

Présentation du logiciel FAIR-DS

Le logiciel **FAIR-DS** (FAIR Dataset) est une application développée par l'INRAE afin de répondre aux besoins croissants en matière de **gestion, d'intégration et de valorisation des données de recherche**. Conçu selon les principes **FAIR** (Findable, Accessible, Interoperable, Reusable), il vise à offrir un cadre technologique robuste et évolutif pour la structuration, la sécurisation et l'exploitation des données expérimentales.

Fonctionnalités principales

- **Gestion des données expérimentales** : création, consultation, modification et suppression via l'interface web ou l'API REST.
- **Interopérabilité** : génération automatique d'API pour l'intégration avec d'autres systèmes et plateformes de données.
- **Sécurité et gouvernance** : authentification centralisée, gestion fine des rôles et droits d'accès.
- **Traçabilité** : suivi des entités et des métadonnées associées, garantissant la qualité et la réutilisabilité des données.

Architecture et technologies

FAIR-DS repose sur une **architecture de microservices** combinant un backend Java/Spring , un frontend en React.js, et une base de données relationnelle MySQL(ou PostgreSQL) . L'authentification et la gestion des rôles sont assurées via le protocole **OpenID Connect** et l'usage de **JWT** (JSON Web Tokens).

Le processus de génération de code suit une démarche **Model Driven Architecture (MDA)**, appuyée par l'outil **FAIR-Gen** développé par la même équipe. À partir d'un **diagramme UML annoté**, FAIR-Gen traduit automatiquement la modélisation conceptuelle en artefacts techniques et génère automatiquement:

- le schéma relationnel de la base de données,
- l'API REST associée,
- les microservices nécessaires,
- et les composantes de l'interface utilisateur.

Ce mécanisme permet de réduire les tâches répétitives de développement, d'assurer la cohérence entre le modèle et le code, et de produire rapidement une application FAIR-DS adaptée aux besoins spécifiques d'une organisation.

Cette démarche permet une forte adaptabilité, chaque projet de recherche pouvant disposer d'une application sur mesure dérivée de son modèle conceptuel.

Exemple de modèle

Le diagramme UML suivant illustre une modélisation simplifiée pour un projet zootechnique :

+-----+ Animal 1 * +-----+	+-----+ Sample 1 * +-----+	+-----+ Measure +-----+
id: String species: String birthDate: Date sex: String +-----+	id: String type: String collectionDate material: Str +-----+	id: String date: Date parameter: Str value: Float +-----+

Ce modèle UML permet de générer automatiquement la structure de la base de données, l'API et l'interface correspondante.

Exemple d'utilisation de l'API REST

- **Lister tous les animaux :**

GET /api/animals

- **Ajouter un nouvel animal :**

POST /api/animals

Content-Type: application/json

```
{  
  "id": "A123",  
  "species": "Pig",  
  "birthDate": "2023-05-10",  
  "sex": "Male"  
}
```

- **Consulter les mesures associées à un animal :**

GET /api/animals/A123/measures

Partage des données

Un des cas d'utilisation de FAIR-DS est de stocker les données de manière précise et structurée dans l'application et, grâce à ses API, de permettre leur diffusion sélective vers la plateforme nationale **Recherche Data Gouv**. Cette articulation permet de combiner la rigueur de la gestion interne (traçabilité, métadonnées, cohérence des formats) avec la mise à disposition publique des jeux de données en open access. Ainsi, les jeux de données publiés sont **décrits de façon homogène et conforme aux principes FAIR**, renforçant leur visibilité et leur réutilisabilité par l'ensemble de la communauté scientifique.

Conclusion

Le logiciel est déjà déployé, au sein de l'UMR **PEGASE**, pour la gestion de données de laboratoire ou il permet de stocker, tracer et partager les données de recherche dans un cadre sécurisé et conforme aux standards de l'open science.

FAIR-DS constitue un environnement logiciel modulaire, facilitant la transition vers des pratiques de gestion de données de recherche alignées sur les principes FAIR. C'est un outil d'intérêt pour les organisations scientifiques souhaitant sécuriser et valoriser leurs données expérimentales.