

Les métadonnées

1- Définition et cycle de vie :

Les métadonnées permettent de décrire plus précisément les données. Ce sont des **données sur les données**.

Il est recommandé de renseigner les métadonnées au fur et à mesure de l'avancée du projet, avec une attention particulière :

- au moment du partage des données
- puis à l'étape de l'archivage pérenne (des métadonnées spécifiques seront à renseigner).

Si on imagine un jeu de données sous la forme d'une boîte de conserve, alors les métadonnées équivalent à l'étiquette qui orne celle-ci et en décrit le contenu (date de fabrication, créateur etc.).



2- Métadonnées embarquées et métadonnées enrichies :

Il existe deux types de métadonnées :

- **les métadonnées embarquées** : elles sont produites **automatiquement** par les appareils (de prise de vue ou de son, de mesure...). C'est typiquement le cas pour la photo ou la vidéo du smartphone. Exemples de métadonnées générées : données GPS, type d'appareil, date, calibrage technique, etc.
- **les métadonnées enrichies** : elles sont ajoutées par l'auteur. Exemples : mots-clés, sujet, auteur, laboratoire ou organisme, nom du projet, licence, etc.

N'oubliez pas de bien **compléter** les métadonnées embarquées par des métadonnées enrichies. L'idéal est de renseigner ces métadonnées au fur et à mesure. Pour cela, il est recommandé d'avoir recours à un **standard** de métadonnées propre à votre discipline ou adapté à vos besoins.

3- La différence entre standard et schéma :

L'idéal est d'utiliser un **standard de métadonnées propre à votre discipline**. S'il n'en existe pas, il est possible de créer un **schéma** de métadonnées.

- **Standard :**

Un standard est un schéma qui a été adopté comme **modèle** par un ensemble d'utilisateurs : il est **reconnu, normalisé et utilisé à grande échelle**. Avoir recours à un standard est donc une **bonne pratique** en accord avec les principes FAIR : cela favorise grandement l'interopérabilité.

- **Schéma :**

Si aucun standard dans votre discipline ne répond à vos besoins, il est possible de créer votre propre schéma de métadonnées.

Une partie des métadonnées étant basiques (titre, date, auteur...), il est conseillé d'avoir recours à un **standard généraliste** comme *Dublin Core* ou *DataCite Metadata Schema*. Il s'agit ensuite de le compléter par un **fichier avec les métadonnées correspondant à vos besoins spécifiques**. Concrètement, cela consiste à lister vos métadonnées sous forme de tableur, avec l'intitulé de la métadonnée d'un côté et le descriptif en face.

De même, si un standard disciplinaire existe mais ne répond que partiellement à vos besoins, vous pouvez l'utiliser et le compléter.

Attention : opter pour un schéma de métadonnées plutôt qu'un standard rend vos données beaucoup moins interopérables et moins FAIR. Seules les métadonnées renseignées selon un standard seront interopérables. Les métadonnées spécifiques le seront moins voire pas du tout.

Pour trouver quels standards sont utilisés dans votre discipline, vous pouvez interroger vos collègues chercheurs mais aussi les informaticiens et professionnels de l'information scientifique et technique (IST) et voir quelles sont les pratiques dans votre domaine.

Pensez également à regarder les informations fournies par les entrepôts de données sur les standards de métadonnées qu'ils utilisent.

Il existe aussi plusieurs répertoires et sites, parmi lesquels le site du [DCC](#) et le site de la [RDA](#).

4- Exemples de standards de métadonnées :

- **Dublin Core** : Il s'agit d'un standard interdisciplinaire pour la description des ressources numériques. <http://dublincore.org/>
- **DataCite Metadata Schema** : standard lié à l'attribution d'identifiants pérennes DOI. <https://schema.datacite.org/>
- **DDI (Data Documentation Initiative)** : standard disciplinaire pour le domaine des sciences sociales, comportementales et économiques. <http://www.ddialliance.org/>
- **MIDAS-Heritage** : standard disciplinaire dans le domaine de l'architecture. <https://historicengland.org.uk/images-books/publications/midas-heritage/>
- **DwC (Darwin Core)** : standard disciplinaire du domaine de la biodiversité. <http://rs.tdwg.org/dwc/>
- **EML (Ecological Metadata Language)** : standard disciplinaire dans le domaine de l'écologie : il a en grande partie été conçu pour décrire des ressources numériques. Il peut également être utilisé pour décrire des ressources non numériques telles que des cartes papier ou d'autres médias. <https://eml.ecoinformatics.org/>

Guide OTELo :

Vous pouvez aussi consulter le chapitre dédié aux métadonnées dans le **Guide de bonnes pratiques : Gestion et valorisation des données de recherche** rédigé par OTELo et l'Inist-CNRS. Il propose un modèle de métadonnées pour apporter une information minimale et suffisante pour comprendre et reproduire ses données.

<https://ordar.otelo.univ-lorraine.fr/record?id=10.24396/ORDAR-1>



Focus : Le Dublin Core – les 15 éléments de base

Le Dublin Core est un standard **international et multidisciplinaire** très utilisé.

Il comporte **15 éléments** qui constituent le **minimum exigé** avec des éléments relatifs au contenu et à la propriété intellectuelle.

Élément	Élément (anglais)	Commentaire
1. Titre (métadonnée)	Title	Nom donné à la ressource
2. Créateur (métadonnée)	Creator	Nom de la personne, de l'organisation ou du service responsable de la création du contenu de la ressource
3. Sujet (métadonnée) ou mots clés	Subject	Thème du contenu de la ressource (mots clés, expressions, codes de classification)
4. Description (métadonnée)	Description	Présentation du contenu de la ressource (résumé, table des matières, représentation graphique du contenu, texte libre)
5. Éditeur	Publisher	Nom de la personne, de l'organisation ou du service responsable de la mise à disposition ou de la diffusion de la ressource
6. Contributeur	Contributor	Nom de la personne, de l'organisation ou du service responsable de contributions au contenu de la ressource
7. Date (métadonnée)	Date	Date de création ou de mise à disposition de la ressource
8. Type	Type	Nature ou genre de la ressource (catégories, fonctions, genres généraux, niveaux d'agrégation du contenu)
9. Format	Format	Manifestation physique ou numérique de la ressource
10. Identifiant de la ressource	Identifier	Référence univoque à la ressource dans un contexte donné (URI, ISBN)
11. Source	Source	Référence à une ressource dont la ressource décrite est dérivée (URI)
12. Langue (métadonnée)	Language	Langue du contenu intellectuel de la ressource
13. Relation (métadonnée)	Relation	Référence à une ressource apparentée
14. Couverture (métadonnée)	Coverage	Couverture spatio-temporelle de la ressource (domaine d'application)
15. Gestion de droits (métadonnée)	Rights	Informations sur les droits associés à la ressource (droits de propriété intellectuelle, copyright, etc.)

5- L'enrichissement des métadonnées :

Il est fortement recommandé d'utiliser des **vocabulaires contrôlés disciplinaires** utilisés couramment par la communauté scientifique du domaine : par exemple des codex de médicaments, des classifications taxonomiques, la nomenclature internationale des formules chimiques....

Exemples :

- **En agriculture/agronomie** : le vocabulaire contrôlé multilingue [AGROVOC](#)
- **En environnement/écologie** : le thésaurus [GEMET](#) et le référentiel taxonomique [TAXREF](#)
- **En médecine/santé** : le thésaurus [MeSH](#)
- **Exemple en archéologie** : le thésaurus [PACTOLS](#)

Les champs de métadonnées sont à renseigner avec ces vocabulaires contrôlés, notamment au moment du dépôt dans un entrepôt pour le partage.

Le recours à des mots-clés, à du vocabulaire connu, reconnu et utilisé par une communauté scientifique augmente ainsi la capacité des données à être combinées avec d'autres données dans le cadre d'une réutilisation.

6- Utilité des métadonnées :

Pour résumer, voici à quoi servent les métadonnées en 5 points :

- 1. Comprendre l'**origine** des données et leur **contexte** de création ou de collecte
- 2. Améliorer le **moissonnage par les machines** (moteur de recherche)
- 3. Garantir l'**interopérabilité**
- 4. Connaître les **conditions de réutilisation et de partage** des données
- 5. Fournir des **informations** très utiles lorsque les données ne peuvent pas être partagées (embargo, accès restreint) ou lors du retrait des données (données obsolètes, etc.).