Implementing FAIR Data Exchange in ACTRIS Switzerland : An Expert-Driven Approach

Juan F. Flórez-Ospina, Leïla H. Simon, Nora K. Nowak, Benjamin T. Brem, Martin Gysel-Beer, and Robin L. Modini

PSI Center for Energy and Environmental Sciences, 5232 Villigen PSI, Switzerland

Motivation

Problem. Scientific datasets in atmospheric science are often large, complex and highly specialized, making them hard to reuse across users, tools, and time.

Adhering to the FAIR data principles promises to resolve long term reuse of data by machines and humans.

Yet in practice, implementations often have the following configuration,

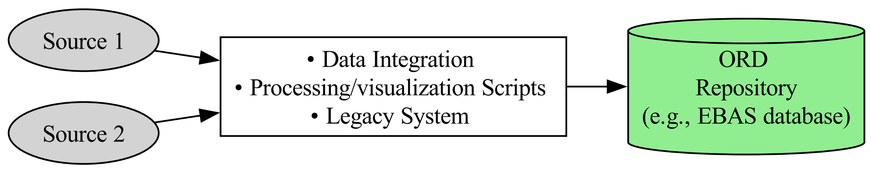


Figure 1. Diagram of tightly coupled system for integration, processing and ingestion of scientific data into ORD repositories.

which in turn exhibits the following

System drawbacks:

- Task-specific outputs, limiting adaptability to new needs or use cases.
- Monolithic design, making the system hard to maintain, reuse or understand internally.
- Embedded domain knowledge in hard-coded parameters, reducing transparency and flexibility.

Expert-driven FAIR data exchange systems

Therefore. *We aim to explore and prototype an expert-driven data exchange system, as conceptualized below.*

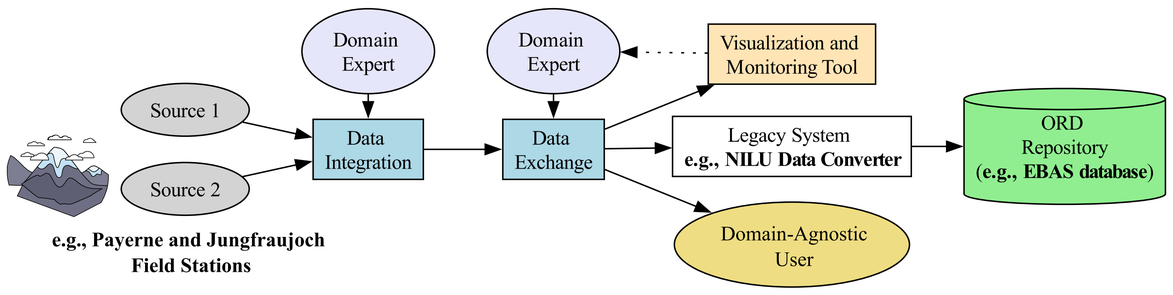


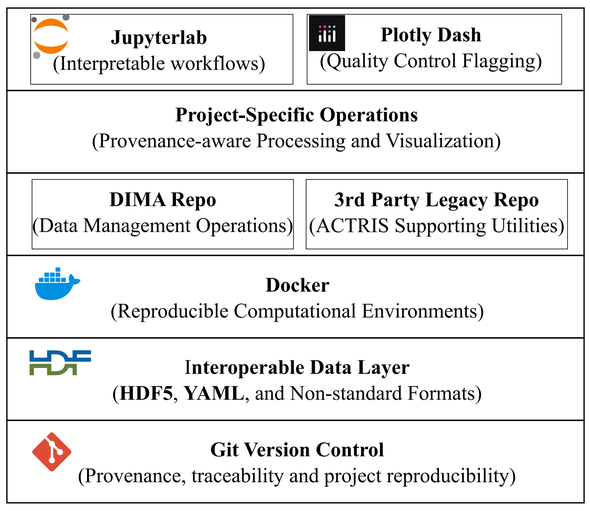
Figure 2. Diagram of the expert-driven FAIR data exchange system.

This system, unlike the above implementation, exhibits the following:

System characteristics:

1. Adaptable to new needs, making it interoperable across users and systems.
2. Modular, allowing components to be reused, replaced, or maintained independently.
3. Configurable with domain knowledge supplied externally, supporting expert feedback loops.
4. Transparent, enabling monitoring of data flows and tracking of data provenance through a clear, structured design.

Practical implementation (Technical stack)



Code Repository: `https://gitea.psi.ch/APOG_public/acsmnode.git`

FAIR data chains in ACTRIS Switzerland—Use case

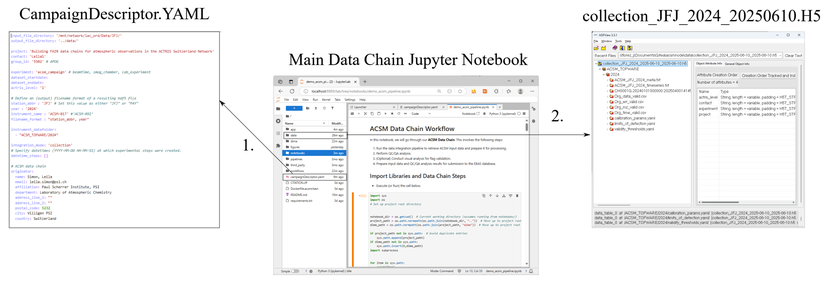
Challenge: Standardized annual submission of aerosol composition observations from field stations (**Payerne** and **Jungfraujoch**) to the **EBAS database**.

We consider raw data described in the Table below.

Data Source	File Formats	Submission Date
Aerodyne Aerosol Chemical Speciation Monitor (ACSM)	TXT, CSV NAS YAML	May 31, 2025

Configurable Data Integration using YAML files

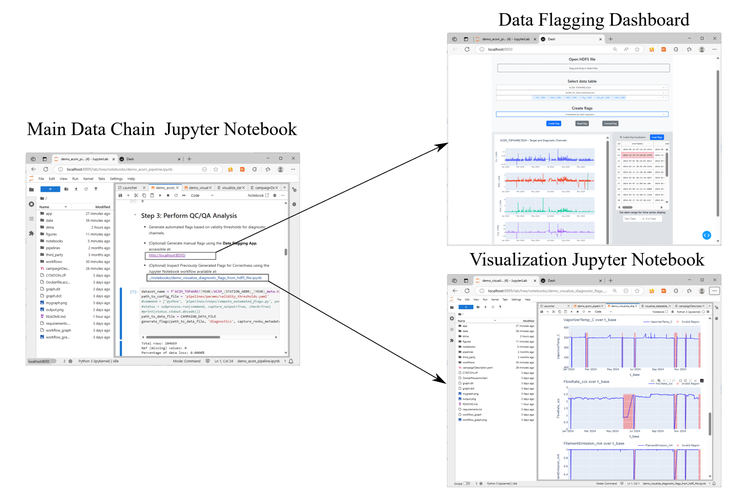
Goal: Search, Retrieve, and Integrate Raw Data in HDF5 format



Code Repository: `https://gitea.psi.ch/5505-public/dima.git`

Interactive Data Annotation for Quality Control

Goal: Apply time-dependent correction factors to observations and generate quality control flags with expert validation, complying with EBAS database.



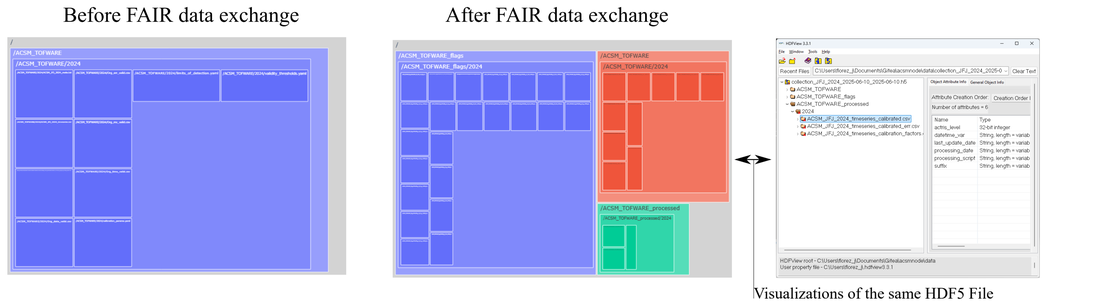
Enables reconciliation of **Automatic Diagnostic Flags**, **Station-specific Flags**, and **Manual Review Flags**.

Accessible Data Products for Domain-Agnostic Reuse

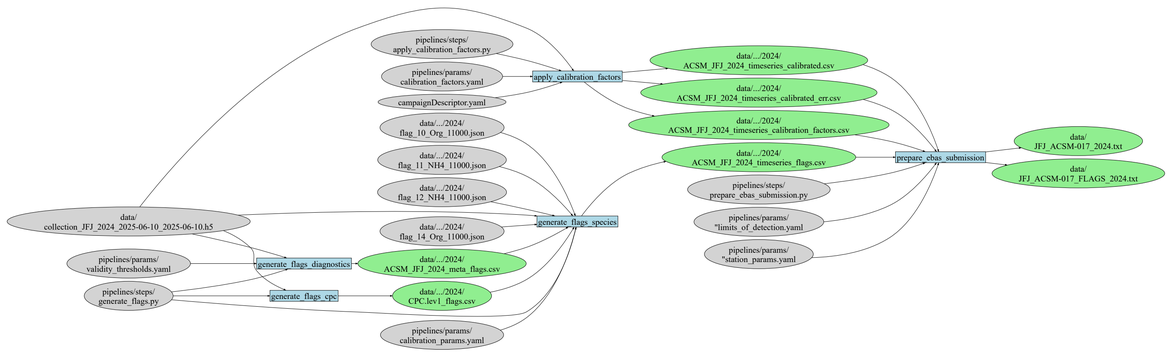
Primary Product: EBAS-ready files conforming to ACTRIS Level 2 quality standards.

Secondary Product: A comprehensive, self-describing file that integrates:

- Original and intermediate data products with rich contextual and provenance metadata, as shown below.



- An auto-generated prospective provenance graph in Renku workflow format, visualized as below



juan.florez-ospina@psi.ch, juanflo16@gmail.com, robin.modini@psi.ch