C'est pourquoi une quarantaine d'éléments plus spécifiques, qui constituent le **Dublin Core étendu**, viennent préciser les éléments simples. L'exemple ci-dessous montre une spécification pour les éléments « Date » et « Relation ».

Élément simple	Définitions	Exemple d'élément spécifique
Title 1	nom de la ressource	
Subject	thème du contenu de la ressource	
Description	résumé, table des matières...	
Creator 2	auteur principal de la ressource	
Publisher	entité responsable de la diffusion de la ressource	4 <i>Date Created, Date copyrighted, Date Valid, Date Available, Date Modified, Date Accepted, Date Submitted, Date Issued</i> 5
Contributor	co-auteurs associés à l'élaboration de la ressource	
Date	date de création ou mise à disposition	
Type 3	nature du contenu : image, son texte...	6 <i>Has Format (Les relations de transformation de format sont celles où une ressource a été dérivée d'une autre à l'aide d'une technologie de reproduction ou de reformatage qui n'est pas fondamentalement une interprétation mais une représentation.)</i>
Format	format ou taille de la ressource	
Identifiant	référence univoque, DOI, URL, ISSN...	
Source	référence à une ressource à partir de laquelle la ressource actuelle a été dérivée ou créée	
Language	langue originale de la ressource	
Relation	référence à une ressource apparentée	<i>isVersionOf (Les relations de version sont celles où une ressource est un état ou une parution historique d'une autre ressource par le même créateur)</i>
Coverage	périmètre spatial et temporel	
Rights	informations sur les droits associés à la ressource	

➔ Utilisation des éléments « Dublin Core » cités ci-dessus dans le site d'exposition de données « nakala ».

La description de base propose des identifiants génériques simples et étendus.

Ajout par liste déroulante de tous les identifiants définis dans Dublin Core. L'utilisateur ajoute « Date Accepted » et « Has Format » pour mieux qualifier son dépôt de données.