



HAL
open science

FAIRe de la curation sur l'entrepôt Recherche Data Gouv

Sonia Bravo, Isabelle Cao, Elias Chetouane, Cécile Delay-Artous, Mouhamed Gassama, Erwan Guillou, Marie-Christine Jacquemot-Perbal, Thomas Jouneau, Sacha Muszlak, Damien Sans, et al.

► **To cite this version:**

Sonia Bravo, Isabelle Cao, Elias Chetouane, Cécile Delay-Artous, Mouhamed Gassama, et al.. FAIRe de la curation sur l'entrepôt Recherche Data Gouv. Printemps de la Donnée 2024, INRAE; Université Haute-Alsace; Université de Strasbourg; INSA; PNDB; AgroParisTech; Université de Lille; Sorbonne Université; Data Terra, 2024, Paris, France. hal-04702516

HAL Id: hal-04702516

<https://hal.inrae.fr/hal-04702516v1>

Submitted on 19 Sep 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 4.0 International License

Bonjour et bienvenue !

Cette classe virtuelle démarrera
à xx h.

Assurez-vous que votre compte
sur le [bac à sable](#) fonctionne.
Vous en aurez besoin pendant la
classe

Le support est téléchargeable ici :



FAIRE DE LA CURATION DANS L'ENTREPÔT RECHERCHE DATA GOUV

Sessions 2024



DANS CET ATELIER

- I. PRÉSENTATION GÉNÉRALE
- II. OBJECTIFS DE LA CURATION
- III. RÔLES DANS L'ENTREPÔT
- IV. BONNES PRATIQUES

Mieux vous connaître

Avez-vous déjà un rôle dans l'entrepôt ?

- Administrateur d'espace institutionnel
- Curateur (Pour un espace ? Pour une sous-collection ?)
- Déposant (Pour vous ? Pour quelqu'un d'autre ?)

Si vous êtes déjà curateur, effectuez-vous de la curation ?

- Curation avant publication
- Curation après publication



RÉPUBLIQUE
FRANÇAISE recherche.data.gouv.fr

*Liberté
Égalité
Fraternité*

PRÉSENTATION DE RECHERCHE DATA GOUV





L'entrepôt Recherche Data Gouv

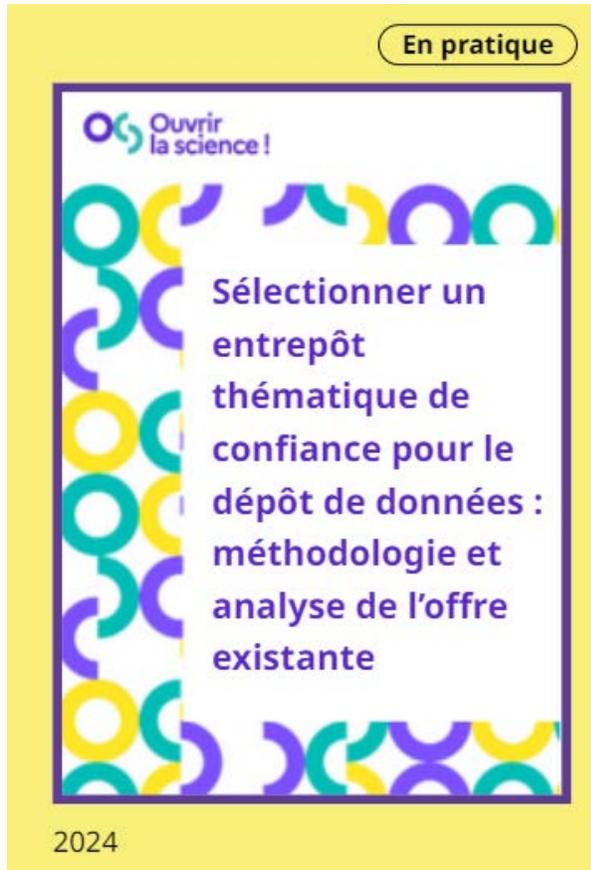
Solution souveraine pour le **partage** et l'**ouverture** des données de recherche produites par les **communautés ne disposant pas d'entrepôt disciplinaire reconnu**

Facilite **la recherche, la découverte, l'accès et la réutilisation des données** de l'établissement dans le respect des bonnes pratiques

Environnement de production

<https://entrepot.recherche.data.gouv.fr>

Liste des entrepôts thématiques de confiance du CoSO



Pour plus d'info :
[Suivre ce lien](https://recherche.data.gouv.fr/fr/entrepots)

- Produite par le **collège Données de la recherche**
- Propose une **méthode d'identification**
- S'appuie sur **des critères**
- Assortie d'une **première liste**
<https://recherche.data.gouv.fr/fr/entrepots>



RÉPUBLIQUE
FRANÇAISE recherche.data.gouv.fr

*Liberté
Égalité
Fraternité*

OBJECTIF DE LA CURATION



Les métadonnées en bref

- Données décrivant les données
- Facilite la découverte
- Pour **comprendre** les données et les conditions de production

Identifiant pérenne ?	doi:10.57745/PM55G5
Date de publication ?	2023-01-26
Titre ?	One-step synthesis of spin crossover nanoparticles using flow chemistry and supercritical CO2
Point de contact ?	Utiliser le bouton de courriel ci-dessus pour joindre la personne-contact. TOULIN, Stéphane (CNRS)
Auteur ?	Daro, Nathalie (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) - ORCID: 0000-0002-3707-9686 Chastanet, Guillaume (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) - ORCID: 0000-0001-6829-4066 Marre, Samuel (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) - ORCID: 0000-0001-8889-187X Aymonier, Cyril (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) - ORCID: 0000-0003-1775-0716 Vaudel, Tony (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) Afindouli, Luc (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France)
Contributeur ?	Data Collector : TOULIN, Stéphane (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) - ORCID: 0000-0002-2087-7403
Producteur ?	Centre National de la Recherche Scientifique (CNRS)  Université de Bordeaux (UBX) 

***Métadonnée = « Ensemble d'informations structurées qui décrit, explicite, localise une ressource informationnelle, dans le but d'en faciliter la recherche, l'usage et la gestion » (source : [DoRANum](#))**

nanoparticles using flow chemistry and supercritical CO2. Chemistry - European Journal 2020 vol.26, n°69, p.16286-16290.
DOI : 10.1002/chem.202002322 / HAL : <https://hal.science/hal-02945200>

Curation dans l'entrepôt Recherche Data Gouv

- **Vérifier** : complétude et qualité des métadonnées* descriptives, la documentation associée et les fichiers déposés.



- **Échanger** avec le déposant, **proposer** des améliorations (cf. [rapport de curation](#))
 - ➔ **Si possible, mettre en place une curation AVANT publication (meilleure réactivité des déposants)**

Objectif : conformités aux principes « FAIR », favoriser la **visibilité** et la **réutilisation**

Principes FAIR



Findable :

Les données sont trouvables de façon efficace et pérenne.



Interoperable

Les (méta)données sont utilisables entre humains et entre machines



Accessible :

Les données sont facilement récupérables avec les conditions d'accès et d'utilisation



Reusable

Les données sont exploitables pour de futures recherches ou d'autres finalités

Contributions de l'entrepôt et des déposants



- **Entrepôt** : DOI ; métadonnées indiquant l'identifiant ; entrepôt permettant la recherche
- **Déposant** : métadonnées riches et utilisant des vocabulaires



- **Entrepôt** : données accessibles avec le DOI ; métadonnées toujours disponibles
- **Déposant** : données ouvertes (éventuellement sous conditions)



- **Entrepôt** : métadonnées standardisées et *machine readable*
- **Déposant** : format de fichiers ouverts et *machine-readable* ; vocabulaires avec URI



- **Entrepôt** : licences disponibles
- **Déposant** : provenance des données ; standards correspondant à la communauté



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

recherche.data.gouv.fr

LES RÔLES DANS L'ENTREPÔT



@RechercheDataGv

Rappel sur les rôles

Rôles  Tous les rôles définis dans votre collection que vous pouvez attribuer aux utilisateurs et aux groupes.

[+ Ajouter un nouveau rôle](#)

Administrateur - Personne disposant de toutes les autorisations sur les collections, les jeux de données et les fichiers. 

[AjoutDeCollection](#) [AjoutJeuDeDonnées](#) [ConsultationDeCollectionNonPubliée](#) [ConsultationJeuDeDonnéesNonPublié](#) [TéléchargementFichier](#) [ÉditionDeCollection](#) [ÉditionJeuDeDonnées](#)
[GestionDesAutorisationsDeCollection](#) [GestionAutorisationsJeuDeDonnées](#) [GestionAutorisationsFichier](#) [PublicationDeCollection](#) [PublicationJeuDeDonnées](#) [SuppressionDeCollection](#)
[SuppressionVersionProvisoireJeuDeDonnées](#)

Contributeur - Pour les jeux de données, personne pouvant modifier la licence et les conditions, puis les soumettre pour révision. 

[ConsultationJeuDeDonnéesNonPublié](#) [TéléchargementFichier](#) [ÉditionJeuDeDonnées](#) [SuppressionVersionProvisoireJeuDeDonnées](#)

Créateur de collections - Personne qui peut ajouter des sous-collections dans une collection. 

[AjoutDeCollection](#)

Créateur de collections et de jeux de données - Personne qui peut ajouter des sous-collections et des jeux de données dans une collection. 

[AjoutDeCollection](#) [AjoutJeuDeDonnées](#)

Créateur de jeux de données - Personne qui peut ajouter des jeux de données dans une collection. 

[AjoutJeuDeDonnées](#)

Curateur - Pour les jeux de données, personne pouvant modifier la licence et les conditions, éditer les autorisations, et publier des jeux de données. 

[AjoutDeCollection](#) [AjoutJeuDeDonnées](#) [ConsultationDeCollectionNonPubliée](#) [ConsultationJeuDeDonnéesNonPublié](#) [TéléchargementFichier](#) [ÉditionJeuDeDonnées](#)
[GestionAutorisationsJeuDeDonnées](#) [GestionAutorisationsFichier](#) [PublicationJeuDeDonnées](#) [SuppressionVersionProvisoireJeuDeDonnées](#)

Membre - Personne qui peut consulter les collections et les jeux de données non publiés. 

[ConsultationDeCollectionNonPubliée](#) [ConsultationJeuDeDonnéesNonPublié](#) [TéléchargementFichier](#)

Téléchargeur de fichiers - Personne qui peut télécharger un fichier publié. 

[TéléchargementFichier](#)

Le créateur de la collection est administrateur

L'administrateur choisit les rôles des utilisateurs et peut nommer des curateurs

En l'absence de curateur, l'administrateur est en charge de la curation



Engagement des curateurs

Charte des curateurs

Mis à jour : 08/08/2022

Curateur : toute personne qui assure la curation de jeux de données en effectuant une vérification des métadonnées et des fichiers de données déposés dans la plateforme *Recherche Data Gouv*. Elle peut proposer des modifications pour améliorer la qualité documentaire des jeux de données, dans le respect de la charte des curateurs.

Les curateurs d'espaces institutionnels s'engagent, pour chaque collection dont ils ont la charge, à :

1. Veiller au respect de l'application des règles établies par les conditions générales d'utilisation de l'Entrepôt-Catalogue *Recherche Data Gouv* ;
2. Assurer la curation selon le niveau défini dans la collection, conformément au guide du curateur ;
3. Traiter les jeux de données envoyés en révision dans les meilleurs délais possibles et assurer le suivi des modifications en lien avec le déposant jusqu'à la publication ;
4. Ne pas modifier un jeu de données, en dehors de corrections mineures (syntaxe et forme), sans l'accord du déposant ;
5. Traiter rapidement les demandes d'accès aux données ;
6. Contribuer, avec l'administrateur, au repérage des jeux de données non publiés au delà d'un délai de 12 mois afin d'éviter l'utilisation de la plateforme comme un espace de stockage et non de partage des données de la recherche ;
7. Participer au réseau des curateurs de la plateforme *Recherche Data Gouv* afin d'harmoniser les pratiques et d'améliorer la qualité des données.

[Lien vers la charte des curateurs](#)





RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

recherche.data.gouv.fr

LES BONNES PRATIQUES DANS L'ENTREPÔT



@RechercheDataGv

Focus sur certaines métadonnées générales

- **Description**
- **Auteurs / Contributeurs**
- **Mots-clés**
- **Publications et jeux de données associés**
- **Financement et projet**
- **Fichiers**
- **Licence**

Exemple d'un jeu de données avant curation

Identifiant pérenne ?	doi:10.15454/DCVDZN
Date de publication ?	2022-12-15
Titre ?	Dataset of an experimental evolution on the effect of fungicide-resistance composition of populations on the durability of anti-resistance strategies on the wheat pathogen <i>Zymoseptoria tritici</i>
Point de contact ?	Utiliser le bouton de courriel ci-dessus pour joindre la personne-contact. BALLU, Agathe (INRAE) Walker, Anne-Sophie (INRAE)
Auteur ?	BALLU, Agathe (INRAE) - ORCID: 0000-0001-7104-6493 Anne-Sophie Walker (INRAE) - ORCID: 0000-0002-1246-246X Florence Carpentier (INRAE) - ORCID: 0000-0001-6204-2220 Anne Dérédec (INRAE)
Description ?	The data and the R scripts that support the findings of this study English (2021-11-22)
Langue ?	English
Sujet ?	Earth and Environmental Sciences; Me
Mot-clé ?	experimental evolution fungicide resistance
Type de données ?	Dataset
Origine des données ?	experimental data; analysis data
Logiciel ?	R
Publication associée ?	doi:
Déposant ?	BALLU, Agathe
Date de dépôt ?	2022-04-15

Ballu, Agathe; Walker, Anne-Sophie; Carpentier, Florence; Dérédec, Anne, 2022, "Dataset of an experimental evolution on the effect of fungicide-resistance composition of populations on the durability of anti-resistance strategies on the wheat pathogen *Zymoseptoria tritici*", <https://doi.org/10.15454/DCVDZN>

Exemple d'un jeu de données après curation

Identifiant pérenne	doi:10.15454/DCVDZN
Date de publication	2022-12-15
Titre	Dataset of an experimental evolution on the effect of fungicide-resistance composition of resistance strategies on the wheat pathogen <i>Zymoseptoria tritici</i>
Point de contact	Utiliser le bouton de courriel ci-dessus pour joindre la personne-contact.
Auteur	BALLU, Agathe (INRAE) Walker, Anne-Sophie (INRAE)
Description	The data and the R scripts that support the findings of this study English (2021-11-22)
Langue	English
Sujet	Earth and Environmental Sciences; Medicine, Health and Life Sciences; Agricultural Sciences
Mot-clé	experimental evolution fungicide resistance
Type de données	Dataset
Origine des données	experimental data; analysis data
Logiciel	R
Publication associée	doi:
Déposant	BALLU, Agathe
Date de dépôt	2022-04-15

Version avant curation

Identifiant pérenne	doi:10.15454/DCVDZN
Date de publication	2022-12-15
Titre	Dataset of an experimental evolution on the effect of fungicide-resistance composition of populations on the durability of anti-resistance strategies on the wheat pathogen <i>Zymoseptoria tritici</i>
Point de contact	Utiliser le bouton de courriel ci-dessus pour joindre la personne-contact.
Auteur	Ballu, Agathe (INRAE) Walker, Anne-Sophie (INRAE) Ballu, Agathe (INRAE) - ORCID: 0000-0001-7104-6493 Walker, Anne-Sophie (INRAE) - ORCID: 0000-0002-1246-246X Carpentier, Florence (INRAE) - ORCID: 0000-0001-6204-2220 Dérédec, Anne (INRAE)
Contributeur	Data Collector : Ballu, Agathe (INRAE) Data Collector : Ugazio, Claire (INRAE) Data Collector : Noly, Alicia (INRAE) Data Collector : Duplaix, Clémentine (INRAE) Data Collector : Wullschlegel, Juerg (Syngenta)
Localisation de la production	Thiverval-Grignon (78), France
Distributeur	Entrepôt-Catalogue Recherche Data Gouv https://entrepot.recherche.data.gouv.fr
Description	Data and R scripts supporting the analysis of both the qualitative and quantitative evolution, through experimental evolution, of fungicides resistances within three different population of distinct composition. The population genetics, population size and resistance phenotypic profiles were considered to bring new insights for the management of resistance (in particular multiple resistance) through selection pressure diversity, in front of various population composition. Also, some additional analysis on mfs1 expression were carried out to potentially explain some resistance phenotypic profile (enhanced efflux) English (2021-11-22)
Langue	English
Sujet	Earth and Environmental Sciences; Medicine, Health and Life Sciences; Agricultural Sciences
Mot-clé	experimental evolution fungicide resistance http://opendata.inrae.fr/thesaurus/INRAE/c_6458 (INRAE Thesaurus) https://consultation.voculaires-ouverts.inrae.fr/thesaurus-inrae/
Type de données	Dataset
Origine des données	experimental data; analysis data
Logiciel	R Version: 4.1.3
Publication associée	"Preventing multiple resistance above all: new insights for managing fungal adaptation" doi: https://dx.doi.org/10.1101/2022.12.17.520869
Informations sur la subvention	Syngenta INRAE
Informations sur le projet	PhD Agathe Ballu (Evaluating fungicide resistance management strategies by means of experimental evolution : the case of <i>Zymoseptoria tritici</i> , the causal agent of Septoria leaf blotch)
Période couverte	St
Date de collecte	St
Déposant	Ballu, Agathe
Date de dépôt	2022-04-15

Version après curation

Description - Bonnes pratiques

Description * ?

La métadonnée **Description** doit reprendre :

- l'intérêt
- les objectifs
- le contexte de production
- le potentiel de réutilisation

La description est à l'image d'un résumé d'article scientifique.

Possibilité de renseigner plusieurs descriptions (ex. en français et en anglais)



Pour mettre en forme,
utiliser les balises html

Ce champ prend en charge uniquement certaines balises HTML.

Texte * ?

```
<br>Data and R scripts supporting the analysis of both the qualitative and quantitative evolution, through experimental evolution, of fungicides resistances within three different population of distinct composition.
```

```
</br>
```

```
<br>
```

```
The population genetics, population size and resistance phenotypic profiles were considered to bring new insights for the management of resistance (in particular multiple resistance) through selection pressure diversity, in front of various population composition.
```

```
<br>
```

```
<br>Also, some additional analysis on mfs1 expression were carried out to potentially explain some resistance phenotypic profile (enhanced efflux)
```

```
</br>
```

Curation de la description

Version avant curation

The data and the R scripts that support the findings of this study English (2021-11-22)

Curation de la description

Version avant curation

The data and the R scripts that support the findings of this study English (2021-11-22)

Version après curation

Data and R scripts supporting the analysis of both the qualitative and quantitative evolution, through experimental evolution, of fungicides resistances within three different population of distinct composition.

The population genetics, population size and resistance phenotypic profiles were considered to bring new insights for the management of resistance (in particular multiple resistance) through selection pressure diversity, in front of various population composition.

Also, some additional analysis on *mfs1* expression were carried out to potentially explain some resistance phenotypic profile (enhanced efflux).

English (2021-11-22)

Auteurs & contributeurs

L'**auteur** est le responsable du jeu de données. Il est présent dans la citation du jeu de données.

Un **contributeur** est une personne ou un organisme ayant participé à la collecte, la gestion, la distribution ou toute autre contribution au jeu de données.



Le contributeur n'apparaît pas dans la citation du jeu de données.

Auteurs & contributeurs - Bonnes pratiques

Pour chaque auteur et contributeur il est important de préciser :

- l'**affiliation** correspondant aux structures auxquelles elle ou il est affilié au moment de la production des données ;
- un **identifiant** associé à l'auteur ou au contributeur (s'il existe).



L'identifiant ORCID
est recommandé

Curation de l'auteur

Version avant curation

Tang, Cathy (INRAE)
Delay Artous Cecile (Nanterre)
Françoise Bedri (CNRS Personnel des unités)

Curation de l'auteur

Version avant curation

Tang, Cathy (INRAE)
Delay Artous Cecile (Nanterre)
Françoise Bedri (CNRS Personnel des unités)

Version après curation

Tang, Cathy (DipSO UAR 1479 ; INRAE ; France) - ORCID: [0000-0003-4787-1690](https://orcid.org/0000-0003-4787-1690)
Delay-Artous, Cécile (SCD ; Université Paris Nanterre ; France) - ORCID: [0000-0003-1273-9681](https://orcid.org/0000-0003-1273-9681)
Tisserand-Bedri, Françoise (Inist-CNRS ; CNRS ; France) - ORCID: [0000-0002-9019-4106](https://orcid.org/0000-0002-9019-4106)

Data Curator : Sans, Damien (DipSO UAR 1479 ; INRAE ; France) - ORCID: [0000-0002-7621-4748](https://orcid.org/0000-0002-7621-4748)
Data Collector : Jarrige, Virgile (DNum ; Université de Strasbourg ; France) - ORCID: [0000-0003-0140-7024](https://orcid.org/0000-0003-0140-7024)
Project Manager : Sowinski, Claire (Inist-CNRS ; CNRS ; France) - ORCID: [0000-0002-1048-8041](https://orcid.org/0000-0002-1048-8041)
Supervisor : Desiles, Camille (LAM UMR ; Université de Bordeaux ; CNRS) - ORCID : [0009-0007-5740-9893](https://orcid.org/0009-0007-5740-9893)

Mots-clés

Mot-clé* = « Mot ou groupe de mots caractérisant un objet numérique et/ou son contenu » (source : [Thesaurus INRAE](#))

- Pose une convention sur l'usage d'un terme (par ex. au singulier) ;
- Crée des **ponts sémantiques** ;
- Permet d'éviter les problématiques de **bruit** ou de **silence** lors de la recherche de contenus.

Mots-clés - Bonnes pratiques

Pour permettre l'interopérabilité il est recommandé d'utiliser un **vocabulaire de référence** dans la discipline concernée s'il en existe

- [AgroVoc](#),
 - [ANAEE](#),
 - [GEMET](#),
 - [MeSH](#),
 - [Pactols](#),
 - [Thésaurus INRAE](#),
 - [Loterre](#),
 - [PhySH](#),
- En connaissez vous d'autres ?



**Il est recommandé de
mettre 4-5 mots-clés**

Curation des mots-clés

Version avant curation

Decision support system
wireworm
risk assessment
insect pest
Pest Management

Curation des mots-clés

Version avant curation

decision support system
wireworm
risk assessment
insect pest
pest management

Version après curation

decision support systems http://aims.fao.org/aos/agrovoc/c_49868
(Agrovoc) <https://agrovoc.fao.org/>
wireworms http://aims.fao.org/aos/agrovoc/c_f34f3fc0 (Agrovoc)
<https://agrovoc.fao.org/>
risk assessment http://opendata.inrae.fr/thesaurusINRAE/c_4522
(INRAE Thesaurus) <https://thesaurus.inrae.fr/>
pest insects http://aims.fao.org/aos/agrovoc/c_5729 (Agrovoc)
<https://agrovoc.fao.org/>
pest management http://aims.fao.org/aos/agrovoc/c_13262
(Agrovoc) <https://agrovoc.fao.org/>

Exemples de mots-clés

Version avant curation

EPR
ABC transporter

Version après curation

Electron Spin Resonance Spectroscopy

<https://meshb.nlm.nih.gov/record/ui?ui=D004578> (MeSH Thesaurus)

<https://www.nlm.nih.gov/mesh/meshhome.html>

ATP-Binding Cassette Transporters

<https://meshb.nlm.nih.gov/record/ui?ui=D018528> (MeSH Thesaurus)

<https://www.nlm.nih.gov/mesh/meshhome.html>

Drug Resistance, Bacterial

<https://meshb.nlm.nih.gov/record/ui?ui=D024881> (MeSH Thesaurus)

<https://www.nlm.nih.gov/mesh/meshhome.html>

Multidrug resistance ABC transporter ATP-binding/permease protein

BmrA <https://www.uniprot.org/uniprotkb/O06967/entry> (UniProt)

<https://www.uniprot.org/>

Exemples de mots-clés

Version avant curation

Oxydation
Pyrolyse
Acetone

Version après curation

oxidation <http://data.loterre.fr/ark:/67375/37T-PDFDL1P2-1>
(Chemistry Loterre) <https://skosmos.loterre.fr/37T/en/>
pyrolysis <http://data.loterre.fr/ark:/67375/37T-J9R0T3W3-9>
(Chemistry Loterre) <https://skosmos.loterre.fr/37T/en/>
acetone <http://data.loterre.fr/ark:/67375/37T-Z5VRVHW7-G>
(Chemistry Loterre) <https://skosmos.loterre.fr/37T/en/>
chemical engineering <http://data.loterre.fr/ark:/67375/QJP-L5QQJ6XQ-B> (Engineering sciences vocabulary Loterre)
<https://skosmos.loterre.fr/QJP/en/>
flame <http://data.loterre.fr/ark:/67375/37T-TQF1V4LK-L> (Chemistry Loterre) <https://skosmos.loterre.fr/37T/en/>

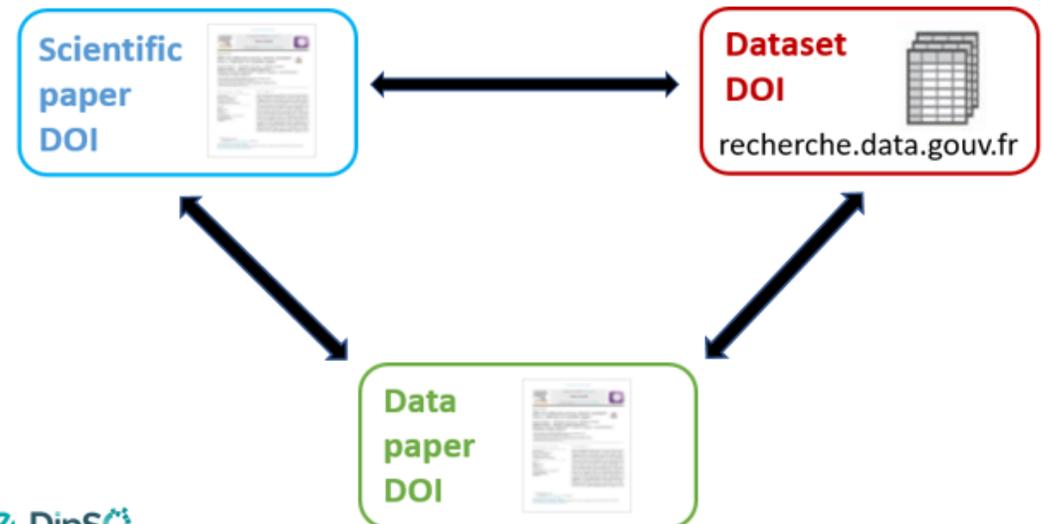
Références aux publications et jeux de données

Renseigner toute publication réutilisant les données dans **Publication associée**

Renseigner tout jeu de données dont les données sont liées dans **Jeu de données associé**

Pour améliorer la **visibilité** des différents objets, il est conseillé de faire des liens en mentionnant leur citation et DOI.

 La première publication associée sera sous la citation, la liste complète sera affichée dans l'onglet « Métadonnées »



Curation des références

Version avant curation

"A decision support system based on Bayesian modelling for pest management: application to wireworm risk assessment in maize fields". Published in : Smart Agricultural Technology

Curation des références

Version avant curation

"A decision support system based on Bayesian modelling for pest management: application to wireworm risk assessment in maize fields". Published in : Smart Agricultural Technology

Version après curation

Julien Roche, Manuel Plantegenest, Philippe Larroudé, Jean-Baptiste Thibord, Le Cointe Ronan, et al.. A decision support system based on Bayesian modelling for pest management: Application to wireworm risk assessment in maize fields. Smart Agricultural Technology, 2023, 4, pp.100162. ⟨hal-04087700⟩ doi: 10.1016/j.atech.2022.100162 <https://doi.org/10.1016/j.atech.2022.100162>

Informations sur le financement et le projet

Afin de donner des éléments de **contexte** pour la production des données, il est important d'indiquer lorsqu'ils existent :

- les informations sur la **subvention** du jeu de données
- le **projet** à l'origine de la production des données.



- Pour un projet européen, les informations sont publiées sur le site : <https://cordis.europa.eu/projects/fr>
- Saisie facilitée pour un projet financé par l'ANR

Curation du financement et du projet

Version avant curation

**Aucune
information
renseignée**

Version après curation

Informations sur la subvention

CNRS: ARED-2019

Agence de l'eau Seine Normandie: AESN LS187322 n°1083739

Région Bretagne: ARED SUCHY LS 192033 coH19003

Informations sur le projet

LEARN (IMPACT DE L'Arasement des barrages sur les flux hyporhéiques et les échanges nappe rivière de la Sélune) <https://programme-selune.com/fr/projets-en-cours/flux-hyporheiques-echanges-nappes-rivieres/>

Fichiers - Ouverture des données

Vérifier que les fichiers ont vocation à être partagés :

- vérifier qu'il n'y a pas de données sensibles
- vérifier l'anonymisation des données personnelles ou identifiantes

En cas de doute, dire au déposant de contacter son service juridique ou son délégué à la protection des données



Voir les [restrictions éventuelles sur les données](#)
[les recommandations CNIL](#)

Fichiers - Bonnes pratiques

Promouvoir les **formats ouverts** ou largement utilisés.

Le **titre du fichier** doit :

- être clair et permettre de comprendre le contenu du fichier
- ne pas contenir de caractères spéciaux comme & * % \$ £ ° { ! @. Préférer le caractère souligné ou le tiret au lieu du point ou de l'espace.
- éviter les accents et majuscules



- **Taille limite par fichier : 50 Go**
- **Si déposés ailleurs, renseigner la métadonnée **Lien vers les données****
- **Si fichier modifié, supprimé ou ajouté après publication, alors il faut republier le jeu de données.**

Fichiers tabulés

Dans Recherche Data Gouv, les fichiers tabulés (xlsx, csv, tsv...) sont **ingérés** sous la forme d'un **fichier .tab** qui est un format ouvert et compatible avec la préservation à long terme.

Ce traitement d'ingestion permet d'**extraire des métadonnées** supplémentaires sur les variables du fichier et de proposer des fonctionnalités de **prévisualisation** et d'**exploration**.



Vous pouvez vous apercevoir que votre fichier a été mal ingéré, soit par un message d'erreur, soit par un nombre de variables incohérent

Recommandations pour une bonne ingestion

- **encodage en UTF-8** des fichiers contenant des caractères spéciaux,
- le **format csv** doit utiliser la **virgule comme séparateur**,
- si les cellules contiennent des virgules, le texte doit être **entre guillemets**,
- **pas d'en-tête vide ni de cellule manquante**,
- chaque **en-tête de colonne** doit avoir un **nom différent**,
- **pas de cellule fusionnée**,
- **pas de légende**,
- **pas plus de 1024 colonnes**,
- chaque **classeur de données** ne doit contenir qu'un **onglet/feuille**.



Attention : s'il y a plusieurs onglets, seul le premier est ingéré par le logiciel Dataverse et sera pris en compte dans l'affichage, l'exploration et l'export au format tabulé.

Fichier de documentation (README) - Bonnes pratiques

La documentation d'un jeu de données doit être suffisante pour permettre à n'importe quel réutilisateur de comprendre et d'évaluer sa qualité.

Le fichier README fournit des **informations complémentaires et accessibles** (lisibles par les humains). Si applicable, il peut aussi contenir le **dictionnaire des variables** et la **convention de nommage**.

Privilégier les formats texte plein (.txt), ou markdown (.md)



Extrait du modèle README

```
# INFORMATIONS METHODOLOGIQUES

## Conditions environnementales / experimentales :

## Description des sources et méthodes utilisées pour collecter et générer les données :
<Si applicable, décrire les standards, les informations de calibration, les instruments utilisés, etc.>

## Méthodes de traitement des données :
<Si applicable, décrire le traitement des données et inclure tout détail pouvant être important pour réutiliser ou reproduire les données. Commenter chaque étape.
Par exemple, inclure les méthodes de nettoyage et d'analyse ; les codes et/ou algorithmes ; les procédés d'anonymisation ou de pseudonymisation pour les données sensibles concernant les humains ou des espèces menacées>

## Procédures d'assurance-qualité appliquées sur les données :

## Autres informations contextuelles :
```

Licence - Bonnes pratiques

Recommandation : utiliser une [licence autorisée](#) par la **Loi pour une République numérique** pour le type de fichier.

Aide pour choisir une licence CC : <https://chooser-beta.creativecommons.org/>

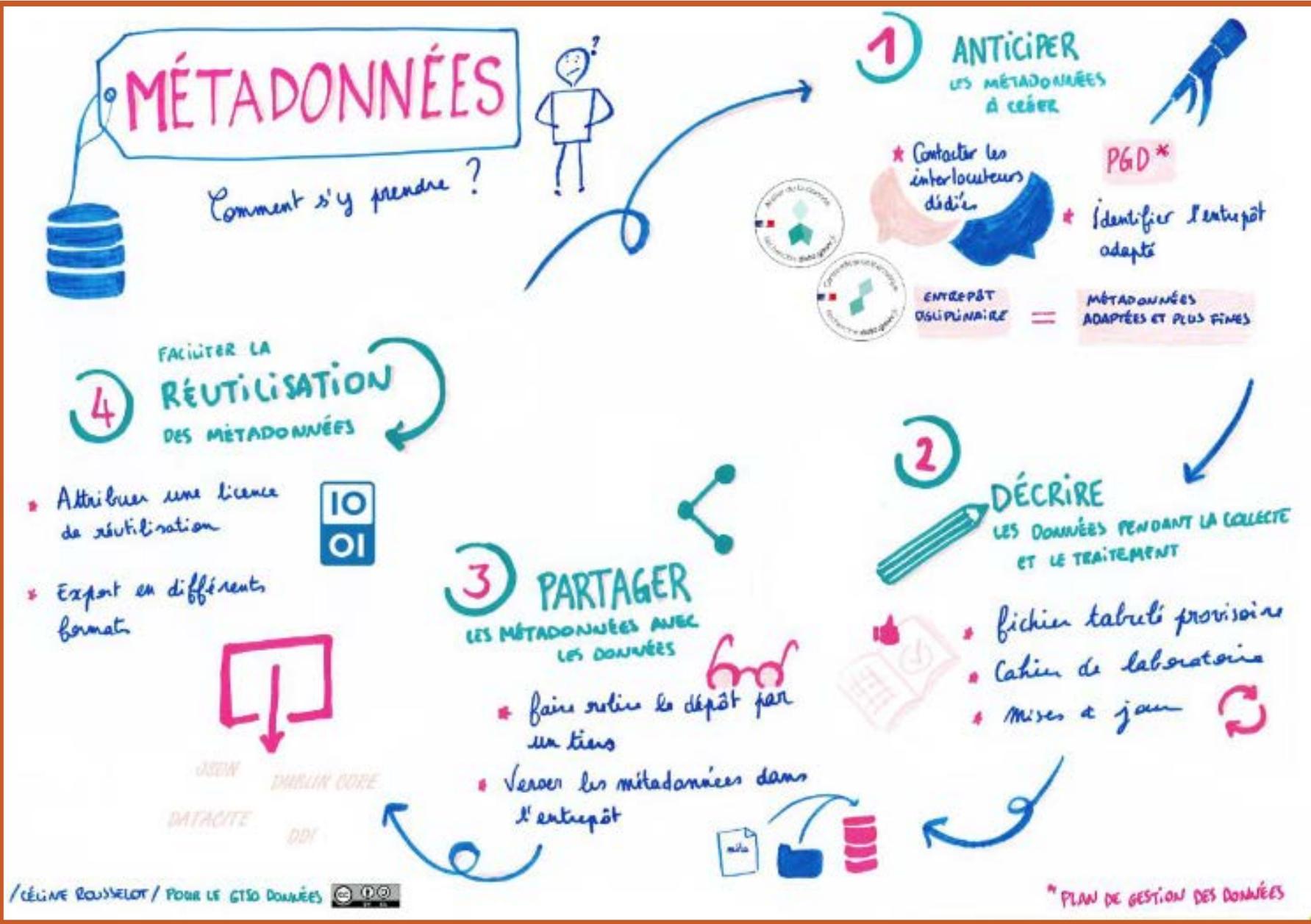
Par défaut, la licence attribuée à un jeu de données est la [licence ouverte etalab2.0](#).



**Pour mettre en forme,
utiliser les balises html**

Et il existe d'autres blocs de métadonnées

- géospatiales
- sciences humaines et sociales
- science de la vie
- revue
- ressources sémantiques
- astronomie et astrophysique
- workflow de calcul



Bracco, L., Brenel, M., Du Pasquier, D., Gauvrit, G., Heude, C., Lemeltier, D., Rousselot, C., & Zuindeau, L. (2023, septembre 25). Décrire ses jeux de données dans les règles du FAIR : accompagner les chercheurs à l'utilisation des métadonnées. Zenodo. <https://doi.org/10.5281/zenodo.8376581>

Présentation des ressources proposées pour l'entrepôt

- [Guide du déposant](#)
- [Guide de saisie des métadonnées](#)
- [Modèle de rapport de curation](#)
- [Modèle fichier README](#)



RÉPUBLIQUE
FRANÇAISE recherche.data.gouv.fr

Liberté
Égalité
Fraternité

ATELIER CURATION



Ce bac-à-sable de Recherche Data Gouv est dédié aux tests et démonstrations. Toutes les données sont effacées le dernier lundi de chaque mois et peuvent l'être avant cette date.

Bac à sable



Lien d'accès à la collection « Atelier curation » :
<https://demo.recherche.data.gouv.fr/dataverse/atelier-curation-2006>

 Les données sont effacées tous les mois.

Consigne

Travailler sur un jeu de données et identifier ce qui pourrait être amélioré :

- Soit l'un de vos jeux de données
- Soit l'un des jeux de données test créé spécifiquement pour l'atelier
 - Open Science Ontology
 - Science Ouverte - Web, Littérature et Législation
 - Gestion optimale des accès à la science ouverte
 - Données de réplication pour :



Vous pouvez vous aider soit du guide de saisie des métadonnées, soit du modèle de rapport de curation



Cliquer sur « **Envoyer** » après chaque réponse

Open Science Ontology

<https://digistorm.app/p/1190645>

Science Ouverte - Web, Littérature et Législation

<https://digistorm.app/p/7868552>

Gestion optimale des accès à la science ouverte

<https://digistorm.app/p/7888195>

Données de réplication pour :

<https://digistorm.app/p/8343117>

Si vous travaillez sur l'un de vos jeux de données :

<https://digistorm.app/p/9436690>



Atelier - Vos retours

- Avez-vous trouvé facilement les informations dans les documents d'aide ?
- Avez-vous repéré facilement des éléments à améliorer dans le jeu de données ?

JDD1 – Open Science Ontology

- pas de fichier, pas de lien vers les données
- tous les auteurs dans le même champs
- spécifier le type de contribution pour les contributeurs
- mots-clés :
 - en FR et EN : proposer titre et description dans les 2 langues
 - associer mot-clé à un vocabulaire ouvert
- bloc des métadonnées de ressource sémantique : mentionner Type de ressource = Ontology
- sujet (Business and Management) incohérent avec thématique du jeu de données
- licence personnalisée non précisée

JDD 2 Science Ouverte - Web, Littérature et Législation

- auteur sans affiliation et sans ORCID (alors que le type d'identifiant est renseigné)
- un seul auteur et pas de contributeur : voir si d'autres personnes ont contribué
- mot-clé : lien renvoie à la page d'accueil du thésaurus et non à l'URI du terme
- date de collecte : date de début > date de fin
- subvention ANR mentionnée dans la description —> à renseigner dans la métadonnée adéquate
- Fichiers zippés : pas de fichier README
- Licence non ouverte : copyright

JDD3 Gestion optimale des accès à la science ouverte

- Auteurs :
 - mise en forme incorrecte (cf point de contact)
 - les liens ORCID ne fonctionnent pas (type d'identifiant non précisé)
- Pas de balise pour la mise en forme de la description
- Publication et jdd associé sans lien et sans citation des auteurs
- Pas de mots-clés
- Licence personnalisée non précisée

JDD4 Données de réplication pour :

- Titre trop générique
- Description insuffisante
- Mots-clés :
 - URL Vocabulaire = URI terme
 - Mélange FR et EN
- Publication associée : DOI indiqué au lieu de idHAL
- Source de données pourrait être complété
- Titre des fichiers pas assez précis et contenant des points.
- Fichier csv mal ingéré



RÉPUBLIQUE
FRANÇAISE
*Liberté
Égalité
Fraternité*

recherche.data.gouv.fr

POUR SUIVRE TOUTES LES ACTUALITÉS

<https://recherche.data.gouv.fr/fr>



@RechercheDataGv



<https://www.linkedin.com/company/recherche-data-gouv/>

Support réalisé par Sonia Bravo (INRAE) ; Isabelle Cao (CNRS-INIST) ; Elias Chetouane (UGA) ; Cécile Delay-Artous (Univ. Paris-Nanterre) ; Mouhamed Gassama (INRAE) ; Erwan Guillou (CNRS-INIST) ; Marie-Christine Jacquemot-Perbal (CNRS-INIST) ; Thomas Jouneau (Univ. de Lorraine) ; Sacha Muzslak (INRAE) ; Damien Sans (INRAE) ; Maria-Grazia Santangelo (UGA) ; Claire Sowinski (CNRS-INIST) ; Dimitri Szabo (INRAE) ; Cathy Tang (INRAE) ; Françoise Tisserand (CNRS-INIST)



@RechercheDataGv