swissuniversities

# Swiss Open Research Data Grants (CHORD): Track C 2nd Call
# List of Approved Projects

Call Deadline: 19.07.2023

Decision by the Delegation Open Science: 25.10.2023

## Overview Approved Projects Track C 2nd Call

Submissions: 17 Approved projects: 8 Funding rate: 47%

| Short title | Full title | Leading institution | Project Manager | Total costs (CHF) | Funds Requested (CHF) |
|---|---|---|---|---|---|
| tag-Xplore | tag-Xplore - a tool and method for the assessment of tag usage within and across multiple digital editions | Universität Zürich | Jürgen Bernard | 100 000 | 50 000 |
| EyeStore+ | Opening EyeStore for exploration: Providing a secure, searchable, well-documented and easy-to-use platform for scientific | Universität Zürich | Lena Jäger | 100 000 | 50 000 |
| EASI-SPSP | Establishing Automated data Submission Interfaces for FAIR clinical bacterial data, integrating also phenotypic antimicrobial resistance data | Universität Bern | Stephen Leib | 100 000 | 50 000 |
| STORIAEDU-CH | Innovation in history of education research through www.storiadelleducazione.ch | Scuola universitaria professionale della Svizzera italiana | Wolfgang Sahlfeld | 15 000 | 7 500 |
| Swiss-AL OER | Strengthening ORD competences in the Applied Sciences: Open Educational Resources for Swiss-AL | Zürcher Hochschule für Angewandte Wissenschaften ZHAW | Julia Krasselt | 100 000 | 50 000 |
| SleePyLand | SleePyLand: A python library to analyse the large amount of NSRR sleep data via deep learning algorithms | Scuola universitaria professionale della Svizzera italiana | Luigi Fiorillo | 100 000 | 50 000 |
| DERMAI | Advancing Dermatology Imaging: Curating and Open-Sourcing Representative Datasets for Inclusive AI Development | Universität Basel | Alexander A. Navarini | 100 000 | 50 000 |
| SUOHSA | SUPSI OPEN HERITAGE SAMPLES ARCHIVE | Scuola universitaria professionale della Svizzera italiana | Francesca Piqué | 90 000 | 45 000 |

## Short Summaries of the Project

*Abstracts by the Applicants:*

tag-Xplore

***tag-Xplore - a tool and method for the assessment of tag usage within and across multiple digital editions***

As a promising approach for standardization, multiple digital editions have agreed on using the TEI-XML format. Many individual digital editions in Switzerland use a set of up to 585 pre-defined TEI tags for a fine-grained annotation of editions with thousands of documents. However, many editions so far rather exist as islands on the Web, and do not offer much connectivity, i.e., much potential for researchers interested in relating, contextualizing and reusing the content of multiple editions remains lost.
A stepping stone in the process of standardization and unification may be the systematic analysis and assessment of tags and tag usage across editions. The fine-grained tag information in multiple thousands of documents across editions forms a most promising exploration space, for the identification of structural patterns and (in-)consistencies within and across editions.
To exploit the potential of tags in digital editions, we propose tag-Xplore, a method and interactive tool that incorporates a diverse assembly of digital editions in Switzerland. tag-Xplore will contribute a novel interactive dashboard for the assessment of tag usage within and across multiple editions. Insights gained with tag-Xplore will help to improve standardization and reuse of data in digital collections, and help to improve the workflow of setting up a new digital edition or for long-term archiving. As a consortium, we will establish a new culture of ORD in the community of digital editions by organizing workshops and developing standards.

EyeStore+

***Opening EyeStore for exploration: Providing a secure, searchable, well-documented and easy-to-use platform for scientific***

We propose substantial extensions to the EyeStore project (CHORD grant approved - Phase B, Track C, 1st call) with the purpose of building an open repository for multilingual eye-tracking-while-reading data in a standardized format.

Through the original project grant we have been able to

1) Provide some storage space for eyetracking data.
2) Define the initial data formats and processes.
3) Set up a framework for documentation and guidelines.
4) Establish preprocessing procedures.

In the current application, we propose to enhance EyeStore by

1) Ensuring EyeStore is a secure space with high data quality standards: Strengthening the safety and privacy aspects of the storage space as well as defining and implementing a comprehensive test suite for the quantitative assessment of data quality
2) Making EyeStore searchable
3) Providing a user interface for documentation and hands-on tutorials
4) Integrating the preprocessing pipeline to ensure transparency and reproducibility

EASI-SPSP

***Establishing Automated data Submission Interfaces for FAIR clinical bacterial data, integrating also phenotypic antimicrobial resistance data.***

The Swiss Pathogen Surveillance Platform (SPSP) has collected, curated, processed and shared openly more than 160’000 SARS-CoV-2 genomic datasets and associated metadata, making Switzerland the 5th world-wide contributor of SARS-CoV-2 open data. It is now time to make SPSP also available for clinical bacterial data.
SPSP is now part of a Swiss-wide project bringing together all the University clinical microbiology laboratories (SPHN NDS IICU). Within this project, data providers will submit to SPSP bacterial genomic data from positive blood cultures of Intensive Care Unit (ICU) patients and bacterial MALDI-TOF profiles from all patients of Swiss University Hospitals ICUs, to be then shared to open data repositories. While the former represents a few thousand genomic datasets per center and per year, the latter corresponds to a few thousands datasets per center and per week. In this context, entries into SPSP should be automated in a robust fashion, similar to the data sharing mechanisms already in place with anresis.ch for phenotypic resistance data.
Here, we aim to (I) set up automated interfaces to allow University clinical microbiology laboratories to robustly prepare high quality data by automatically extracting the required fields from their internal information systems and submitting them to SPSP by a standardized procedure; and (II) implement a bidirectional interface between anresis and SPSP for automatic data exchanges between these platforms, thereby enabling interoperable data integrations and making unique FAIR datasets accessible to the research community.

STORIAEDU-CH

***Innovation in history of education research through www.storiadelleducazione.ch***

The aim of this project is to improve the already existing multilingual knowledge portal www.storiadelleducazione.ch | www.bildungsgeschichte.ch | www.histoiredeleducation.ch as a research tool for crosslinguistic and nationwide data research about history of schooling and education in Switzerland. The search engine offers a central interface for full text and metadata search across a wide variety of historical publications, images, transcripts and other source materials and data outputs provided by research projects and archives alike through the internet. The portal, which is broadly supported by the scientific community and run by an association, is currently undergoing a transition from prototype status to an operating phase. This involves a necessary rebuild of the website in order to improve ORD practices from a technical stance and to further promote data sharing and reuse in the field of history of education. As a side effect of this re-implementation, some data sets prepared for the prototype have to be re-indexed. This project will cover the re-integration of data from three major projects with a significant relevance to the history of education in Switzerland that, although publicly available, cannot be retrieved via standard open access interfaces. It will include specifically for each data set the analysis of the data structure, the definition of the retrievable units and the harvesting, mapping and importing of the data into the index of the search engine. The know-how gathered in this process will provide a core asset for future expansions of the inventory.

Swiss-AL OER

***Strengthening ORD competences in the Applied Sciences: Open Educational Resources for Swiss-AL***

Recent developments of platforms for Open Research Data (ORD) have led to a high demand for training materials that enable researchers to successfully use FAIR data. That is particularly true for platforms offering access to large collections of language data and the application of data-driven and distant-reading methods that are not (yet) established in many non-linguistic disciplines. The following proposal builds on the language data platform Swiss Applied Linguistics (Swiss-AL), developed by the ZHAW Digital Discourse Lab and part of the CLARIN-CH ecosystem of linguistic infrastructures. We argue that a successfully implemented and sustainable ORD strategy needs to include the training and empowerment of the research community that is addressed by ORD resources. In the proposed project, we will develop open educational resources for Swiss-AL (OER) consisting of explanatory videos, user stories, statements by discipline specific testimonials, screencasts, wikis, and hands-on exercises. Resources will address four selected target disciplines for language data (public health, social work, organizational communication, journalism). Based on exemplary research scenarios, the resources will show how Swiss-AL can be successfully used as an ORD resource outside (applied) linguistics. OER will be developed in collaboration with representatives of the four target disciplines and experts in media-based learning and design (ZHAW OER Competence Centre, LerNetz, FrameEleven). The resources will be integrated in Swiss-AL and into the CLARIN-CH infrastructure and published as OER.

SleepPyLand

***SleePyLand: A python library to analyse the large amount of NSRR sleep data via deep learning algorithms***

Polysomnography (PSG) is used in sleep medicine as a diagnostic tool to objectively analyze sleep quality. Sleep scoring is the procedure of extracting sleep cycle information from whole-night electrophysiological signals. The sleep recordings are usually scored by human sleep experts according to the American Academy of Sleep Medicine (AASM) manual. The scoring procedure requires up to two hours of work per whole-night, and it is highly biased by the inter- and intra-scorer variability. A wide variety of machine learning based algorithms have been proposed to automatize the sleep scoring task, reaching very good results. However, none of these algorithms has been ever introduced into the daily clinical routine. One undeniable reason is that there is no tool, to date, to fairly compare and evaluate the available sleep scoring algorithms on the same datasets and metrics. The main goal of the project is to release SleePyland, a freely accessible Python based software, able to:

1. Manage, pre-process, analyze and extract overall sleep parameters/statistics from a large amount of sleep studies, i.e., tens of thousands of PSG recordings from open access data repository (i.e., NSRR).
2. Provide for each sleep recording a structured report with a completely automated ensembled sleep scoring analysis. The idea is to include in a single open-source repository the recent high-performing deep learning based sleep scoring algorithms. The algorithms will be trained and fairly evaluated on common-ground datasets, to then output the ensemble prediction.

DERMAI

***Advancing Dermatology Imaging: Curating and Open-Sourcing Representative Datasets for Inclusive AI Development***

Our project addresses key gaps in the field of dermatology imaging by curating and open-sourcing a diverse and representative dataset. We aim to enhance data interoperability by converting labels to ICD-11 coding and implementing rigorous data cleaning processes. Specifically, we focus on filling the gap in representative data for richly pigmented skin populations by open-sourcing 80% of the data collected worldwide through the PASSION project. The remaining 20% will be used as a private test set, ensuring unbiased evaluation. Researchers will have the opportunity to test their open-source algorithms on the curated dataset, provided they remain freely available to the public. By fostering sustainable practices and engaging with the community, our project aims to advance the field of dermatology imaging, promote inclusivity, and facilitate collaboration and innovation.

SUOHSA

***SUPSI OPEN HERITAGE SAMPLES ARCHIVE***

Research institutions involved in the study of cultural heritage often maintain archives of samples collected from valuable and unique artefacts. Like the object they were taken from, these samples hold value for further research and education. Unfortunately, access to these collections is often impeded by poor archiving and difficult access and therefore the long-term preservation and full use is not guaranteed. In addition, physical samples not properly stored might age and deteriorate. SUPSI_IMC has an important collection of more than 1,000 samples taken as part of research projects in many different cultural heritage sites, but currently organized in separate groups and not under a common unified index. The SUPSI Open Heritage Sample Archive (SUOHSA) project aims to physically reorganize the collection of samples and to create a digital open-access catalogue that will allow FAIR management of the heritage samples data. The project will define standards for cataloguing that will allow easy registration when new samples are collected as part of future research projects. Once created, the database could serve as model by other research institutions within the SwissCRC and beyond for archiving built heritage samples. The SUPSI catalogue will be included in the "International Registry of Heritage Sample Archives", an open web-based platform created as the result of ICCROM-led initiative of which SUPSI is the only Swiss partner.