

**swissuniversities**

**Call for proposals B3.2: ORD Funding for Existing Research Infrastructures of Strategic Relevance  
List of Approved Projects**

Call Deadline: 22.02.2024

Decision by the Delegation Open Science: 06.06.2024

## Overview Approved Projects Call B3.2

Submissions: 10

Approved projects: 6

Funding rate: 60%

Short title	Full title	Leading institution	Partner institution(s)	Project Manager	Total costs (CHF)	Funds Requested (CHF)
<a href="#">ORD-FORHET-CAT</a>	ORD for Heterogeneous Catalysis: Implementation and Dissemination of Standardized Data Processing	ETHZ		<a href="#">Paco Laveille</a>	304 000	152 000
<a href="#">Surgical Data Science OR-X</a>	Advancing Surgical Data Science with Open Research Data Repositories in the OR-X	UZH	ZHAW, Universitätsklinik Balgrist (Digital Medicine Unit)	<a href="#">Philipp Fürnstahl</a>	988 600	457 000
<a href="#">FAIR-CPB</a>	FAIRification of pathogen bioinformatics resources under the Centre for Pathogen Bioinformatics	ETHZ	UNIBAS, Swiss-TPH, SIB	<a href="#">Tanja Stadler</a>	1 000 000	500 000
<a href="#">FAIR-FI-LD</a>	Moving towards a national FAIR-compliant ecosystem of Federated Infrastructure for Language Data	UZH	LiRI, CLARIN-CH, ZHAW, USI	<a href="#">Cristina Grisot</a>	365 000	182 000
<a href="#">Bgee-access</a>	Powerful and integrative access to FAIR gene expression data in the Bgee infrastructure	UNIL		<a href="#">Marc Robinson-Rechavi</a>	485 072	220 000
<a href="#">BioMoQA</a>	Biodiversity Monitoring with Question-Answering	HES-SO	UNINE	<a href="#">Patrick Ruch</a>	362 024	181 012

## Short Summaries of the Projects

*Abstracts by the applicants:*

ORDFORHETCAT

### ***ORD for Heterogeneous Catalysis: Implementation and Dissemination of Standardized Data Processing***

Catalytic technologies are playing a key role in our society and are used in over 90% of all chemical processes such as the production of pharmaceuticals, fertilizers, or sustainable fuels. Many processes still use fossil feedstocks and the shift towards a carbon-neutral world puts pressure for industry to use alternative feedstocks. Thus, cross-disciplinary efforts involving the field of natural science and engineering are required to develop new catalysts and processes.

To speed-up discovery and development of new catalysts, the ETH domain has built a new technology platform for catalysis research in Switzerland, Swiss Cat+, combining automated high-throughput experimentation with artificial intelligence. Such automated facilities generate massive amounts of high-quality experimental data where their capturing, processing and standardization are fully digitalized, and thus increases reproducibility and traceability of experiments. When processed properly and in a standardized way, these data can be further used for design of experiments and training of machine learning models in catalysis.

While open access platforms to upload such research data in the field of catalysis exist (e.g. Zenodo, ETH Research Collection), the proper and standardized data processing to comply with FAIR data principle limits the full utilization of such open access platforms. Therefore, this proposal aims at developing open access codes/software to streamline and standardize data processing for the research community. The streamlined data processing workflow will bridge the missing link between experimental data generation and available opensource platforms and therefore facilitate open research data implementation in the field of catalysis. All the proposed actions will enhance the transparency of catalysis research data management and make them more amenable for applying data science methods to speed up research and discovery of new catalysts for a sustainable world.

Surgical Data Science OR-X

### ***Advancing Surgical Data Science with Open Research Data Repositories in the OR-X***

This Collecting high-quality intraoperative data is paramount for advancing surgical data science and innovating surgical technologies for the in-vivo treatment. This data also facilitates objective and evidence-based assessment of treatment quality, efficacy, or deficits across interventions and surgeons. However, data collection in real operating rooms face obstacles such as proprietary interfaces, workflow and regulatory constraint and safety concerns, impeding comprehensive intraoperative data acquisition. The Operating Room-X (OR-X, <https://or-x.ch/>) stands as a distinctive research infrastructure implementing a perfect copy of a real operating room in which ex-vivo surgeries are efficiently conducted. Supported by a robust backbone network, the OR-X enables the seamless integration of Augmented Reality (AR), surgical robotics, and Artificial Intelligence (AI) technologies. This integration accelerates the implementation of novel surgical innovations and yields invaluable insights through continuous analysis of on-site experiments.

The overarching goal of the ORD project is to develop and seamlessly integrate ORD components into the OR-X, thereby facilitating the generation and publication of large-scale multi-modal surgical databases (digitalization of surgeries). The planned ORD components will include 1) the integration of all relevant surgical devices and technologies into the high performance network of the OR-X, 2) the automatic fusion and structuring of multimodal data to establish surgical data models and 3) the efficient population of these data as public datasets.

As a primary outcome of the ORD project, a novel Data Collection Framework will be established, automating the collection, structuring, and standardization of surgical data. This framework aspires to set new benchmarks in surgical data science within the research community. Furthermore, a public Surgical Data Cloud Platform will be launched that is aligned with Open Science best practices including FAIR principles. Showcasing open research datasets of surgical procedures in highest quality, this platform will function as a collaborative space to support advancements in surgical data science and contribute to national initiatives like the Swiss Personalized Health Network.

FAIR-CPB

***FAIRification of pathogen bioinformatics resources under the Centre for Pathogen Bioinformatics***

The COVID-19 pandemic has highlighted the importance of genomic surveillance of rapidly changing pathogens but also exposed shortcomings in the way such data are processed, shared, and linked. The Centre for Pathogen Bioinformatics has the goal to provide key tools, expertise, and infrastructure to process, analyse, store and share genomic data of pathogens in a way that respects the interests of data generators and privacy concerns, while allowing actionable inference, maximising data reuse, and open-ness.

We maintain several renowned resources (Swiss Pathogen Surveillance Platform (SPSP), Nextstrain, Nextclade, GenSpectrum, CoVariants, V-pipe, COJAC and LolliPop) that are used individually in strategically important national or international pathogen surveillance projects. Here, we propose to make these tools interoperable and exploit their complementarities as part of our ambition to provide state-of-the-art solutions for pathogen genomics.

To enable tool integration and data sharing, we will build upon the first prototype instance of our open-source database Loculus such that the database will couple data sub-mission to analysis and quality control via Nextclade and accept waste-water data from COJAC and LolliPop. GenSpectrum will interface with Loculus to allow flexible querying and CoVariants for variant frequency analysis. SPSP will implement a SPARQL endpoint for API access and an interface to Loculus. Loculus will allow for restricted data sharing and will release data to INSDC databases eventually.

Together, these data portals and analysis platforms are core in enabling reproducible analysis and real-time surveillance of pathogens and already play important roles at the national and international levels. By further strengthening this ecosystem through increased interoperability and by providing an integrated system for analysis and sharing of pathogen genomic data, we aim to catalyse genomic epidemiology globally.

FAIR-FI-LD

***Moving towards a national FAIR-compliant ecosystem of Federated Infrastructure for Language Data***

In the last 5-10 years, Swiss HEIs have been working on building national services for language data. They include, up to now, the Linguistic Research Infrastructure (LiRI-UZH), the Swiss-AL Platform for Applied Sciences (ZHAW), a national repository for the publication and long-term preservation of language data LaRS@SWISSUbase (UNIL, UZH), and various smaller tools and services. These units however are not all interoperable, which reduces the potential for collaboration and data reuse. In addition, fields such as interactional linguistics or second language acquisition lack adequate infrastructure.

With the foundation of the CLARIN-CH consortium in 2020 (9 HEIs and the SAGW), the HEI's efforts took a new direction: work together to build a FAIR-compliant, sustainable and expandable CLARIN-CH ecosystem of federated infrastructure to answer the needs of researchers and professionals using language data in Switzerland and beyond; an ecosystem that must be interoperable at the national and European levels. The present project aims at realizing important steps towards this mid- and long-term goal, in compliance with the Swiss ORD strategy, by prototyping

- interoperable underlying software using NLP techniques and exploratory AI techniques
- harmonized metadata between the existing Swiss infrastructure components and the European CLARIN infrastructure
- CLARIN federated content search (FCS) to query each component of the infrastructure
- a FCS multilingual landing page hosted on the CLARIN-CH website
- a frontend of the VIAN-DH@LiRI environment to visualize, query and analyze multimodal talk-in-interaction data, hosted at USI,

by producing documentation and training to support the use of the infrastructure and inform about legal and ethical issues related to language data in the context of Open Science, and by planning the future collaboration with further stakeholders and aggregation of further tools and services.

Bgee-access

### ***Powerful and integrative access to FAIR gene expression data in the Bgee infrastructure***

Gene expression is a core concept of the central dogma in biology, responsible for the transfer of information from DNA to effector proteins, through the production of short-lived RNA molecules. In the past decades, tremendous progress has been achieved in studying this transcriptional process, and current technologies allow for studying gene expression at the single-cell level, and even at the single-molecule level. These advancements open new doors in biomedical science, e.g., to treat cancer, or in fundamental research, e.g., to study cell identity and evolution. This project aims at unravelling these new types of data through the Bgee infrastructure, the only Swiss resource providing FAIRification of public transcriptomics data.

Distributing these new data, such as single-cell RNA-Seq (scRNA-Seq) to quantify gene expression at the single-cell level, comes with a new level of complexity. Hundreds of thousands of cells can be studied within a single experiment, exponentially increasing the data volume to manage, and the data curation work to make these data FAIR. The large majority of these data currently live unstructured in public repositories, hampering the ability of researchers to reuse or integrate them to discover new knowledge.

The leading house for this project, developing Bgee, is currently receiving funding through the open science programme "CHORD - Track B" (project "scFAIR"), to create a consortium of transcriptomics resources to agree on the metadata schemas and tools to make these data FAIR. This current project aims to support implementing the conclusions of scFAIR and to provide powerful access to open transcriptomics data, standardized between experiments and species. This will position Switzerland at the forefront of transcriptomics data FAIRification, knowledge graph construction such as used in pharmacology, and Machine-Learning (ML) algorithm training for, e.g., discovering cell identity.

BioMoQA

### ***Biodiversity Monitoring with Question-Answering***

Gene For over a decade, the Swiss Institute of Bioinformatics and HES-SO/HEG Geneva have maintained the SIB Literature Services (SIBiLS, Gobeill et al 2020). The services are part of a broader initiative, led by SIB, the Swiss BioData ecosystem (SBDe). In 2023, SIBiLS received a significant upgrade with the launch of the Biodiversity PMC digital library. SIBiLS operated the backend of the Biodiversity PMC front-end. It is emerging as a global resource in the field as it complements and addresses limitations of biological scholarly databases such as NLM's PMC or EuropePMC as it is likely the largest digitally-native repository of articles for biodiversity research and related disciplines; thus delivering a broad coverage "One Health" library.

BioMoQA aims to enhance Biodiversity PMC to address key biodiversity questions about the effects of climate change, habitat loss or invasive species on biodiversity and eco-systems, and the implications for human societies. The new services will support experts and researchers in biodiversity, ecology and environmental sciences who need better search & access services to all bio-related publications.

We will enhance the platform's AI-powered analytical services (i.e Question-Answering, SPARQL endpoint) and the acquisition & FAIR-ification of new and original contents as operated by PLAZI and in particular the TreatmentBank database, which contains Open Access taxonomic treatments from Open Access (OA) and non OA articles. We will evaluate how these services can help ecologists from the University of Neuchâtel to monitor biodiversity on island ecosystems. These new services will serve to supply scientific communities interested in biodiversity with a holistic Single Access Point to enhance evidence-based conservation of biodiversity and support the restoration of ecosystems.