swissuniversities

Programme Open Science II

Call for proposals: B3.2 ORD Funding for Existing Research Infrastructures of Strategic Relevance

List of Approved Projects

Call Deadline: 02.03.2025

Decision by the Delegation Open Science: 23.06.2025

Overview Approved Projects Call B3.2

Short title	Full title	Leading institution	Partner institution(s)	Project Manager	Total costs (CHF)	Funds requested (CHF)
Bgee1k	Towards open single-cell data in a thousand of species	UNIL		Marc Robinson-Rechavi	500 000	250 000
SPSP-FAIR-Bact	Fostering bacterial genomic data FAIRification by extending IMMense and SPSP	UZH	SIB Swiss Institute of Bioinformatics	Tim-Christoph Roloff Handschin	392 000	165 500
e-cORD	e-codices – linked Open Research Data	Unifr		William Duba	304 252	151 946

Short Summaries of the Projects

Abstracts by the applicants:

Bgee1k

Towards open single-cell data in a thousand of species

Bgee is an infrastructure allowing the retrieval and comparison of gene expression data in animal species. It integrates several data types, notably single-cell RNA-Seq along with bulk RNA-Seq, in 52 species to date. Bgee has ORD funding until 30.06.2025 for Existing Research Infrastructures of Strategic Relevance, as well as to coordinate scFAIR, a consortium of single-cell resources to define metadata schema and data exchange formats for single-cell data. As a recogni-tion of its contribution to the international life sciences research, Bgee is certified as a Global Core Biodata Resource, i.e. a resource that is “of fundamental importance to the wider biological and life sciences community”. Thanks to its standardization of transcriptomics data, Bgee is also recognized at the European level by Elixir as a Recommended Interoperability Resource, i.e., a resource “that facilitates the FAIR-supporting activities in scientific research”.

While there is an exponential increase in single-cell and bulk transcriptomic data, they are still underexploited because they are difficult to access and to reuse, i.e., not FAIR (findable, accessible, interoperable, reproducible). We propose to extend the Bgee model from 52 to 200 species within the timeline of this pro-ject, paving the way to reach 1000 species integrated ("Bgee1k"), to allow FAIR access to reference gene expression data for all animals with data. These spe-cies for which we will make data structured and easily accessible include key pest or pathogen carriers (e.g. mosquitoes), agronomic species (e.g. bees), and a large coverage of biodiversity by projects such as ERGA (European Reference Genome Atlas).

SPSP-FAIR-Bact

Fostering bacterial genomic data FAIRification by ex-tending IMMense and SPSP

The Swiss Pathogen Surveillance Platform (SPSP) is an infrastructure to store and share genomic data from microbes and is seen as positive example across Europe. Labs, data scientists, authorities and federal offices heavily use SPSP since the COVID-19 pandemic, especially for viruses. The objective of this project is to extend SPSP for bacterial genome sequences. IMMense is the analysis pipeline processing bacterial sequencing data within SPSP and is being developed as an open-source data analysis pipeline at the Institute for Medical Microbiology at UZH. With a steadily growing number of bacterial species being added to SPSP, new sequencing data types becoming relevant and species-specific analyses being requested, IMMense has to be extended and partly redesigned to fulfil all the needs. Reproducibility of analyses within such a clinically relevant pipeline is critical. Therefore, all updates and feature additions require careful quality control (QC) and validation. For this, a validated in-silico reference dataset of all relevant species could be used.

SPSP can currently store bacterial data and share it with other repositories and authorities but so far there is no front-end allowing scientists to explore the data generated by IMMense which limits SPSP’s usefulness.

In this project, we plan to extend IMMense to cover the current top priority analyses needed for SPSP, generate a reference dataset to regularly benchmark the quality of the output of IMMense, and develop a front-end for bacteria for SPSP to visualize the results generated by IMMense.

e-cORD

e-codices – linked Open Research Data

For the last twenty years, e-codices – Virtual Manuscript Library of Switzerland has been one of the leading digital libraries containing complete series of scientific images of medieval manuscripts, linked with scholarly descriptions encoded for reuse. Other projects worldwide have built on the spread of virtual libraries to offer complementary information on their contents, such as musical notation, chant, or specific texts. At the same time, the rise in the study of the materiality of manuscripts has produced data on the composition of texts, from analyses of types of ink to DNA sequencing of the animals whose skin forms the parchment of the pages.

The project e-cORD will upgrade the infrastructure of e-codices to support the next generation of linked Open Research Data (ORD) and establish a common European framework for sharing rich metadata and images concerning medieval manuscripts and related cultural heritage objects. It will strengthen coordination with similar national-level infrastructure projects in other European countries (such as the German Handschriftenportal, the Dutch eCodicesNL) to establish a common means of information exchange. It will enrich the data e-codices shares concerning the manuscripts it publishes and their images. To achieve these objectives, the project will systematically upgrade the Free and Open-Source components and frameworks that constitute the infrastructure of e-codices.