

Décrire son jeu de données dans un fichier '*readme*' avant publication



Tutoriel pas à pas destiné aux membres de l'UMR ECOBIO pour décrire un jeu de données dans un fichier 'readme', avant de le publier sur *data.IndoRES*, l'entrepôt de données du CNRS Ecologie & Environnement (E&E), et de la cataloguer sur *OSURIS*.

Décrire son jeu de données

Décrire son jeu de données avant de le déposer et le publier sur un entrepôt de données (ex: data.InDoRES) et de renseigner le catalogue OSURIS.

- 1) Préconisations de l'Institut CNRS E&E
- 2) A quoi sert le fichier 'readme'? Quel contenu?
- 3) Proposition de contenu
- 4) Exemples de readme
- 5) Qui saisit les métadonnées? Où et quand?

1) Décrire le jeu de données : préconisations (CNRS E&E)

- On suppose que le jeu de données est prêt à être publié (tel qu'il le serait pour être envoyé à une revue scientifique).
 → Pour la constitution, la structuration du jeu de données, le choix du format des fichiers de données, voir les préconisations:

<https://www.indores.fr/index.php/publier-mes-donnees-pas-pas>

- On propose ici d'aider à la préparation un fichier '*readme*' décrivant le contenu du jeu de données (archive ou simple fichier). Ex: description des attributs, variables etc.
- **Préconisations CNRS E&E:** le fichier '*readme*' est **obligatoire** lors du dépôt sur un entrepôt de données sur *Data.InDoRES*.
- De plus, il sera utile pour compléter les métadonnées:
 - ✓ d'une part sur l'entrepôt (par exemple *data.InDoRES*),
 - ✓ d'autre part sur le catalogue (*OSURIS* ou bien *Cat.InDoRES*).

2) A quoi sert le fichier 'readme'? Quel contenu?

- **Le but de ce 'readme'** est de fournir et de rassembler toutes les **informations** et les **références** nécessaires à la compréhension et la réutilisation du jeu de données.
- Son contenu et sa forme dépendent de votre discipline scientifique, de la nature et de la forme des données.
- Il est conseillé de soigner la qualité du 'readme'. Les informations doivent être aussi précises que possible, et d'une qualité compatible avec une revue scientifique. Le 'readme' fait partie intégrante du dépôt sur l'entrepôt de données, avec les fichiers de données. Il suffira alors de transmettre le DOI du jeu de données aux revues lors de publications scientifiques.
- Le fait d'avoir préalablement rassemblé toutes informations utiles dans un 'readme' facilite et accélère la saisie de ces informations lors du dépôt sur l'entrepôt de données.
- Formats préconisés pour ce 'readme': texte (.txt), (.csv)

(*data.InDoRES* accepte les fichiers tableurs; si le jeu de données est au format tableur, on peut donc rédiger le 'readme' sur une feuille de tableur; mais de manière générale les tableurs sont déconseillés pour pérenniser les données, il est préférable d'utiliser des fichiers texte)

- Pour vous guider, une *proposition* de contenu disponible à l'URL:

<https://accueil.osuris.fr/tutoriel/les-metadonnees/> (au format .ods ou .xlsx)

[MetadataPartyZAAr2024 Fichier readme contenu A PREPARER.xlsx](#) (au format .xlsx)

[MetadataPartyZAAr2024 Fichier readme contenu A PREPARER.ods](#) (au format .ods)

3) Proposition de contenu de 'readme' (1/3)

- Titre/intitulé/nom du jeu de données
- Date de création du jeu de données
- Date de publication du jeu de données
- Type de document (papier, numérique, audio, vidéo, tableur, carte etc.)
- Description/résumé du jeu de données
- Etat de finalisation du jeu de données
- Auteur(s) / contact(s)
- Affiliation /organisme
- Identifiant auteur
- Personne(s) ressource(s)

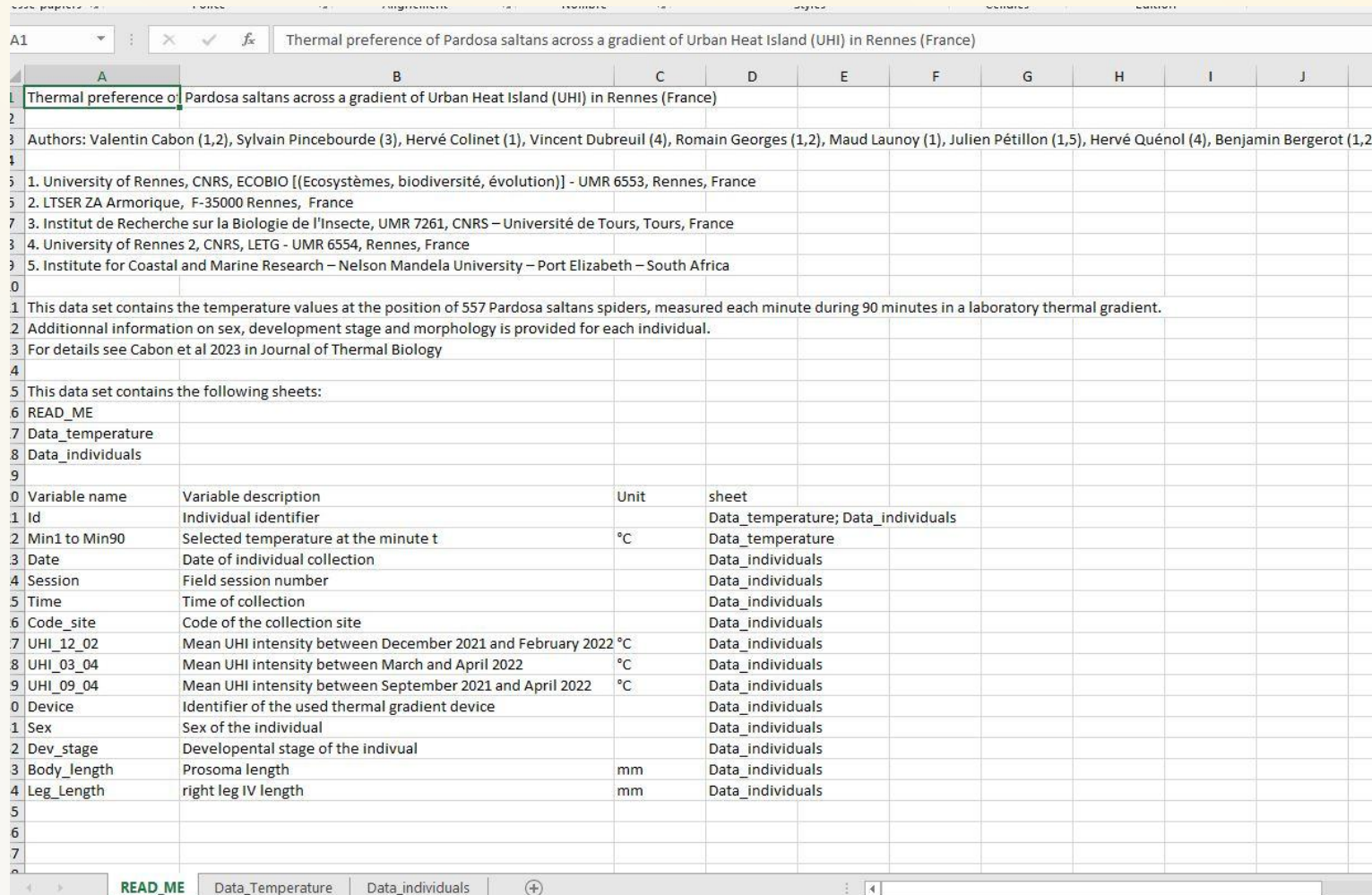
3) Proposition de contenu de 'readme' (2/3)

- Fréquence de mise à jour de votre jeu de données
- Mots-clés libres (ou thésaurus)
- Conditions d'utilisation, contraintes sur la ressource (sauf cas particuliers, licence CC-BY 4.0)
- Type de représentation spatiale (si pertinent)
- Résolution spatiale (si pertinent)
- Langue utilisée pour le jeu de données
- Encodage des caractères pour le jeu de données
- Sujet/grandes catégories: champs disciplinaires auxquels se rattache votre jeu données

3) Proposition de contenu de 'readme' (3/3)

- Etendue temporelle des données
- Emprise géographique des données (si pertinent)
- Format de distribution de la donnée
- Distributeur de la donnée
- Généalogie
- Auteur des métadonnées
- Point de contact pour les métadonnées

4) Exemples de 'readme' : (4-1) exemple 1: onglet de tableur



Thermal preference of Pardosa saltans across a gradient of Urban Heat Island (UHI) in Rennes (France)

Authors: Valentin Cabon (1,2), Sylvain Pincebourde (3), Hervé Colinet (1), Vincent Dubreuil (4), Romain Georges (1,2), Maud Launoy (1), Julien Pétillon (1,5), Hervé Quéno (4), Benjamin Bergerot (1,2)

1. University of Rennes, CNRS, ECOBIO [(Ecosystèmes, biodiversité, évolution)] - UMR 6553, Rennes, France
2. LTSER ZA Armorique, F-35000 Rennes, France
3. Institut de Recherche sur la Biologie de l'Insecte, UMR 7261, CNRS – Université de Tours, Tours, France
4. University of Rennes 2, CNRS, LETG - UMR 6554, Rennes, France
5. Institute for Coastal and Marine Research – Nelson Mandela University – Port Elizabeth – South Africa

This data set contains the temperature values at the position of 557 Pardosa saltans spiders, measured each minute during 90 minutes in a laboratory thermal gradient.
 Additional information on sex, development stage and morphology is provided for each individual.
 For details see Cabon et al 2023 in Journal of Thermal Biology

This data set contains the following sheets:

- READ_ME
- Data_temperature
- Data_individuals

Variable name	Variable description	Unit	sheet
Id	Individual identifier		Data_temperature; Data_individuals
Min1 to Min90	Selected temperature at the minute t	°C	Data_temperature
Date	Date of individual collection		Data_individuals
Session	Field session number		Data_individuals
Time	Time of collection		Data_individuals
Code_site	Code of the collection site		Data_individuals
UHI_12_02	Mean UHI intensity between December 2021 and February 2022	°C	Data_individuals
UHI_03_04	Mean UHI intensity between March and April 2022	°C	Data_individuals
UHI_09_04	Mean UHI intensity between September 2021 and April 2022	°C	Data_individuals
Device	Identifier of the used thermal gradient device		Data_individuals
Sex	Sex of the individual		Data_individuals
Dev_stage	Developmental stage of the individual		Data_individuals
Body_length	Prosoma length	mm	Data_individuals
Leg_Length	right leg IV length	mm	Data_individuals

4) Exemples de 'readme' : (4-2) exemple 2: fichier txt

```

read_me.txt - Bloc-notes
Fichier Edition Format Affichage Aide

This is a data set on some morphological and physiological impacts of salinity on two isolated of
Two non axenic isolated colonial Microcystis aeruginosa strains (PMC1323.21 and PMC1262.20) were u
For these experiments, colonies of both strains were grown in BG11 medium in Erlenmeyer flasks and
For details see Bormans et al 2023 in MicrobiologyOpen

This data set contains the following sheets:
read_me
cell_diameter_1323.21
cell_diameter_1262.20
volume_colonies_7d_1262.20
volume_colonies_3d_1262.20
autofluorescence_1323.21
autofluorescence_1262.20
intercellular_spacing_1262.20

Variable name  Variable description  Unit  sheet
S              salinity                    all
              cell diameter  µm    cell diameter 1323.21, cell diameter 1262.20
              volume colonies µm3  volume colonies 7 d 1262.20, volume colonies 3 d 1262.20
              autofluorescence          autofluorescence 1323.21, autofluorescence 1262.20
              intercellular spacing  µm    intercellular spacing 1262.20
              % dead cells      %
              time            day or hours
EtOH          ethanol 80%      autofluorescence 1323.21, autofluorescence 1262.20

```

5) Qui saisit les métadonnées? Où et quand? (métadonnées = informations contenues dans le readme)

1. Lors du dépôt du jeu de données dans l'entrepôt de données(Data.InDoRES), **le déposant saisit les métadonnées** associées au jeu de données (cf. tuto_data_ecobio_02_deposer)
2. Une fois le dépôt du jeu de données effectué sur un entrepôt (quel que soit l'entrepôt, Data.InDoRES ou un autre), **il faut saisir les métadonnées sur le catalogue de l'OSERen (GeoNetwork d'OSURIS)**. Le déposant s'adresse à SISAE-Data qui pourra générer cette fiche à partir du dépôt sur Data.InDoRES et du fichier 'readme'.

Remarque:

Jusqu'à maintenant, la politique du CNRS-E&E est d'inventorier les jeux de données de l'institut sur un catalogue accessible depuis le catalogue Cat.InDoRES de l'institut. Pour cataloguer les données d'Ecobio, il a été convenu d'utiliser OSURIS, le catalogue de l'OSERen.

D'où l'intérêt de préparer soigneusement le contenu des métadonnées dans un fichier 'readme'. Une application est en cours de développement pour faciliter le transfert des métadonnées de l'entrepôt au catalogue (générer la création d'une fiche de métadonnées GeoNetwork à partir des métadonnées saisies sur Data.InDoRES).

Liens utiles

- Geocatalogue OSURIS, portail des données de l'OSUR: <https://www.osuris.fr/geonetwork/>
- Site web du projet OSURIS: <https://accueil.osuris.fr/>
- Site web de présentation du projet InDoRES: <https://www.indores.fr/>
- Liste des formats préconisés pour les jeux de données.
<https://www.indores.fr/sites/default/files/documents/04FormatsListe.pdf>



« Décrire son jeu de données avant de le déposer sur l'entrepôt *data.InDoRES* et de le cataloguer sur *OSURIS* », Le Moal F., Tutoriels Data Ecobio, 2026.

<https://creativecommons.org/licenses/by/4.0/>