FAIR Data Spaces

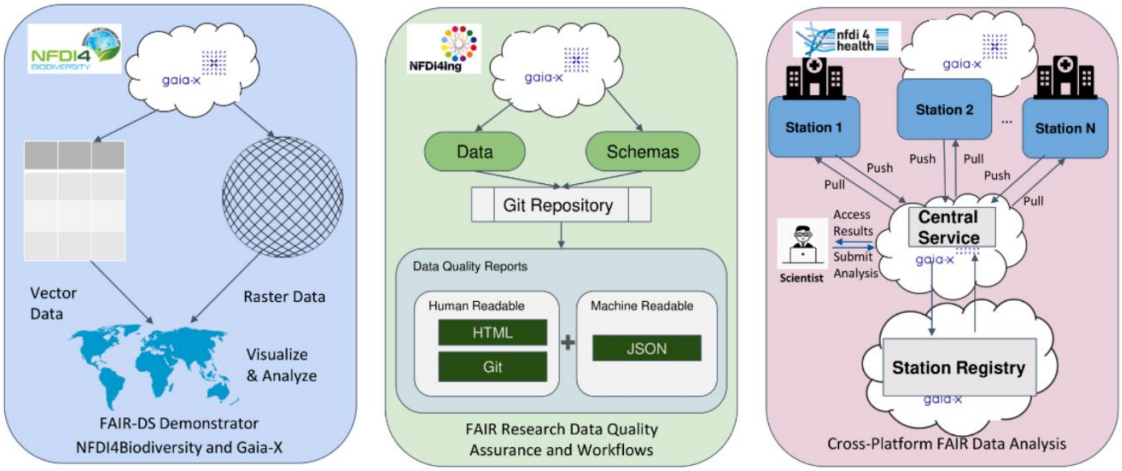
FAIR Data Spaces

A Data Space for Research and Industry

Project FAIR Data Spaces provides a common, cloud-based data space for industry and research in compliance with FAIR Principles.

The project establishes a roadmap for the collaboration of the European Gaia-X federated and secure data infrastructure and the National Research Data Infrastructure (NFDI), clarifies the ethical and legal framework for data exchange between research and industry, establishes a common technical foundation and demonstrates the use of Gaia-X technology for providing and using research data along the FAIR Principles in different fields of science and business sectors.

FAIR Data Spaces demonstrators show the applicability of the structural framework for real use cases in Biology, Geosciences, Health and Data Science.



Geo-Engine

Geo Engine is a cloud-based research environment that connects data sources and provides researchers with the ability to process spatiotemporal data interactively and visually. In the demonstrator, scalable access to data provided by NFDI4Biodiversity, a NFDI consortium, in a cloud is supported. This is done based on the Gaia-X Cloud specifications, which provide both technical and legal frameworks for data exchange. In

the first use case in FAIR Data Spaces, data from industry (satellite data) is combined with data from science (GFBio).

FAIR Research Data Quality Assurance and Workflows

The purpose of the Demonstrator for FAIR Research Data Quality Assurance and Workflows is to show the use of decentralized task runners for automated quality control and data assurance in a widely available or easily deployed environment. In doing so, the demonstrator uses the workflow engine provided by the source code hosting platform GitLab to analyze, transform, and verify research data artifacts. Based on given schema data, the demonstrator analyzes newly added data for compatibility and provides a warning if violated. An incompatible dataset can thus be quickly cleaned up and then smoothly integrated into existing datasets.

Cross-Platform FAIR Data Analysis

The goal of the Demonstrator for cross-platform data analysis is to reuse the current results of NFDI (in particular the NFDI4Health consortium) and the Medical Informatics initiative in terms of medical data structures, formats, and ethical and legal requirements, while also being compatible with Gaia-X specifications. To this end, a cross-platform data analytics infrastructure called Personal Health Train (PHT) will be used. The key elements of the PHT ecosystem are the so-called Trains and Stations, an analogy to trains and stops. Trains encapsulate analytics tasks using container technologies. Trains contain all the requirements to query the data, run the algorithm, and store the results. Stations act as data providers that manage data sets. To analyze the decentralized data, a specific train is transmitted to each station in turn. The train performs the analysis task and computes the results (e.g., statistics) based on the locally available data.

More Information