

National policy and support actions for research data skills: Impact & Experiences

University use cases & experiences on developing RDM & FAIR data skills

Dr. Aude Bax de Keating, Open Science Portfolio Manager @ swissuniversities

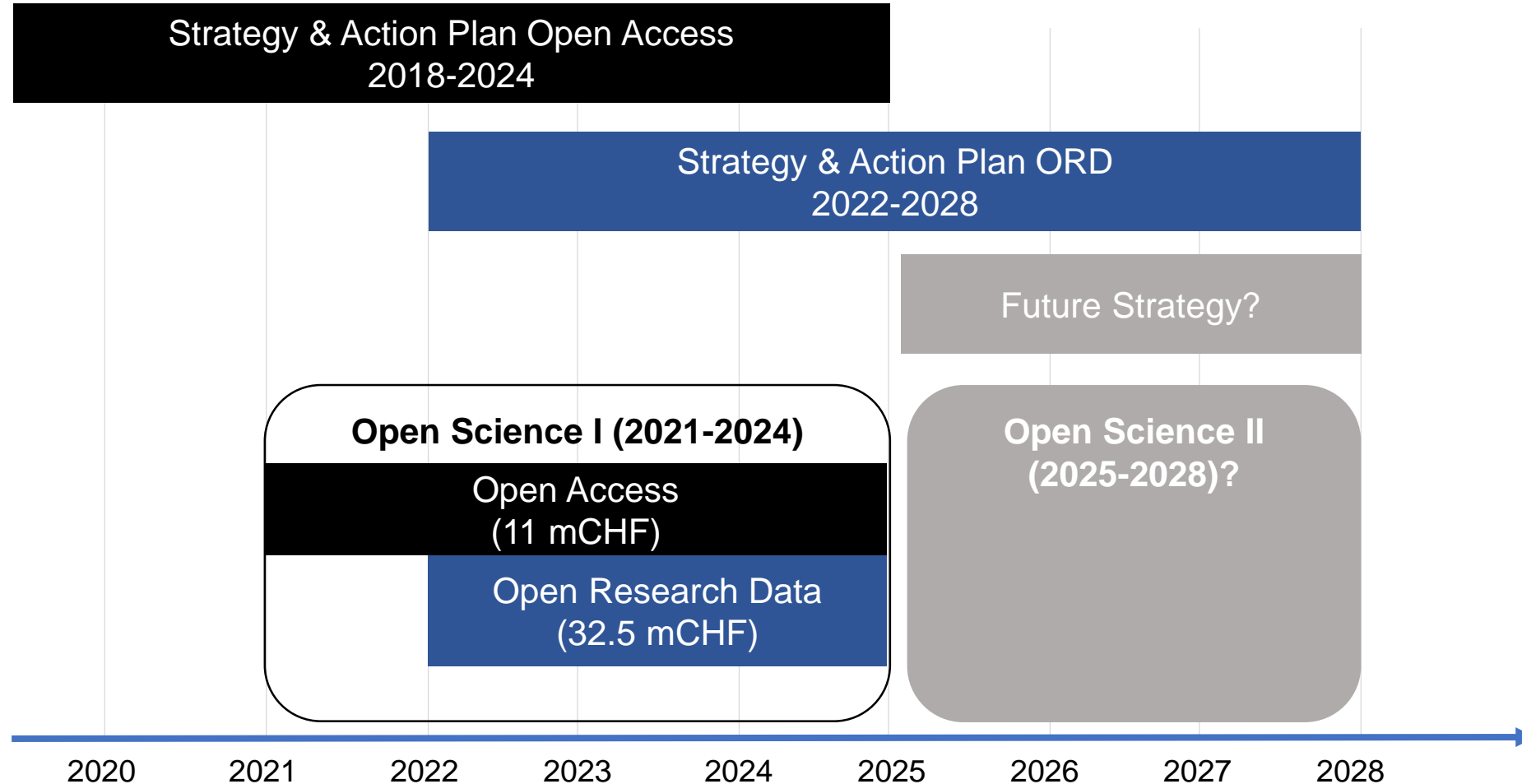
Wednesday April 28, 2021

Online Workshop

RDM & FAIR Data Skills in Switzerland & Beyond

From 3:00 pm to 3:15 pm	
3:00 – 3:02	The Scientific Information & Open Science Programs @ swissuniversities, Dr. Aude Bax de Keating
3:02 – 3:06	Tools & Services for FAIR RDM at ETH Zurich, Dr. Henry Lütcke
3:06 – 3:10	DLCM & OLOS, UNIGE, Dr. Pierre-Yves Burgi
3:10 – 3:15	Developing Research Data Skills in Computational Materials Science: AiiDA & Materials Cloud, EPFL Dr. Giovanni Pizzi

The Open Science Programme





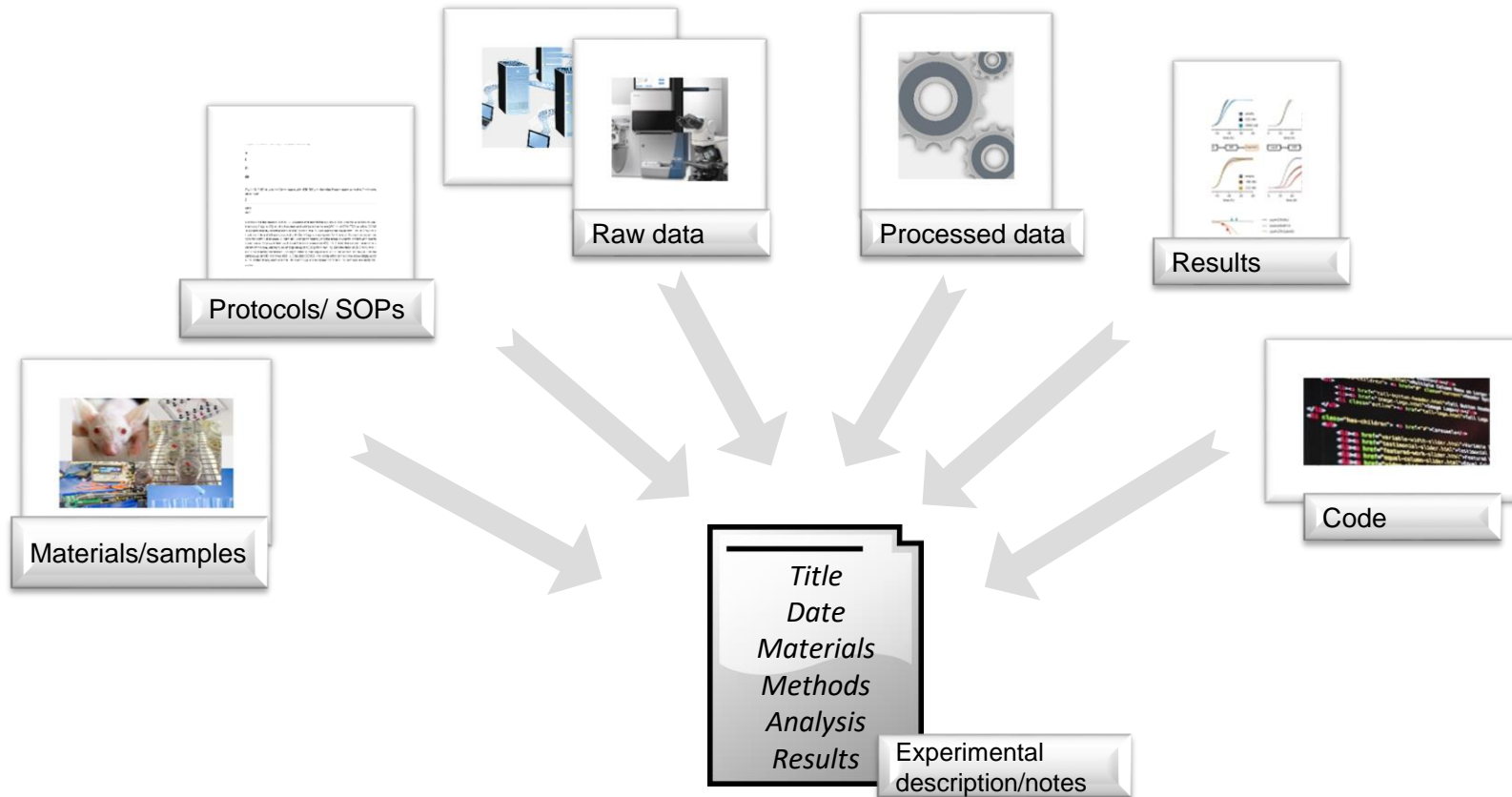
Tools & Services for FAIR RDM at ETH Zurich

28.04.2021

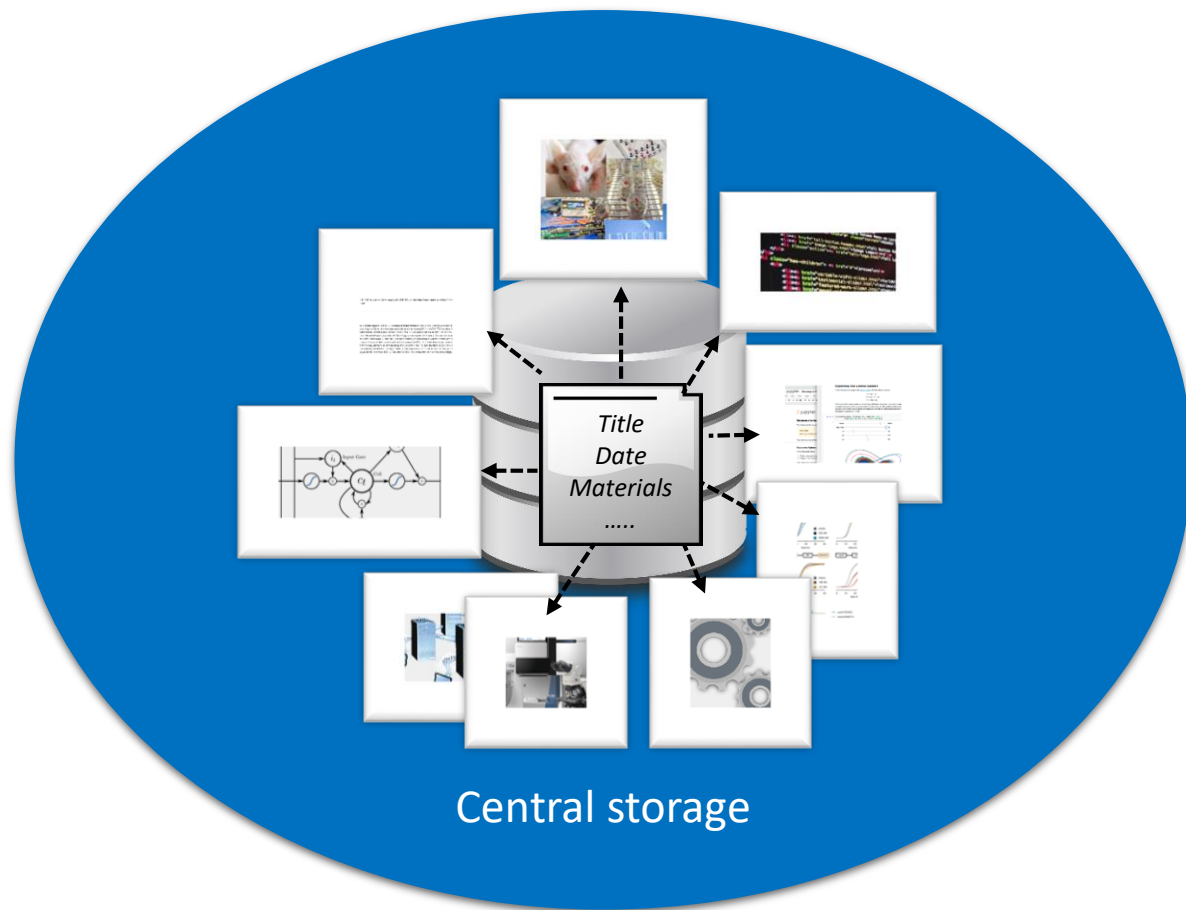
Dr. Henry Lütcke, Scientific Computing Lead
Scientific IT Services, ETH Zurich

FAIR Data Management

- Funders and journals increasingly demand data publication according to the **FAIR**¹ data principles (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable)
- Making data truly **FAIR involves** tracking and linking **different types of information**



openBIS – Data Management System and Electronic Lab Notebook

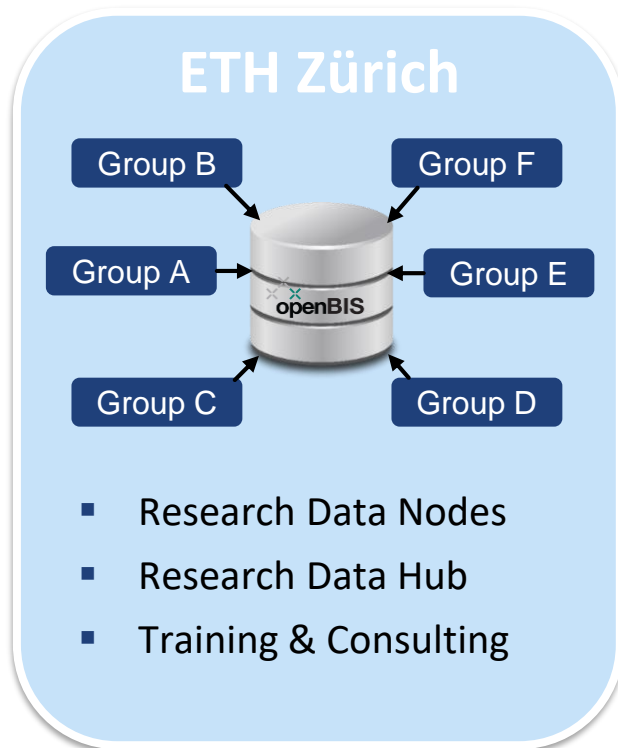


A combined ELN/LIMS can help make data FAIR



RDM Services based on openBIS

- openBIS **source code** is freely available: <https://sissource.ethz.ch/sispub/openbis>
- **Universities and non-profit** research institutes can download **binaries** for free
- ETH SIS offers **support services** based on openBIS
- May include **installation, maintenance**, customizations, data modelling or **training & support**



The national DLCM Project 2015-2020

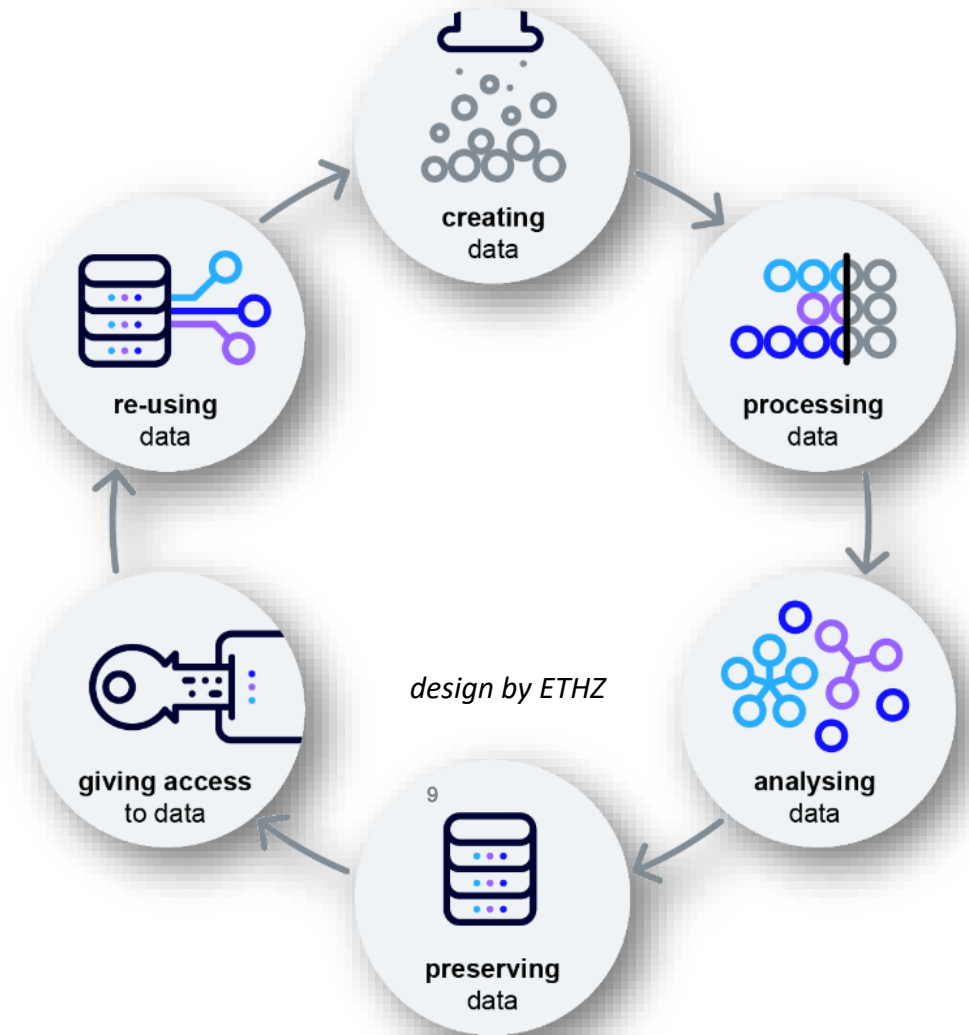
Aug 2015 – Jul 2018: Phase 1
Exploration of the whole life-cycle

Jan 2019 – Dec 2020: Phase 2
From Prototypes to Services

June 2019:  **YARETA** in production
Geneva state's solution

2020:  in production
National solution

2021: OLOS becomes an association



More information at www.dlcm.ch

Partners



EPFL



**UNIVERSITÉ
DE GENÈVE**

Hes·SO

University of Applied Sciences and Arts
Western Switzerland

h e g

Haute école de gestion
Genève



**Universität
Zürich** ^{UZH}

ETH zürich

SWITCH

Zürcher Hochschule
für Angewandte Wissenschaften


zhaw



DLCM

| Mandated by **swissuniversities**



- ✓ Developed as part of the DLCM project
- ✓ Repository for **Swiss researchers** 
- ✓ **Swiss-based** servers (original & copy)
- ✓ Compliant with **FAIR principles** & provides a **DOI**
- ✓ **Any format** across all scientific disciplines

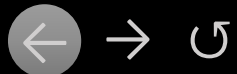
**Project supported by swissuniversities
&
Non-profit Association**



OLOS.swiss

13

20



http://olos.swiss



[Home](#)

| EN



Welcome to Olos

The consultation and archive portal of Switzerland

Search for an archive



Search for an archive

Search for publications in public or private archives



Create a Deposit

First step to archive your data



EASY AND USEFUL FEATURES

- ✓ Control access to your data
- ✓ Define your data retention period, licenses, etc.
- ✓ Choose your workflow, with or without validation
- ✓ Keep your folder structure



COMPLIANT WITH POLICIES

- ✓ Complies with Swiss law
- ✓ Follows FAIR principles
- ✓ Assigns a persistent identifier (DOI)
- ✓ Eligible for funding from the SNSF for data upload
- ✓ Complies with 12 open and international standards, protocols, and registries



SUITABLE TO YOUR NEEDS

- ✓ Try it yourself!
- ✓ Consulting & Training: Receive customized assistance for large datasets, etc.
- ✓ Customize the metadata schema to your field
- ✓ Integrate OLOS to your research environment



Annual program for
training session in
collaboration with
DevPro HESSO



DMP coaching



Tools, guides, practical
sources, documents
customized tools



On demand trainings
in different discipline



RDM Mooc & on demand e-
learning session



RDM Certification

Available by the end of
2021 !



Website <https://www.dlcm.ch/olos>

Contact us > olos.swiss/contact

Integrate with your lab equipment

- DLCM integration guide > <https://sandbox.dlcm.ch/administration/docs/DLCM-IntegrationGuide.html>

Sign up for our newsletter > olos.swiss/newsletter/register



Association OLOS, 2021

Projet DLCM, 2020

Bibliothèque de l'UNIGE, 2021

This document is under Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0): <http://creativecommons.org/licenses/by-sa/4.0/deed.en>



UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG



Hes·so
Haute Ecole Spécialisée
de Suisse occidentale



UNIVERSITÉ
DE GENÈVE

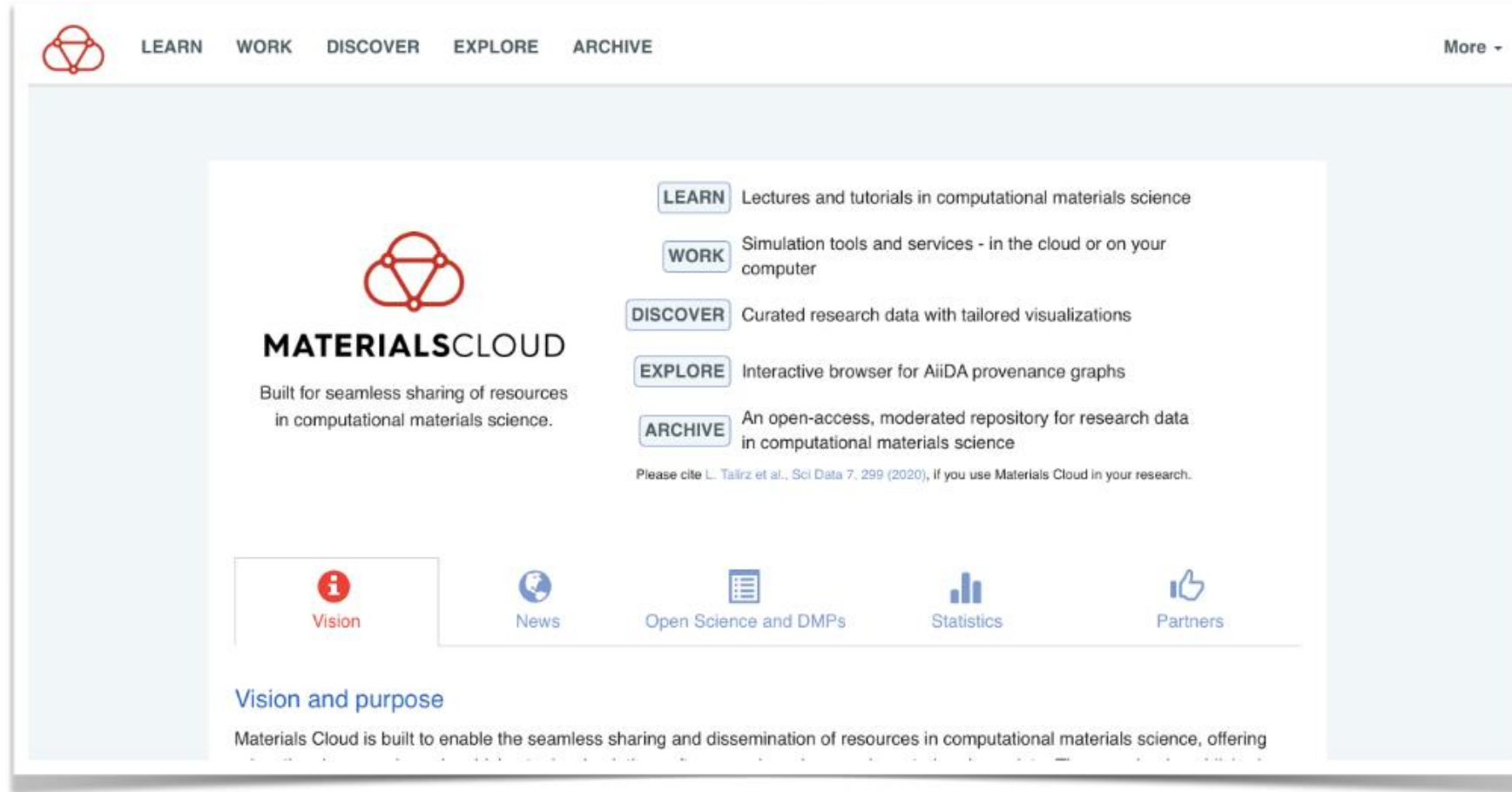
Developing Research Data Skills in Computational Materials Science: AiiDA and Materials Cloud

Giovanni Pizzi

Theory and Simulation of Materials (THEOS)
National Centre for Computational Design and Discovery
of Novel Materials (MARVEL)

EPFL, Switzerland

- A portal for research data dissemination and cloud simulations



L. Talirz et al., Sci Data 7, 299 (2020)

Materials Cloud Archive: a repository for data sharing



Latest records



Semi-local and hybrid functional DFT data for thermalised snapshots of polymorphs of benzene, succinic acid, and glycine

DOI [10.24435/materialscloud:vp-jt](https://doi.org/10.24435/materialscloud:vp-jt)

Edgar A. Engel, Venkat Kapil

Structure prediction for molecular crystals is a longstanding challenge, as often minuscule free energy differences between polymorphs are sensitively affected by the description of electronic structure, the statistical mechanics of the nuclei and the cell, and thermal expansion. The importance of these effects has been individually established, but rigorous free energy calculations, which simultaneously account for all terms, have not been computationally viable. Here we reproduce the experimental stabilities of polymorphs of prototypical compounds -- benzene, glycine, and succinic acid -- by computing rigorous first-principles Gibbs free energies, at a fraction of the cost of conventional methods ...

Latest version: v1
Publication date: Mar 26, 2021

Simulating solvation and acidity in complex mixtures with first-principles accuracy: the case of $\text{CH}_3\text{SO}_3\text{H}$ and H_2O_2 in phenol

DOI [10.24435/materialscloud:2x-7x](https://doi.org/10.24435/materialscloud:2x-7x)

Kevin Rossi, Veronika Juraskova, Raphael Wischert, Laurent Garel, Clemence Cominboeuf, Michele Ceriotti

**Recommended by Nature's
journal "Scientific Data"**

SCIENTIFIC DATA

<https://www.nature.com/sdata/policies/repositories#materials>

Indexed by [Google Dataset Search](#) and by
EUDAT/EOSC's B2FIND

Registered on [FAIRsharing.org](https://fairsharing.org) and re3data.org

- Releases DOIs for data associated to papers, open to the world
- Based on CERN's Invenio v3

**New! Recommended by the new
"Open Research Europe" journal**



Research and Innovation

Open Research Europe

<https://open-research-europe.ec.europa.eu/for-authors/data-guidelines>

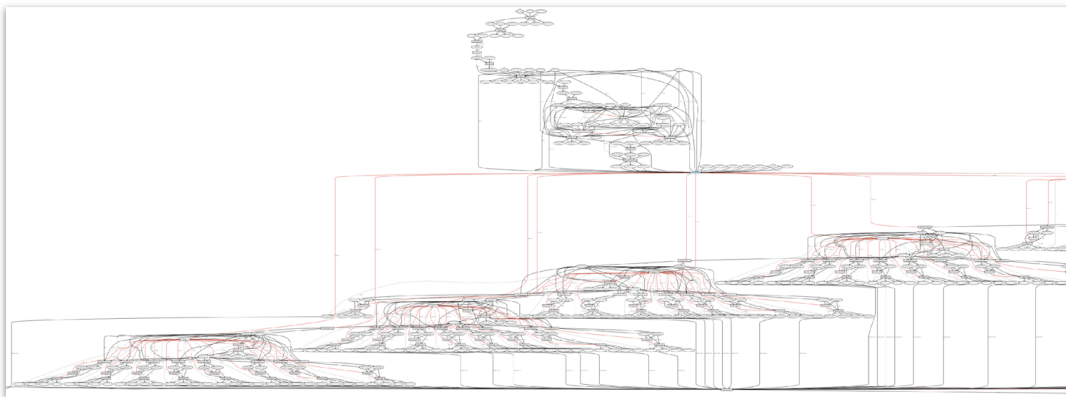
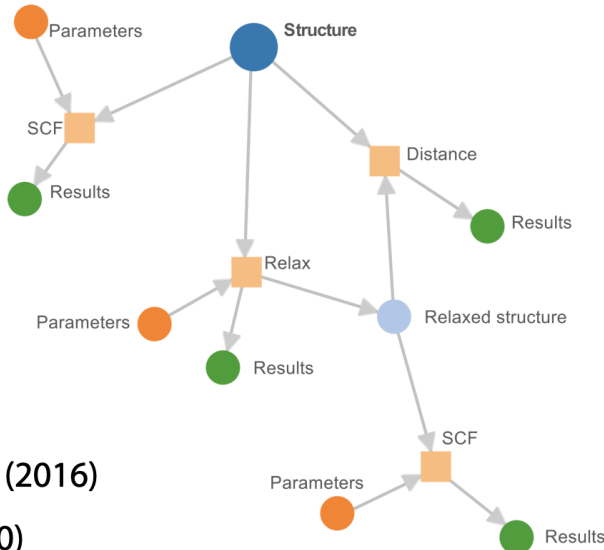
Providing tools for reproducible research




<http://www.aiida.net>

G. Pizzi et al., Comp. Mat. Sci. 111, 218-230 (2016)

S.P. Huber et al., Scientific Data 7, 300 (2020)



Molecular dynamics study of lithium in a solid electrolyte

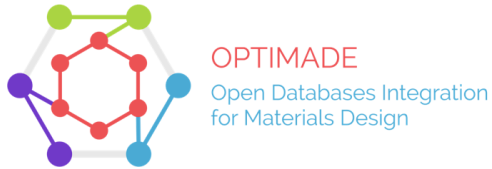
- Computational science infrastructure to automate high-throughput simulations, locally and on HPC resources
- Track automatically **provenance graphs**: full reproducibility
- Plugin interface
- Facilitate managing simulations with appropriate data management
- Work in progress to integrate with  openBIS for experiments



Registered plugin packages: 58

Calculations	98 plugins in 39 packages
Parsers	84 plugins in 40 packages
Data	79 plugins in 24 packages
Workflows	95 plugins in 27 packages
Console scripts	19 plugins in 14 packages
Other	95 plugins in 26 packages

Interoperability, Standards, APIs: OPTIMADE



<http://www.optimade.org>

C. W. Andersen et al., arXiv:2103.02068 (2021)

Standard REST API to search for crystal structures in 10+ different databases

Provider: Materials Cloud data available via OPTIMADE

Client: client available on Materials Cloud: get data from any compliant server

Client connected with cloud simulations:
quickly launch a calculation from a structure

Query

Materials Cloud
Three-dimensional crystals database

« < Showing 1-5 of 5 results > »

Materials Cloud
Materials Cloud: A platform for Open Science built for seamless sharing of resources in computational materials science

Three-dimensional crystals database
Curated set of relaxed three-dimensional crystal structures based on raw CIF data from the experimental databases MPDS, COD, and ICSD.

Apply filters

Basic Raw

Chemistry

Chemical Formula: e.g., (H2O)2 Na

Elements: H, O, Cl, ...

Number of Elements: e.g., 3 or >=5

Cell

Dimensions: 0: Molecule, 3: Bulk, (Not supported: 1: Wire, 2: Planar)

Number of Sites: >20

Provider specific

Provider ID: NBI Will take precedence

Q Search

Results

Ag16Ca6 (id=8286)

« < Showing 1-10 of 23 results > »

Select a format Download

Structure details Sites

Chemical formula (hill): Ag₃Ca

Elements: Ag, Ca

Number of sites: 22

Unit cell: $\begin{pmatrix} 0.00000 & 5.00023 & 5.00023 \\ 5.00023 & 0.00000 & 5.00023 \\ 5.0002291459902 & 5.0002291459902 & 0.0 \end{pmatrix}$

Unit cell volume: 250.03 Å³

<https://www.materialscloud.org/optimadeclient>



Availability on the EOSC Marketplace

Two of the services we provide are available on the EOSC Marketplace to all researchers

Cloud simulations

The screenshot shows the AiiDALab service page on the EOSC Marketplace. The header includes navigation links: Contact us, Portal Home, Catalogue & Marketplace, Providers Dashboard, and Login. The main content area features the AiiDALab logo, a description of its role in materials science, and a rating of 0.0/5 with 0 reviews. A blue button labeled 'Access the resource' is prominent. Below the description, there are links for 'Webpage', 'Helpdesk e-mail', 'Manual', and 'Training information'. A 'SCIENTIFIC CATEGORISATION' section lists 'Natural Sciences' and 'Chemical Sciences'. The footer contains the text: 'AiiDA is a workflow manager for computational science with a strong focus on provenance and performance. Through its flexible plugin infrastructure, it supports a wide range of simulation codes and makes them available for use through the Python programming language, and automatically records the full provenance of your simulation pipeline in a graph. AiiDALab lets you run and manage AiiDA-powered workflows through tailored web applications in the browser. Use the App store to pick and install apps from the application registry or write your own in just a few lines of python using jupyter widgets and appmode. This service is supported by the MaX European Centre of Excellence, the MARVEL National Centre for Competence in Research, the European H2020 MarketPlace project, by swissuniversities as well by a number of other partners.'

<https://marketplace.eosc-portal.eu/services/aiida-lab>

Repository and long-term storage

The screenshot shows the Materials Cloud Archive service page on the EOSC Marketplace. The header includes navigation links: Contact us, Portal Home, Catalogue & Marketplace, Providers Dashboard, and Login. The main content area features the Materials Cloud Archive logo, a description of its role as a long-term FAIR data repository, and a rating of 0.0/5 with 0 reviews. A blue button labeled 'Access the resource' is prominent. Below the description, there are links for 'Webpage', 'Helpdesk e-mail', and 'Manual'. A 'SCIENTIFIC CATEGORISATION' section lists 'Natural Sciences' and 'Physical Sciences'. The footer contains the text: 'The Materials Cloud Archive is an open-access, moderated repository for research data in computational materials science, that allows researchers worldwide to upload and publish their data free of charge. The repository guarantees long-term storage (for at least 10 years) of data records and associated metadata, their findability via persistent identifiers (including a DOI), and their accessibility via standard protocols. Submissions are moderated to ensure their completeness, adherence to the FAIR principles, and relevance to the field of materials science. This service is supported by a number of partners.'

<https://marketplace.eosc-portal.eu/services/materials-cloud-archive>


Simplifying researchers's life: Data Management Plan templates

Data Management Plan

In order to support researchers using the Materials Cloud to prepare grant applications or to comply with agencies' requirements, we provide here below templates for data management plans (DMPs) that use the Materials Cloud.

Dissemination can be as simple as depositing data freely (and in any format) on the Archive, or, if using AiiDA, disseminating entire workflows, raw and curated data in the Explore or Discover sections. AiiDA plugins and workflows can also be distributed through the AiiDA plugin repository, while simulation services can be exposed through AiiDALab, either in a virtual machine environment (the Quantum Mobile) or on the cloud (e.g. on the European Open Science Cloud).

Feel free to [contact us](#) with any questions regarding the use of the Materials Cloud Archive as part of your data management plan.

Funding Body	DMP template (using  AiiDA)	DMP template (no AiiDA)
SNF	.docx .odt .pdf	.docx .odt .pdf
H2020	.docx .odt .pdf	.docx .odt .pdf

Please also note the [resources provided by EPFL](#), including extensive DMP templates for many different project types (SNSF, ERC, H2020, ...).

Materials Cloud fulfils the requirements for data repositories

The Materials Cloud Archive fulfils the [SNF requirements for data repositories](#):

- **Non-commercial:** Yes, all backing institutions are non-profit
- **Usage of globally unique and persistent identifiers:** Yes, uses DOI as permanent identifier system

H2020 DMP Template

This is a data management plan (DMP) template for a H2020 project in computational materials science using the [Materials Cloud Archive](#) as a data repository and [AiiDA](#) to collect the full provenance of the calculations performed. If you have any feedback or questions feel free to contact us at info@materialscloud.org.

Further H2020 guidelines on DMPs can be found [here](#).

Please adapt the template to your project!

1. Data Summary

- What is the purpose of the data collection/generation and its relation to the objectives of the project?
- What types and formats of data will the project generate/collect?
- Will you re-use any existing data and how?
- What is the origin of the data?
- What is the expected size of the data?
- To whom might it be useful ('data utility')?

The *ab initio* calculations performed in this project are expected to generate large volumes of data, most of it is in the form of intermediate binary files, which can be straightforwardly regenerated from the original inputs.

In this project, calculations will be performed through the *Automated Interactive Infrastructure and Database for Computational Science* (AiiDA), a python framework for high throughput calculations and provenance tracking (www.aidata.net). AiiDA automatically stores all information required to reproduce the result of each calculation. On the output side, AiiDA is designed to strike a balance between the cost of storage and the cost of recomputing a piece of data. For example, by default, total energies, electronic band structures and log files are stored, while Kohn-Sham wave functions and charge densities are not.

Before submitting a calculation through AiiDA, all inputs (+ metadata, see 2.1) are automatically stored in a local database. AiiDA then transmits the necessary information to the target computer, which can be a remote supercomputer, a local cluster or the workstation of the researcher. AiiDA adds the calculation to the computer's job queue, monitors the status, and retrieves the results once the calculation is finished.

In this model, all data is generated on the target computer, while only the data intended for preservation is transferred back to the workstation of the researcher.

<https://www.materialscloud.org/dmp>

Training events (tutorials, workshops)

<https://www.aiida.net/events/>

Events

Upcoming events

- (Tentative) [Week of July 6th, 2021] Online AiiDA tutorial for all timezones

Past events

- [10 Feb 2021] AiiDA tutorial at the Federal University of ABC, São Paulo – Brazil, with ~20 participants.
- [2 Dec 2020] AiiDA tutorial at the BIG-MAP workshop with ~80 participants ([event site](#))
- [10 July 2020 – 4 pm CT] Materials Science Q&A @ SciPy 2020 ([event site](#))
- [9 July 2020 – 5 pm CT] AiiDA talk @ SciPy 2020 ([event site](#))
- [7-10 July 2020] AiiDA virtual tutorial week, sponsored by MARVEL, MaX, CECAM, swissuniversities and MarketPlace. Introductory presentations and hands-on workshops for ~80 participants ([event site](#), [tutorial material](#), [report](#))
- [Oxford (UK) 25-27 March 2020] Introduction to AiiDA, aiida-wannier90 plugin, and the workflows to compute automatically Wannier functions within the School "Wannier90 v3.0: new features and applications" with ~40 participants ([event site](#)) ([slides](#), [material](#) and [video recordings](#))
- [CINECA Bologna (IT) 17-21 February 2020] Hackathon on plugin and workflow development for AiiDA with ~25 participants ([event site](#))
- [ISSP University of Tokyo (JP) 19-20 December 2019] AiiDA tutorial ([tutorial site](#))
- [Fiesch (CH) 9-13 December 2019] Fourth annual AiiDA coding week with ~15 participants ([event site](#))
- [IIT Mandi (IN) 9-11 October 2019] Writing reproducible workflows for computational materials science using AiiDA ([tutorial site](#))
- [SINTEF Oslo (NO) 23-27 September 2019] VASP & AiiDA workshop ([tutorial site](#))
- [Ljubljana, 16-20 September 2019] AiiDA tutorial within the [summer school on Advanced Materials and Molecular Modelling with Quantum ESPRESSO \(QE-2019\)](#).
- [Xiamen University (CN) 6 September 2019] Introduction to AiiDA 1.0 (as part of a workshop on [computational chemistry and machine learning](#))
- [EPF Lausanne (CH) 21-24 May 2019] Writing reproducible workflows for computational materials science, using AiiDA 1.0 (with >50 participants)
- [EPF Lausanne (CH) 25-29 Mar 2019] AiiDA plugins migration workshop, to support

- Training on tools to facilitate research data generation
- Specific sessions on research data management
- Inclusive, duplicated to cover all timezones



Results from July 2020 virtual tutorial