



# {méta}données de biodiversité

1.1 définitions

1.2 Contexte et enjeux

COMPRENDRE



PARTAGER



UTILISER

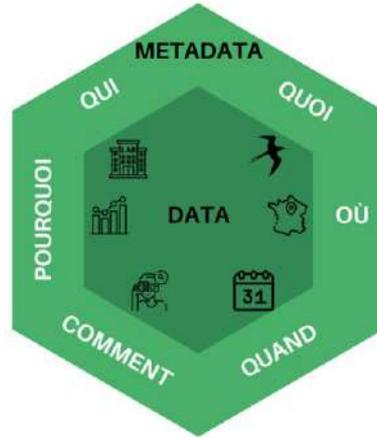


# # Comprendre les données par les métadonnées : définitions

*"Les données de recherche sont définies comme des enregistrements factuels utilisés comme sources primaires pour la recherche scientifique, et qui sont généralement acceptés dans la communauté scientifique comme nécessaires pour valider les résultats de la recherche."*

OCDE, 2007. <https://www.oecd.org/sti/inno/38500813.pdf>

OECD Principles and Guidelines for Access to Research Data from Public Funding



QU'EST-CE QUE  
**LE TRAVAIL  
SCIENTIFIQUE  
DES DONNÉES?**

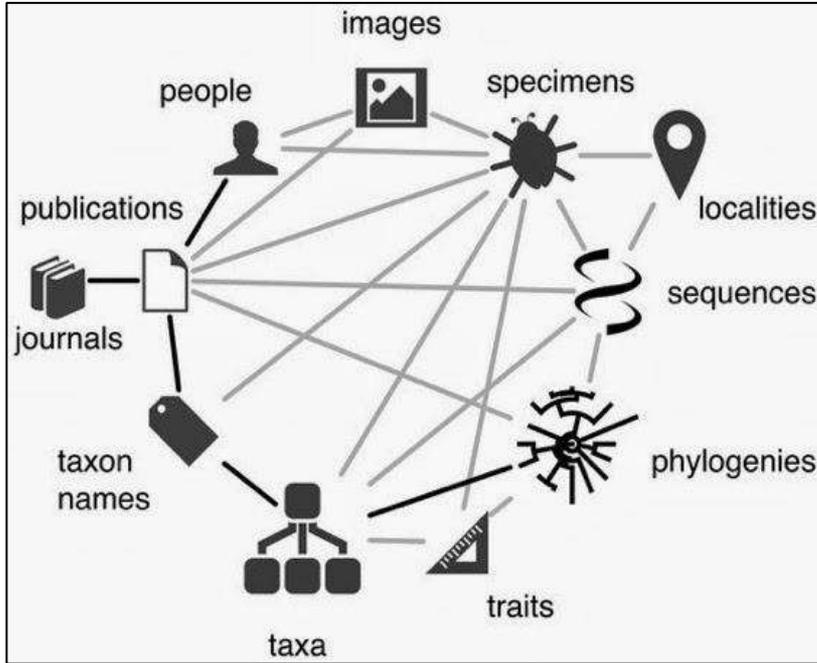
*Big data, little data, no data*

CHRISTINE L. BORGMAN

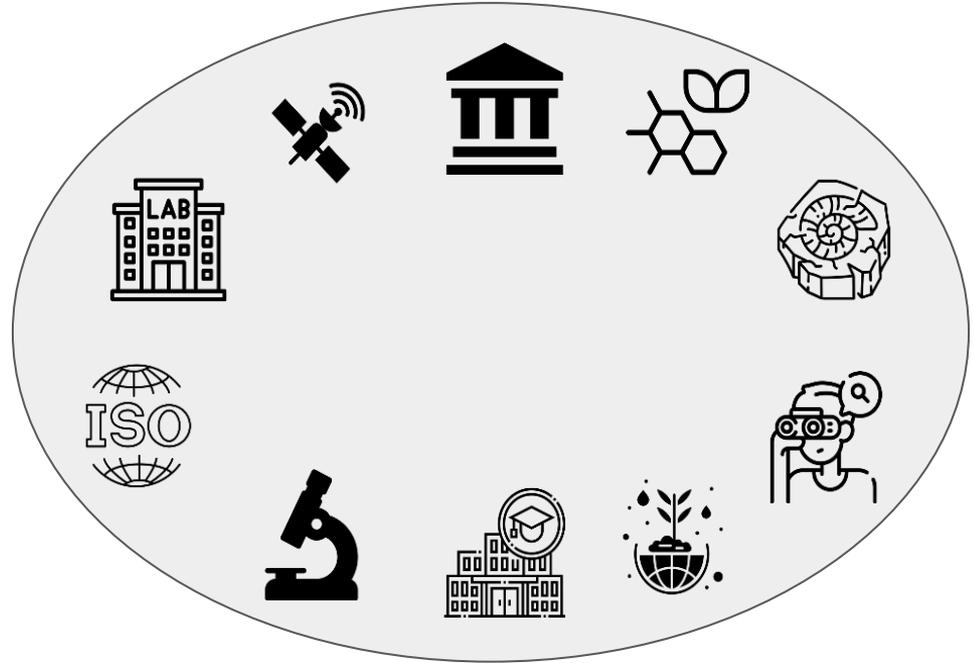
*"Les métadonnées, que l'on peut définir simplement comme « des données sur les données », sont un moyen de nommer les choses et de représenter les données et leurs relations."*

Christine L. Borgman, 2020. <https://books.openedition.org/oep/14692>

## # Comprendre les données par les métadonnées : *contexte et enjeux*



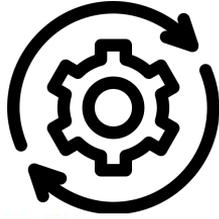
R. Page, 2016 : doi: 10.3897/rio.2.e8767



**Hétérogénéité** (types de données, provenance, standards) & **diversité "d'objets"** à relier entre eux



# # Comprendre les données par les métadonnées : *contexte et enjeux*



Ecology and Evolution

Open Access

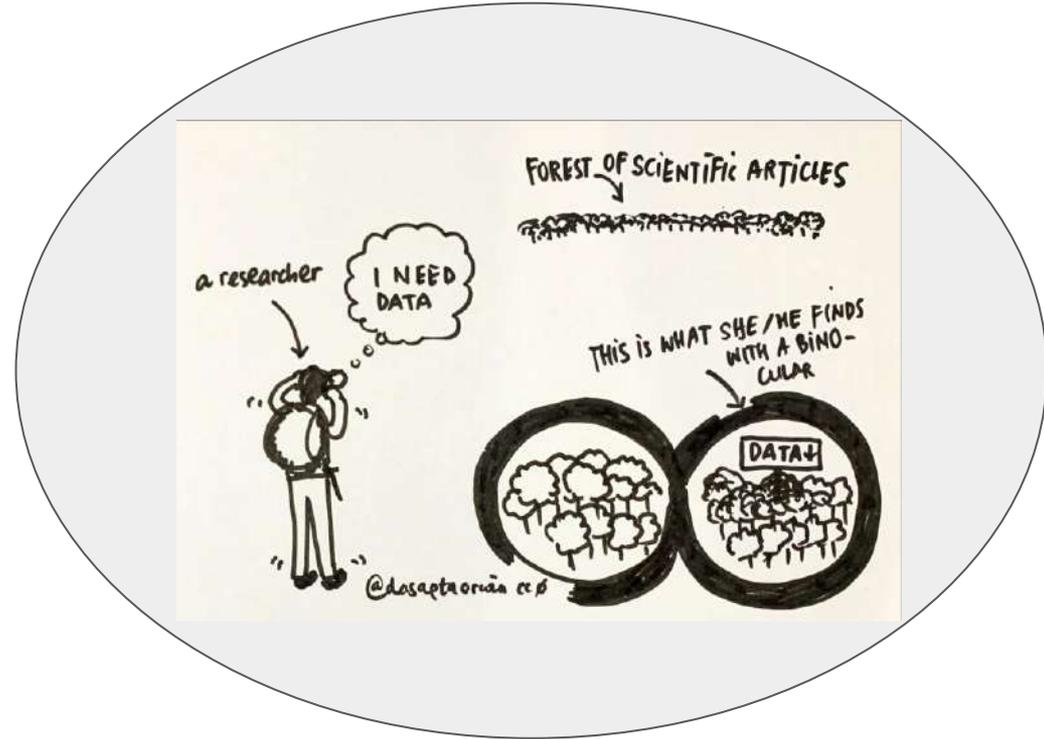
VIEWPOINT |  Open Access  

## Reproducibility in ecology and evolution: Minimum standards for data and code

Gareth B. Jenkins , Andrew P. Beckerman, Céline Bellard, Ana Benitez-López, Aaron M. Ellison, Christopher G. Foote, Andrew L. Hufton, Marcus A. Lashley, Christopher J. Lortie, Zhaoxue Ma, Allen J. Moore, Shawn R. Narum, Johan Nilsson, Bridget O'Boyle, Diogo B. Provete, Orly Razgour, Loren Rieseberg, Cynthia Riginos, Luca Santini, Benjamin Sibbett, Pedro R. Peres-Neto  
... See fewer authors 

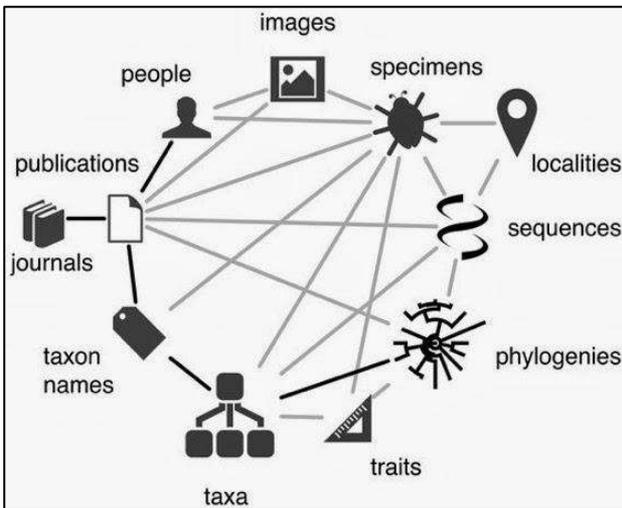
First published: 10 May 2023 | <https://doi.org/10.1002/ece3.9961>

G. B. Jenkins et al., 2023 doi.org/10.1002/ece3.9961

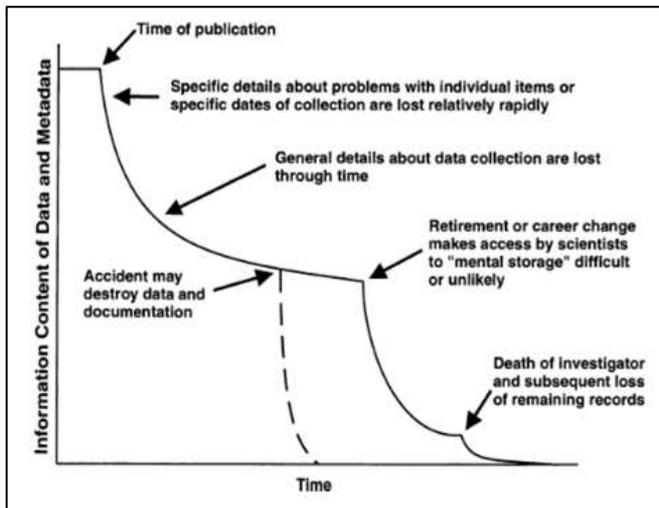


Reproductibilité

# # Comprendre les données par les métadonnées : *contexte et enjeux*



R. Page, 2016 : doi: 10.3897/rio.2.e8767



W.K. Michener et al., 1997 doi.org/10.1890/1051-0761(1997)007[0330:NMFTE5]2.0.CO;2

Ecology and Evolution Open Access

VIEWPOINT | [Open Access](#) | [CC BY](#) | [DOI](#)

Reproducibility in ecology and evolution: Minimum standards for data and code

Gareth B. Jenkins [✉](#), Andrew P. Beckerman, Céline Bellard, Ana Benítez-López, Aaron M. Ellison, Christopher G. Foote, Andrew L. Hufon, Marcus A. Lashley, Christopher J. Lortie, Zhaoxue Ma, Allen J. Moore, Shawn R. Narum, Johan Nilsson, Bridget O'Boyle, Diogo B. Provete, Orly Razgour, Loren Riesberg, Cynthia Riginos, Luca Santini, Benjamin Sibbett, Pedro R. Peres-Neto  
... See fewer authors [^](#)

First published: 10 May 2023 | <https://doi.org/10.1002/ece3.9961>

G. B. Jenkins et al., 2023 doi.org/10.1002/ece3.9961

**Hétérogénéité** (types de données, provenance, standards) & **diversité** "d'objets" à relier entre eux

**Perte d'informations dans le temps**

**Reproductibilité**

# # Contexte et enjeux en bioinformatique : *introduction*



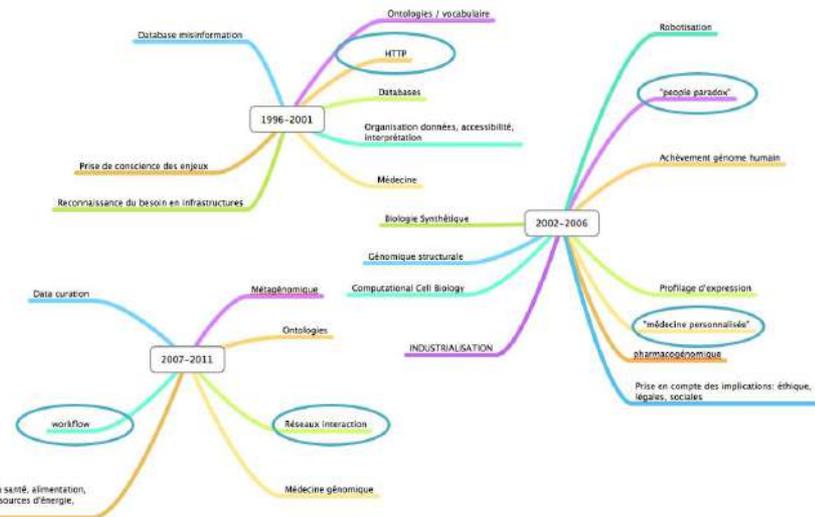
## • Wikipedia

- Concepts et techniques nécessaires à l'interprétation informatique de l'information biologique
- Champ de recherche multi-disciplinaire

- Volet recherche (élaboration de nouvelles méthodes) et volet service (analyse de données assurée par/sur les plates-formes).

- Intimement liée à la Biologie Moléculaire
- Mais également aux progrès en traitement de l'information
- <http://www.dipity.com/cistronic/BioInformatics/>

Ouzounis. Rise and demise of bioinformatics? Promise and progress. PLoS Comput Biol (2012) vol. 8 (4) pp. e1002487



# # Contexte et enjeux en bioinformatique : *acteurs*

Bio-informaticien

Informaticien

Biologiste



Produit les données pour répondre à une question biologique

Bio-analyste



Analyse les données produites en amont

Bio-informaticien



Développe de nouvelles méthodes ou de nouveaux outils

Développeur logiciel

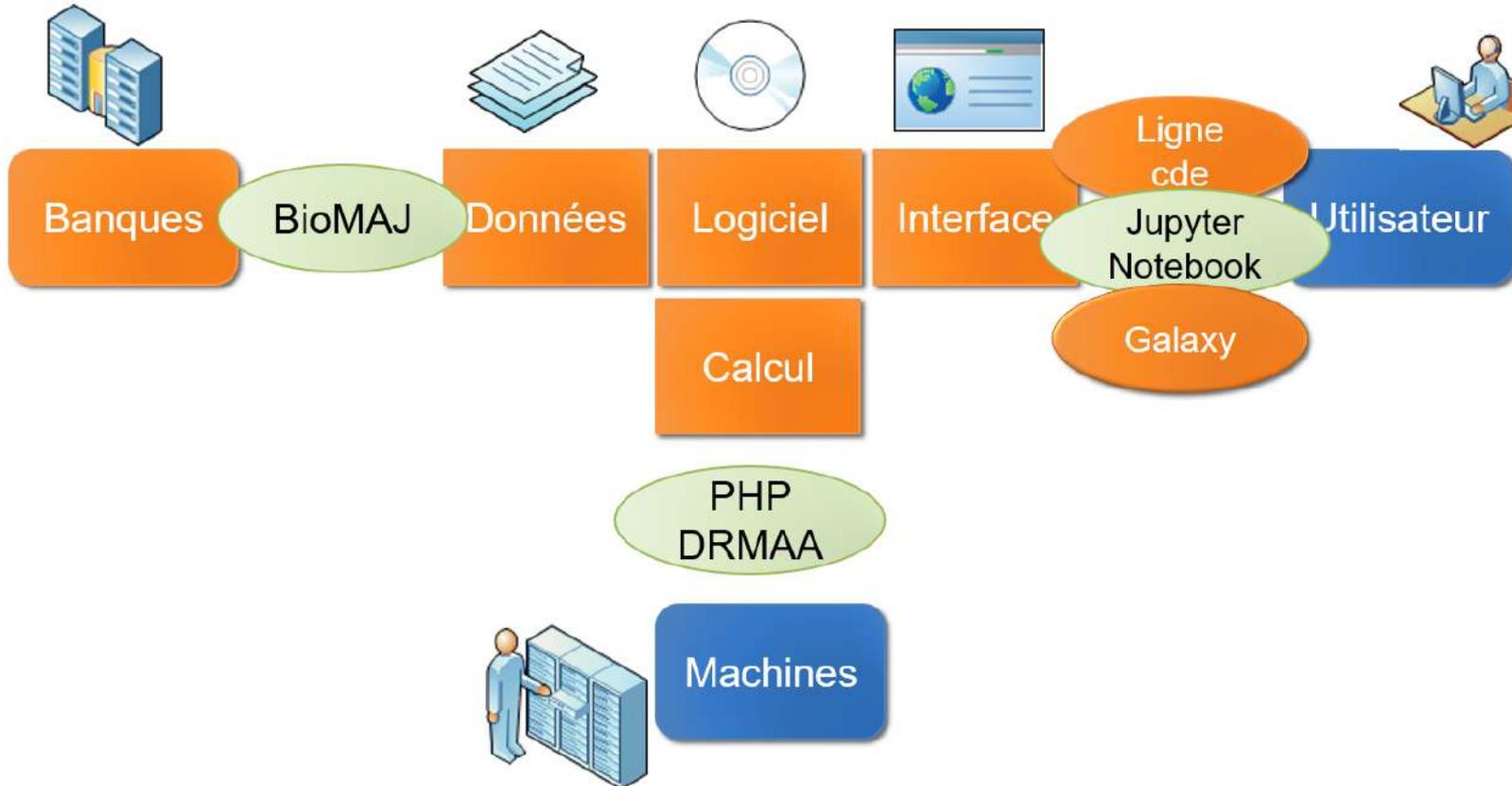


Administrateur système



Installe les outils pour l'analyse et maintient les systèmes de calcul

## # Contexte et enjeux en bioinformatique : *environnement*



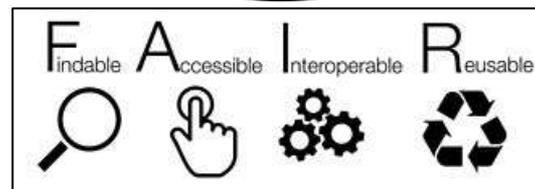
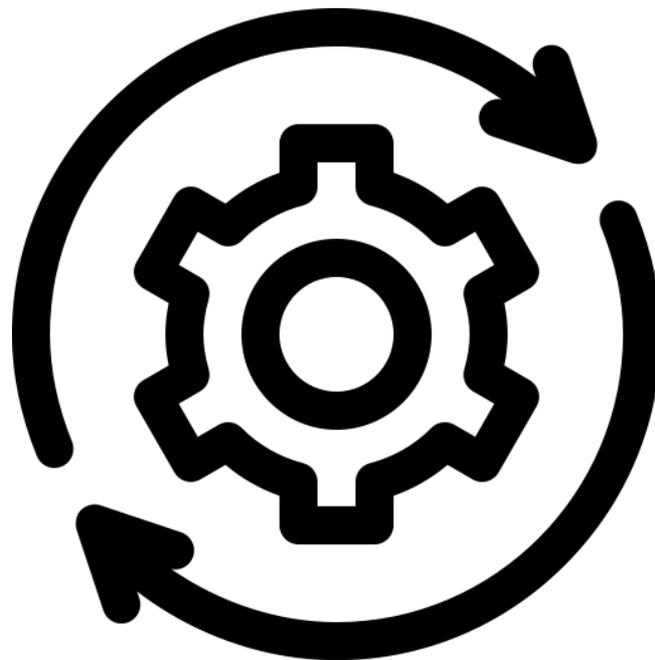
## 2. Cadre de travail & bonnes pratiques

2.1 Le cycle de la donnée

2.2 Générer et partager des  
métadonnées

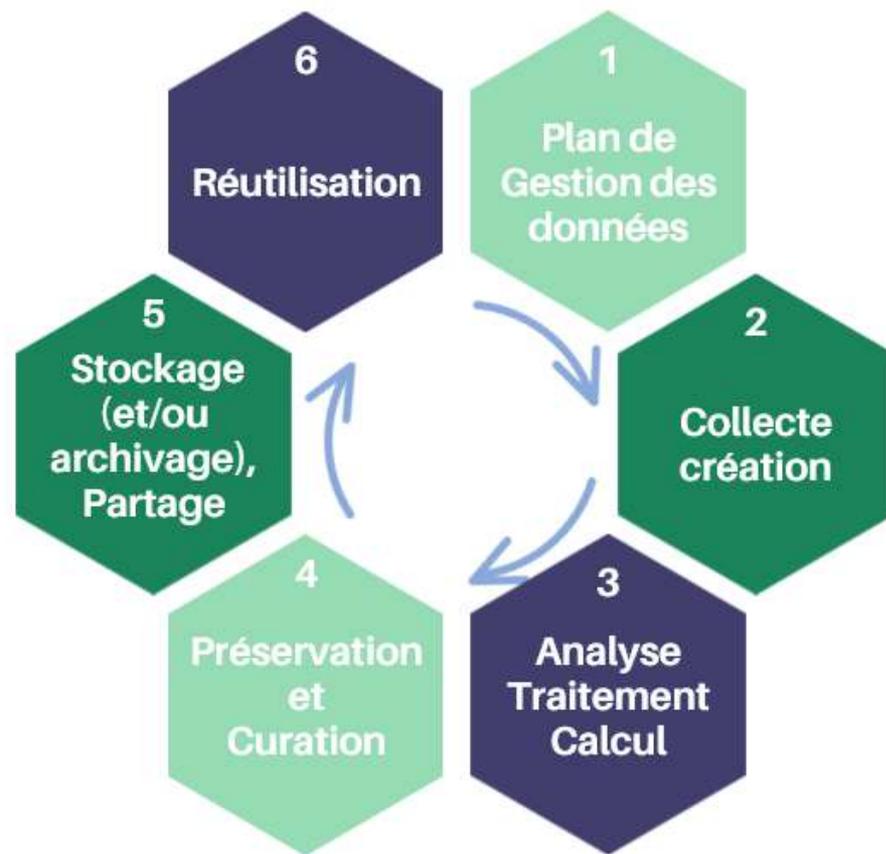
2.3 L'approche FAIR

2.4 Flux et Stocks des {méta}données

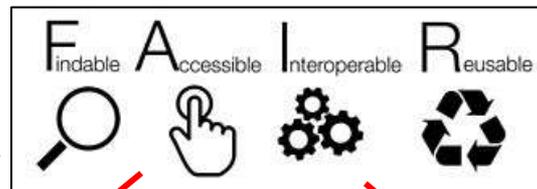


## # Cadre de travail et bonnes pratiques : le cycle de la donnée

Le cycle de vie des données est l'ensemble des étapes de gestion, de conservation et de diffusion des données de recherche, associées aux activités de recherche.



## # Cadre de travail et bonnes pratiques : *les principes FAIR*



### Facile à trouver

Métadonnées détaillées dont renseignements des attributs

### Accessible

Licence ouverte, téléchargement direct des données

### Interopérable

Standards internationaux, ontologies, thésaurus, outils open source

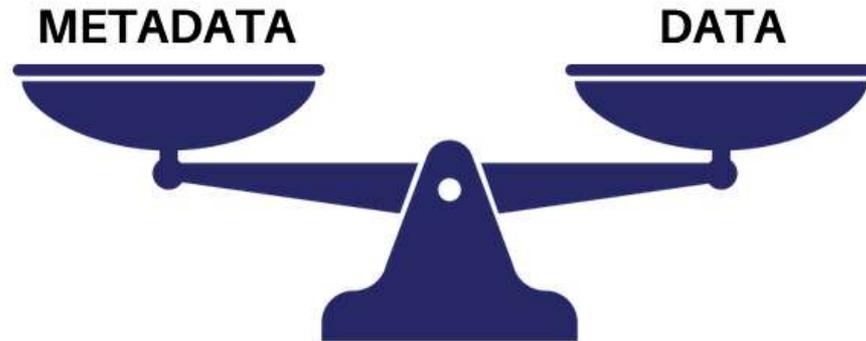
### Réutilisable

Réutilisation par mon "futur moi" et par les autres

Un ensemble de **principes directeurs** pour gérer les données de la recherche visant à les rendre **faciles à trouver, accessibles, interopérables et réutilisables** par l'**humain et la machine**.

*"This is the way !"*

## STANDARDISER LES {MÉTA}DONNÉES



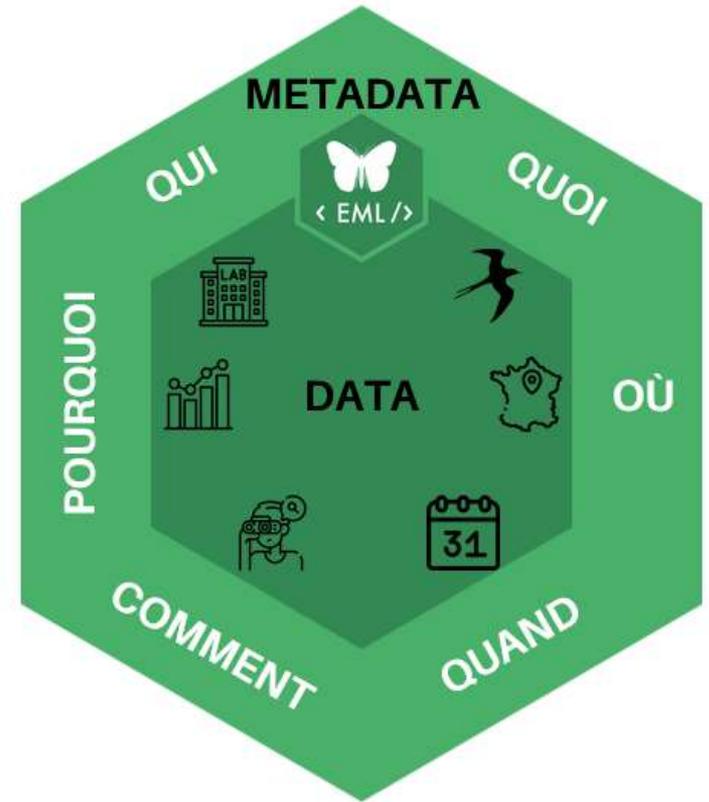
*Informations brutes vs. spécifiques et dérivées*  
*Connaissance du standard, son formalisme, ses restrictions*  
*Nécessité de connaître le temps pour la standardisation*  
*Diversité des types de données*

**COMPLÉMENTARITÉ DES DEUX APPROCHES**

---

# # Cadre de travail et bonnes pratiques : générer et partager les métadonnées

L'Ecological Metadata Language - EML- est un standard pivot mondialement reconnu et qui a plus de 25 ans de retours d'expériences par les écologues & écoinformaticiens



Standardiser par la métadonnée permet la description fine, l'inférence, l'identification et l'interopérabilité des données

# # Flux et stocks des {méta}données : processus recommandé

1

Séquences génétiques, occurrences d'espèces, liste sp, fossiles, données satellites, flux de carbone...

**Caractérisation de mon jeu de données (=métadonnées) via un accompagnement complémentaire aux services existants**

- Qui est le "data owner" ?
- De quoi parle mon jeu de données ?
- Quel est le cadre d'acquisition (recherche, politiques Publiques, ...) ?
- Où déposer ses données ?
- Quel est le format et standard utilisé ?
- Publication d'un data paper en parallèle ?

2

Exemples de Systèmes d'Informations de biodiversité dont entrepôts et catalogues de données des infrastructures

recherche.data.gov.fr

**Centres de ressources & ateliers de la donnée**

**Centres de références thématiques**

**DATA TERRA**

+ Référents entrepôts données

**outils et services pour la description, l'accès, la validation, l'analyse et la réutilisation des données de biodiversité**

**{méta}données ouvertes : moissonnage, conversion/standardisation et enrichissement (description attributs et termes issus de ressources terminologiques) des métadonnées + liens directs de téléchargement**

4

Exemples de Systèmes d'informations et catalogues de données internationaux

**{Méta}données et workflows analytiques visibles sur catalogues/instances internationales**

3

e-Infrastructure de recherche et pôle de données

# Contexte et enjeux des données en bioinformatique et écoinformatique : *exemple du*



CC BY 4.0 Olivier Norvez



Merci !  
Questions ?

Atelier Technique Odatis - mars 2024 - Bioinformatique

par **Olivier Norvez** (PNDB - Data Terra; FRB & MNHN-PatriNat), **Yvan Le Bras** (PNDB - Data Terra; MNHN-PatriNat), 2024



@PNDBiodiv & www.pndb.fr



DATA  
TERRA

