



HAL
open science

Maggot, outil de catalogage des données de recherche

Daniel Jacob, François Ehrenmann

► **To cite this version:**

Daniel Jacob, François Ehrenmann. Maggot, outil de catalogage des données de recherche. Printemps de la Donnée 2024, INRAE; Université Haute-Alsace; Université de Strasbourg; INSA; PNDB; AgroParisTech; Université de Lille; Sorbonne Université; Data Terra, Jun 2024, Bordeaux, France. hal-04682198

HAL Id: hal-04682198

<https://hal.inrae.fr/hal-04682198>

Submitted on 30 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Daniel Jacob
INRAE UMR 1332 BFP
&
François Ehrenmann
INRAE UMR BIOGECO

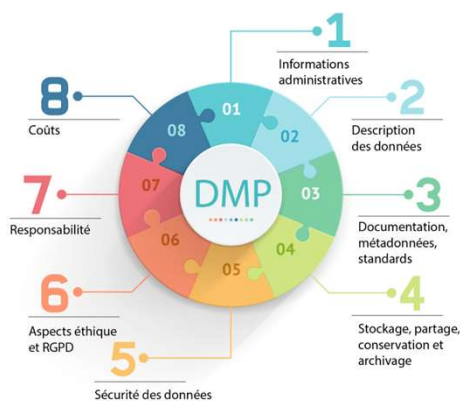


Une politique de gestion des données ?

- **De qui ?** Plans Nationaux de la Science Ouverte (2018-2021, 2021-2024)
- **Pour quoi ?** Les données sont considérées comme des produits de la recherche, au même titre que les publications, et nécessitent qu'une vraie politique de donnée soit mise en place, c.à.d avec des ressources et des actions à mener
- **Par qui ?** des activités reconnues par le DU (ex : Mission RDO à l'INRAE, Responsable Qualité dans les unités, etc...) mais à laquelle tout personnel scientifique doit participer

-> Mise en place d'un "plan de données" à l'échelle du collectif

concerne notamment, mais pas seulement, les PGD de structure et de projets, la démarche Qualité appliquée à la Donnée, Outils de Gestion de données, Métrologie, ...



Un outil de gestion des données ? Quelles motivations ?

Imaginez...

- Que vous puissiez avoir une **visibilité de ce qui est produit au sein de votre collectif**
⇒ ensembles de données, analyses, logiciels, scripts, images, sons, vidéos, ...
- Que vous puissiez **trouver facilement un jeu de données** proche de vos préoccupations sur l'espace de stockage de l'unité
⇒ Collaboration, échanges, synergies, ...
- Que ce jeu de données soit **facilement compréhensible** et **exploitable**
⇒ Gain en efficacité donc de temps
- Que vous puissiez **facilement publier vos données** sur un entrepôt de donnée de votre choix
⇒ Partage, visibilité
- Que vous puissiez **enrichir vos données** avec d'autres données externes
⇒ Apport de connaissance, collaboration, ...



Un outil de gestion des données ? Quelles motivations ?

Imaginez...

- Que vous puissiez avoir une **visibilité** de ce qui est produit au sein de votre collectif
 - ⇒ ensembles de données, analyses, logiciels, scripts, images, sons, vidéos, ...
 - ⇒ **Concerne toutes les données, pas seulement celles qui seront publiées - Éviter l'archéologie**
- Que vous puissiez **trouver facilement un jeu de données** proche de vos préoccupations sur l'espace de stockage de l'unité
 - ⇒ **Collaboration, échanges, synergies, ...**
 - ⇒ **Saisie et consultation simples sans compétences particulières sur les technologies de l'Open Data**
- Que ce jeu de données soit **facilement compréhensible** et **exploitable**
 - ⇒ **Gain en efficacité donc de temps**
 - ⇒ **Gestion de métadonnées sur le temps long du cycle de vie de la données et principes FAIR**
- Que vous puissiez **facilement publier vos données** sur un entrepôt de données de votre choix
 - ⇒ **Partage, visibilité**
 - ⇒ **Pouvoir diffuser ses métadonnées dans un ou des entrepôts en évitant de ressaisir quoi que ce soit**
- Que vous puissiez **enrichir vos données** avec d'autres données externes
 - ⇒ **Apport de connaissance, collaboration, ...**
 - ⇒ **Partage de bonnes pratiques / de standards entre communautés**





Plan de Gestion de Données (PGD)

Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour garantir cette confidentialité ?

Confidentialité

Type de structure et son affiliation ?
Responsables ?
Quelle politique en matière de gestion des données ?
Qui est responsable de la gestion des données ?

Information sur la structure et sur le plan de gestion

Quelles type de données sont produites et/ou utilisées (format, volume et augmentation...) ?
Comment seront-elles produites, puis traitées ?

Présentation générale des données

Qui sera le propriétaire des données produites ?
Comment seront-elles utilisées ?

Droits de propriété intellectuelle

Comment, où et par qui les données seront-elles stockées, sauvegardées et sécurisées ?

Stockage et sécurité des données

Qui sera en mesure d'accéder aux données ? Les données seront-elles partagées ? publiées ? Avec qui ?
Comment ?
Pendant combien de temps ? Sous quelle licence ?

Partage des données

Comment les données seront-elles identifiées, décrites ? Quelles normes de métadonnées seront utilisées ?
Comment les métadonnées seront-elles générées ?

Organisation et documentation des données

Quel est le plan pour l'archivage et la conservation à long terme ?

Archivage et conservation des données

=> Mise en œuvre / Actions concrètes



Gestion des métadonnées (et des données)



Pouvoir décrire facilement ses données (métadonnées descriptives)

- avec son vocabulaire métier
- sans saisies fastidieuses
- sans avoir à ressaisir à chaque fois les mêmes informations
- en y associant des ressources externes (*liens*)



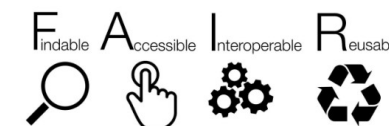
Pouvoir gérer facilement ses données

- Limiter la perte de données (*après le départ du personnel temporaire*)
- Retrouver des données (*à partir des métadonnées*)
- Ne partager que les métadonnées
- Pouvoir donner accès aux données si besoin
- Pouvoir les diffuser sans avoir à tout ressaisir



Garantir que les métadonnées suivent les principes FAIR

- Respecter un standard (schéma de métadonnées)
- Utiliser du vocabulaire contrôlé conforme à son domaine (thésaurus, ontologies)
- Être à *minima* « Trouvables, Accessibles & Interopérables »





Gestion des données

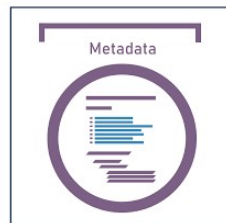
Le fil d'Ariane ...

Organiser

son espace de travail
(stockage, sauvegarde,
nommage, ...)



Décrire ses données
(métadonnées)



Partager & Rechercher des
métadonnées / données



Publier ses
métadonnées
... avec les données





Gestion des données

Le fil d'Ariane ...

Organiser

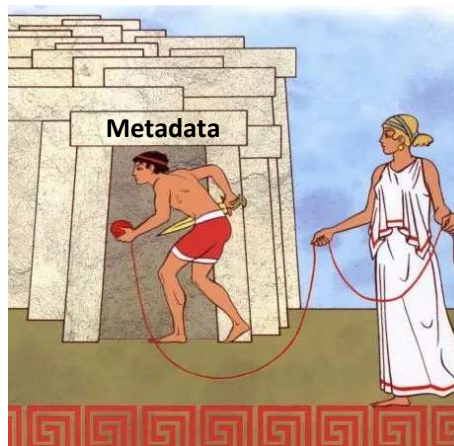
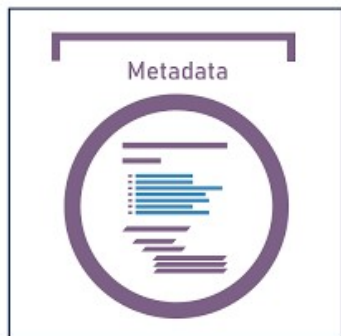
son espace de travail
(stockage, sauvegarde,
nommage, ...)

Partager & Rechercher des
métadonnées / données

Publier ses
métadonnées
... avec les données



Décrire ses données
(métadonnées)





Gestion des données

Le fil d'Ariane ...

Organiser

son espace de travail
(stockage, sauvegarde,
nommage, ...)

Partager & Rechercher des métadonnées / données



Publier ses métadonnées ... avec les données



Décrire ses données (métadonnées)

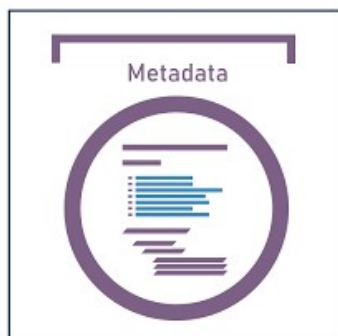


Schéma de Métadonnées

Contenant : quoi décrire

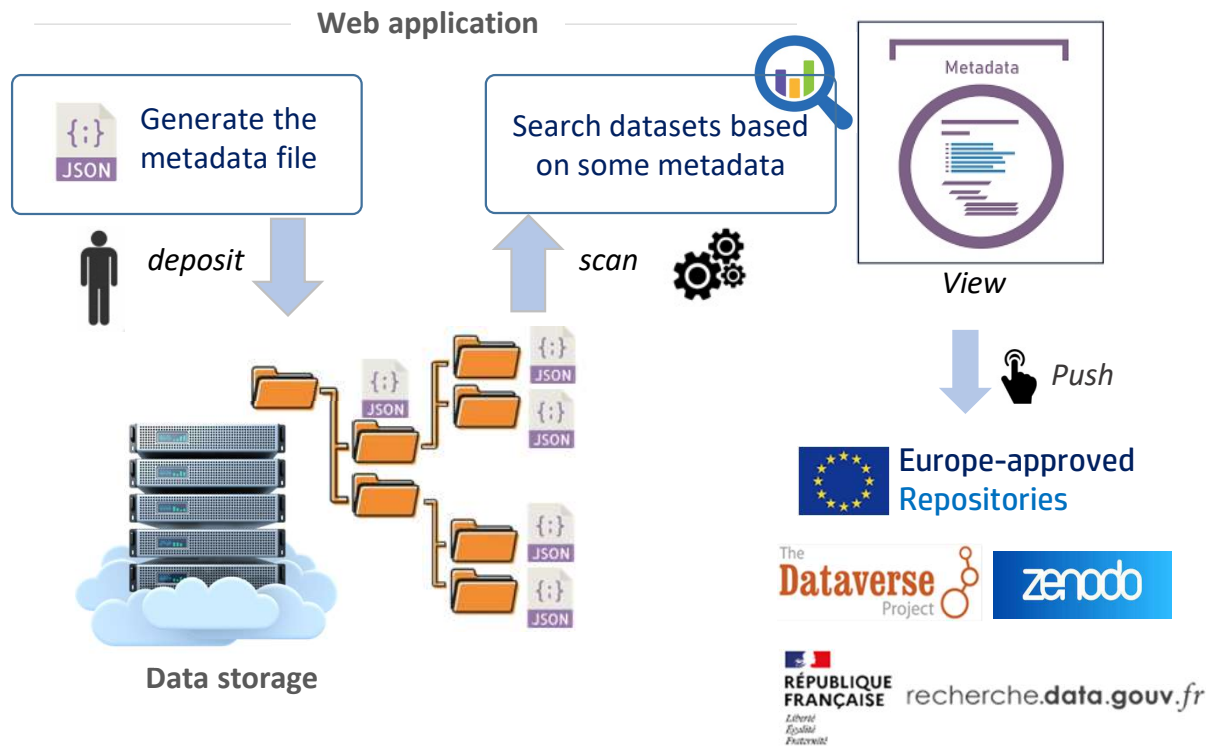


Vocabulaire Contrôlé

Contenu : comment le décrire



Un écosystème pour la gestion des métadonnées



Décrire des métadonnées
avec un vocabulaire contrôlé

Partager & Rechercher
des métadonnées / données


Publier ses métadonnées
... avec les données





Un écosystème pour la gestion des métadonnées

Démonstration

 <https://pmb-bordeaux.fr/maggot/>





L'outil Maggot permet à un collectif de :

- **Avoir une visibilité** de ce qui est produit au sein du collectif
 - ensembles de données, logiciels, bases de données, images, sons, vidéos, analyses, codes, ...
 - => **partager les métadonnées**
- **Sensibiliser** les nouveaux arrivants et les étudiants à une meilleure description de ce qu'ils produisent.
 - Limiter la perte de données (*après le départ du personnel temporaire*)
- **Favoriser le FAIR** au sein du collectif
 - particulièrement dans le cadre d'une démarche qualité

**Merci de votre
attention**



<https://inrae.github.io/pgd-mmdt/>