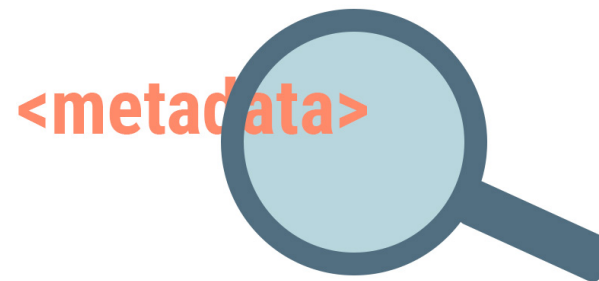


# La description d'un jeu de données

Facundo Muñoz  
facundo.munoz@cirad.fr



*» Votre collaborateur les plus important êtes vous-même dans le passé. Mais il ne répond pas à vos é-mails.*

# Principes

1. Pouvoir **transmettre** des données avec toute les infos nécessaires pour travailler avec
2. Éviter des **erreurs** d'interprétation

# Méta-données

- Ficher Readme / Lisez-moi

Description succincte du projet, la source des données et méthodes de collecte, le contexte, les objectifs, références, contacts, etc.

- Description de variables (*data-dictionary*)

- Le nom de la variable tel qu'il apparaît
  - Version du nom adaptée pour visualisations
  - Type de variable (catégorielle, quantitative, etc.)
  - Unités de mesure
  - Rang possible de variation, valeurs possibles,
  - Description

# Dictionnaire de données

C'est en soi un jeu de données !! E.g. :

	A	B	C	D
1	name	plot_name	group	description
2	mouse	Mouse	demographic	Animal identifier
3	sex	Sex	demographic	Male (M) or Female (F)
4	sac_date	Date of sac	demographic	Date mouse was sacrificed
5	partial_inflation	Partial inflation	clinical	Indicates if mouse showed partial pancreatic inflation
6	coat_color	Coat color	demographic	Coat color, by visual inspection
7	crumblers	Crumblers	clinical	Indicates if mouse stored food in their bedding
8	diet_days	Days on diet	clinical	Number of days on high-fat diet

# Conclusions

- Très **vagues suggestions**, plusieurs choix et options
- Peu importe le **système** ou les outils :
  - Fichiers de texte associés
  - Dépôt avec des métadonnées intégrées (e.g. Dataverse)
  - Formats spécifiques
  - Plan de Gestion de Données
- Choisir un et respecter les **principes**

# Références

- Karl W. Broman & Kara H. Woo (2018). Data organisation in Spreadsheets. *The American Statistician*, 72:1, 2-10, DOI: [10.1080/00031305.2017.1375989](https://doi.org/10.1080/00031305.2017.1375989)

# Merci!

Diapositives créées à l'aide du package R **xaringan**.

En s'appuyant sur **remark.js**, **knitr**, et **R Markdown**.



Ce(tte) œuvre est mise à disposition selon les termes de la **Licence Creative Commons Attribution - Partage dans les Mêmes Conditions 4.0 International**.