



Les métadonnées dans un DMP

Marie Puren

ANF – INIST – Vandoeuvre-Les-Nancy – 6 juillet 2017



Objectifs

A la fin de cet atelier, les participants sauront:

- Définir les métadonnées
- Comprendre l'importance des métadonnées dans le cadre de la rédaction des DMP

Introduction

Cela va vous aider, ainsi que ceux qui souhaiteraient réutiliser vos données, à :

- Localiser,
- Comprendre,
- Interpréter vos données.

=> Pratique utile durant le processus de recherche et étape indispensable pour partager ses données

Un peu d'étymologie

- « Métadonnées » : mot « données » + préfixe « méta » = au-delà, après
- Signifie « au-delà des données », « qui dépasse les données », « qui englobe les données »

Des données sur les données

Dans son sens le plus basique:

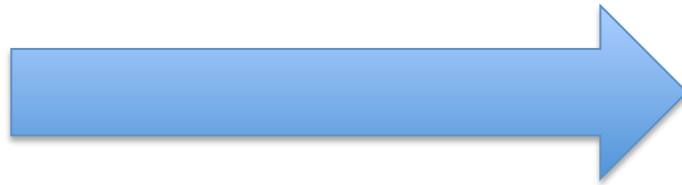
- Métadonnées = données sur les données
- Décrivent certaines caractéristiques des données comme:
 - qui a créé les données,
 - ce que le fichier de données contient,
 - quand les données ont été générées,
 - où les données ont été générées,
 - pourquoi les données ont été générées,
 - comment les données ont été générées.

Une production continue

- Même si le terme est assez récent, il y a toujours eu production de métadonnées.
- On produit des métadonnées lorsque **l'on catalogue, indexe, décrit, classe ou produit un instrument de recherche.**
- Avec ces activités, les bibliothécaires et les archivistes produisent des métadonnées “métier”.

Des métadonnées au quotidien

| Nutrition Facts | |
|---------------------------------------|----------------------|
| Valeur nutritive | |
| Per 250 mL / par 250 mL | |
| Amount | % Daily Value |
| Teneur | % valeur quotidienne |
| Calories / Calories 110 | |
| Fat / Lipides 0 g | 0 % |
| Saturated / saturés 0 g | 0 % |
| + Trans / trans 0 g | 0 % |
| Cholesterol / Cholestérol 0 mg | |
| Sodium / Sodium 0 mg | 0 % |
| Potassium / Potassium 470 mg | 13 % |
| Carbohydrate / Glucides 27 g | 9 % |
| Fibre / Fibres 0 g | |
| Sugars / Sucres 23 g | |
| Protein / Protéines 2 g | |
| Vitamin A / Vitamine A | 0 % |
| Vitamin C / Vitamine C | 120 % |
| Calcium / Calcium | 2 % |
| Iron / Fer | 0 % |
| Folate / Folate | 25 % |



Fournit des informations sur



Des métadonnées au quotidien

Notice bibliographique

Notice Au format public

Auteur(s) : [Colette \(1873-1954\)](#) ☰

Titre(s) : Sido [Texte imprimé] ; suivi de Les vrilles de la vigne / Colette ; préf. d'Alain Brunet

Publication : Paris : Librairie générale française, 2000

Impression : impr. en Allemagne

Description matérielle : 224 p. : couv. ill. en coul. ; 18 cm

Collection : Le livre de poche ; 373

Lien à la collection : [Le Livre de poche](#) ☰

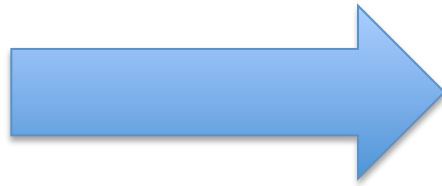
Autre(s) auteur(s) : [Brunet, Alain \(1950-2005\)](#). Préfacier ☰

Numéros : ISBN 2-253-00523-1 (br.) : 26 F

Notice n° : FRBNF37108727



Cette notice appartient à [l'univers jeunesse](#)



Fournit des
informations sur



Des métadonnées au quotidien

Raw hyperspectral imaging data of Baltic Sea algae cultures

Eskelinen, Matti Aleksanteri; Salmi, Pauliina

Researcher(s)

Pölönen, Ilkka; Kremp, Anke; Eskelinen, Matti; Salmi, Pauliina

This file archive contains the raw data from hyperspectral imaging of Baltic sea algae cultures performed on 16th of August, 2018 at the hyperspectral imaging laboratory of the Faculty of Information Technology, University of Jyväskylä, Finland.

The dataset contains images of cultures of the following algal species in various dilutions and mixes:

- *Diatoma tenuis* DTTV-1401
- *Melosira arctica* MATV-1402
- *Scorpiella hangoei* (aka *Apocalathium Malmogjense*) SHTV-1
- *Kryptopenidium foliaceum* KFF-1001
- *Monoraphidium* TV70
- *Chlorella pyrenoidosa* TV216

In addition, the dataset includes images of pure water samples, empty petri dishes and millimeter paper useful for transmittance calculations and size measurement.

The imaging setup consisted of living samples pipeted on glass Petri dishes, with a halogen light source illuminating the dish from the bottom towards the camera on top.

The signal in each image contains slight fluctuation in the spectral dimension due to the AC current light source used.

OpenAIRE

Publication date:

July 3, 2017

DOI:

[10.5281/zenodo.822338](https://doi.org/10.5281/zenodo.822338)

Keyword(s):

Baltic Sea algae Hyperspectral imaging

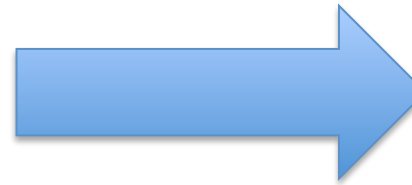
License (for files):

[CC BY](https://creativecommons.org/licenses/by/4.0/) Creative Commons Attribution 4.0

Versions

Version 1 [10.5281/zenodo.822338](https://doi.org/10.5281/zenodo.822338) Jul 3, 2017

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.822337](https://doi.org/10.5281/zenodo.822337). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)



Files (11.0 GB)

Name

[baltic_algae_raw.tar.gz](#)

[md5:3ab7170c696af3eb6e889daf77f01059](#) ⓘ

Fournit des informations sur

Carte d'identité des données

- Métadonnées = **carte d'identité** des données, ressources, documents...
- Les métadonnées rendent plus facile, pour leur créateurs et pour des tiers,
 - la recherche d'informations,
 - la réutilisation et l'identification des données.
- Par exemple, une suite de chiffres dans un tableau devient incompréhensible si l'on ne sait pas à quoi correspondent les abscisses et les ordonnées, dans quel but le tableau a été créé et à quelle date.

A quoi servent les métadonnées?

Quand vous créez des métadonnées, vous documentez vos données. Cela vous permet de:

- **retrouver et réutiliser** facilement vos propres données;
- découvrir, évaluer, et réutiliser les **données produites par d'autres**;
- **aider les autres** à découvrir, reproduire, réutiliser, et citer vos données;
- faciliter la **préservation** des données numériques, alors que les logiciels et les formats ne cessent d'évoluer au fil du temps.

Quelles informations donnent-elles?

Les métadonnées permettent de décrire les données. Elles donnent des informations sur:

- le **contenu intellectuel** : titre, résumé, domaine de recherche, mots-clés, type de ressource;
- le **contexte de production** : date, créateur, projet, financeur, procédés de création;
- les **caractéristiques techniques** des fichiers et des données: formats, taille, organisation;
- **les propriétés et droits d'usage**: détenteurs des droits, conditions d'accès, conditions de partage, conditions d'usage, outils spécifiques pour accéder/lire les données

Différents types de métadonnées

Plusieurs types de métadonnées existent:

- **les métadonnées de description**, qui donnent une idée précise du contenu de la ressource
- **les métadonnées de gestion** qui permettent d'accéder à la ressource
- **les métadonnées de préservation**, qui garantissent un accès pérenne au document et sa compréhension dans le long terme.

Les métadonnées descriptives ou de description

Grâce à elles, on organise le savoir. Elles permettent:

- d'identifier,
- de classer
- de hiérarchiser l'information contenue dans une ressource numérique.

Par exemple : titre, dates, auteurs...

Dublin Core : schéma de métadonnées descriptives, simple et générique.

Les métadonnées administratives ou de gestion

Métadonnées qui vont participer à la gestion de la vie de l'objet numérique.

- **Les métadonnées d'identification** identifient de manière univoque les objets décrits, si possible avec un identifiant pérenne.
- **Les métadonnées de provenance et de contexte** donnent une vision précise de la vie de l'objet numérique:
 - D'où vient-il?
 - Comment les données ont-elles été collectées? Par quel moyen technique? (satellite, numérisation)
 - Des traitements ont-ils été effectués? Par qui? Pour quelle(s) raisons? Quand?
 - Pourquoi a-t-on créé, collecté ou reçu ces données?

Les métadonnées de préservation

- Les **métadonnées techniques** sont indispensables pour la pérennité des données et pour faciliter leur réutilisation.
 - Elles servent à identifier, caractériser, définir l'environnement techniques des objets numériques.
 - Elles doivent identifier:
 - le type de format,
 - la version du logiciel utilisé,
 - l'encodage,
 - la compression,
 - le système d'exploitation, etc

Les métadonnées de préservation

- Les **métadonnées de structure** donnent des informations sur les fichiers qui composent un document et précisent les relations qui existent entre eux. Il y a deux niveaux de structure:
 - le niveau logique : il s'agit des liens entre des éléments pertinents pour l'utilisateur comme des numéros de page, des plages audio, des titres de chapitres...etc.
 - le niveau physique, c'est-à-dire la manière dont les documents sont enregistrés : fichier, répertoire, support...

Les métadonnées de préservation

- **Les métadonnées de droit** englobent les droits d'usage et les droits d'accès, et respectent les contraintes légales en matière de propriété intellectuelle, de droits d'auteur ou encore les droits de propriété.
 - En découlent la définition d'une politique de diffusion, parfois matérialisées par une licence : comment peut-on accéder à l'objet numérique? qu'a t-on le droit de faire avec cet objet? (imprimer, copier, modifier...)

A quoi servent ces différents types de métadonnées?

- Les métadonnées **descriptives** permettent aux utilisateurs de **trouver les objets numériques** qui les intéressent, et d'évaluer leur intérêt.
- Lors de **l'archivage des données**, on s'appuiera largement sur les **métadonnées de préservation** (métadonnées techniques, de structure et de droit)

Pense-bête (1)

Source : [Bibliothèque de Cornell](#)

| Type | But | Exemples |
|--------------------------|--|---|
| Métadonnées descriptives | <p>Décrire et identifier les ressources:</p> <ul style="list-style-type: none">• Au niveau local pour permettre la recherche et l'identification (ex. rechercher une image d'animal dans une collection d'images)• Au niveau « Web », permettre aux utilisateurs de découvrir des ressources (ex. rechercher sur Internet de la poésie numérisée) | <ul style="list-style-type: none">• Identifiants uniques• Attributs physiques (media, dimensions, conditions d'utilisation et d'accès)• Attributs bibliographiques (titre, auteurs/ créateurs, langue, mots-clés) |

Pense-bête (2)

| Type | But | Exemples |
|--------------------------|--|---|
| Métadonnées de structure | <p>Faciliter la navigation et la présentation des ressources électroniques::</p> <ul style="list-style-type: none">• Fournit des informations sur la structure interne des ressources, comme le numéro de page, la section, le chapitre, les index, et la table des matières• Décrit les relations entre les matériaux (ex. la photographie B était incluse dans le manuscrit A)• Relie les fichiers et scripts liés entre eux (ex. le fichier A et le fichier B de l'image archivées) | <p>Etiquettes (tags) de structures comme le titre de la page, la table des matières, les chapitres, les parties, les errata, les index, les relations entre sous-objets (ex. photographie issue d'un journal)</p> |

Pense-bête (3)

| Type | But | Exemples |
|-----------------------------|--|---|
| Métadonnées administratives | <p>Faciliter la gestion sur le court et le long terme et le traitement des collections numériques:</p> <ul style="list-style-type: none">• Inclut des données techniques sur la création et le contrôle qualité• Inclut la gestion des droits, le contrôle de l'accès et les exigences d'utilisation• Information concernant les actions de préservation | <p>Les données techniques comme le type de scanner, la résolution, la définition, l'espace de couleur, la source de lumière, le propriétaire, la date de copyright, les limitations en matière de copie et de distribution, la licence, les activités de préservation (migration, etc.)</p> |

Le cycle de vie des métadonnées

Il faut être conscient que **le cycle de vie des métadonnées est plus grand que le cycle de vie des données.**

- Les métadonnées peuvent être créées **avant** la création ou la capture des données, afin de donner des informations sur les données qui seront produites dans le futur.
- Les métadonnées doivent être **conservées**, même si les données ont été supprimées, de façon à donner des informations sur les données qui ont été déclassées ou retirées.

Le cycle de vie des métadonnées

1. Créer

La création des métadonnées peut être réalisée de manière semi-automatique .

- les propriétés des documents générés par les logiciels de bureau (date de création)
- les informations spatiales et temporelles enregistrées par des caméras et des capteurs (GPS)
- les informations liées au processus de publication, comme une URL

Toutefois, une intervention humaine est nécessaire pour préciser:

- l'objet de la ressource
- l'utilisation de la ressource (lien vers une licence)
- les informations complémentaires sur la ressource.

Le cycle de vie des métadonnées

2. Entretien

Deux cas de figure peuvent se présenter:

- Les données sont stables, les métadonnées le sont aussi.
- Les données changent régulièrement, et les métadonnées doivent suivre ces évolutions de manière quasi instantanée.

Le cycle de vie des métadonnées

3. Mettre à jour

Les métadonnées doivent être mises à jour, parce que:

- les organisations changent, et les responsabilités peuvent être transmises;
- les données peuvent être utilisées différemment;
- les normes, les vocabulaires et les technologies évoluent;
- les outils et les systèmes se transforment.

Mais tout cela se fait en fonction du budget alloué, et du temps disponible...

Le cycle de vie des métadonnées

4. Stocker

- Les métadonnées peuvent être intégrées aux données elles-mêmes (exemple : documents bureautiques) ou stockées séparément dans une base de données.
- Dans ce cas, la gestion des données est facilitée avec des liens vers les fichiers correspondants.

Le cycle de vie des métadonnées

5. Gérer la suppression des données

- Il arrive très souvent que les métadonnées doivent être conservées, même après la suppression des données.
- Dans ce cas, les métadonnées vont indiquer que les données sont été supprimées ou archivées. Si elles ont été archivées, il faut indiquer comment et où.

Le cycle de vie des métadonnées

6. Publier

Dans le cadre des données ouvertes, on privilégie une publication ouverte, c'est-à-dire un accès direct par les URIs.

Horizon 2020 FAIR Data Management Plan (DMP) template

- [Guidelines on FAIR Data management in Horizon 2020](#)
- Dans l'annexe 1, section 2. « FAIR data », sous-section 2.1. « Making data findable, including provisions for metadata » :
« What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how »

Les métadonnées dans les DMP

Dans un DMP, il faut:

- Expliciter **comment vous allez décrire** les données,
- Inclure les **standards** applicables au contenu et au format des métadonnées,
- Préciser **les procédures, les outils, les logiciels** utilisés pour recueillir et éditer les métadonnées.

Métadonnées et DMP Opidor

Intital DMP

Standards and metadata

B *I* ☰ ☰ 🔗 📄

Save

Not answered yet

Guidance Share note

CE Guidance +

Digital Curation Centre guidance on
Metadata -

Questions to consider:

- How will you capture / create the metadata?
- Can any of this information be created automatically?
- What metadata standards will you use and why?

Guidance:

Metadata should be created to describe the data and aid discovery. Consider how you will capture this information and where it will be recorded e.g. in a database with links to each item, in a 'readme' text file, in file headers etc.

Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place. The DCC offers a [catalogue of disciplinary metadata standards](#).

Créer un plan de gestion des données

Mid-term Review DMP

Scientific research data should be easily: 1. Discoverable (1 question, 0 answered)

3. Assessable and intelligible (1 question, 0 answered)

4. Usable beyond the original purpose for which it was collected (1 question, 0 answered)

Des métadonnées sont créées avec les données de recherche

Lors de la création de données de la recherche, des métadonnées sont créées en même temps.

Les carnets de laboratoire ou les fichiers de recherche associés, comprennent (presque) toutes les informations suivantes:

- nom du chercheur
- date
- projet
- détails sur l'expérience/ l'analyse en cours, objectifs et méthodes inclus
- sources des autres données utilisées durant l'expérience/ l'analyse

La clé est de recueillir toutes les informations nécessaires (métadonnées) lors du travail de production des données, et de lier ces métadonnées aux fichiers des données.

Comment collecter ces métadonnées?

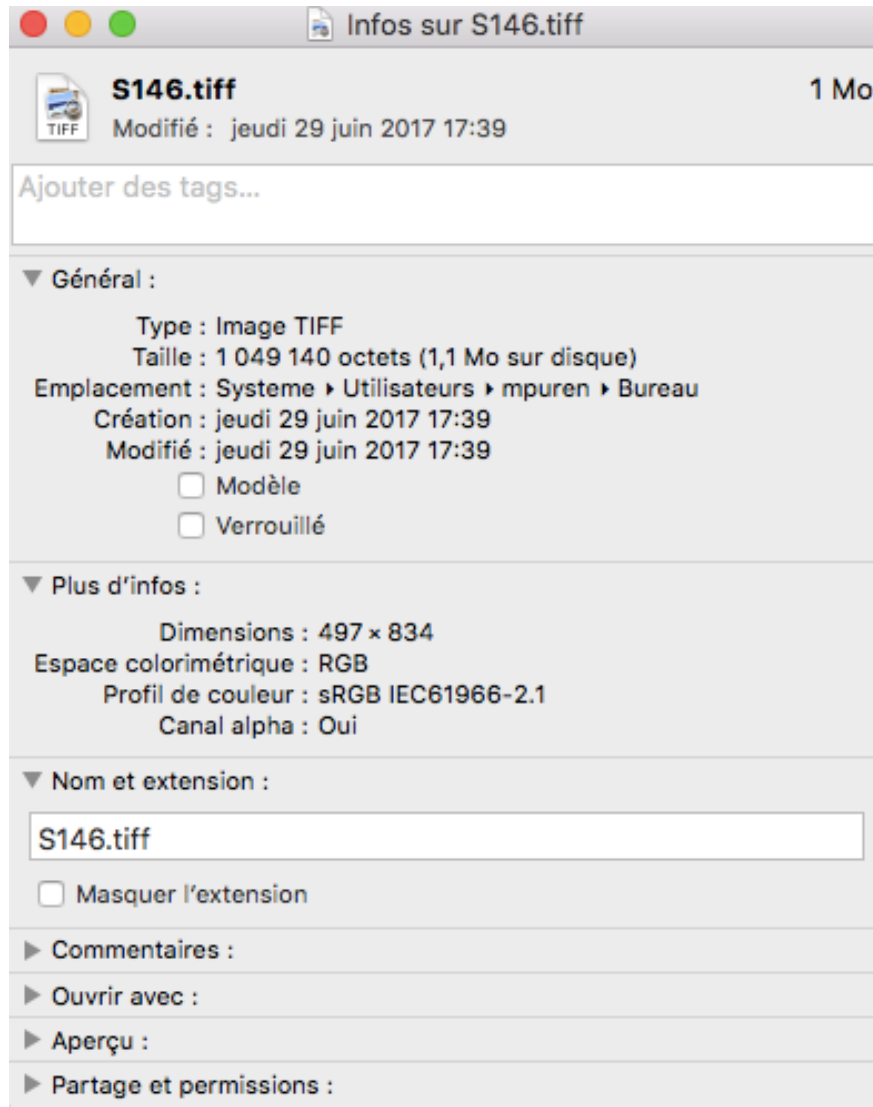
Pour recueillir ces métadonnées, on peut utiliser différentes méthodes:

- indiquer dans un carnet les informations concernant les projets, en notant les emplacements et les noms des fichiers numériques associés aux expériences individuelles;
- indiquer dans un carnet numérique ces informations avec des liens vers les fichiers de données pertinents;
- inclure une note dans chaque fichier de données qui indique l'emplacement des métadonnées;
- dans chaque dossier qui contient des données, inclure un document qui décrit le contenu des fichiers, en incluant les explications des abréviations et des en-têtes de colonnes dans les fichiers. On peut aussi y inclure les références des publications qui décrivent les données.

Des métadonnées sont créées automatiquement

- A chaque fichier informatique sont associées des informations issues du système de gestion des fichiers. Certaines d'entre elles peuvent être pertinentes en tant que métadonnées.
- Choisissez par exemple n'importe quel fichier sur votre ordinateur et affichez les informations qui lui sont attachées.

Des métadonnées sont créées automatiquement



Des métadonnées sont créées automatiquement

- Vous allez pouvoir trouver les informations suivantes:
 - le nom du fichier,
 - les dates de création et de modification
 - sa taille, son emplacement...
- ⇒ Il est possible de récupérer automatiquement ces informations.

L'extraction automatique de métadonnées

Il existe des outils qui vous permettent:

- d'extraire des métadonnées en masse de fichiers numériques :

[Metadata Extraction Tool](#)

- d'extraire des métadonnées de PDF ([Cermin](#), [GROBID](#))

Définir des responsabilités

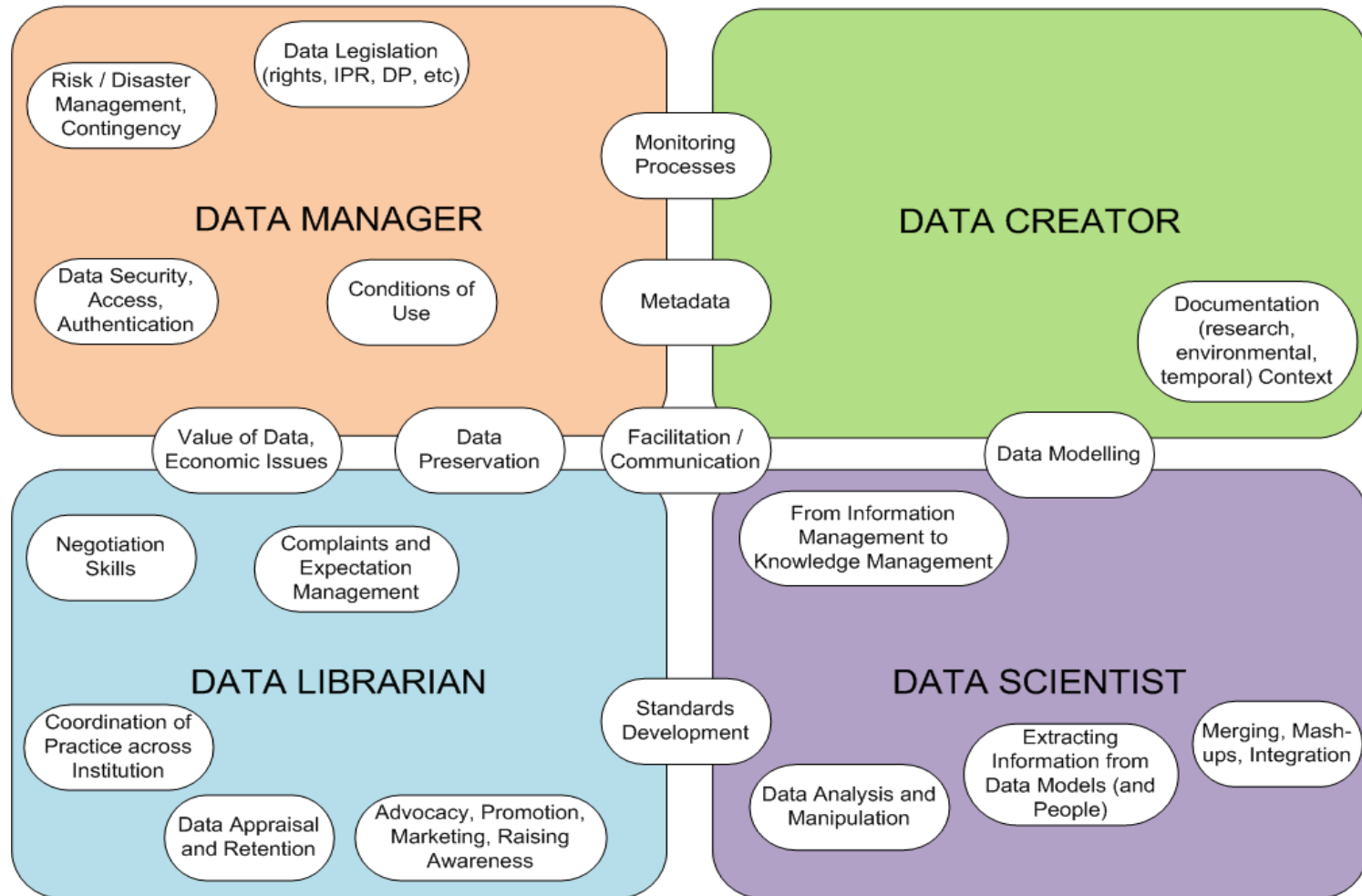
Pour créer les métadonnées, il faut s'appuyer sur:

- les chercheurs;
- les documentalistes, bibliothécaires ou archivistes;
- les informaticiens.

Définir des responsabilités

CORE SKILLS FOR DATA MANAGEMENT

A follow-up from the second DCC Research Data Management Forum (November 2008)



<http://www.dcc.ac.uk/resources/roles>

Les standards de métadonnées

- Des standards existent pour les métadonnées, et les entrepôts de données dans lesquels vous allez déposer les données peuvent exiger de respecter un standard particulier.
- On parle souvent de “schémas” pour désigner les structures des métadonnées. Le schéma va comporter un ensemble défini de caractéristiques pour décrire les données. Les métadonnées utilisent souvent un langage lisible par ordinateurs comme le XML.

Les standards de métadonnées

Un schéma de métadonnées est “un système de labellisation, marquage et codage utilisé pour l'enregistrement d'information de catalogage ou pour structurer des notices descriptives. Un schéma de métadonnées établit et définit les éléments de données et les règles d'application pour l'usage de ces éléments de données pour décrire une ressource”.

Source : Open Data Support,
[Introduction à la gestion des métadonnées](#)

Les standards de métadonnées

A noter

- Si vous utilisez un standard peu connu et difficilement accessible, il est important de préserver le schéma lui-même et sa documentation, avec les données et les métadonnées, de façon à permettre la compréhension et la réutilisation des données dans le futur

Des standards interdisciplinaires

Dublin Core

- Le **Dublin Core** est certainement le standard de métadonnées le plus utilisé, et il constitue un bon point de départ lorsque l'on commence à travailler sur les métadonnées.
- Le Dublin Core comporte 15 éléments constitutifs (version française : <https://who.rocq.inria.fr/Anne-Marie.Vercoustre/METADATA/DC-fr.1.1.html>)

Des standards interdisciplinaires

Les 15 éléments du Dublin Core

| Elements | Description |
|-------------|--|
| Contributor | Entité responsable de contributions à la ressource |
| Coverage | Couverture spatiale ou temporelle de la ressource, périmètre d'application de la ressource, ou juridiction sous laquelle la ressource est pertinente |
| Creator | Principale entité responsable de la création de la ressource |
| Date | Point ou période de temps associé(e) à un événement dans le cycle de vie de la ressource |
| Description | Description/ compte-rendu de la ressource |
| Format | Le format de fichier, le support physique, ou les dimensions de la ressource |
| Identifier | Référence non équivoque à la ressource au sein d'un contexte donné |
| Language | Langue de la ressource |

Des standards interdisciplinaires

Les 15 éléments du Dublin Core

| Elements | Description |
|-----------|---|
| Publisher | Entité responsable de la mise à disposition de la ressource |
| Relation | Ressource liée |
| Rights | Informations concernant les droits liés à la ressource |
| Source | Ressource apparentée dont la ressource décrite est dérivée |
| Subject | Sujet de la ressource |
| Title | Nom donné à la ressource |
| Type | Nature ou genre de la ressource |

Des standards interdisciplinaires

Version XML du Dublin Core : un exemple

```
<?xml version="1.0"?>
  <metadata
    xmlns="http://example.org/myapp/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://example.org/myapp/ http://
example.org/myapp/schema.xsd"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
    <dc:title>
      UKOLN
    </dc:title>
    <dc:description>
      UKOLN is a national focus of expertise in digital information
management. It provides policy, research and awareness services
to the UK library, information and cultural heritage communities.
UKOLN is based at the University of Bath.
    </dc:description>
    <dc:publisher>
      UKOLN, University of Bath
    </dc:publisher>
    <dc:identifier>
      http://www.ukoln.ac.uk/
    </dc:identifier>
  </metadata>
```


Des standards interdisciplinaires

Un aperçu

- [MARC](#) est un format de métadonnées pour le catalogage, très utilisé par les bibliothèques. Il est basé sur le format MARC (MAchine-Readable Cataloging).
- [EAD](#) (Encoded Archival Description) est utilisé la description des documents d'archives.
- [DataCite Metadata Schema](#) fournit une liste de propriétés choisies pour l'identification précise et cohérente des données en vue de leur citation et de leur découverte.
- [Project Open Data Metadata Schema](#) : schéma utilisé pour décrire les jeux de données publiés sur [agency.gov/data](https://www.data.gov/). (données produites dans le cadre de projets de recherche financés par le gouvernement américain).

Des standards interdisciplinaires

Un aperçu

- Metadata Object Description Schema (MODS) : un ensemble d'éléments bibliographiques qui peut être utilisé à des fins diverses, particulièrement pour les descriptions bibliographiques.
- Metadata Encoding and Transmission Standard (METS): variation de MODS qui permet de décrire des ressources numériques.
- PREMIS (Preservation Metadata: Implementation Strategies) fournit un ensemble de métadonnées que la plupart des entrepôts d'objets numériques auraient besoin d'enregistrer et d'utiliser pour préserver ces objets sur le long terme.

Des standards disciplinaires Pour les SHS

- [DDI - Data Documentation Initiative](#)
- [OLAC](#) (Open Language Archives Community)
- [CMDI](#)
- [MIDAS-Heritage](#)
- [CARARE metadata schema](#)
- [CESSDA MLI - Council of European Social Science Data Archives Minimum Level of Information](#)

Des standards disciplinaires Pour la biologie

- [Darwin Core](#)
- [EML - Ecological Metadata Language](#)
- [Genome Metadata](#)
- [ISA-Tab](#)
- [OME-XML - Open Microscopy Environment XML](#)
- [PDBx/mmCIF – Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework](#)

Des standards disciplinaires Pour les sciences de la terre

- Federal Geographic Data Committee's standard (FGDC)
- AgMES - Agricultural Metadata Element Set
- AVM - Astronomy Visualization Metadata
- DIF - Directory Interchange Format

A noter : Les normes ISO de la série 19000 couvrent le domaine géographique (ISO 19115 – métadonnées, ISO 19119 – services, ISO 19101 – modèle de référence...).

Des standards disciplinaires Pour les sciences physiques

- [AVM - Astronomy Visualization Metadata](#)
- [CSMD-CCLRC Core Scientific Metadata Model](#)
- [International Virtual Observatory Alliance Technical Specifications](#)
- [NeXus](#)
- [Observations and Measurements](#)

Par types et formats de données

- Pour les images : [IPTC](#) (International Press Telecommunications Council)

Pour trouver le standard qu'il vous faut

- Ressources interdisciplinaires
 - [RDA Metadata Standards Directory Working Group](#)
 - [Digital Curation Center](#)
- [Huma-Num](#) pour les SHS
- [BioSharing](#) pour la biologie

Comment choisir un standard de métadonnées?

- En fonction des ressources décrites par les métadonnées : type, format
- En fonction de la discipline concernée
- En fonction des types de métadonnées que vous voulez créer:
 - Métadonnées descriptives: Dublin Core, MARC
 - Métadonnées de préservation : PREMIS, METS
 - Métadonnées de structure: EAD... etc.

Un certain nombre d'éléments „standards“ (1)

| | Champ | Description |
|------------------------|------------------------------------|--|
| Informations générales | Titre | Nom du jeu de données |
| | Créateur | Noms et institutions de rattachement des créateurs des données |
| | Dates | Dates-clés associées aux données, telles que la période de temps couverte ou la date de création |
| | Agences de financement et périodes | Organisations ou agences de financement du projet, et période de financement |

Un certain nombre d'éléments „standards“ (2)

| | Champ | Description |
|------------------------|----------------------------|---|
| Informations générales | Mots-clés | Mots-clés ou phrases décrivant le sujet ou le contenu des données |
| | Identifiant | Chaîne de caractères numériques alphanumériques unique |
| | Couverture (si applicable) | Couverture géographique |
| | Langue (si applicable) | Langue du contenu intellectuel |

Un certain nombre d'éléments „standards“ (3)

| | Champ | Description |
|-------------------------|-----------------------|--|
| Informations d'accès | Accès et restrictions | Où et comment d'autres chercheurs peuvent accéder aux données |
| | Droit d'auteur | |
| Informations techniques | Formats des fichiers | |
| | Liste des fichiers | |
| | Nombre de fichiers | |
| | Méthodologie | Comment les données ont été générées : équipement ou logiciel utilisés, protocole expérimental |

Utiliser des ontologies

- En plus d'utiliser un standard de métadonnées, il est possible d'utiliser des ontologies ou des vocabulaires contrôlés pour créer des métadonnées.
- L'initiative Dublin Core recommande par exemple d'utiliser l'[Internet Media List](#), un vocabulaire contrôlé, pour l'élément "Format". On recommande également d'utiliser un vocabulaire contrôlé pour entrer les termes "Subject", mais il n'y a pas de recommandations particulières.

Utiliser des ontologies

- Les ontologies sont des vocabulaires partagés, utilisés pour décrire les les composants d'une discipline particulière, et les relations entre ces composants.
- L'utilisation d'ontologies permet de faciliter la compréhension des données par des tiers. D'autre part, les vocabulaires contrôlés sont des listes de termes prédéfinis et autorisés.
- Un vocabulaire contrôlé est une liste prédéfinie de valeurs. Ces valeurs vont être utilisées dans un schéma de métadonnées pour décrire des propriétés spécifiques. L'utilisation de vocabulaires contrôlés permet d'assurer l'interopérabilité des données.

Quelques exemples

- [Bioportal](#).
- [Gene Ontology](#)
- [Medical Subject Headings \(MeSH\)](#)
- [Web Ontology Language \(OWL\)](#)
- [Getty Thesaurus of Geographic Names \(TGN\)](#).
- [RFC4646](#) (vocabulaire utilisé pour indiquer la langue d'un objet)
- [Chemical Entities of Biological Interest \(ChEBI\)](#)
- [Microarray Gene Expression Society Ontology \(MGED\)](#) [Internet Media List](#)
- [Environmental Ontology \(EnvO\)](#) [Name Reaction](#), [Chemical Methods](#)
- [Molecular Processes Ontologies](#)

Gérer les métadonnées sur le long-terme

Il faut gérer les métadonnées de façon à s'assurer de leur:

- **disponibilité**: les métadonnées doivent être stockées de façon à être consultées, et indexées pour être trouvées.
- **qualité** : la qualité des métadonnées doit être assurée de manière constante de façon à ce que les utilisateurs aient confiance en celles-ci.
- **maintenance**: les métadonnées doivent être régulièrement maintenues, de façon à s'assurer de leur pérennité sur le long terme.
- **ouverture** : les métadonnées devraient être disponibles sous une licence ouverte pour permettre leur réutilisation.

Evaluer la qualité des métadonnées

- La qualité et l'exhaustivité des métadonnées vont directement influencer sur la visibilité et la réutilisation des données décrites par les métadonnées.
- Comment évaluer la qualité des métadonnées? Les métadonnées sont-elles:
 - précises?
 - accessibles dès maintenant et sur le long terme?
 - exhaustives?
 - conformes avec les normes acceptées (standards)?
 - cohérentes? (pas de contradictions entre les informations)
 - crédibles et sûres?
 - lisibles par la machine?
 - pertinentes?
 - mises à jour?

Des métadonnées ouvertes?

- Ouvert = « Quiconque peut librement avoir accès, utiliser, modifier et partager pour quelque motif que ce soit. »
- Cette définition couvre toutes les données: les métadonnées sont elles-mêmes des données, elle prend donc en compte ces dernières.

Des métadonnées ouvertes?

Revenons sur les 3 principes constitutifs de cette définition:

- disponibilité et accès
- réutilisation et redistribution
- participation universelle

Attribuer une licence à ses données, c'est la première étape pour que celles-ci deviennent des données ouvertes.

Quelles licences pour vos métadonnées?

Pour une réutilisation entièrement ouverte:

- Public Domain Mark
- Creative Commons Zero Public Domain Dedication
- Open Data Commons Public Domain Dedication and license (PDDL)

Pour une réutilisation ouverte, mais avec la nécessité de tracer la chaîne d'attributions:

- CC-BY
- ODC Attribution
- ISA Open Metadata licence 1.1
- Open Government License
- Licence ouverte/ Open License

Quels résultats?

En produisant des métadonnées ouvertes,

- vous favorisez la découverte de vos données,
- vous rendez vos données plus visibles,
- vous facilitez l'ajout, l'amélioration, la liaison des données à d'autres ressources.

Dans ce cadre,

- il faut faire le choix de licences les plus ouvertes possibles pour les métadonnées,
- une licence du domaine public favorise la réutilisation des métadonnées,
- une licence d'attribution garantit d'être crédité, mais peut favoriser l'accumulation des attributions si les données sont partagées à plusieurs reprises.

Des avantages et des inconvénients

S'il y a des avantages, il y a aussi des inconvénients lorsque l'on choisit une licence domaine public.

- Les utilisateurs ne pourront pas savoir qui a créé les métadonnées de départ. Mais ce choix facilite les éventuelles améliorations apportées par les utilisateurs.
- Vous perdrez le contrôle sur vos métadonnées, puisque vous n'aurez pas connaissance des déclarations faites à leur propos. Mais la qualité pourra être assurée par la communauté (cf. Wikipedia)
- Les utilisateurs ne sont pas sûrs que les métadonnées soient à jour et fiables. Le suivi de la qualité pourra toutefois être fait par les partenaires du réseau.
- Des erreurs peuvent se glisser dans les ajouts et modifications. Mais il y a moins de chance que cela se produise si on se base sur vos métadonnées d'origine.

Bibliographie (1)

Généralités

- Banat-Berger Françoise, Huc Claude, [Section 9 - Métadonnées](#), v2, 2011
- CINES, [Les métadonnées de pérennisation](#), 2016
- Dekkers Makx, De Keyzer Michiel, Loutas Nikolaos and Goedertier Stijn, [Introduction à la gestion des métadonnées](#), Open Data Support, 2013.
- Dekkers Makx, De Keyzer Michiel, Loutas Nikolaos and Goedertier Stijn, [Licences pour les données et les métadonnées](#), Open Data Support, 2013.
- Riley Jenn, [Understanding metadata. What is metadata, and what is it for?](#), NISO, 2017.

Bibliographie (2)

DMP et métadonnées

- European Commission, [H2020 Programme Guidelines on FAIR Data Management in Horizon 2020](#), v3.0, 25 July 2016.
- Inist, [Tutoriel 4 : DMP - Standards et métadonnées](#), 2015
- MIT Libraries, [Documentation & Metadata](#)
- NCSU Libraries, [Metadata for Data: Data Management Plan](#)
- Stanford Libraries, [Creating metadata](#)
- University of Oregon Libraries, [Metadata & Data Documentation](#)

Bibliographie (3)

Standards et métadonnées

- Cherhal Elizabeth,
[Présentation des standards : le Dublin Core \(DC\)](#), 2004
- DCC, [Digital curation standards](#)
- Dublin Core - [site officiel](#)
- Metadata Standards Directory Working Group, [Metadata Standards Directory](#)