

**INRAE**

**➤ FAIR-DS**

Démonstration du 08 - 04 - 2024

# Les templates

## FAIR-DS – un outil pour la qualité

- Des champs communs – les métadonnées
  - Opérateur : annotateur de l'observation, technicien de la mesure  
→ Je peux retrouver mes mesures / observations.  
  
Je peux ainsi retrouver les analyses que j'ai déjà produites et donc les échantillons déjà analysés  
  
Quelqu'un peut se référer à moi pour des remarques concernant l'analyse. L'échantillon était-il convenable ?
  - Date de ma mesure  
→ Je peux retrouver toutes les mesures d'un atelier produites un certain jour, ou sur une période donnée  
  
Les mesures ont-elles la même qualité l'été et l'hiver ?



# Les templates

## FAIR-DS – un outil pour la qualité

- Des champs communs – les métadonnées
  - Expérience
    - Je peux retrouver mes expériences au travers de différents types de mesures
    - Lien facile avec le projet et les études réalisées – Possibilité d'ajouter ces variables au fichier de résultat ?
  - SampleId
    - Je peux retrouver les analyses effectuées sur tel ou tel échantillon provenant de tel animal ou tel aliment
    - Je pourrais, par la suite, faire le lien avec l'historique de l'animal et faire des analyses poussées avec de nombreuses métadonnées.



# Les templates

## FAIR-DS – un outil pour la qualité

- Des champs communs – les métadonnées
  - Mode Opérateur
    - Je peux retrouver plus précisément le type de mesure effectuée sur mon échantillon
    - Je peux ainsi comparer facilement des modes opératoires sur des analyses de mêmes échantillons
    - Je peux savoir de quel manière à été produite mon analyse
  - Machine / Logiciel
    - Je peux retrouver les mesures effectuées par tel ou tel outil
    - Possibilité de comparer la qualité des mesures par comparaison des outils...



# Les templates

## FAIR-DS – un outil pour la qualité

- Des champs communs – les métadonnées
  - Réactif – référence / lot
    - Je peux retrouver la référence des produits utilisés pour les manip
    - Le fournisseur a-t-il signaler une anomalie a posteriori?
    - Je peux moi-même le savoir en comparaison avec d'autres réactifs



# Les templates

FAIR-DS – un outil pour la qualité

- Des variables fixes et génériques
  - Savoir ce qui est déposé dans la base de données
    - Recherches facilités
      - Je veux obtenir les données ou le résultat est tant.
        - Difficile si j'ai différentes syntaxes
          - Result, RESULTAT, Résultats, ...reslta
        - Je sais que le champ résultat existe
        - Je n'ai pas besoin de demander au collègue s'il a déposé des données et champ de recherche utiliser
    - Contrôle automatique de la qualité possible
      - Intervall de valeur, présence/absence de la valeur, pattern...
    - ...



# Les templates

FAIR-DS – un outil pour la qualité

- Des variables supplémentaires
  - Non recommandé, pour les raisons cités
  - Parfois utile tout de même (insérer un calcul ?)
  - Parfois nécessaire (données comportementales)



# Les templates

## Exemples

- **Données d'observation comportementales**
  - Merci à Caroline Clouard
- **Données de mesure d'azote et de matière sèche**
  - Merci à Rémy Delagarde et Nicole Huchet qui ont aussi participer aux tests
- **Données de mesure de créatinine**
  - Merci à Clément Ribas qui a aussi participer aux tests



# Les templates

1<sup>er</sup> exemple - Comportement

- Deux fichiers de données d'observations
  - Selon une métadonnées spécifiée
    - Ici : ???
- Métadonnées commune complétées, selon pertinence
- Des champs fixes : à voir
- Des champs variables : toutes les observations



# Les templates

1<sup>er</sup> exemple - Comportement

- Comment déposer ces fichiers sur le serveur ?
- Puis-je récupérer ces fichiers tels quels par la suite?
- Puis-je faire une synthèse des deux fichiers ?
- Admettons qu'on ait un troisième fichier, je veux récupérer la synthèse, uniquement des fichiers 1 et 2.



# Les templates

## 1<sup>er</sup> exemple - Comportement

- Des données peu explicites...
  - Je veux récupérer uniquement les données de tel projet et tel essai
  - Je veux récupérer les données de tel vidéo et tel animal, car je sais que cet animal a un historique spécifique
    - Les échantillons et l'association mesure - échantillon
- Un dossier personnel ?
  - Mais je souhaite avoir un répertoire projet et partager quelques données seulement avec d'autres,
  - Aussi, je ne veux pas associer tel essai avec mon répertoire personnel



# Les templates

2eme exemple – Azote et matière sèche

- Deux types de mesures, des mêmes échantillons
  - Puis-je retrouver toutes les mesures effectuées sur mes échantillons
    - Téléversement
      - Mesures
      - Echantillons
      - Associations mesures – échantillons
    - Recherche ‘semi-manuelle’ → Recherche automatique dans le futur
      - L’architecture est présente
  - Je veux un fichier de synthèse de plusieurs mesures issues de types d’échantillons différents.
    - Manuel pour le moment
    - Possibilité dans le futur ?



# Les templates

## 3eme exemple – Mesure de la créatinine

- Que de données ...
  - Et pourtant, je souhaite bien séparer mes différentes données, selon la bande, le jour de prélèvement et selon que l'alimentation soit conventionnelle ou de précision
    - Soit, je découpe mon gros fichier en plusieurs fichiers
      - Je téléverse chaque fichier et peut associer les métadonnées correspondantes...  
par exemple...
    - Autre possibilité non implémentée
      - J'annote chaque ligne de mon fichier selon le découpage souhaité (10 minutes de travail ici)



# Mais à quoi ça sert au juste ?

Quelques raisons ...

- **Centraliser les données**
  - Veut dire
    - Retrouver facilement ses données
    - Partager facilement ses données dans l'unité
    - Montrer l'existence de telles données
  - Ne veut pas dire
    - Partager ses données de manière systématique
- **Gérer la qualité**
  - Par des contrôles automatiques, selon des champs référencés
  - Par des métaanalyses (voir premières diapos)
- **Regrouper données et métadonnées**
  - Analyses multifactorielles / métaanalyses



# Et après ?

Quelques raisons ...

- Je peux analyser mon fichier de données
  - Par exemple sur Excel
  - Par exemple sur Rstudio
  - Voire python
- Je peux déposer mon fichier sur data gouv
  - ...



# Calendrier

## Des déploiements

- JUIN
  - V1
    - Actuel
    - Prise en compte des demandes
- Décembre
  - V2
    - Système de recherche
    - Contrôle qualité
- Plus tard
  - Ajout des ontologies
  - Ajouts des liens vers l'animal?
  - Amélioration des métadonnées

