ACCURIDS

Deliverable Report for Milestone 1

FAIR Data Spaces Demonstrator for Cross-Organizational Governance and Usage of Persistent Identifiers in Healthcare & Pharma for faster drug development

Date: 2025-01-21

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

Contents

ACCURIDS

Introduction

The FAIR Data Spaces program is exciting and ground-breaking, because it helps to connect data from research and industry according to FAIR data principles to advance innovation and the digital transformation in various sectors.
We describe two use cases for a FAIR Data Space Demonstrator that can realize a data exchange between research and industry with the ACCURIDS Master Data Management Platform.

Overview of the use cases

We have performed interviews with different people in a selected German pharmaceutical company covering roles in IT/Data, Clinical, Regulatory and Early Research. Based on these interviews two use case descriptions have been created that highlight challenges of cross-organizational collaboration and data exchange between a pharma company and clinical (Use Case 1), and between a pharma company and research (Use Case 2).

ACCURIDS

Use Case 1: Clinical Trials – Connecting Pharma with a University Hospital

Challenge: The execution of a clinical trial is a race against time to bring novel medicines to the market. Over three phases, a pharmaceutical company must scientifically prove the efficacy of a new treatment for a target patient group. This process generates vast amounts of data, collected by diverse research organizations, including university hospitals. Currently, this data is often submitted to pharma companies in the form of numerous Excel sheets through contract research organizations (CROs), lacking standardized formats, making integration and reliable identification and traceability of each data point almost impossible.

Solution Outline: We propose a connector for the FAIR Data Spaces with ACCURIDS to combine a trusted infrastructure with our existing Data Registry software that is already used within big pharma. The solution will provide following benefits:

1. *ACCURIDS Registry Instance:* Each participant operates an ACCURIDS registry, linked to the FAIR Data Spaces infrastructure for standardized, interoperable data sharing.
2. *Unique Data Identifiers:* Every data point has a globally unique, persistent identifier, ensuring traceability and integration.
3. *Standardized Metadata Sharing:* Metadata follows industry standards, ensuring consistent and reliable data exchange.
4. *Controlled Data Granularity:* Data originators control how much detail is shared, ensuring security and customization.
5. *Seamless Data Integration:* Pharma companies integrate external and internal data automatically, reducing manual work.
6. *Data Traceability and Quality:* Full traceability and automatic quality checks maintain data integrity and reliability.
7. *Real-Time Data Analysis:* Enables immediate analysis, helping pharma companies adjust trials faster and accelerate drug delivery.

Result: The Demonstrator shows the feasibility of faster, cheaper and better execution of clinical trials. This allows a pharma company to bring novel medicines faster to the market.

ACCURIDS

Use Case 2: Early Research on Cell Cultures – Connecting Pharma with University Research in Biotech.

Challenge: Early-stage research for novel medicines relies on *in vitro* testing of RNA constructs to identify viable candidates. However, collaboration between academia and industry faces following challenges limiting collaboration efficiency and speed:

- *Increased number of different Cell Cultures:* Advances in biotechnology have led to an explosion in cell culture variants and modifications.
- *Diverse Analytical Techniques:* Increasingly complex methods make it more difficult to interpret results effectively.
- *Data Exchange Issues:* Current practices use Excel sheets and ad hoc numbering or naming, creating ambiguities in identifying cell cultures, RNA constructs, and plasmids.
- *Knowledge Silos:* Reliance on individual expertise, e.g., a PhD student conducting the joint research project, risks losing critical insights during transitions.
- *Plasmid Complexity:* Plasmid designs involve intricate metadata, yet current systems poorly handle this complexity.

Solution Outline: We ensure the unique identification of key research objects and standardize their metadata. This has the following benefits:

1. *Standardized Taxonomy:* Classify key research objects like cell cultures (e.g., by lineage, phenotype) and plasmids (e.g., by design and function) for consistency across institutions.
2. *Minimal Metadata Standards:* Define essential attributes for cell cultures (e.g., genetic modifications, growth conditions) and plasmids (e.g., sequence details, modification history) to ensure traceability.
3. *Unique Identification:* Use globally unique IDs based on standardized metadata agreed normalization functions and public references (e.g., GenBank, UniProt) for unambiguous tracking of cell cultures, RNA constructs, and plasmids.
4. *Seamless Metadata Integration:* Enable structured metadata sharing to automate integration into pharma systems, reducing manual work and ensuring data consistency.
5. *Real-Time Analysis and Governance:* Leverage integrated data for immediate analysis to refine research focus, accelerate RNA construct validation, and establish governance for scalable, standardized collaborations.

Target Result: The solution enhances collaboration, accelerates research timelines, and improves reproducibility for faster drug discovery.

Demonstration of Use Case 1: Clinical Trails

The demonstrator implements healthcare and pharma use cases showcasing how a pharma company and a hospital can exchange data about human bio samples in context of a clinical trial sponsored by the pharma company in an automated manner allowing to speed up clinical trial data analysis and decision making.

- *Pharma:* It captures the basic clinical trial information and requirements for collecting human bio sample data in trials performed at clinical research institutions such as a university hospital.

- *Hospital:* It manages patient information and collection of bio samples as part of the clinical trial perform for the sponsoring pharma company.

Setup of the Demonstrator

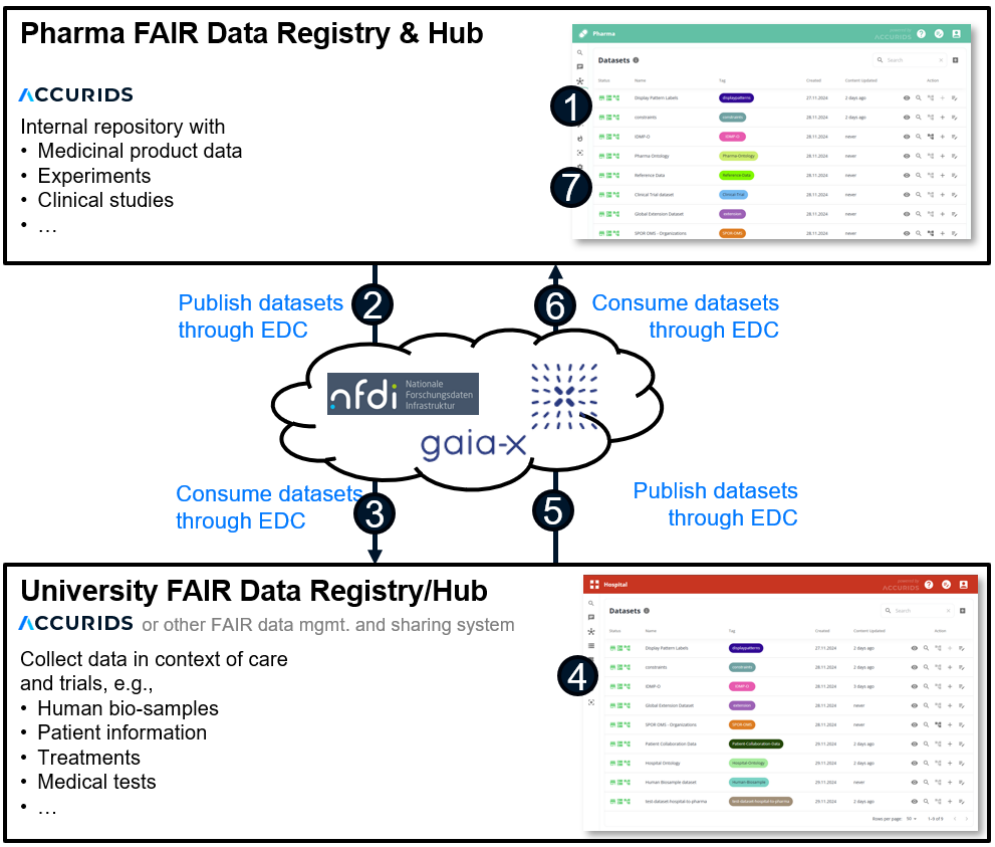

Figure 1: Demonstrator setup for Use Case 1 with Pharma and University instances of ACCURIDS connected to the FAIR Data Spaces through EDC connectors

The diagram above provides an overview of the high-level setup of the demonstrator. We implement a collaborative data management process between pharma and hospital as a stepwise workflow:

1. Manage FAIR clinical trial data in pharma instance
2. Provide (meta-)data standards through EDC

3. Consume (meta-)data standards through EDC
4. Capture trial related bio samples in hospital instance
5. Provide bio samples datasets through EDC
6. Consume bio samples datasets
7. Analyze clinical trial data

The Accurids Data Registration Software is used as a distributed governance platform to register and resolve data objects managed in different organizations (pharma, university hospitals) within a secure environment. The publication of semantic data schema for different type of objects, e.g., clinical trial or human bio samples with corresponding laboratory tests are shared between pharma and hospital data management platforms together with corresponding reference vocabularies.

The demonstrator implements the FAIR Data Spaces system and architecture to ensure interoperability with other FAIR Data spaces demonstrators.