Local and national Swiss initiatives towards FAIR data

Caterina Barillari

Scientific IT Services, ETH Zurich

NETTAB 2018, Genova
Overview

1. What does it take to have FAIR data?
2. ETH Zurich initiatives towards enabling FAIR data
3. Swiss-wide initiatives
Overview

1. What does it take to have FAIR data?

2. ETH Zurich initiatives towards enabling FAIR data

3. Swiss-wide initiatives
Research workflow in experimental & computational labs
The “Data Iceberg”

- Funding agencies and journals increasingly demand FAIR published data.
- Published data is only the “tip of the iceberg”.
- Published data can be FAIR only if all collected data that led to the publication is properly managed from the start.
What does it take to manage research data?

Complex process that requires tracking and linking different types of information.

- Raw data
- Model
- Processed data
- Results
- Protocols/SOPs
- Materials/samples
- Code
- Analysis notebooks

Title
Date
Materials
Methods
Analysis
Results
Experimental
description/notes
A common scenario @ ETHZ
What is needed

Education

Tools

Support in using tools
Overview

1. What does it take to have FAIR data?

2. ETH initiatives towards enabling FAIR data

3. Swiss-wide initiatives
Data life cycle support @ ETHZ

Active data management
Annotate, store, backup data while it is produced

Long term data management
Annotate, share/publish, preserve data at the end of a project / publication
Education services at ETH

- ARDM workshops
- Git workshops
- Code clinics
- Running workflows on clusters
- ...

- RDM workshops
- New RDM summer school in 2019

- Short RDM workshops (German)
- ETH Research Collection trainings
- Custom trainings
- Trainings in reference management
- ...

- Topics of common RDM workshop: intro to RDM, policies, DMPs, active RDM, data sharing, data publication.
- Workshops are aimed at PhD students, post docs, staff.
- Regular workshops run twice/year. From next year, RDM sessions (3 half-days in 3 weeks).
- Workshops on demand.
Tools and related services at ETH

- openBIS for ARDM
- DMP templates for openBIS
- Infrastructure for sensitive data
- Infrastructure for computation (HPC)
- Workflow analysis development

- DMP templates
- ETH Research Collection (publication platform for paper and data)
- ETH Data Archive (long term preservation)
openBIS in a nutshell

- openBIS is a DM solution for research groups.
- openBIS is open source and free of charge for academic institutions and non-for-profit organizations.
- @ETHZ:
  - Central instance (ETH RDH)
  - Private group instances

https://labnotebook.ch/
**FAIR data in openBIS**

<table>
<thead>
<tr>
<th>F</th>
<th>Unique identifier (PermID) for objects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard metadata fields provided. Can be customized by users.</td>
</tr>
<tr>
<td></td>
<td>Data uniquely attached to objects, which contain the metadata.</td>
</tr>
<tr>
<td></td>
<td>All items stored in the openBIS database are indexed and searchable.</td>
</tr>
<tr>
<td>A</td>
<td>Data and metadata accessible via https and FTP.</td>
</tr>
<tr>
<td></td>
<td>APIs</td>
</tr>
<tr>
<td></td>
<td>Authentication/authorization required.</td>
</tr>
<tr>
<td></td>
<td>Data can be deleted data and metadata is preserved. The history of deletions is stored in the database. If metadata is deleted, this is still retrievable from the database.</td>
</tr>
<tr>
<td>I</td>
<td>Semantic annotations of types.</td>
</tr>
<tr>
<td></td>
<td>JSON-RPC2 API: methods to retrieve and annotate data and data models.</td>
</tr>
<tr>
<td>R</td>
<td>Default metadata for life sciences available, further customizable.</td>
</tr>
<tr>
<td></td>
<td>Authentication needed. Audit trail.</td>
</tr>
</tbody>
</table>
The ETH Research Collection

- Data repository provided by the ETH Library
- Acknowledged by SNF (non-commercial, FAIR)
- 2nd most popular repository in Switzerland
- For publications and research datasets
- Web upload, DOI-reservation and registration, ORCID, export to OpenAire
- Long term preservation in the ETH Data Archive
- Possibility to control access (but metadata always visible)

https://www.research-collection.ethz.ch
### FAIR data in the Research Collection

<table>
<thead>
<tr>
<th>F</th>
<th>DOI for Open Access / Embargoed Open Access / ETH only items; Handles for Closed Access / Selected Users items.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Metadata schema based on DataCite.</td>
</tr>
<tr>
<td>I</td>
<td>All items indexed and searchable (including items not in Open Access).</td>
</tr>
<tr>
<td>R</td>
<td>Data and metadata accessible via https and via OAI-PMH interface.</td>
</tr>
<tr>
<td></td>
<td>REST interface for authorized users.</td>
</tr>
<tr>
<td></td>
<td>When data are deleted, metadata is always accessible (DOI points to landing page).</td>
</tr>
<tr>
<td></td>
<td>The following metadata standards are used via the OAI-interface: Dublin Core following the OpenAIRE Guidelines, Qualified Dublin Core or DataCite-Metadata Scheme</td>
</tr>
<tr>
<td></td>
<td>Versioning of items and referencing between interdependent items.</td>
</tr>
<tr>
<td></td>
<td>CC licences per item.</td>
</tr>
<tr>
<td></td>
<td>Metadata schema based on DataCite.</td>
</tr>
</tbody>
</table>
Long-term preservation – ETH Data Archive

- **Digital preservation** solution for ETH Zurich, operated by ETH Library
- Data storage **longer than 10 years**
- Records are described by **meaningful metadata**
- Open, well documented file formats
- Automatically archives **content from Research Collection** and also **heritage content** from ETH University Archives and ETH Library
- More information: https://www.library.ethz.ch/Digital-Curation
Support services at ETH

- openBIS support
- Code management support
- Data analysis support
- ARDM consulting

DMP consulting

- RDM consulting
- ETH Research Collection support
- ETH Data Archive support
- “Book a librarian” service
Data Management Planning

- **Document** describing the **data produced** in a research project and strategies for **data handling** during the projects and **after it ends**

- DMPs are increasingly **mandated by funding agencies** (SNF, H2020, NSF)

- Guidance for ETH researchers on completing SNSF DMPs: [https://documentation.library.ethz.ch/display/DD/Guidance+for+ETH+researchers+on+filling+out+SNSF+Data+Management+Plans](https://documentation.library.ethz.ch/display/DD/Guidance+for+ETH+researchers+on+filling+out+SNSF+Data+Management+Plans)

- Info events on SNSF DMPs by ETH Library and Scientific IT Services

- Templates for SNF DMPs

- Consulting service
Research data website

www.ethz.ch/researchdata
Overview

1. What does it take to have FAIR data?

2. ETH Zurich initiatives towards enabling FAIR data

3. Swiss-wide initiatives
The Data Life Cycle Management (DLCM) Project

August 2015 – July 2018

- Aim: provide sustainable and tangible solutions at a national level to implement research data life-cycle management
DLCM outcomes

- Trainings in different aspects of RDM given at the different partner institutions
- Consultancy in different aspects of RDM given at the different partner institutions
- DM checklist
- SNF DMP template
- DMP generic template
- DMP Canvas generator (SNF)
- DLCM policy tool
- ELN/LIMS solutions
- Swiss national data repository
DLCM follow ups

- August 2018 - July 2020

1. openRDM.swiss: ETHZ (UZH, ZHAW)
   - openBIS as cloud service to Swiss academic community
   - openBIS trainings
   - User support
   - Support with local openBIS installations

2. DLCM 2: University of Geneva, HSE-SO
   - Swiss national data repository
   - National training platform
Take home messages

- **RDM** is the basis for **FAIR** data
- Researchers need **education, tools, and support** to successfully manage their data
- RDM requires **investments** and **commitment from Universities:**
  - @ETHZ services are shared by **SIS** and the **ETH Library:**
    - DMP support
    - Trainings
    - Tools and support for active RDM
    - ETH Research Collection & ETH Data Archive for publication and long term preservation
- Importance of RDM requires **recognition from funding agencies:**
  - @swissuniversities has invested in national projects to provide national RDM services and tools to Swiss academics
Acknowledgements

Bernd Rinn
Henry Lütcke
Thomas Wüst
Matthew Baker
Cristian Scurtescu
Juan Fuentes
Antti Luomi
Franz-Josef Elmer
Yves Noirjean
Swen Vermuel
Diana Coman Schmid

Matthias Töwe
Ana Sesartic Petrus
Malin Ziehmer
Barbara Hirschmann

Contacts

caterina.barillari@id.ethz.ch

https://sis.id.ethz.ch/

https://www.library.ethz.ch/

https://twitter.com/ETH_SIS

https://twitter.com/ETHBibliothek