

## ETH ORD Programm

### Contribute call 1 - deadline 1st April 2022

Number of projects submitted : 12 / Number of projects financed : 7 / 58% success rate

Project title	Abstract	Applicant	Institution involved
'XYT', a python package to analyze activity-travel behaviors, organization and scheduling	Today, more and more digital data are generated by urban dynamics. Yet, the generated data is extensive and heterogenous. Datasets are large, multi-sourced, often noisy, and come in various formats and standards. In addition, a particularity of urban data is its mix in terms of level of restriction. While geolocation data is private and sensitive, public transit schedules are open data. In this context, there is a need (i) to provide a framework to re-unify the multiformity of urban dynamics data, (ii) to articulate open and restricted data, (iii) to cohere offer and demand data, and (iv) to keep track of a privacy metric. This project proposes to develop and release an open Python package to address these four needs and therefore contribute to Urban Mobility Open Research Data practices.	Schultheiss, Marc-Edouard	EPFL
A standardized database framework for synthetic carbon-based solar fuels	In this proposal, we seek to significantly extend the recently created Solar Fuels Database where we developed a systematic machine-readable framework for solar to fuel devices and developed an online interface for data entry and visualization, with the ultimate aim of accelerating the development of such technologies. This project would extend the database to further chemistries and methodologies beyond currently captured in the database (photo-electrochemical water splitting). Specifically, we plan to extend the solar fuels considered to include carbonmonoxide, syngas, formic acid, methane, ethanol etc. and extend the technological pathways to include thermochemical redox cycles. We will engage with international research communities to ensure a judicious selection of metadata is captured by the database. These reporting guidelines will facilitate standardization in data reporting, and inclusion in the database will provide an incentive for authors as inclusion of research could lead to improved findability and greater dissemination of results. In line with the open-science goals of this project, it is envisaged the database will become continually maintained through community submission of new papers. This greatly improved and openly accessible resource will consolidate previous work and provide the overview required to gain novel insight in to the field and thereby identify promising future research trends required to move solar fuels towards wide-scale implementation.	Holmes-Gentle, Isaac	EPFL

Project title	Abstract	Applicant	Institution involved
DISDRODB: A global database of raindrop size distribution observations	The raindrop size distribution (DSD) describes the concentration and size distributions of raindrops in a volume of air. It is a crucial piece of information to characterize microstructural properties of rainfall, to quantitatively interpret remotely sensed signals, and to properly represent rainfall in numerical atmospheric models. Disdrometers are instruments used to collect observations of the DSD and have been deployed in many regions around the globe. But those data remain scattered in various places, with disparate formats, and specific processing tools. It is hence difficult at present to investigate the spatial and temporal variability of the DSD on large scales. DISDRODB aims to address this challenge by promoting common standards for data format, quality-control and processing. To do so, we will build a database and processing framework to store raw measurements in an easy-to-share format, process those data to obtain clean DSD data and finally derive DSD related products (rain rate, kinetic energy, mean size...). The ultimate goal is to provide a much-needed tool to the scientific communities dealing with DSD (for process understanding, remote sensing and modelling applications).	Berne, Alexis	EPFL
Open Workshops on Image Data Best Practices	Image data are an essential part of many fields of scientific inquiry ranging from astronomy to microbiology. Thanks to advances in technology we are able to generate massive, rich image datasets from which insights for many studies could be generated if they are distributed in an open manner. There have been several successful and useful centralized efforts to host and distribute open image data. Unfortunately due to the size, variety of vendor and field-specific formats, and complexity of modern analysis methods, it remains difficult for many scientific researchers to produce and prepare their data in a findable, accessible, interchangeable, reproducible (i.e., FAIR) manner that is suitable for existing open research data (ORD) resources. Thus, we propose to contribute an open source training resource for learning how to process and prepare image data for distribution with ORD best practices. We will produce an online handbook that teaches image ORD best practices and curates existing image ORD resources. Further, we will create and present workshops based on our online handbook to both train researchers and build community consensus on image ORD best practices. We anticipate these training materials will both help empower ETH domain researchers to make the most of existing ORD resources and new ones being developed.	Yamauchi, Kevin	ETH Zürich
PhenoMast - Integration of standardized tree seed mast observations into existing phenology monitoring networks	Most of the dominant tree species in Switzerland do not produce a constant amount of seeds from year to year but rather show a regular pattern of massive seed production every few years followed by years with no seed production, i.e. they exhibit a masting behaviour. Seed mast dynamics have a strong influence not only on tree regeneration but also on population dynamics across the food web. The timing and interval of seed masts are controlled by diverse environmental triggers and are changing in the recent warming decades. Given the ecosystem-wide impacts of seed mast it is not surprising that in recent years, the processes underlying forest regeneration have gained considerable attention in the scientific literature. Nevertheless, many networks and field surveys focusing on phenology often ignore seed mast. We plan to organize a workshop bringing together managers from different existing phenological networks in order to contribute and integrate an ORD protocol for collecting seed mast data obeying elaborated standards within these observation networks (programs). Based on this standardized and simple method, seed mast data can be collected within existing phenology networks to benefit researchers from different fields, including modellers, physiologists and ecologists. The collected data will be curated and made publicly available on the ORD platform of MastWeb facilitating collaborations with other European- or worldwide phenology networks.	Scherrer, Daniel	WSL

Project title	Abstract	Applicant	Institution involved
Seismological Software Stack, "Portable self-contained software environments enabling reproducible seismological research"	<p>In this project, I will create containers for vital pieces of software that are used in the seismological community. These softwares are used by researchers around the globe, but often have complicated requirements and dependencies, which makes it difficult for inexperienced users to compile the softwares. The containers that I will create in the context of this proposal run in a virtual environment so that all the dependencies and requirements can be guaranteed and this environment will allow for the interaction between the various pieces of software. The containers are portable as well, which will allow researchers to share their applications using the softwares with greater ease, improving the reproducibility of science. The project has two intended contributors, a postdoctoral researcher and a doctoral student, who will work on the creation of the containers. In order to ensure the longevity of the project, Continuous Integration and Continuous Development pipelines will be employed, to maintain the containers on an open access repository. For these pipelines, a physical workstation will be built with GPU-accelerated hardware. Three quarters of the way through the project, a workshop will be organized together with the developers of the software in order to collect feedback about the project and to increase awareness of the containers. The feedback will be incorporated into the project in the final three months.</p>	Brackenhoff, Johannes	ETH Zürich
Seizing the treasure: making long-term environmental data available for eLTER and beyond	<p>The European Long-term Ecosystem Research (eLTER) facility Stillberg was established in 1975 in a treeline ecosystem near Davos, Switzerland. In almost 50 years of research, we have collected a wealth of environmental and ecological data at this unique alpine site. These include monitoring data of a large-scale treeline afforestation experiment, meteorological data, time series of plant responses to free-air carbon dioxide enrichment, soil warming and nutrient addition, as well as plant-snow, plant-soil interactions, and drivers of tree seedling recruitment. In this project we aim at contributing these important ecological datasets from Stillberg as open research data (ORD). By curating the existing datasets before the upload to national and international ORD platforms (EnviDat and DEIMS-SDR) we will improve their visibility, quality, and availability. The provisioning of this long-term environmental data will foster transdisciplinary and global research syntheses and meta-analyses, improving our understanding of long-term ecosystem processes in mountain regions and supporting the development of adequate adaptation strategies.</p>	Frei, Esther	WSL