

**Architecture Document for the ACCURIDS FAIR Data Spaces  
Demonstrator**

**Date:** 2025-01-20

**Contacts:** Heiner Oberkampff (CEO), Helge Krieg (Senior Data Engineer)

Introduction .....2

Station 1 : Pharma.....3

Station 2: Hospital .....4

Demonstrator Manager Service .....5

Eclipse Dataspace Components (EDC) Connectors .....6

    Provider .....6

    Consumer .....6

Architecture Diagram.....7

## Introduction

The technical implementation of the FAIR Data Space Demonstrator consists of the following parts:

- Two ACCURIDS platforms that were added as instances to the Accurids cluster in the Accurids cloud.
- An EDC launcher similar to one of the [examples of the Eclipse Dataspace components](https://github.com/eclipse-edc/Connector/tree/main/launchers/sts-server) (<https://github.com/eclipse-edc/Connector/tree/main/launchers/sts-server>).
- A Python script that manages the communication with the International Dataspace (IDS) protocol.

We have prepared an [Architecture Diagram](#) that helps to understand the components of the technical implementation.

## Station 1 : Pharma

This is an instance of the Accurids Master Data Management Platform.

It is available at: [pharma accurids.com](https://pharma accurids.com).

The platform consists of six service components:

1. The Backend component is a Java Spring Boot application. It takes care of all aspects of data management like data processing, data exchange, storage, state management, user management, security and access management. The component offers a REST API and a GraphQL API.
2. The Frontend component is a React-based web interface. The communication between frontend and backend happens via the REST API and the GraphQL API.

The pharma frontend allows pharma users to manage users, groups and roles. Users can manage information content, define information structure, and specify the type of information registration e.g., with persistent identifiers to ensure FAIR data.

3. A PostgreSQL database manages the application state, security configuration, and can store the audit trail of changes e.g., required for data management in a regulated environment.
4. An RDF triple store persists all information content, as well as the underlying information schema. The triple store has multiple databases and datasets. In case of the Pharma instance, we have following datasets:
  - a. Patient collaboration data  
The RDF dataset contains anonymized patient information like gender and year of birth.
  - b. Clinical Trial  
The RDF dataset contains the information content about a clinical trial that is relevant to the Pharma company for investigating the efficacy of their drug.
  - c. Hospital 1  
The purpose of this RDF dataset is information exchange with Hospital 1 that takes part in clinical studies of the Pharma company. In our demonstrator, we take the example of human bio samples that are analyzed in Hospital 1, and the measurement result is shared with the Pharma company.
  - d. Hospital N  
There can be many hospitals contributing to clinical studies. This RDF dataset is an example that demonstrates how users at the Pharma company can integrate the information content from Hospital 1 and Hospital N because they share the same FAIR data information structure.
  - e. Data Schema  
This RDF dataset contains the specification of the information structure

expressed as constraints following the Shapes Constraint Language (SHACL). The Accurids platform processes the constraints and can determine the required information structure. We can dynamically create user interfaces and APIs from the constraints. They also enable the validation of information content to ensure that all data fulfills the FAIR data requirements.

- f. Master Data  
It is a placeholder highlighting the fact that there can be many more master data domains. Our demonstrator focuses on the single one of clinical trials.
  - g. Vocabularies  
All RDF datasets describe information in a standardized way following the Resource Description Framework (RDF) as specified by the World wide Web Consortium (W3C). Datasets with master data like patient collaboration contain e.g., explicit values of date of birth of a patient. The vocabularies dataset adds an information level on top of this. We can define exactly e.g., what “patient” means in the context of a clinical trial, we can add further labels like “client”, or in different languages. Accurids uses the W3C standards of the Web Ontology Language (OWL) and the Resource Description Framework (RDFS) for the description of vocabularies. The combination of standardized vocabulary with standardized schema ensures that we can achieve FAIR master data.
  - h. IDMP-O  
The RDF dataset contains the ontology for the identification of medicinal products (IDMP-O). It is an example of a vocabulary that is useful for the standardized representation of information in the domain of clinical trials and drug development. IDMP-O is a technical implementation of an ISO standard.
  - i. SPOR  
The RDF dataset contains standardized reference data from the European Medicines Agency (EMA). It helps to manage information about substances, products, organizations and referentials (SPOR) in a standardized way to fulfill the requirements of EU regulatory authorities.
- 5. The ElasticSearch service provides an index of all information content persisted in the RDF triple store. We can achieve a good user experience with responsive user interface and fast answers due to the additional search index.
  - 6. The source code of the Accurids platform is under version control in a GIT repository. We have a CI/CD pipeline that builds the application from the source code and publishes a Docker image.

## Station 2: Hospital

This is an instance of the Accurids Master Data Management Platform.  
It is available at: [hospital.accurids.com](https://hospital.accurids.com).

The platform consists of the same six service components as described above. In case of the Hospital instance, we have a couple of datasets that are different from the Pharma ones:

1. Patient Master Data

The RDF dataset contains detailed information about the patients of the hospital. It is sensitive data. Accurids ensures that only authorized accounts have access.

2. Patient Collaboration

The RDF dataset is an anonymized view on the Patient Master Data. We extract certain anonymous aspects from the patient data that are relevant to the collaboration between the hospital and the pharma company with respect to the clinical trial. For example, the hospital shares the information of gender and year of birth of a patient with the pharma company.

3. Clinical Trial

The RDF dataset contains the information from the pharma company about clinical trials that the hospital takes part in.

4. Human Biosample

The RDF dataset contains measurement results of bio samples like tissue samples or blood samples that were taken at the hospital. A user from the hospital can connect the results of a bio sample to a clinical trial.

5. Data Schema

This RDF dataset contains the specification of the information structure expressed as constraints following the Shapes Constraint Language (SHACL). The hospital instance receives constraints from the pharma company that specify exactly what information is required for a clinical trial, and how the information must be represented.

## Demonstrator Manager Service

It is a Python program that manages the synchronization of datasets between the Pharma instance and the Hospital instance. The source code is available at <https://github.com/FAIR-DS4NFDI/accurids-demonstrator>.

The synchronization process has 21 steps. The first steps configure the FAIR Dataspace for the data exchange, the remaining steps execute repeatedly for every dataset update.

- 1) Create Assets

- a) Create an asset for the Human Biosample dataset at the hospital
- b) Create an asset for the Human Biosample dataset at the pharma company
- c) Create an asset for the Accurids update process at the pharma company
- d) Create an asset for the Patient Collaboration at the hospital
- e) Create an asset for the Patient Collaboration at the pharma company
- f) Create an asset for the Clinical Trial dataset at the pharma company
- g) Create an asset for the Clinical Trial dataset at the hospital

- h) Create an asset for reference data at the pharma company
- i) Create an asset for the data schema at the pharmacy
- j) Create an asset for the data schema at the hospital
- 2) Create policy
- 3) Create contract
- 4) Fetch the data catalog
- 5) Negotiate a contract for data retrieval
- 6) Find the agreement of the contract negotiation
- 7) Start the data transfer (retrieval)
- 8) Find the endpoint data reference to download the data
- 9) Retrieve the data content from the source Accurids instance
- 10) Fetch the data catalog
- 11) Negotiate a contract for data upload
- 12) Find the agreement of the contract negotiation
- 13) Start the data transfer (upload)
- 14) Find the endpoint data reference to upload the data
- 15) Push the data content to the target Accurids instance
- 16) Fetch the data catalog
- 17) Negotiate a contract for data update
- 18) Find the agreement of the contract negotiation
- 19) Start the data transfer (update)
- 20) Find the endpoint data reference to update the data
- 21) Update the dataset at the target Accurids instance

## Eclipse Dataspace Components (EDC) Connectors

### Provider

We prepare the Provider connector as a Jar file from the source code repository of the Eclipse Dataspace Component Connector. It is possible to prepare a launcher that uses set of EDC modules as required for our demonstrator. Our launcher configuration is available at: <https://github.com/FAIR-DS4NFDI/accurids-demonstrator/tree/main/launcher/accurids>.

The Provider connector manages the setup of the dataspace i.e., creating assets, policies, contracts, and provides a catalog of data assets.

### Consumer

We prepare the Consumer connector as a Jar file from the same source code as for the Provider connector. The Provider connector takes care of the contract negotiation, the agreement for data exchange, initiating data transfers, and providing endpoint data references to access the content of a data asset.

# Architecture Diagram

## Architecture Diagram

### ACCURIDS FAIR Data Spaces Demonstrator

