



Les métadonnées : pourquoi, pour quoi et comment?

Partager les données de la recherche : pourquoi, comment ?
Guadeloupe, du 25 au 27 Novembre 2019

Agenda

- ❖ Qu'est-ce qu'une métadonnée?
- ❖ Pourquoi les métadonnées sont importantes?
- ❖ Les métadonnées en pratique

Qu'est-ce qu'une métadonnée?

Bonjour chers marmitonnateurs !

Aujourd'hui est un grand jour, puisque j'ai enfin décidé d'essayer de deviner ce que contient ma boîte. Pour cela, j'ai besoin de votre aide :

Dans mon placard, j'ai découvert une **boîte de conserve** dont le papier a malencontreusement disparu. Surtout ne me demandez pas pourquoi, ni comment, ni rien d'autre, dans une grande famille, c'est bien connu, même Sherlocks Home y perd son latin et ne peut trouver le responsable.

Toutes ces considérations ne font pas avancer le Schmilblick. Comment puis-je savoir ce que contient ma boîte de **conserve** 1/4 ?

Je vous vois venir, vous allez me proposer de l'ouvrir. Trop facile.

J'ai senti : pas d'odeur. J'ai secoué : un vague effet de liquide en mouvement. J'ai regardé : c'est doré.

Sur le dessus de la boîte, il est indiqué : TN (serait-ce du thon au naturel ?)

21-12-2010 (ça, c'est la date, j'ai de la marge)

L-278-N



Définition des métadonnées

❖ La métadonnée est une donnée sur une autre donnée

❖ Trois dimensions importantes

- Une dimension descriptive
 - Quoi?
 - Où?
 - Qui?
 - Quand?
 - Comment?
- Une dimension structuration
 - La finesse et la complexité du niveau de structuration dépendent des objectifs visés (par exemple collecter des données à des fins d'analyse vs modéliser des connaissances) et des informations qu'on veut capturer
- Une dimension sémantique
 - Utilisation de vocabulaires



Les métadonnées descriptives

Category	Sub-category	Terms
Descriptive metadata	What	abstract , accrualMethod , accrualPeriodicity , accrualPolicy , alternative , audience , bibliographicCitation , conformsTo , coverage , description , educationLevel , extent , format , hasPart , isPartOf , identifier , instructionalMethod , isRequiredBy , language , mediator , medium , relation , requires , spatial , subject , tableOfContents , temporal , title , type
Provenance	Who	contributor , creator , publisher , rightsHolder
Provenance	When	available , created , date , dateAccepted , dateCopyrighted , dateSubmitted , issued , modified , valid
Provenance	How	accessRights , hasFormat , hasVersion , isFormatOf , isVersionOf , license , isReferencedBy , isReplacedBy , references , replaces , rights , source

Extrait de <https://www.w3.org/TR/2013/NOTE-prov-dc-20130430/>

Les métadonnées de structure

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Year	Cropping s	Replicate	Plot	Sample	Crop speci	Date	Stage	Aboveground orga	Abovegrou	Abovegrou	N content (kg.ha ⁻¹)	
20	2009	No-Pest	1	2	1	Maize	28/09/2009	Maturity	stalk + cob	5,4382	26,64718		
21	2009	No-Pest	1	2	2	Maize	28/09/2009	Maturity	stalk + cob	4,50175	22,058575		
22	2009	No-Pest	1	2	3	Maize	28/09/2009	Maturity	stalk + cob	5,187325	25,4178925		
23	2009	No-Pest	1	2	4	Maize	28/09/2009	Maturity	stalk + cob	6,035475	29,5738275		
24	2009	No-Pest	1	2	5	Maize	28/09/2009	Maturity	stalk + cob	6,155675	41,2430225		
25	2009	No-Pest	1	2	6	Maize	28/09/2009	Maturity	stalk + cob	6,9675	46,68225		
26	2009	No-Pest	1	2	7	Maize	28/09/2009	Maturity	stalk + cob	6,431525	43,0912175		
27	2009	No-Pest	1	2	8	Maize	28/09/2009	Maturity	stalk + cob	5,58355	37,409785		
28	2009	No-Pest	1	2	9	Maize	28/09/2009	Maturity	stalk + cob	6,7823	37,98088		
29	2009	No-Pest	1	2	10	Maize	28/09/2009	Maturity	stalk + cob	6,794025	38,04654		
30	2009	No-Pest	1	2	11	Maize	28/09/2009	Maturity	stalk + cob	6,4624	36,18944		
31	2009	No-Pest	1	2	12	Maize	28/09/2009	Maturity	stalk + cob	5,618925	31,46598		
32	2009	No-Pest	1	2	1	Maize	28/09/2009	Maturity	seed	4,906025	78,9870025		
33	2009	No-Pest	1	2	2	Maize	28/09/2009	Maturity	seed	3,999575	67,5928175		
34	2009	No-Pest	1	2	3	Maize	28/09/2009	Maturity	seed	4,78355	86,1039		
35	2009	No-Pest	1	2	4	Maize	28/09/2009	Maturity	seed	5,7058	91,86338		
36	2009	No-Pest	1	2	5	Maize	28/09/2009	Maturity	seed	5,6235	95,03715		
37	2009	No-Pest	1	2	6	Maize	28/09/2009	Maturity	seed	6,14475	103,846275		
38	2009	No-Pest	1	2	7	Maize	28/09/2009	Maturity	seed	6,3385	107,12065		

La dimension sémantique

- ❖ Création de dictionnaires de données
- ❖ Utilisation de vocabulaires standards qui proviennent de
 - Listes contrôlées
 - Thésaurus
 - Ontologies

Nom de la variable tel que utilisé dans le fichier de données	Nom de la variable tel qu'on veut l'afficher en visualisation	Description	Valeur minimale	Valeur maximale	Méthode	Protocole
Grain	Grain	Nombre de grains	1	10000	Comptage manuelle	

Exemple : données d'annotation de séquences génomiques

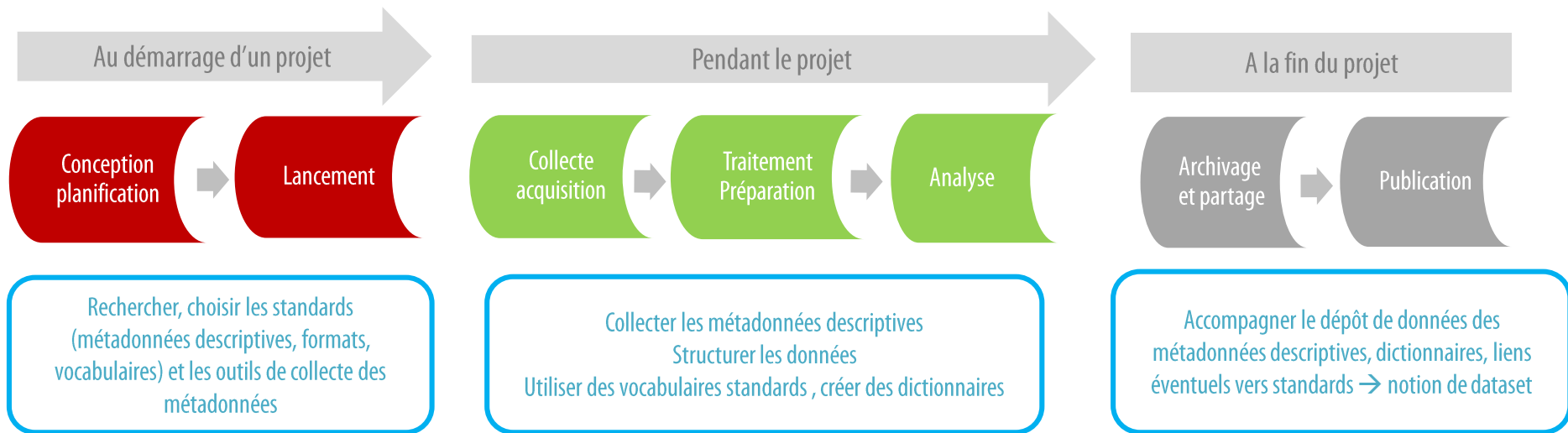
- ❖ Dublin Core pour les métadonnées descriptives
- ❖ Structuration → GFF3
- ❖ Ontologie : Gene Ontology et Sequence Ontology pour l'annotation fonctionnelle

Rôle des métadonnées

❖ Décrire et structurer le contenu dans le but de :

- Faciliter la découverte, l'accès, l'intelligibilité
 - Mots clés, localisation, contextualisation et la validation
- Favoriser la réutilisation
 - Potentiel de réutilisation, citation, termes d'usage
- Faciliter l'interopérabilité et le traitement analytique

Les métadonnées en pratique



Planifier

Search by Discipline



Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

❖ Rechercher et choisir des standards

- <http://www.dcc.ac.uk/resources/metadata-standards> ou <http://rd-alliance.github.io/metadata-directory/subjects/> ou <https://fairsharing.org/standards/>, <https://data.inra.fr/dataverse/lovinra>, <http://agroportal.lirmm.fr/>

❖ Trouver, choisir des outils / supports de collecte des métadonnées

- Bases de données
- Cahiers de laboratoire, notamment électroniques
- Carnets de terrain
- Etc.

FAIRsharing.org
standards, databases, policies



LOVInra Linked Open Vocabularies @ Inra

Voir les vocabulaires Proposer un vocabulaire Comment publier ? A propos

Recherche

Collecter les métadonnées au cours du projet

- ❖ Commencer tôt, dès l'acquisition et tout au long du cycle de vie de la donnée
- ❖ Au cours du projet
 - Décrire les plans d'expérience et/ou méthodes, les protocoles, les instruments et/ou logiciels utilisés, les conditions de production, couverture géographique, etc.
 - Tracer et documenter les modifications successives
 - Automatiser autant que possible la collecte
- ❖ Structurer les données
 - Utilisation d'un standard, création d'un dictionnaire, etc.
 - Utiliser des outils adaptés

Documentation des données lors de la mise à disposition

- ❖ Accompagner le dépôt du jeu de données des métadonnées collectées
 - les protocoles, plans d'expérience et/ou méthodes utilisés ;
 - les enregistrements de l'utilisation et de la vérification des instruments et/ou logiciels utilisés ;
 - les références aux cahiers de laboratoire et tout autre document (fiche projet, compte-rendu...) reprenant
 - les précédents points ainsi que les réflexions et décisions prises au cours du déroulé des expériences ;
 - la liste des échantillons utilisés dans les expérimentations ;
 - toute autre information jugée pertinente pour la compréhension et la reproduction des expérimentations aux données;
- ❖ Compléter avec
 - les métadonnées permettant la citation : auteurs, titre, année de production, DOI
 - les conditions de réutilisation (licence standard, conditions particulières)

Structuration des données

❖ De la consistance avant tout

- Dans le codage de certaines données (ex : genre) ou des valeurs manquantes
- Dans le nommage des variables d'un fichier à l'autre
- Dans le nommage des fichiers
- Dans le formatage des dates
- Dans les unités de mesures

❖ Mais aussi

- Éviter l'implicite (ex : coloriage)
- Faire attention à la granularité des informations
- Etc.

Merci de votre écoute