

FAIR

Des principes pour une gestion durable des données de la recherche

INTRODUCTION : Reproductibilité



Données

+

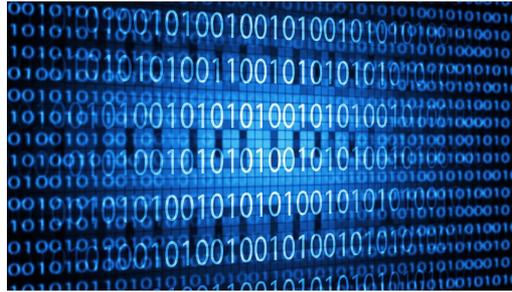


Codes



Résultats

INTRODUCTION : Reproductibilité



Données

+

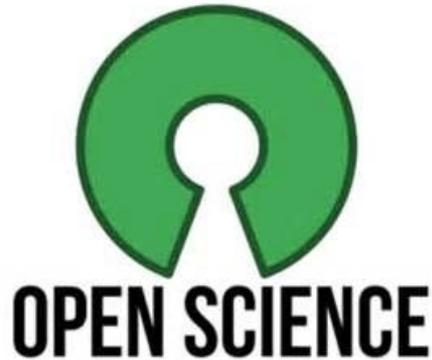


Codes



Résultats

INTRODUCTION : Les données



INTRODUCTION : Les principes FAIR

F
Findable



A
Accessible



I
Interoperable



R
Reusable



FINDABLE

Findable = Trouvable = Découvrable = Facile à trouver
(Par un humain et/ou une machine)

F A I R

3 RÈGLES

Findable

- Utilisation d'un identifiant pérenne
- Fourniture des métadonnées riches
- Publication dans un entrepôt



IDENTIFIANT PÉRENNE



DOI = *Digital Object Identifier*

~ “ISBN” des données

ISBN 978-1-234-56789-7



9 781234 567897

Digital Object Identifier

DOI <https://doi.org/10.15454/BNEUSM>

IDENTIFIANT PÉRENNE

DOI : Consortium DataCite
Agence DOI Fr : INIST CNRS



Outils pour obtenir un DOI

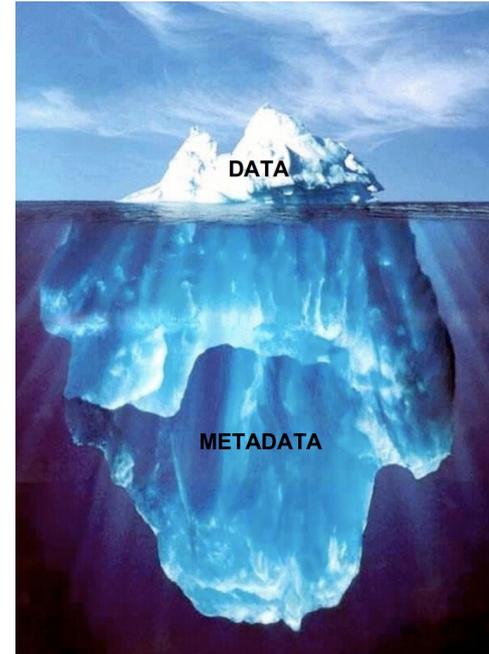
→ API DOI (Data INRAe, Zenodo, Geonetwork ...)



METADONNÉES

Métadonnée : données à propos de données

RICHERS & NORMALISÉES



DONNÉES SANS METADONNÉES



METADONNÉES

RICHES

=

Ne pas se contenter du strict minimum

Bibliothèque Centrale Provinciale
HAINAUT

8 A
no. 420

FC

Auteur(s) *Marken, Chris*
Titre *Le coem net*

I. P. - 63 - C. - 1411 - 9000

	Sorti	Signature du lecteur
1	19 OCT. 1962	<i>G. M.</i>
2	9 NOV. 1963	<i>G. M.</i>
3	10 JAN. 1964	<i>S. J.</i>
4	14 FEV. 1964	<i>M. L.</i>
5	125 SEP. 1965	<i>M. Degives</i>
6		
7		
8		
9		

METADONNÉES

RICHERS



- Fournir des mots-clés en utilisant un thésaurus
- Expliquer la provenance des données
- Fournir un maximum de “relations”
- Fournir un contact à jour
- Fournir un dictionnaire de données si possible

METADONNÉES

NORMALISÉES

= Utiliser une norme

= Se conformer à des champs descriptifs communs



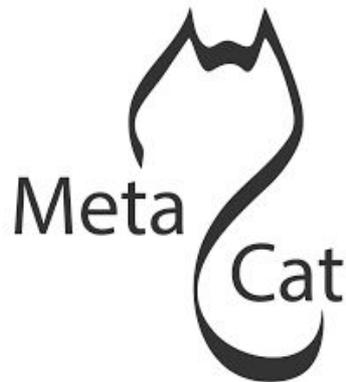
METADONNÉES

Outils pour créer des métadonnées

- Interfaces “clés en main”
- Workflow et API



GeoFlow



ENTREPÔT

Hébergement pérenne des données et métadonnées



Dataset Search

Rechercher des ensembles de données



ENTREPÔT

Visibilité & pérennité



géo catalogue



Outils pour identifier le bon entrepôt

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

ACCESSIBLE

Accessible = Facile à récupérer, selon des conditions explicites
(Par un humain et/ou une machine)

F **A** **I** **R**

2 RÈGLES

Accessible

- (Méta)données récupérables via leur identifiant et un protocole normalisé
- Métadonnées toujours accessibles même lorsque les données ne le sont plus

Accès normalisé

- Via identifiant pérenne 
 - Pas de protocole propriétaire
 - Aussi ouvert que possible,
aussi fermé que nécessaire
- ⇒ processus d'authentification si besoin



FAIR ≠ OPEN



Métadonnées toujours accessibles

- Garantissent la connaissance de l'existence du jeu de données



ACCESSIBLE

Outil pour évaluer l'accessibilité

[Intranet sur la protection des données personnelles](#)



ACCESSIBLE

Outil pour garantir l'accessibilité



INTEROPERABLE

Interoperable = Facile à combiner avec d'autres données
(Par un humain et/ou une machine)

F A I R

3 RÈGLES

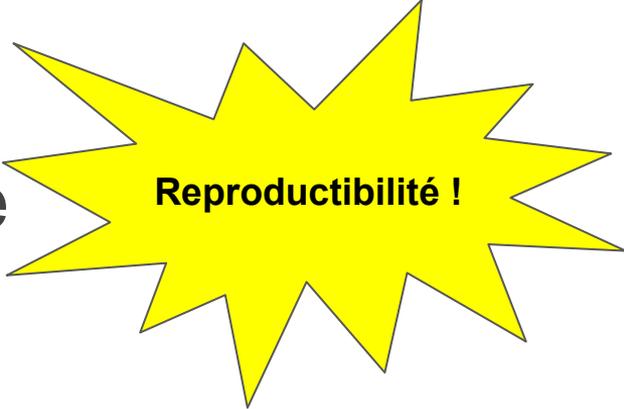
Interoperable

- Utilisation d'un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances
- Utilisation de vocabulaires qui respectent les principes FAIR
- Liens vers d'autres (méta)données



Langage partagé

- Utilisation de formats standards, libres et ouverts
exemple : CSV vs XLS
- Mise à disposition du code source
et de la documentation



Reproductibilité !



⇒ interopérabilité syntaxique

Vocabulaires contrôlés

- Utilisation de thésaurus
 - dictionnaires de mots clés partagés
 - ⇒ interopérabilité sémantique

EnvThes - Environmental Thesaurus

AGROVOC
backbone of information services



Références à d'autres (méta)données

- Contextualisation des données
 - références à d'autres documents
 - références à des articles liés





Outils pour l'interopérabilité

- Pipeline d'annotations sémantiques



- OpenSilex



REUSABLE

Reusable = Facile à réutiliser
(Par un humain et/ou une machine)

F A I R



Reproductibilité !

3 RÈGLES

Reusable



- Mise à disposition selon une licence explicite et accessible
- Informations de provenance
- Conformité avec les standards des communautés visées

Licence

- Apposer une licence sur son jeu de données

⇒ Expliciter les droits et devoirs des réutilisateurs



LICENCE OUVERTE

OPEN LICENCE

Provenance

- Expliciter la généalogie de la donnée

Quel cheminement (projet, protocole, logiciel ...) a conduit à la création du jeu de données ?



Standards disciplinaires

- Adapter la structure des (méta)données et des codes aux standards disciplinaires

Notion de dictionnaire de données

Valable également pour les codes sources

Biodiversity
Information
Standards
TDWG



EN RÉSUMÉ

Findable **A**ccessible **I**nteroperable **R**eusable



Doc (très) utile

[Datapartage INRAe](#)

[Guide des bonnes pratiques de gestion des données de la recherche](#)

[Coop IST Cirad](#)

[Go FAIR](#)

