swissuniversities

**Swiss Open Research Data Grants (CHORD): Track A 3rd Call**
**List of Approved Projects**

Call Deadline: 02.06.2023

Decision by the Delegation Open Science: 25.10.2023

**Overview Approved Projects Track A 3ʳᵈ Call**

Evaluated projects: 15      Approved projects: 7, of which 1 with conditions      Withdrawals: 1      Funding rate: 53%

| Short title | Full title | Leading institution | Partner insti-tution(s) | Project Manager | Total costs (CHF) | Funds Requested (CHF) |
|---|---|---|---|---|---|---|
| TechnoPortal | An ontology hosting portal for enhancing data sharing in wind energy | Eastern Switzerland University of Applied Sciences (OST) | HES-SO VS | Sarah Barber | 149'800 | 74'900 |
| PILOT-IAQ | Development of the First Pilot Global Indoor Air Quality Database | École Polytechnique Fédérale de Lausanne (EPFL) | | Dusan Licina | 150'000 | 75'000 |
| ORoDaPT | Open Rock Glacier Data Production Tools | University of Fribourg (UniFR) | | Cécile Pellet | 150'000 | 75'000 |
| PCaMiner | Prototyping the harmonization and the inte-grated mining of prostate cancer RNA sequenc-ing data sets | Università della Svizzera italiana (USI) | UniBE, EPFL, UniBas, PSI | Jean-Philippe Theurillat | 157'500 | 75'000 |
| DAMOCLES | Data Management Concept for Underground Laboratories | ETH Zurich | | Marian Hertrich | 145'309 | 72'001 |
| OpenSWICE | Open research data practices for the interdisci-plinary consortium SWICE | University of Fribourg (UniFR) | EPFL | Julien Nembrini | 119'100 | 59'400 |

## Short Summaries of the Projects

*Abstracts by the applicants:*

TechnoPortal

***An ontology hosting portal for enhancing data sharing in wind energy***

A lack of data sharing is one of the largest hurdles to innovation in the wind energy sector. To successfully share data and implement open science principles, domain-specific ontologies and other semantic artefacts need to be adopted by the community. Most existing conceptual models in wind energy and related domains are not published following the linked data principles or FAIR guidelines. Therefore, the aims of this project are to (1) prototype a public ontology hosting portal for the technology sciences domain, and (2) use it to build up a community of users in the wind energy sector. First, the packaged On-toPortal Virtual Appliance will be forked and used to host a new public "TechnoPortal". A community of users will then be built up using WeDoWind, a framework developed for creating mutually beneficial collaborations using the idea of "challenges". The WeDoWind Wind Energy Ecosystem consists of 300+ people actively collaborating and sharing data around multiple "challenges". For this part, existing semantic artefacts related to wind energy will first be evaluated, chosen and hosted on the TechnoPortal. Then, WeDoWind "challenges" will be run in order to bring the community together to develop new ontologies and conceptual data models for specific tasks, such as describing coordinate systems on a wind turbine. Finally, the new ontologies will be applied and tested on real data shared via WeDoWind "challenges". The activities will be communicated as part of the IEA Wind Task 43 Metadata Challenge webinar series and connected training activities for sustainability.

PILOT-IAQ

***Development of the First Pilot Global Indoor Air Quality Database***

The advent of smart building sensors is enabling rapid expansion of indoor air quality (IAQ) monitoring and controls aiming to secure healthier, and more comfortable indoor spaces. Deployment of these technologies can be informed by vast amounts of existing IAQ measurements, linked to human health, satisfaction, and productivity. This project aims to compile, catalog, and openly disseminate IAQ datasets from decades of indoor air quality field investigations. We will adopt a unique cost-effective method that makes use of the existing data to create new value available to everyone. Through a robust network of internationally recognized experts and incentive-based outreach to data contributors, we will collect and standardize a vast amount of published scientific data from recent indoor air quality field campaigns. Indoor air pollution data will be accompanied with comprehensive data about building, climatic and geographical characteristics, and whenever available, individual-level exposure, and outdoor air data. The collected data will be subject to rigorous quality control, statistical treatments, and will be made available online and as open-source, to allow end users to access, download and explore the gathered data. Finally, the open-source database will be promoted to wide communities around the world.
This pilot open-source IAQ database (PILOT-IAQ) is expected to enable stakeholders to retrieve and analyze data from it, to produce new knowledge, and improve and undertake new and existing initiatives and policies for effective IAQ management.

ORoDaPT

***Open Rock Glacier Data Production Tools***

As a component of the cryosphere, permafrost is one of the most sensitive environmental systems to climate change. Its observation at the global scale with comparable data products is thus critical, and ice/debris landforms generated by permafrost creep, such as rock glaciers, can be used as a proxy to assess the impact of climate change on mountain permafrost. Since 2018, the Rock Glacier Inventories and Kinematics (RGIK) Action Group has been developing community-based standards and guidelines for compiling Rock Glacier Inventories (RoGI) and calculating Rock Glacier Velocity (RGV) time series from a climate-oriented perspective. However, there is currently a lack of resources to extend the work of the RGIK community, to support Open Research Data (ORD) practices and to meet the requirements for scientific reproducibility.
This project intends to bridge this structural and operational gap by (i) developing prototype tools tailored for the collection, processing and analysis of RoGI and RGV data products and (ii) setting up a Data Management Structure (DMS) beta version for the archiving, visualization and dissemination of rock glacier data products. We start by understanding the research community's needs and determining the expected functionalities and requirements. We will then develop open-access, easy to use prototype tools and a DMS beta version using open source widely-used software to ensure their long-term operability and further development. Both outputs will be complemented by training and outreach materials to facilitate and broaden their use within the permafrost community in particular and the geosciences in general.

PCaMiner

***Prototyping the harmonization and the integrated mining of prostate cancer RNA sequencing data sets***

Next-generation sequencing (NGS) has revolutionized biomedicine by enabling us to shed light on disease-specific molecular alterations. Because NGS raw data is frequently protected and the analysis labor-intensive, specific web portals have emerged that enable integrated mining of various DNA data sets (e.g., cbioportal). In contrast, the cross-study comparison of RNA sequencing data is significantly more challenging because of the artifacts introduced by different analysis pipelines, referred to as "batch effects". We recently overcame these hurdles and generated a harmonized prostate cancer RNA sequencing atlas from different data sets of early primary to late-stage metastatic disease. This enabled us to describe the roadmap to tumor progression in an unprecedented quantitative manner.
The goal of this project is to update the atlas with newly released data sets and establish a user-friendly web tool for researchers and clinicians in academia and health care to accurately investigate gene expression changes related to prostate cancer progression and therapy resistance in a customized and multidimensional manner. Researchers will use the data to identify drug targets to prevent disease progression or develop biomarkers related to prognosis or drug response. Clinicians may use the tool to determine specific therapeutic opportunities for their patients.
As such, the harmonized RNA sequencing atlas and the related web tool will extract relevant information from various protected data sets and make them findable, accessible, reusable, and interoperable (FAIR Principles). More broadly, the deliveries may establish a blueprint for RNA sequencing Open Data Research in other cancer types or more largely other diseases.

DAMOCLES

***Data Management Concept for Underground Laboratories***

Recent advances in open data science are initiating procedures that allow past, recent, and future data to be fully reproducible, fulfill best scientific and technical practice and thus allow results and analyses to be comprehensible and citable for scientific publications and technical reports.
With the present project the BedrettoLab an infrastructure of ETH Zürich, will develop a concept to get underground laboratories ready for this challenge. Key aspects of the proposal are a systematic data workflow (acquisition, transfer, storage, and dissemination) and a governance structure to produce exchangeable and comparable data.
The BedrettoLab is a young research facility with increasing projects and partners making the establishment of a scientific data structure essential. With this proposal we have the opportunity to develop a prototype workflow for a state-of-the-art data management structure which is applicable in other underground laboratories. Ideally, the results of the project will provide the basis for a standardization of data workflows in the different underground laboratories.
With this objective in mind we will focus on responding to questions corresponding to the ultimate goal of developing a data management concept towards FAIR principles acknowledging the fact that many of these questions have been addressed in the past but failed in the implementation. The data management concept in development in this proposal shall address the data governance and the data coordination between all partners and stakeholders.

OpenSWICE

***Open research data practices for the interdisciplinary consortium SWICE***

The SWICE project (sweet-swice.ch), supported by the Swiss Federal Office for Energy's SWEET program, aims to accelerate innovations for Switzerland's Energy Strategy 2050. With a diverse consortium comprising 10 Swiss Higher Education Institutions and 4 Research and Consultancy Companies, along with over 25 cooperation partners from the public and private sectors, coordinating and harmonizing open research data (ORD) practices presents significant challenges. These challenges are crucial for effective collaboration among project stakeholders and the transparent publication of sources for federal-level political decisions under the SWEET program.
To tackle these challenges, the OpenSWICE project focuses on establishing and coordinating ORD practices within the existing SWICE consortium. The project follows a three-phase plan. First, surveys and workshops within the consortium collect the needs and requirements of research partners regarding data types, anonymization techniques, and infrastructure. Second, a platform is proposed to simplify the formatting and dissemination of ORD based on FAIR principles, aligning with the consortium's identified requirements. Lastly, workshops are conducted to facilitate the adoption and establishment of the developed ORD practices among consortium members, as well as to adapt data collection platforms to evolving needs.
The overarching goal of the OpenSWICE project is to address coordination challenges related to ORD within the SWICE consortium. By establishing standardized practices, ensuring data accessibility, and promoting sustainable implementation, the project aims to contribute to the success of Switzerland's energy strategy. It strives to enhance collaboration, enable transparency, and support informed decision-making in the energy sector through effective management and utilization of open research data.