

Quelles métadonnées pour les logiciels de recherche ?

Printemps de la donnée

Université de Strasbourg

Jérôme Pansanel, le 21.06.2024



Cet exposé est distribué sous licence [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

Sommaire

- ♦ Les métadonnées
- ♦ Application aux logiciels
- ♦ Démonstration

Les métadonnées

Les métadonnées

- ♦ Des données pour décrire précisément les données de la recherche
- ♦ Intégrables de deux manières :
 - ❖ Embarquée (dans le fichier)
 - ❖ Externe (un fichier accompagnant un jeu de données)
- ♦ Format ouvert et structuré (json, xml, ...)
- ♦ Basées sur un schéma (Dublin Core, Darwin Core, ...) :
<https://rdamsc.bath.ac.uk/>
- ♦ Définition du titre, auteurs, date, etc

Quels intérêts ?

- ♦ Intégrer la stratégie FAIR de la Science Ouverte
- ♦ Aider à comprendre le contenu d'un jeu de données (recueil / génération, structure et contexte)
- ♦ Permettre leur compréhension en utilisant une ontologie éprouvée et en facilitant la création du PGD
- ♦ Rendre les données plus facilement trouvables et moissonables
- ♦ Permettre l'interopérabilité et la traçabilité
- ♦ Faciliter la réutilisation des données (format, licence)
- ♦ Développer une science reproductible

Les métadonnées

Un format interdisciplinaire : Dublin Core

<https://www.dublincore.org/>

Élément simple	Définition
Title	Nom du jeu de données
Subject	Thème du jeu de données
Description	Résumé
Creator	Auteur principal
Publisher	Entité responsable de la diffusion
Contributor	Co-auteurs associés
Date	Date de création ou de mise à disposition
Type	Nature du contenu (image, son, ...)
Format	Format ou taille du jeu de données
Identifiant	Identifiant univoque (DOI, URL, ...)
Source	Données à partir desquelles ce jeu de données est dérivé
Rights	Informations sur les droits associés
Relation	Référence à une ressource apparentée

Les métadonnées

Exemple de métadonnées

```
<metadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns="http://dublincore.org/documents/dcmi-terms/">
  <dcterms:title>
Experimental gamma-ray data recorded with a LaBr3 and digital acquisition.
</dcterms:title>
  <dcterms:identifiant>https://doi.org/10.57745/D20CFB</dcterms:identifiant>
  <dcterms:creator>HENNING, Greg</dcterms:creator>
  <dcterms:publisher>Recherche Data Gouv</dcterms:publisher>
  <dcterms:issued>2023-02-08</dcterms:issued>
  <dcterms:modified>2023-06-26T14:02:06Z</dcterms:modified>
  <dcterms:description>
This package publishes a data set of gamma-ray recorded
using an LaBr3 and digital acquisition.
</dcterms:description>
  <dcterms:subject>Physics</dcterms:subject>
  <dcterms:subject>Radiation detectors</dcterms:subject>
  <dcterms:language>English</dcterms:language>
  <dcterms:date>2022-03-02</dcterms:date>
  <dcterms:contributor>HENNING, Greg</dcterms:contributor>
  <dcterms:dateSubmitted>2023-01-31</dcterms:dateSubmitted>
  <dcterms:temporal>2022-03-02</dcterms:temporal>
  <dcterms:temporal>2022-03-02</dcterms:temporal>
  <dcterms:type>Dataset</dcterms:type>
  <dcterms:license>etalab 2.0</dcterms:license>
</metadata>
```

Bonnes pratiques sur les métadonnées

- ♦ Mettre en place une stratégie pour les gérer en corrélation avec vos objectifs
- ♦ Aligner cette stratégie avec leur usage prévu
- ♦ Commencer par en ajouter aux données les plus importantes pour stimuler leur utilisation
- ♦ Impliquer l'ensemble des acteurs
- ♦ Identifier la norme connue la plus adaptée pour les décrire et contrôler son respect
- ♦ Sensibiliser sur leur importance

Application aux logiciels

Problématique

- ♦ La recherche s'appuie fortement sur les logiciels scientifiques
- ♦ Une part importante et croissante des chercheurs développe des logiciels dans le cadre de leurs propres recherches
- ♦ Écosystème très vaste et hétérogène pour la gestion des codes (archivage, documentation, indexation, partage, découverte et citation)
- ♦ Difficulté de citation et de reproductibilité

Application aux logiciels de recherche

Le logiciel FAIR : six recommandations

- ♦ Utiliser une forge logicielle accessible au public intégrant un outil pour la gestion de versions (par ex. git)
- ♦ Diffuser le logiciel sous une licence libre
- ♦ Utiliser un système de contrôle qualité pour le logiciel :
<https://www.bestpractices.dev/en>
- ♦ Enregistrer le code dans un annuaire :
<https://github.com/NLeSC/awesome-research-software-registries>
- ♦ Mettre en place un système pour la citation du logiciel
- ♦ Ajouter des métadonnées !

Application aux logiciels de recherche

Quels intérêts ?

- ♦ Citation des logiciels
- ♦ Reproductibilité des analyses
- ♦ Valorisation des contributions
- ♦ Faciliter la découverte grâce aux catalogues internationaux (Zenodo, DataOne, figshare)
- ♦ Affichage et mise en valeur des dépendances

Application aux logiciels de recherche

Un standard : Codemeta

- ♦ Un schéma de métadonnées minimal
- ♦ Format JSON et XML
- ♦ Un vocabulaire conceptuel pour normaliser l'échange de métadonnées logicielles
- ♦ Dernière version v3.0 : <https://w3id.org/codemeta/v3.0>
- ♦ Fichier à placer à la racine de votre projet (codemeta.json)
- ♦ Indexation par Software Heritage / HAL : <http://doc.hal.science/deposer/deposer-le-code-source/>

Application aux logiciels de recherche

Exemple

```
{
  "@context": "https://doi.org/10.5063/schema/codemeta-2.0",
  "@type": "SoftwareSourceCode",
  "license": "https://spdx.org/licenses/GPL-2.0+",
  "codeRepository": "https://github.com/mychem/mychem-code.git",
  "dateCreated": "2010-07-06",
  "datePublished": "2010-07-06",
  "dateModified": "2021-02-21",
  "downloadUrl": "https://github.com/mychem/mychem-code/archive/refs/tags/1.0.1.tar.gz",
  "issueTracker": "https://github.com/mychem/mychem-code/issues",
  "name": "Mychem",
  "version": "1.0.1",
  "identifiant": "https://doi.org/10.5281/zenodo.4557895",
  "description": "Mychem is a chemoinformatics extension ...",
  "applicationCategory": "Chemistry",
  "programmingLanguage": [
    "C",
    "C++"
  ],
  "operatingSystem": [
    "Linux",
  ],
  "softwareRequirements": [
    "OpenBabel"
  ],
  "author": [
    {

```

Application aux logiciels de recherche

Quelle application à l'Université de Strasbourg ?

- ♦ Mise en place d'un catalogue des logiciels libres
- ♦ Cadre du comité de la Science Ouverte de l'Université
- ♦ Disponible sur le site Science Ouverte dès que possible :
<https://scienceouverte.unistra.fr>
- ♦ Valorisation des développements des laboratoires et des services de l'Université
- ♦ Basé sur des outils simples et fonctionnels (maintenabilité, réutilisation, etc)
- ♦ Indexation basée sur le format Codemeta !

Démonstration

Démonstration

Des métadonnées faciles à générer !

- ♦ Un site permet de générer les métadonnées selon le standard Codemeta :

<https://codemeta.github.io/codemeta-generator/>

Questions ?