

# Open Research Data Program of the ETH Board

proposal for an open metadata dataset

Lars Schöbitz

Global Health Engineering - ETH Zurich

May 12, 2025

# Measure 1: Calls for Field-Specific Actions

- 11 calls for proposals
- 96 funded projects (60 Contribute, 34 Explore, 2 Establish)
- X.X million CHF in funding

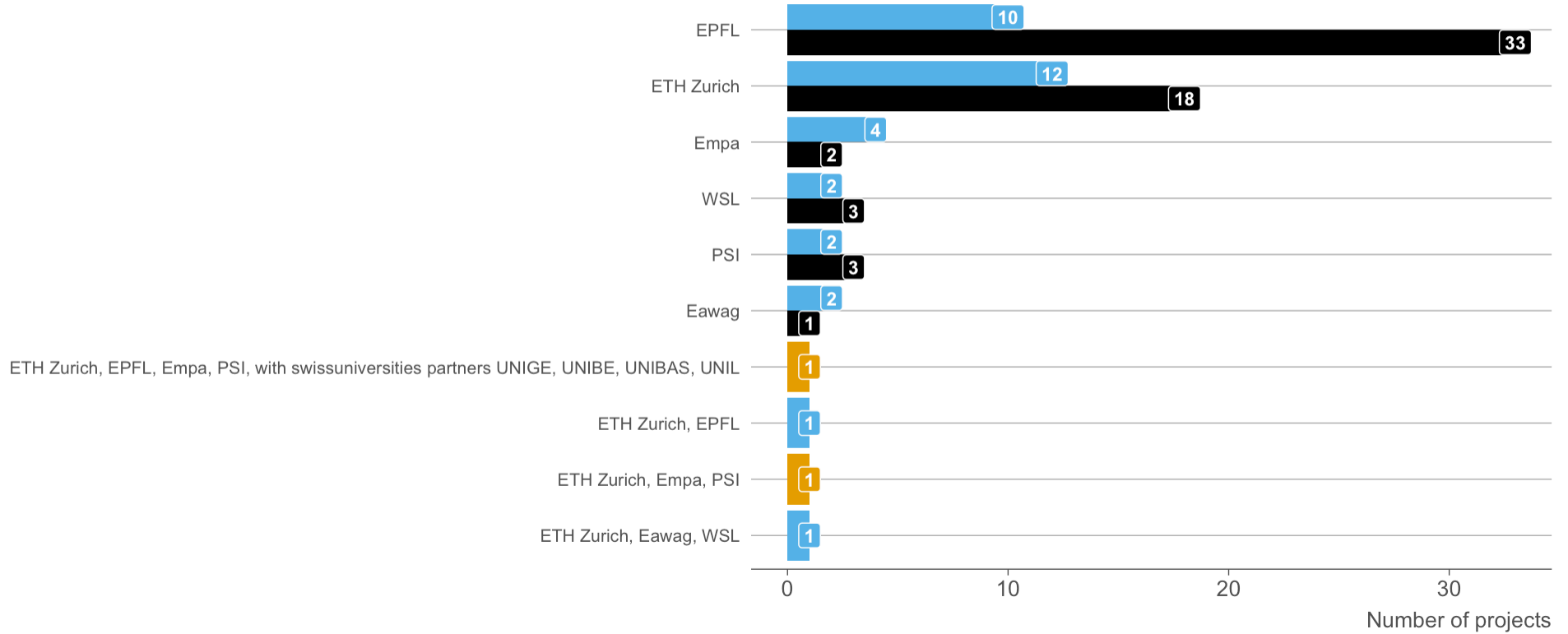
# Program evaluation

- Who are the researchers that received funding? 🍷
- Which institutions are they employed at? 🍷
- Which disciplines do they represent? 🍷 kind of

## Open Research Data Program of the ETH Board

Number of funded projects per institution and project category

Project category: ■ Contribute ■ Establish ■ Explore





# Program evaluation

- Who are the researchers that received funding? 👍
- Which institutions are they employed at? 👍
- Which disciplines/fields do they represent? 👍 kind of
- What is their scientific role (e.g. Professor, Post Doc, PhD, etc.)? 👤
- How many publications are derived from these projects? 👤
- How many ORD datasets have been published? 👤
- How where budgets distributed among their cost categories? 👤
- How many proposals were submitted to each call? And how many received funding? 👤
- etc.

# Problem statement

A metadata dataset that brings together all the information from the ETH ORD programme is currently not available.

# Proposed solution

Project relevant internal data sources (e.g. submitted proposals, final scientific reports, MS Excel files, Email exchange, online research) will be used to prepare a set of metadata datasets.



# Make it Open and FAIR

In the spirit of ORD, the metadata datasets will be shared as open research data, following FAIR principles for data sharing.

# SNFS Grants database

Not FAIR, but open:

`data.snf.ch/stories/snsf-open-data-who-gets-  
funding-en.html`

# **SNSF open data: Who gets funding? How to calculate figures yourself**

30.11.2022

The SNSF makes data on its research funding publicly available. Here's  
~~how you can use these data to calculate useful figures easily and quickly~~

# Datasets

At minimum, the following datasets are suggested, which can be combined by unique identifier (proposal\_id):

Dataset: proposals

- call\_id
- call\_category
- proposal\_id
- proposal\_url (if proposal is public somewhere)
- acronym
- title
- abstract
- funding\_requested
- funding\_received
- panel\_score\_excellence\_reviewer1
- panel\_score\_implementation\_reviewer1
- panel\_score\_impact\_reviewer1
- panel\_score\_excellence\_reviewer2
- panel\_score\_implementation\_reviewer2

# Dataset: **proposals** - Sources

- Project proposals (accepted & rejected)
- Internal MS Excel files for scores
- Email exchange
- Online research

**Note:** If access to rejected proposals is not possible, a fourth dataset **calls** is suggested.

Dataset: **projects**



- proposal\_id
- project\_start
- project\_end
- no\_cost\_extension
- project\_website
- budget\_cost\_category\_senior\_staff
- budget\_cost\_category\_postdocs
- budget\_cost\_category\_students
- budget\_cost\_category\_other
- budget\_cost\_category\_publications
- budget\_cost\_category\_conferences\_workshops
- budget\_cost\_category\_other
- budget\_sub\_contracting\_costs
- etc.

# Dataset: **projects** - Sources

- Project proposals
- Scientific reports

# Dataset: outputs

- proposal\_id
- category (e.g. website, publication, dataset, software, etc.)
- repository\_catalogue
- etc.

# Dataset: **outputs** - Sources

- Scientific reports, list of outputs
- Email exchange

# Dataset: **calls**

A summary table:

- call\_id
- call\_category
- outcome\_total (e.g. accepted, rejected)
- value

# Thanks!

Slides created via revealjs and Quarto:

<https://quarto.org/docs/presentations/revealjs/>

Data (Nicoló Massari) available at:

<https://raw.githubusercontent.com/Global-Health-Engineering/ethord/refs/heads/main/data/data.csv>

Access slides as PDF on GitHub