

**INRAE**

**> FAIR-DS**

**Un logiciel de gestion de mesures expérimentales  
ouvert et flexible**

**AG PEGASE**

**08/04/2024**

François Moreews

Gwendal Virlet

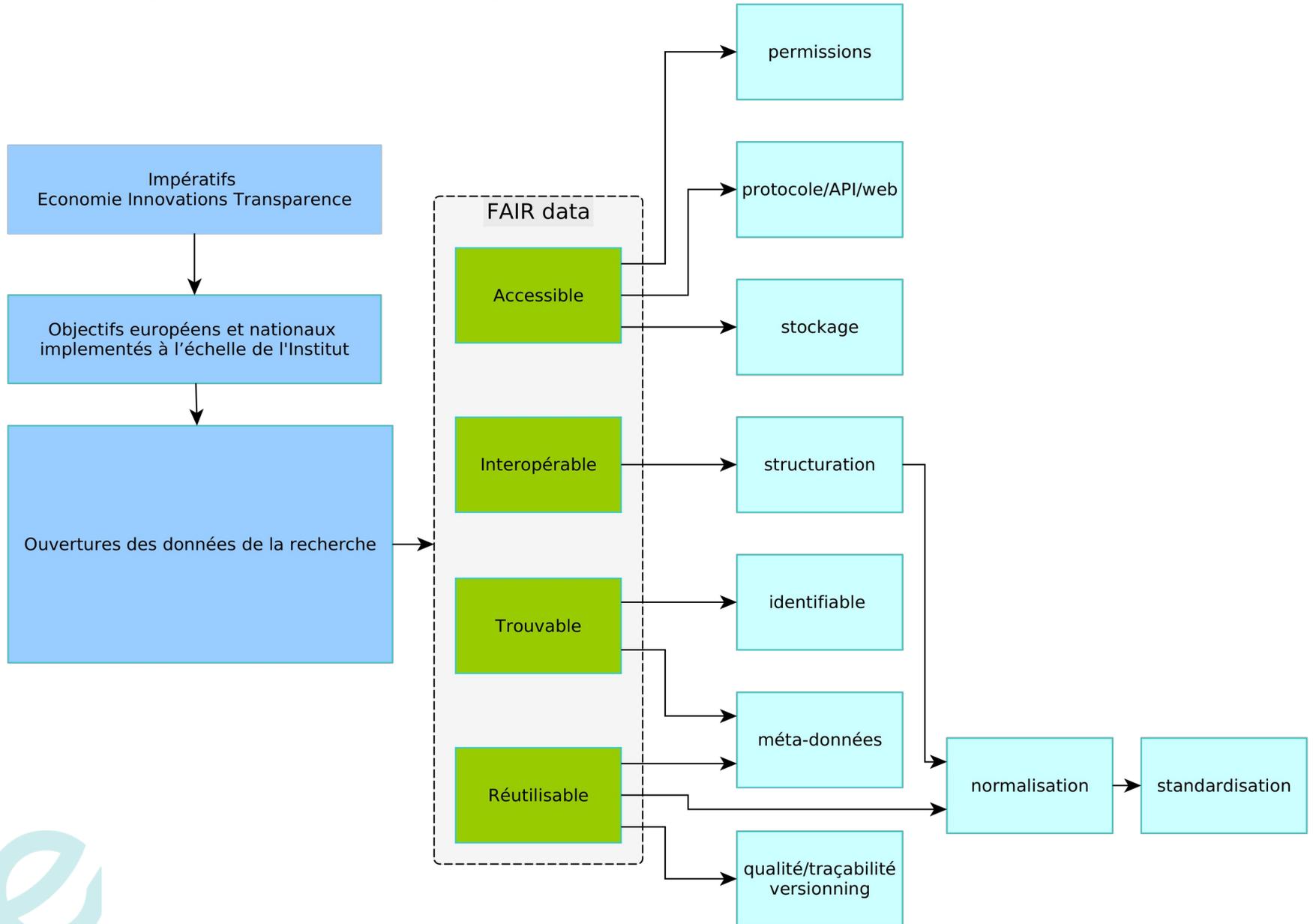
# Sommaire

FAIR-DS – un logiciel de gestion de mesures expérimentales ouvert et flexible

- ◆ Enjeux et objectifs
- ◆ L'organisation du projet
- ◆ Phasage
- ◆ Présentation des fonctionnalités générales / ouverture / FAIR



# Enjeux et objectifs



# Enjeux et objectifs

FAIR-DS – un logiciel de gestion de mesures expérimentales ouvert et flexible

## ◆ Objectifs du logiciel

- ◆ Ouverture des données produites à l'échelle du laboratoire conformément aux directives
  - ◆ Publication (DataVerse)
- ◆ Réutilisation des données de la recherche
  - ◆ En interne
  - ◆ Avec les partenaires externes
  - ◆ Avec la communauté scientifique
- ◆ Garantir la qualité et la traçabilité des données de la recherche
- ◆ Garantir la confidentialité et la propriété des données
  - ◆ “aussi ouvert que possible, aussi fermé que nécessaire”



# Organisation du projet

FAIR-DS – un logiciel de gestion de mesures expérimentales ouvert et flexible

- ◆ Les parties prenantes
  - ◆ Comité de pilotage :
    - ◆ DU et scientifiques, groupe data
  - ◆ Groupe Données:
    - ◆ F. Herault, C. Mustiere, E. Bernard, F. Moreews , G. Virlet
  - ◆ Equipe de développement :
    - ◆ G. Virlet , F. Moreews + stagiaires



# Organisation du projet

Historique : normalisation des données et qualité

- ◆ 2019 :
  - ◆ Phase initiale, état des lieux,
  - ◆ Inventaire des données
  - ◆ Création du diagramme des types de données expérimentales (UML)
  - ◆ Création des templates (modèles Excel) : début de structuration
- ◆ 2022 :
  - ◆ Début du développement de FAIR-DS , application web et base de données
- ◆ 2024 :
  - ◆ Mise en production d'une V1 et retours utilisateurs



# Résultat et présentation de l'outil

**Accessible** : Sécurité – Gérer l'accès aux données finement tout au long de leur cycle de vie

Principe : 'as open as possible, as closed as necessary', (politique d' Horizon Europe et INRAE)

- ◆ Base sécurisée par le LDAP
- ◆ Permissions propriétaire. Chacun peut régler l'accès des jeux de données qu'il gère.  
creation de groupe pour le partage



# Résultat et présentation de l'outil

**Interopérable** : Des formats interopérables – Du dépôt des données à l'ouverture

- ◆ Téléversement de fichiers types complétés (Excel, puis csv, libre office)
- ◆ Téléchargement des fichiers téléversés (csv, puis Excel, libre office)
- ◆ Téléchargement de mesures (filtrées)



# Résultat et présentation de l'outil

**Findable** : Des données identifiables et 'trouvables'

- ◆ Une vue par type de mesure
- ◆ Un lien URI par enregistrement
- ◆ Une recherche simple possible par vue
- ◆ Par la suite, un moteur de recherche avancé avec indexation du contenu



# Résultat et présentation de l'outil

Reusable : Recherche, réplication et interprétation des données

- ◆ Des métadonnées de projet de recherche (hors template)
  - ◆ Projet → Investigation → Etude → Essai
- ◆ Des métadonnées de mesure
  - ◆ Longueur d'onde utilisée (biochimie)
- ◆ Des métadonnées d'échantillon
  - ◆ Aliment, Tissu prélevé – animal
  - ◆ Liens potentiels avec barcode
  - ◆ Liens vers des fichiers lourds
- ◆ Sémantisation/standardisation:
  - ◆ Perspectives: Introduction de vocabulaire contrôlé



# Résultat et présentation de l'outil

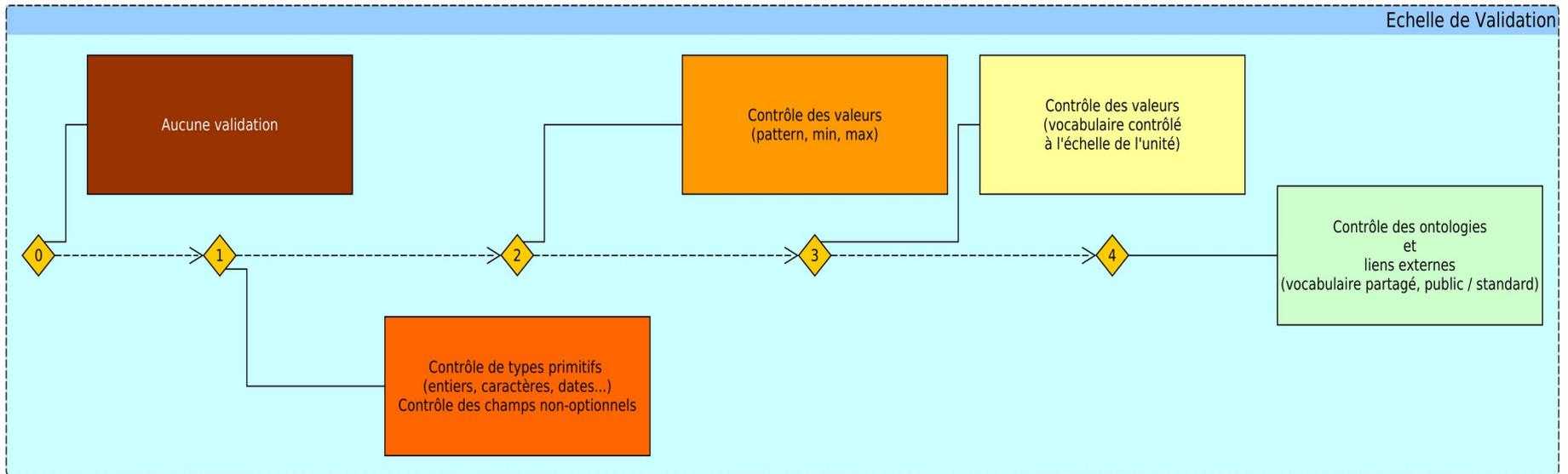
Reusable : Cohérence des données

- ◆ **Objectif**
  - ◆ Versionner les modifications des données
- ◆ **Comment ?**
  - ◆ Permettre des retours en arrière, à une date donnée
    - ◆ Globalement sur la base de données
    - ◆ Sur une recherche
- ◆ **Pourquoi ?**
  - ◆ Gestion des erreurs
  - ◆ L'utilisateur peut connaître la version des données analysées
  - ◆ Traçabilité



# Résultat et présentation de l'outil

Reusable : Qualité des données - Validation



# Résultat et présentation de l'outil

Ouverture des données

- ◆ Partage dans l'unité / INRAE
  - ◆ Système de permissions
- ◆ Envoi à [data.gouv.fr /dataverse](https://data.gouv.fr/dataverse)
  - ◆ Par téléchargement des données
- ◆ Partage avec le CATI SICPA



# FAIR-DS

- ◆ Gestion des données de mesure
- ◆ Type de données re-définissable
- ◆ Outil de suivi qualité
- ◆ Outil pour la FAIRisation des données
  - ◆ Au long du cycle de vie de la donnée
- F ++++
- A ++++
- I ++++
- R ++++

