



Ecole Thématique DATA SDUE

**Guide de Survie dans la jungle des données
en Sciences de l'Univers et de l'Environnement (SDUE) :
Comment gérer les données pour les valoriser?**

**Session « Choix et utilisation de formats de fichiers
standards, ouverts et libres »**

Atelier Format CSV

Juliette Fabre

Atelier Format CSV

Des bonnes pratiques pour permettre de plus facilement :

- comprendre
- ré-utiliser
- partager
- analyser
- comparer
- ..

Atelier Format CSV

- **pas de lignes / colonnes qui ne servent à rien**
 - sauf éventuelles lignes de métadonnées en entête préfixées par #
- 1ère ligne = entête (noms de colonnes)
- **noms des colonnes :**
 - pas de caractère spécial, pas d'espace. utiliser des _ (ex water_temperature)
 - compréhensibles (ex : wt vs water_temperature)
- **1 colonne = 1 paramètre (= variable)**
- **1 ligne = 1 observation / individu / date**
- **1 format par colonne :** numérique, texte, date
 - ne pas mettre du texte dans une colonne qui contient du numérique (ex : hauteur = "+ ou - 2")
 - si besoin ajouter une colonne "commentaire"
- **1 valeur par cellule**
- **valeurs manquantes :** vide ou NA
- **dates :**
 - format classique et homogène, sans texte (ex : 04/06/2008 ou 2008-06-04 01:00:00)
 - si la précision est le mois, on peut ajouter une colonne "précision" = "mois" (mais ne pas mettre "nov. 2020")
 - fournir en UTC de préférence pu préciser le fuseau horaire (time zone)
- **variables qualitatives / facteurs :** utiliser des codes compréhensibles
- **encodage :** UTF-8
- **séparateur :** , ou ; ou tab (tsv)