Università
della
Svizzera
italiana

Data Management Plan drafting for SNSF

Outline

1. Research Data Management and FAIR data

2. SNSF policy on Open Research Data

3. SNSF Data Management Plan

4. Concluding remarks

5. Calls for projects

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

Research Data Management and FAIR data

Research data and the Data Management Plan

"Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form"

- A research Data Management Plan (DMP) describes the organization, treatment and dissemination of research data during the whole data lifecycle
- The great majority on funding agencies now require a DMP in order to complete the funding request and release funds
- Like it or not, you would better get used to it!
- More in general, think of a DMP as a tool to build best-practices in Research Data Management (RDM)

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

RDM concerns the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results
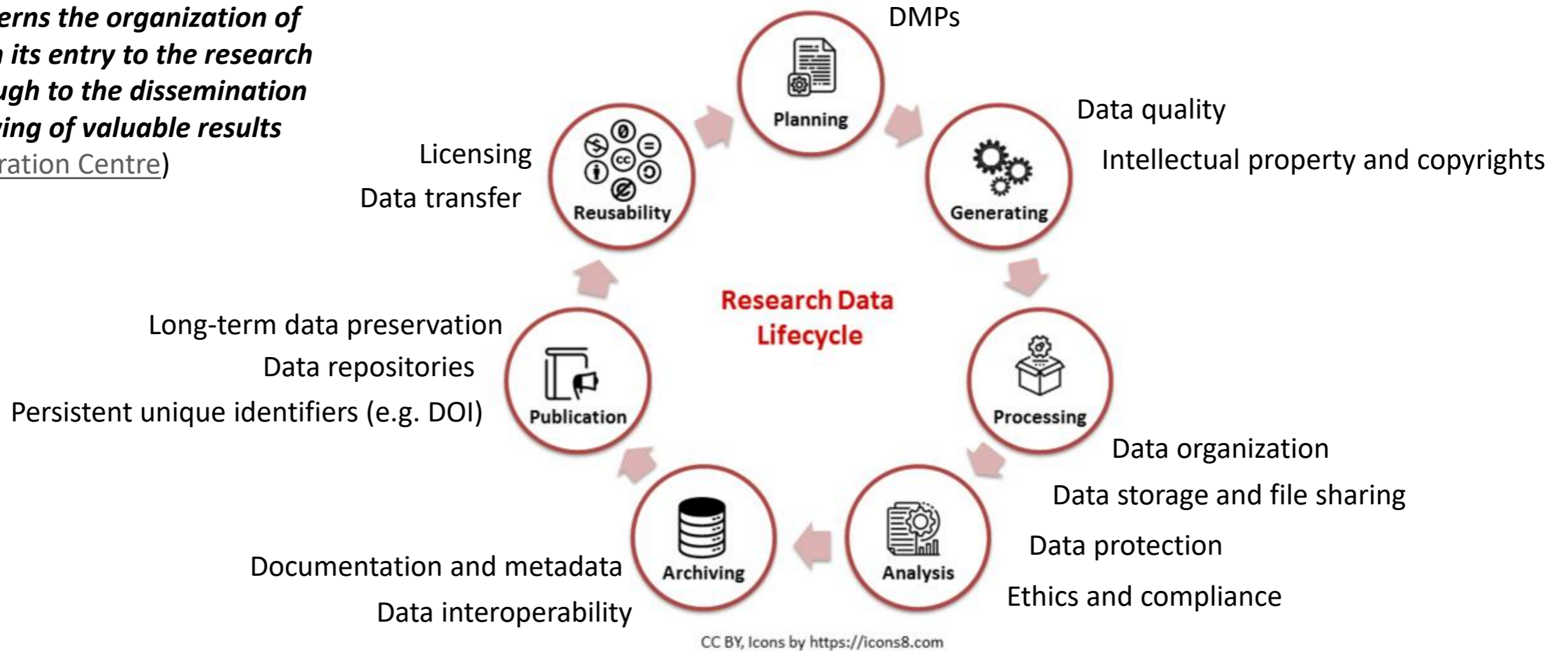([Digital Curation Centre](#))

DMPs

Data quality

Intellectual property and copyrights

Data organization

Data storage and file sharing

Data protection

Ethics and compliance

Documentation and metadata

Data interoperability

Long-term data preservation

Data repositories

Persistent unique identifiers (e.g. DOI)

Licensing

Data transfer

Research Data Lifecycle

Planning

Generating

Processing

Analysis

Archiving

Publication

Reusability

CC BY, Icons by https://icons8.com

Image source: FID-move

FAIR principles

"*There is an urgent need to improve the infrastructure supporting the reuse of scholarly data*"

- Guidelines to make data:
  - ✓ Findable: data and metadata should be easy to find for both humans and computers
  - ✓ Accessible: third-users need to know how data can be accessed. Accessible does not mean open!
  - ✓ Interoperable: data need to be easily integrable with new data and interacted with applications or workflows for analysis, storage, and processing
  - ✓ Reusable: data and metadata should be well-described so that they can be replicated and/or combined in different settings

Image source: Australian National Data Service [ANDS]

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

SNSF policy on Open Research Data

SNSF Open Research Data Policy

- The SNSF adopts an Open Research Data policy

  - Research data should be freely accessible to everyone (scientists and the general public)

- SNSF believes data should be carefully curated, stored and shared as openly as possible

- Therefore, funded researchers are expected to:

  - ✓ store the research data they have worked on and produced in their research work

  - ✓ share data with other researchers, unless they are bound by legal, ethical, copyright, confidentiality or other clauses

  - ✓ deposit data and metadata onto existing public repositories in formats that anyone can find, access and reuse

- SNSF grants public funds for research and expects that research artefacts become public!

SNSF Data Management plan

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

SNSF Data Management Plan guidelines

- SNSF requires a DMP when submitting a project for most of the funding schemes
  - The DMP is an integral part of the proposal submission

- The DMP <u>does not enter</u> (yet) in the scientific evaluation of the project and is not sent out for external review BUT submission of a plausible DMP is a requirement for any transfer of funding

- The DMP is a living document that is supposed to be amended as the project evolves

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

SNSF Data Management Plan content

1. **Data collection and documentation**

    1.1 What data will you collect, observe, generate or reuse?

    1.2 How will the data be collected, observed or generated?

    1.3 What documentation and metadata will you provide with the data?

2. **Ethics, legal and security issues**

    2.1 How will ethical issues be addressed and handled?

    2.2 How will data access and security be managed?

    2.3 How will you handle copyright and Intellectual Property Rights issues?

3. **Data storage and preservation**

    3.1 How will your data be stored and backed-up during the research?

    3.2 What is your data preservation plan?

4. **Data sharing and reuse**

    4.1 How and where will the data be shared?

    4.2 Are there any necessary limitations to protect sensitive data?

    4.3 I will choose digital repositories that are conform to the FAIR Data Principles.

    4.4 I will choose digital repositories maintained by a non-profit organization.

1. Data collection and documentation

1.1 What data will you collect, observe, generate or reuse?

- Create a list and identify each and every dataset or collection(s) of datasets

- Briefly describe the data you will collect, observe, generate and/or reuse:

    - content

    - role in the project

    - method of collection, observation or generation

    - potential evolution of data during the project

- If you reuse existing data (yours and/or third-party), specify clearly the original source

- Mention the general datatype (text, tabular data, images, movies, audio files, ...) the file formats you plan to use or in which data are available (.csv, .jpeg, .txt, …) and an estimate of the total volume of data (order of magnitude is perfectly fine in the first draft)

1. Data collection and documentation

1.1 What data will you collect, observe, generate or reuse?

- Example:

   "We will collect:
   - *primary data in interviews (audio and text files) and surveys (tables);*
   - *secondary data on the actors (professionals of accountability practices) and the units of analysis (city government websites) of our project.*

   We will also generate "reference datasets", which will be metadatasets.

   In total, we plan to collect, generate and store approximately 100 to 150 gigabyte of data."

1. Data collection and documentation

1.2 How will the data be collected, observed or generated?

- Describe how you plan to control and document the consistency and quality of the collected data (calibration processes, data recording standards, usage of controlled vocabularies, …)

  - Standards, methodologies and processes

  - Codes of good research practice

  - Quality control procedures

- Describe data organization and structure during the project (file naming conventions, version control, folder structures and hierarchy, …)

  - Consider the interaction of different project members

  - Consider data volume and data structure

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

1. Data collection and documentation

1.2 How will the data be collected, observed or generated?

- Example:

  "For the 1st phase of the project, we will collect and analyse academic and technical studies produced by third-party researchers and institutes or by us. All studies will be referenced in the final project report.

  For the 2nd phase, we will gather audio and (possibly) video files. Data will be collected with the consent of all interviewees. We will also generate field notes as .txt files.

  Data originated from each phase will be stored in dedicated folders, numbered in progressive order and sub-structured with specific folders for each data format. Successive versions of the same file will be named as the original file with a progressive number accounting for the number of the version."

1. Data collection and documentation

1.3 What documentation and metadata will you provide with the data?

- Metadata is data about data

  - They should always be available even if actual data are not openly available

  - Include common metadata elements!

  - If possible, adopt community standards and controlled vocabularies

- General metadata: basic details allowing other users to find, sort, index, access and propagate data (PID for each file, who and when, conditions to access data)

- Scientific metadata: information required for users to read, interpret, assess, replicate, build upon and reuse the data in the future

- Explain how you plan to generate the documentation and if any community standard and guidelines will be applied to annotate metadata

1. Data collection and documentation

1.3 What documentation and metadata will you provide with the data?

- Example:

  "Every image will be enclosed within a file including the name and a persistent identifier for each file, the name and ORCid of the person who collected or contributed to the data, the date of collection and the conditions to access the data.

  The collection of images will be accompanied by a README.txt file which will describe the directory hierarchy and the file naming scheme. Each directory will contain an INFO.txt file describing the experimental protocols used in that specific experiment. Produced images store the following metadata: field size, magnification, lens phase, zoom, gain, pinhole diameter.

  The researchers who will be in charge of each specific experiment will be responsible for checking that metadata are properly compiled according to the agreed standard, stored and accessible."

2. Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

- Outline all ethical issues in your project and the corresponding measures in data management
  - Personal data / sensitive data → ethical issues!
- Consider and describe:
  - the identifiers you plan to collect
  - necessary permission to obtain, process, preserve and share personal/sensitive data (ethics committee, informed consent form, non-disclosure agreement, …)
  - methods you will use to ensure the protection of personal/sensitive data (data anonymization, data pseudonymization, data encryption, …)
  - relevant protection standards for your data

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

2. Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

- Examples:

    "The research project does not present challenges related to ethical issues as we do not plan to collect, generate or reuse personal data."

    "Protocols used in this study will be approved by local Ethical Committee for Clinical Research. Only eligible subjects will be enrolled. Research subjects are entitled to withdraw from the study at any stage and to require a complete removal of their data from the study database upon request. Even though research subjects will be anonymised, consent of research subjects is mandatory in order to enrol them in the study and collect their data. No personal data or data that may easily identify subjects will be shared."

2. Ethics, legal and security issues

2.2 How will data access and security be managed?

- Very important in the case of personal/sensitive data

- Mention:

    - main concerns regarding data security and associated measures

    - data access rights and permissions/restrictions

    - measures for safe data storage and transfer of personal/sensitive data

- USI grants storage space and security measures to protect data

    - When dealing with personal/sensitive data avoid third-party cloud services, personal laptops and/or hard drives as much as possible

    - Data transfer can be performed via <u>Switch</u>

2. Ethics, legal and security issues

2.2 How will data access and security be managed?

- Standard for USI:

    "Project data will be stored on file servers managed by USI IT service in folders with limited and managed access permissions. USI servers can be accessed via secure VPN connection.

    Restricted access will be implemented, the principal investigator will be in charge of communicating to the USI IT team which researchers can be granted access to the data folders.

    When necessary, data will be transferred via SWITCH after being previously encrypted. Encryption keys will be securely stored and accessible only by project members. Partner research groups will adopt the same security protocols as USI according to the respective IT services."

2. Ethics, legal and security issues

2.3 How will you handle copyright and Intellectual Property Rights issues?

- Mention:

    - Data ownership (copyright, Intellectual Property Rights)

    - Specific permissions required should third-party data be re-used

    - Licenses, non-disclosure agreements, any other contract restricting usage and/or dissemination of the data

- If you think your data or source code might be commercially exploitable at the end of the project, mention it (<u>USI Technology Transfer</u>)

- If your project involves groups from other institutions, make sure that there is an agreement about a common policy on the dissemination of data and results

- Important: third-party ownership of rights does not preclude articles authorship

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

2. Ethics, legal and security issues

2.3 How will you handle copyright and Intellectual Property Rights issues?

- Examples:

  "Data collected within the different archives remain property of their respective owners. They will not be shared without the owners' permission. Data generated by the project (interviews and content analysis results) are property of USI."

  "Part of the data is obtained from data providers (Datastream, Orbis), and is subject to a nondisclosure agreement. Other data (such as, some macro series) are from central banks and other organizations (such as the World Bank) and are freely available. The research is not expected to lead to patents."

3. Data storage and preservation

3.1 How will your data be stored and backed-up during the research?

- Mention the needs in terms of data storage and the storage locations you plan to use
  - In general, data storage on laptops or hard drives is risky. Storage through IT teams is safer
  - USI grants project-specific storage space that can be shared with project collaborators
  - If you use your own or third-party resources, consider a procedure to regularly move data to institutional storage
- Specify back-up procedures
  - If you use your own or third-party resources, plan your own back-up routine or make sure that one exists

3. Data storage and preservation

3.1 How will your data be stored and backed-up during the research?

- Standard for USI:

    "The project data will be stored and processed on file servers managed by the IT service of USI, which are protected by backup copies of data made every night of the working days."

- Example:

    "The project data will be stored and processed on file servers managed by the IT service of USI, which are protected by backup copies of data made every night of the working days.
    Cloud services are not intended to be used. Researchers will be trained to transfer data from the recording devices to the file servers as soon as possible to avoid risks of data loss."

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

Università della Svizzera italiana

3. Data storage and preservation

3.2 What is your data preservation plan?

- Specify which data will be retained, shared and archived

  - Preserve all relevant data generated or used by research projects

- Explain the corresponding data selection procedure considering:

  - long-term value of data

  - potential value for reuse of data

  - obligations to destroy data and/or data sharing restrictions

- Outline a long-term preservation plan for the datasets beyond the lifetime of the project

  - Consider open/standard file formats (.txt, .csv, .jpeg, .flac, .tiff, …) and the use of community standards (if any) to grant (distant future) reusability

  - No formal rule, best practice is at least 5 years

  - Preserving data does not mean necessarily sharing it!

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

3. Data storage and preservation

3.2 What is your data preservation plan?

- Example:

  "At the end of the project, the entire database will be archived in a reusable format. Archives encompass all raw data, metadata, transformed data, transformation operations, deviations, version history, and audit trails. Redeployment of the entire database is therefore possible whenever needed, there is no legal obligation to destroy data.

  The comprehensive archive will be maintained during a minimum period of 10 years, after this term data will be re-evaluated in order to determine what to preserve.

  Data will be provided to authorised third parties as much as possible in non-proprietary formats (.txt, .csv, .xml, .pdf)."

4. Data sharing and reuse

4.1 How and where will the data be shared?

- Plan in advance which online repository to share data on

- If possible, privilege a discipline-specific data repository for your kind of data

    - Registry of Research data Repositories (re3data.org)

    - Otherwise, pick a generalist one (SNSF examples)

    - Consider elements such as PIDs, long-term preservation plan, licensing, metadata availability and format

- Consider how the reuse of your data will be valued and acknowledged by other users

- Depositing data grants more visibility to your research and eases the link to articles

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

4. Data sharing and reuse

4.1 How and where will the data be shared?

- Example:

    "Due to its humanistic nature, sharable data of this research will be preserved on repositories addressing a broader audience. Wikidata – even if it is a peculiar repository and not a typically academic one – complies with the FAIR Data Principles and it is noncommercial. The research content will also be published on an internal website managed by the university with texts under CC BY-SA.
    Copyright-covered data will eventually be preserved on DaSCH which grants archival resource keys (ARKs) as persistent identifiers and grants access control. Third-users data access will be subject to formal and motivated request."

4. Data sharing and reuse

4.2 Are there any necessary limitations to protect sensitive data?

- Define conditions to make data available (timing of data release, motivated embargo)

- Data have to be shared as soon as possible, but at the latest at the time of publication of the respective scientific output

- Making data accessible does not mean they have to be open!

    - Mention what data you plan not to share and why (legal, ethical, copyright, confidentiality or other clauses)

    - Consider if a non-disclosure agreement gives sufficient protection for confidential data

- As open as possible, as closed as necessary!

Università della Svizzera italiana

4. Data sharing and reuse

4.2 Are there any necessary limitations to protect sensitive data?

- Examples:

  "Legal restrictions to reuse of the data are based on the Federal Act on Research involving Human Beings (HRA), individual research subjects' data cannot legally nor ethically be made available to non-authorised people. Only the sponsor, the investigation team, reviewers, auditors and inspection authorities are entitled to access such data."

  "This historical research in the field of architecture does not present challenges related to sensitive data. Images under copyright will be available only by agreement with the rights holders. The discussion about how to release images will be addressed during the project."

4. Data sharing and reuse

4.3 I will choose digital repositories that are conform to the FAIR Data Principles.

- *Yes/no*

- FAIR-compliant repositories (<u>SNSF's explanation</u> of the FAIR Data Principles)

- If no FAIR repositories in your discipline, deposit a copy of data on a generalist platform

- If no data can be shared, this is a statement of principles

4.4 I will choose digital repositories maintained by a non-profit organization.

- *Yes/no*

- SNSF supports the use of non-commercial repositories for data sharing

- In case, explain why you cannot share your data on a non-commercial digital repository

- Checklist for non-commercial, FAIR-compliant repositories: <u>section 5 of the guidelines</u>

Concluding remarks

PROs and CONs of a Data Management Plan

- PROs for drafting a DMP (and sharing data):
  - ✓ Minimize the risk of incurring into certain issues
  - ✓ Use the DMP as a checklist to control the project schedule
  - ✓ Build and implement Research Data Management best practices
  - ✓ Save time for publishing, sharing and archiving data at the end of the project
  - ✓ Increase transparency and reproducibility of results
  - ✓ Generate research efficiencies, data reuse, new collaborations, projects and networks
  - ✓ Increase exposure of work and citations
- CONs for drafting a DMP:
  - ✓ It needs collaboration efforts and good will to draft a good-quality document
  - ✓ Time consuming and extra work (but this can payoff as the project progresses)
  - ✓ Needs knowledge on how to share data and make it interoperable
  - ✓ No concrete incentives yet

General considerations and recommendations

- DMP is an awareness tool, it makes you think and consider data taken in a broader context!

  - DMPs make researchers realize if and how they are aware about what they are doing

- Make an effort to visualize as many details as possible since the beginning of the project

- DMPs must not be overextended - stick to the questions!

- The SNSF DMP schema is pretty good and well structured, get used to it and it will be helpful also for other funding agencies

- Regard data as an important resource that must be carefully managed (just like funds)

- Regard DMPs as an investment and an opportunity, not as a mere requirement

- Write DMPs for yourself and your team, not for the SNSF!

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

Calls for projects

Open Research Data calls for projects

- USI promotes a call for projects on Open Science and Open Research Data

  - Projects with the objective to foster the participation of USI in the development of competences and resources

  - Projects size between CHF 10'000 and 50'000, duration between 3 and 18 months

  - Deadline: 30 June 2023 (USI webpage)

- swissuniversities has published some preliminary calls for project on ORD

  - Measure A1 - Provide bottom-up support to promote excellence in ORD practices

  - Next deadline: 2nd August 2023 (swissuniverisites webpage)

  - Further calls for project are expected in 2023, news will be shared via USI communication channels

Data Management Plan drafting for SNSF - Dr. Igor Sarman (Research Data Manager), 7th March 2023

Università della Svizzera italiana

Useful links

- SNSF resources:
  - SNSF Data Management Plan content
  - SNSF Open Research Data
- swissuniversities resources:
  - Swiss National Open Research Data Strategy and action plan
  - swissuniversities ORD calls
- FAIR principles (extensive illustration)
  - Knowledge clip
- Data repositories:
  - Registry of Research data Repositories (re3data)
  - SNSF examples of generalist data repositories
- Research Data Management at USI

Università
della
Svizzera
italiana

Contact

Igor Sarman, Research Data Manager

USI SRIT - Research and Transfer Service

igor.sarman@usi.ch

058 666 4790

USI Research Data Management webpage

Q&A

Thank you!