ETH ORD Programm

Explore call 1 - deadline 14th April 2022

Number of projects submitted : 48 / Number of projects financed : 13 / 27% success rate

Project title	Abstract	Applicant	Institutes	Amount CHF
Initiative for primary bio-NMR open research data	Nuclear Magnetic Resonance (NMR) spectroscopy, one of the major techniques in structural biology, suffers from the lack of an open database for its primary data, the multidimensional NMR spectra. These spectra are the basis for studies in atomic detail of the structure, dynamics, interactions, and functions of proteins, but are so far generally not made available to other researchers. Our NMRprime initiative for primary bio-NMR open research data will establish a database following the FAIR principles for biomolecular NMR spectra and explore its coordination with existing open databases for related, derived data, such as the Protein Data Bank (PDB) for protein structures and the Biological Magnetic Resonance Bank (BMRB) for chemical shift assignments. NMRprime will provide the framework for upload of NMR spectra in all commonly used formats, automatic annotation using machine learning techniques, search facilities, and open access to the archived spectral data. The long-term goal, to be achieved together with PDB, BMRB, and journal representatives, is to make the deposition of the underlying NMR spectra an integral part of bio-NMR projects published in peer-reviewed journals in the same way as the requirement to deposit protein structures in the PDB. We are singularly suited for the NMRprime initiative because of decades of experience in bio-NMR and the possibility to build NMRprime on the re-use of central components of our recent fully automated NMR spectra analysis system.	**Riek Roland** Güntert Peter	ETH Zürich	150,000
Development of Open Research Data Analysis Services supplementing astronomical Open Research Data	Over the last decade, astronomers have been developing a new field of Multi-messenger astronomy that combines radio-to-gamma-ray electromagnetic signals with neutrino and gravitational wave signals to get new insights into physics of astronomical sources. Diversity of data sets (mostly available as ORD) involved in multi-messenger data analysis poses a challenge of reproducibility of multi-messenger analysis results and their traceability to raw observational data. Without thoughtful stewardship of the data reduction and elaboration methods, the key principle behind the ORD concept, the “Findable-Accessible-Interoperable-Reusable” (FAIR) principle, is not fully implemented: the results derived from ORD are not reproducible, the raw observational data are not reusable. The project proposes a solution to this problem through a setup of cloud-based ORD data analysis services (ORDAS) that supplement the ORD and explicitly assure reproducibility of results and reusability of the ORD. We propose to develop such services for two major next-generation facilities, gamma-ray observatory Cherenkov Telescope Array (CTA) and radio observatory Square Kilometre Array (SKA). We will demonstrate the power of ORDAS for CTA and SKA by engaging the community of multi-messenger astronomers in ORD practices through a Multi-Messenger Online Data Analysis (MMODA) platform for the study of transient multi-messenger astronomical sources for which on-the-fly analysis of multi-messenger ORD is vital.	**Kneib Jean-Paul** Tolley Emma Neronov Andrii Savchenko Volodymyr	EPFL	149,277

Project title	Abstract	Applicant	Institutes	Amount CHF
Explore AiiDA for Regional Inverse Modelling of Greenhouse Gases	Inverse modelling of greenhouse gas emissions is a powerful tool combining atmospheric observations and trans port simulations to provide policy-relevant, real-world emission estimates. Inverse modelling requires computational workflows that deal with multiple atmospheric model simulations, pre- and post-processing steps, inversion codes and collection/aggregation of results across an ensemble of individual realisations. These steps are not handled in a traceable, repeatable and user-friendly way. Here, we propose to use the 'Automated Interactive Infrastructure and Database for Computational Science' AiiDA system to develop prototype workflows and implement plugins for automation of the above steps. AiiDA is a well-established tool in the field of computational material science within the ETH-domain. We will build on its flexible structure to transfer its powers to the atmospheric modelling community, specifically that of inverse modelling. We will interact with our Swiss and European partners to design generic workflows that can be easily adopted by other researchers working on inverse modelling. Our implementation will improve the traceability, repeatability, findability, and sharing of emission estimates that are crucial in the assessment of climate change policy and legislation. We expect to carry the achievements beyond the use in our own group to other researchers with whom we cooperate closely in the context of international and European programmes and projects.	**Henne Stephan**	EMPA	147,000
The Imaging Plaza: A Curated Online Catalog of FAIR Imaging Software.	This project proposes the creation of The Imaging Plaza, a curated online catalog of FAIR imaging software for scientists of the ETH domain. The concept stems from the pressing need for a much greater visibility of, and accessibility to, the research output of the Swiss imaging community. The Imaging Plaza aims to address this issue by lowering the entry cost for non-experts to find and adequately use the imaging code produced by their peers. Contrarily to existing options, its purpose is not to host code, but to allow users to navigate with confidence the sea of available options, providing extra incentives, guidance and support. The end goal is to accelerate the dissemination of new imaging code within the ETH domain. The project is a joint endeavor between the EPFL Center for Imaging and the SDSC. It involves deployment test sites at EPFL, PSI, and ETHZ. Experts from the EPFL Center for Imaging will actively curate the platform to ensure that the listed software is easily reusable and well-documented. SDSC engineers will provide readily shareable and reusable runtime environments for published codes. Users will be able to easily navigate the catalog through a specially crafted search system and the definition of a common language (ontology) that describes the vast and heterogeneous world of imaging. Finally, a new “FAIR levels” framework will indicate how easily an untrained user can access and reuse a particular code.	**Unser Michael** Donati Laurène Riba Grognuz Oksana	EPFL	150,000

Project title	Abstract	Applicant	Institutes	Amount CHF
A FAIR metadata standard for observation data from urban drainage systems	Open Research Data (ORD) is at the heart of open science. Unfortunately, in the environmental sciences publicly available observations from research groups which can readily be used by others are lacking. The field of urban drainage research is a particularly relevant case, because i) due to the high infrastructure costs, the society can expect that better knowledge resulting from ORD in this field can lead to substantial cost savings in the future, and ii) observations come from very diverse sensors, such as weather radars or online submerged spectrometer s, for which detailed meta-data, e.g. on calibration protocols are lacking. This makes it difficult to interpret monitoring data and assess whether they are fit for a certain use case. To improve using ORD in the field of urban drainage, we suggest transferring existing information models, such as InfraML and WaterML2.0, to urban drainage observations. In addition, we will improve the interpretability and reusability by providing standardized dashboards of datasets, which can be used to quickly explore the provided information. Third, we will engage with the international urban drainage community by applying the concept on three different datasets from an emerging Europe an urban drainage community. For the ETH domain, this will create added value by harvesting synergies with on going activities at SLF and WSL, as well as other EXPLORE projects.	**Rieckermann Jörg** Leitao Joao Paulo	EAWAG	150,000
Open WASH data by building Open Science Competencies and Community	Poor data management practices hold back progress in the Water, Sanitation, and Hygiene (WASH) sector. WASH professionals are not taught the competencies needed to manage the vast quantities of data they collect. Specifically, the activities and strategies related to the storage, organization, description, and sharing of data and other materials (i.e. data analysis code) following FAIR data principles are underused. We will approach this gap by building the openwashdata community to support members in this process. We use and teach open-source computational tools, enabling community members to further competencies aligned with FAIR data principles. We expect to publish a large number (> 50) of previously unpublished data sets shared by community members who will contribute to the process throughout the program. The result will be a growing network of WASH professionals passionate about applying FAIR data principles to their work, which will benefit everyone in the entire sector.	**Tilley Elizabeth** Schöbitz Lars Bannert Matthias	ETH Zürich	149,990
An ecosystem for community driven scanning probe microscopy research and development.	Scientific instruments these days are often sophisticated, intelligent systems consisting of complex hardware, electronics, and software components. To innovate in the technique, one must get access to all levels; instrument, hardware, software. Unfortunately, most commercial instrument manufacturers keep their technology proprietary, and provide very little access to the hardware and software. Before one can improve things, one has to either hack into the commercial system, or completely re-engineer the instrument. In our research field (scanning probe microscopy) this has become painfully noticeable as a marked decrease in innovation. In this project we propose to bring together the community to share and standardise their hardware and software developments in the field of SPM. The goal is to reach an open development ecosystem in which new and existing research projects can be embedded without having to first reinvent the wheel. We plan to leverage our existing open hardware AFM projects and their early adopters to generate an OpenSPM ecosystem consisting of the Open Hardware tools (hardware and software), an extensive knowledge base and development resources, a world-wide user community, and a sourcing platform that helps people adopt the open hardware.	**Fantner Georg**	EPFL	150'000

Project title	Abstract	Applicant	Institutes	Amount CHF
Next-generation structural biology: An open atlas of in-situ protein structural dynamics	Knowledge of protein structure is key for studying mechanisms of protein function, and the Protein Data Bank (PDB), the main open repository of protein structures, has been of tremendous value since its inception 50 years ago. While PDB structures are mostly static, the functions of proteins almost always involve motion and conformational changes, and a complete picture of protein function requires an understanding of structural dynamics. The development of a structural proteomics tool called LiP-MS at ETHZ enabled probing the dynamic structural changes of proteins in situ in response to a variety of perturbations and on a proteome-wide scale. The approach found a broad range of applications in biology, biomedicine and drug development. We propose to set up the first open research data infrastructure for storage and dissemination of these unique dynamic and in situ protein structural data. The infrastructure will make LiP-MS data findable, accessible, and re-usable by the scientific community. Specifically, we will establish ORD guidelines for LiP-MS-based dynamic structural data, as a joint effort of data producers and data consumers (Aim 1), will develop a web-based, open-access repository of dynamic structural data according to FAIR principles (Aim 2) and will populate this repository with dynamic structural data for ~25,000 proteins from six organisms. We believe that this resource will be highly relevant for a large number of users from different communities.	**Picotti Paola** Beltrao Pedro	ETH Zürich	150,000
Interfacing Natural Language Processing (NLP) Tools with Open Access Publications	Scientific publications are among the most valuable achievements of mankind. Thanks to open research data (ORD) such as open access publications, the world's scientific output is increasingly accessible. But access to this massive trove of data is not enough—we must be able to assimilate the knowledge it contains and put it to good use. We need accessory technologies to harness knowledge for the world's benefit. Search engines and their keyword-driven interfaces are very limited in their ability to bring scientific knowledge to end users, i.e. researchers with specific questions and authors of scientific manuscripts. Tasks such as discovering, reviewing, summarizing, and generating discussions from the scientific literature are easier to accomplish as our Natural Language Processing (NLP) techniques advance. We propose to radically increase the value of the vast trove of open access scientific content by aggregating an OR database of more than 140 Million papers and by providing interfaces to state-of-the-art NLP tools. We intend to publish the code and content of this database, to provide access via an application programming interface (API), and to operate a web-based application that uses NLP algorithms to help authors of scientific manuscripts to assimilate the scientific literature. Our efforts will help scientists discover, characterize, and harness the store of knowledge in open access publications and streamline the scientific discovery and writing processes.	**Hahnloser Richard**	ETH Zürich	150,000

Project title	Abstract	Applicant	Institutes	Amount CHF
Instant and versatile data visualization during the current dark period of the life cycle of FAIR research.	This proposal tackles urgently needed changes to current data treatment exploring and prototyping processes to foster teamwork, effective communication and improve data visualization to the benefit of the entire atmospheric chemistry research community. Motivation comes from the ongoing move towards interdisciplinary collaborations and strong team focus as well as the fast-growing amount of data, while it remains a challenge to effectively explore and visualize raw data in the early stages of the FAIR research cycle. The latter is what we call the “dark period”, i.e. the data acquisition phase of a science project, where scientists, in particular early career researchers, are faced with myriads of non-standardized data formats and highly diverse, raw data output from scientific instruments. The objective of this proposal is to explore FAIR ORD procedures based on different FAIR services, openBIS (https://openbis.ch/) and Renku (https://renkulab.io/), that are actively developed by our service-partners within the ETH-domain. The goal is to prototype procedures for quick data visualization by any team member (junior, senior, scientist, technician) or any interested scientist to foster collaborative approaches in data analysis and discussion.	**Bartels-Rausch Thorsten** Imad El Haddad Julia Schmale (EPFL)	PSI	150,000
A Toolbox for Providing Open Data on Chemical and Material Uses	Open data on the uses are key for the safety assessment of existing chemicals and materials, and for the design of novel safer ones. However, the availability and accessibility of use data in the public domain is limited. For many chemicals and materials, use data have been reported, but they cannot be readily extracted and used by scientists, regulators and the industry on a large scale. This is because they are scattered over the public domain, and are often reported using different terminologies and in a non-machine-readable/processable format. Against this background, this project aims to enable and engage the scientific and regulatory communities to build up open FAIR data on the uses of chemicals and materials by developing and disseminating a toolbox of the missing tools. In particular, the toolbox will include a standard and templates for future reporting, as well as a set of novel cheminformatics and natural language processing-based workflows and tools for automated extraction and harmonization of existing reporting from different types of sources, to be used by researchers within the ETH-Domain, as well as external researchers including industrial scientists, regulators, and civil society organisations. The project will take a participatory approach, with stakeholders within and outside the ETH-Domain to be engaged in the testing, finalization and dissemination of the toolbox.	**Wang Zhanyun** Nowack Bernd	EMPA	150,000
COmmunity Needs of Open Research Data PractIces iN FibEr-Optic Sensing - Leading by Example	In recent years, seismological monitoring and observation has been revolutionized by the use of fiber optic sensing technologies. Distributed Acoustic Sensing (DAS) uses pulses of light and backscattered signatures to measure strains along fiber optic cables at an unprecedented spatial resolution. Within the Seismology and Wave Physics group at ETH, we have been pioneering DAS experiments around the world, including records of urban seismic activity, on glaciers, monitoring of avalanches, volcanic activity, and more. Given that DAS systems can produce hundreds of GBs of information in a single day, these projects have resulted in a correspondingly unprecedented increase in the amount of data that needs to be stored, processed, and shared. As more and more research institutions around the world explore this new technology, clear standards need to be established to further encourage interdisciplinary collaboration. Such standards include data and meta-data information according to FAIR Open Research Data (ORD) and Open Science (OS) principles. Whereas the scientific community around DAS is starting to tackle parts of the entire ORD and OS lifespan of geophysical experiments and data, examples and references of full solutions are not yet available. With our proposed ETH ORD project, we plan to establish ORD and OS practices throughout the full scientific lifespan of (geo-)physical projects - in close collaboration with the community, and leading by example.	**Bowden Daniel**	ETH Zürich	150,000

Project title	Abstract	Applicant	Institutes	Amount CHF
Cloud and web based platform for dissemination of computational solid mechanics	Academic knowledge is traditionally disseminated by academic journals. However, nowadays the production of scientific data in any given project exceeds by a vast amount what can be contained in a few journal pages. Reproducible scientific data and publications must be associated to boost scientific collaborations and discoveries. Open-science aims at publicly distributing the production of scientists. To be successful platforms simplifying workflows are required. Our project shall provide such a platform for the vast computational solid mechanics (CSM) community. It will allow to describe input, code and output of a simulation, therefore enabling storage on a repository. Facilities will be included for mainstream software and high-performance computing calculations. Results access and analysis will be web-based. When ready to be published, simply clicking a submission button to an open-access repository will be the only requirement. There are already existing open-source software for some of the bricks composing our proposal, but not for CSM needs. We propose to implement a distributed platform, based on well-established initiatives such as Meshio and Renku projects, yet focusing on such needs.	**Anciaux Guillaume** Molinari Jean-François Nicolas Richart	EPFL	131,777