Data Management Planning

C2SM Data Management Information Event and Workshop
5th March 2018, ETH Zurich

Dr. Ana Sesartic, Digital Curation Office
Research Data Management

- **Organise your data** in a way
  - which not only *allows yourself to track back* what you have done some time ago
  - but also *facilitates sharing and publishing* relevant data
  - and ensures its *comprehensibility* over time

- **Data** means
  - text documents, image files, audio files, software
  - “A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing.”
Roles and Responsibilities according to ETH

- **Project Members:**
  - adhere to the principles of good scientific practice and the guidelines for Research Integrity at ETH.
  - All steps of treatment of primary data must be **documented** and results must be **reproducible**.

- **Project Manager:**
  - responsible for execution of a scientific project and **data management** (data collection, storage, data access, compliance with data protection requirements...).
  - Ensures that all research project participants are aware of the guidelines.
  - Determines together with the professor, which departed project members should retain access to the primary data or materials.

From: **Guidelines for Research Integrity**, [https://doi.org/10.3929/ethz-b-000179298](https://doi.org/10.3929/ethz-b-000179298)
What is a Data Management Plan (DMP)?

A brief plan written at the start of a project and updated during its course to define:

- **What data** will be collected or created?
- **How the data will be documented and described**?
- **Where** the data will be **stored**?
- **Who will be responsible** for data security and backup?
- **Which data will be shared and/or preserved**?
- **How the data will be shared** and with whom?

DMPs are e.g. demanded by:

**SNSF from October 2017 on**
[http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx](http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx)

**Horizon2020 EU funding programme**
SNSF and Data Management
SNSF policy on Open Research Data

- Goal of the SNSF: **Research data should be freely accessible** to everyone – for scientists as well as for the general public
  - **Minimum requirement:** make underlying data of publications accessible
- A data management plan (DMP) is just one of the tools to reach this goal
  - **The DMP does not force researchers to open up all their data**
  - but to **make and document informed decisions** what to share and keep
- Data should be **FAIR:** Findable – Accessible – Interoperable – Re-usable
  - These are requirements both to researchers (document data appropriately) and to data repositories and their functionalities
How to submit a DMP to SNSF

- A proposal can only be submitted if a DMP was created
- A DMP for SNSF must be created online in mySNF
- You cannot upload a DMP created outside of mySNF – except in lead agency process
- DMP should be updated during project. Final version at end of project, published in p3.snf.ch database
- Contents of DMP:
- Instructions and examples for ETH Zurich:
  - http://www.library.ethz.ch/en/Media/Files/DLCM-template-for-the-SNSF-Data-Management-Plan

Interested? Ask for a workshop or consultation!

https://www.mysnf.ch
Assessment of the DMP

- The DMP is assessed by SNSF staff for **plausibility** and **compliance** with its Open Research Data policy
- It is not sent to external reviewers
- Applicants can be assigned «tasks» in the funding decision for enhancing their DMP
- DMP Guidelines for researchers, FAQ and more from the SNSF
  [http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx#](http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx#)
- Please contact the SNSF via [ord@snf.ch](mailto:ord@snf.ch) for comments, suggestions, questions
What to do for other funders?

- **Data Management Checklist**
  - Supports you in the creation of a DMP or in discussing data management in general, even if you don’t need to do it to comply with funders
  - [https://documentation.library.ethz.ch/display/DD/Data+Management+Checklist](https://documentation.library.ethz.ch/display/DD/Data+Management+Checklist)

- **DMPOnline**
  - A tool by the UK Digital Curation Centre that helps you create *Horizon 2020* compliant data management plans, by answering a questionnaire
  - [https://dmponline.dcc.ac.uk](https://dmponline.dcc.ac.uk)

**Collection of DMP examples:**
A few data management tips in practice
File organisation tips

- Sort files by **topic**
- Keep **path names** short
- **Unique/comprehensible** file names
- **No special characters** (ASCII only)
- Dates according to ISO standard: **YYYY-MM-DD** (a.k.a. Big-Endian)

Further information:

http://www.data.cam.ac.uk/data-management-guide/organising-your-data
http://datalib.edina.ac.uk/mantra/organisingdata/
Version control

What?
- Management of changes to documents and computer programs

How?
- Naming convention with numbers (filename_v1, filename_v2, filename_FINAL, etc)
- Use versioning tools

Hosted at ETH Zurich:
- GitLab
  https://gitlab.ethz.ch
- SharePoint (For documents only!)

Cloud based – use with caution:
- Bitbucket
  https://bitbucket.org
- GitHub
  https://github.com
- Subversion
  https://subversion.apache.org

ID-SIS provides hands-on trainings on git for code management (info @ https://sis.id.ethz.ch/consulting)
Preferences for file formats

- **Open standards** (non proprietary)
  - If proprietary, convert or if not possible include data viewer
- **Well documented**
- **Widely used and supported** by many tools
- **Uncompressed** (or at least losslessly compressed)
- **Unencrypted**
- When in doubt, *keep original and create a copy* in an open or exchange format
- Don’t rely on *file extensions*
- Consider that **data might be used in different operating systems**
ETH Services
Services at ETH Library

- **Share and publish Research Output** according to SNSF guidelines for FAIR data: ETH Research Collection ([https://www.research-collection.ethz.ch](https://www.research-collection.ethz.ch))
  - Publications, **Research Data**
  - Web upload, **DOI-reservation** and registration, ORCID, Export to OpenAire
  - Long term preservation in **ETH Data Archive** ([http://www.library.ethz.ch/Digital-Curation](http://www.library.ethz.ch/Digital-Curation))
- **ORCID** ([http://www.library.ethz.ch/en/ORCID](http://www.library.ethz.ch/en/ORCID)) - add your ORCID ID to your nethz-account
IT services and ETH transfer

**IT Services**

- Storage provisioning (usually via your IT Support Group)
- Active research data management support, for data management during the course of research [www.sis.id.ethz.ch/researchdatamanagement](http://www.sis.id.ethz.ch/researchdatamanagement)
- openBIS Electronic Lab Notebook & Laboratory Information Management System [https://openbis-eln-lims.ethz.ch](https://openbis-eln-lims.ethz.ch)

**Versioning**

- Gitlab - [gitlab.ethz.ch](http://gitlab.ethz.ch) (hosted by IT services)
- SharePoint - [mysite.sp.ethz.ch](http://mysite.sp.ethz.ch) (only for documents, free up to 1 GB)

**ETH transfer** [https://www.ethz.ch/en/the-eth-zurich/organisation/staff-units/eth-transfer.html](https://www.ethz.ch/en/the-eth-zurich/organisation/staff-units/eth-transfer.html)

- Software disclosure workflow with ETH Data Archive
- Advice on Intellectual Property, Patents, Licensing of Software etc.
Trainings

- Training courses on information research, reference management, data management, scientific writing and open access by the ETH Library:
  
  http://www.library.ethz.ch/en/Services/Training-courses-guided-tours

- Comprehensive workshop on data management offered by ETH Library in collaboration with Scientific IT Services: see link above or ask for additional dates!

- SIS trainings and courses (e.g. openBIS, Python, bioinformatics, etc, etc):
  
  https://sis.id.ethz.ch/consulting/

- Courses offered by the ETH Information Center for Chemistry/Biology/Pharmacy:
  
  http://www.infozentrum.ethz.ch/en/whats-up/events/

- Further topics on demand
Thank you for your attention!  

Dr. Ana Sesartic  
ana.sesartic@library.ethz.ch  

Digital Curation Office  
ETH Library  
ETH Zurich  
http://www.library.ethz.ch/Digital-Curation