Details

DataPLANT Research Data Management and Publication Platform (ARChub) (ARChub)

The DataPLANT Research Data Management and Publication Platform (ARChub) is a research data management service built on cloud and storage infrastructures that supports collaboration ranging from individual researchers and laboratory teams to interdisciplinary research groups through the Annotated Research Context, functioning as a universal science gateway. It combines the open-source frameworks GitLab and InvenioRDM, adapted for research data management (RDM), with various (bioinformatics) workflow systems such as Galaxy. Through integrated Git versioning, issue tracking, and CI/CD, and workflow integration ARChub enables comprehensive provenance tracking, reliable sharing of research data, and straightforward project management covering the entire data lifecycle (generation, annotation, versioning, publication). Integration with underlying stable storage systems provides a scalable and secure foundation for large datasets, while the integrated REST and web hooks allow seamless integration into existing research workflows and automated data pipelines. InvenioRDM is a widely used data publication platform developed at CERN and serves as the foundation for datasets published with persistent identifiers (DOIs). The cloud-based platform is designed to be open and extensible, allowing workflow systems beyond Galaxy to be integrated as well.

Address: Hermann-Herder-Str. 10
79104 Freiburg
Baden-Württemberg
Deutschland
To website

Host Institution

Universität Freiburg
Friedrichstraße 39
79098 Freiburg
Baden-Württemberg
Deutschland
https://uni-freiburg.de/

Scientific Domain

Primary Subjects:
- Biology
- Agriculture, Forestry, Horticulture and Veterinary Medicine

Secondary Subjects:
- Medicine
- Geosciences (including Geography)

Category

Research data repositories

Scientific Services

1. Provision of a stable and long-term storage platform for (large-scale) data volumes in the terabyte++ range 2. Operation of a data management and data publication platform with access via (community) AAI 3. Assignment, registration, and resolution of long-term persistent identifiers (DOIs) (for ARC RO-Crates) 4. Provision of APIs for advanced search, AI integration, and data federation 5. Integration with workflow systems such as Galaxy, and support for connecting additional infrastructures 6. Education and training: * Introduction to the use of the platform * Training in research data management based on the Annotated Research Context * Consulting on data management, data publication, and data reuse * Guidance on data preparation, packaging, data minimization, and licensing 7. Support services: * Creation and long-term maintenance of project-specific working environments * Integration of laboratory- and research group-specific workflows * Integration of new analysis tools (within Galaxy or other workflow systems) * Support for data quality management and publication * Use and further development of the ARC schema for custom applications

Scientific Equipment

- bwSFS storage system: netzwork shares and Storage-Grid (S3)
- de.NBI infrastructure (server for Galaxy workflow execution)
- bwCloud (virtual machines for (micro-) services)
- Virtualisation Cluster
- HPC system

Keywords

- data management platform
- data publication platform
- Galaxy workflows
- Annotated Research Context
- ARC RO-crate
- bioinformatic tools

Networks

NFDI consortium DataPLANT
https://nfdi4plants.org/

de.BI - German Network for Bioinformatics Infrastructure
https://www.denbi.de/

Galaxy Europe
https://usegalaxy.eu/

Users per annum

Internal Users: ~ 30
External Users in total: ~60
External Users: 50+
External Users in the EU: ~5
External Users outside of EU:

© DFG