Data Management Plans

The why, what, and how!

Sven Trelle, CTU Bern

Sep 28, 2022 (CTU Lecture); with credits to OpenScience Team UB, Sarah Jones (DCC Glasgow)

The next 20-30 minutes

The menu

- Definitions
- Data management requirements by funding agencies
  - Swiss National Science Foundation
  - HORIZON Europe (2020)
  - US National Institutes of Health
- Principles and open research data
- The Data Management (and Quality) Plan (DM(Q)P) in clinical research (trials)

- HORIZON 2020 Online Manual
- Good Clinical Data Management Practices by the Society of Clinical Data Management

Definitions

To ensure we understand each other ☺

– **Research data**
Information (particularly facts or numbers) collected to be examined and considered, and to serve as a basis for reasoning, discussion or calculation. [numbers, text/strings, images, recordings etc. in digital form not all artifacts]

– **Data set**
Data needed to validate the results presented in scientific publications.

– **Open access**
Set of principles and a range of practices through which research outputs are distributed online and free of barriers/unrestricted (→ but see later)

HORIZON 2020 Manual; Wikipedia & UNESCO

Data management requirements

Funding agencies

- Data management has become more and more important in various funding programs (SNSF, EU and others e.g., NIH)
- Funders have realized how important proper management of research data is over the whole lifecycle (planning → conclusion) including risk management
- Empirical study on experiences (EU HORIZON 2020: 840 projects/DMPs; survey with 108 responders)
  - 45% unaware of Data Management Plans
  - 51%at Project Management Work Package, 22% dissemination WP, 18% dedicated WP
  - 53% consider it a useful tool

Swiss National Science Foundation

UNIVERSITÄT BERN

– Investigator Initiated Trials Call and project funding

After completion of the study, appropriately anonymised datasets must be made available for further analysis wherever possible.

Other requirements set by the SNSF concerning the accessibility of research data must be met during the submission of the application in mySNF. This holds in particular for the submission of a data management plan (DMP).

Horizon 2020

Open Research Data

UNIVERSITÄT BERN

- Since 2017: Research data open per default but opt out is possible
- DMP as a deliverable within the **first 6 months of the project**, updated regularly
- Deposit your data & associated metadata in a data repository of your choice

- DMPonline: input template with H2020 DMP
- Horizon Europe Template

National Institutes of Health (NIH)

Data Sharing & Data Management Plan

UNIVERSITÄT BERN

- **Before25 January 2023**
- Data sharing plan in research proposals seeking $500,000 or more
- Release and share the data
- Report any progress made on data sharing
- Genomic Data Sharing (GDS) policy may also apply to your research

After 25 January 2023

Submit DMS Plans and budget requests as part of the funding application or proposal.

Peer Review will not see or review DMS Plans, but will consider any related budget items.

NIH staff will review the DMS Plan for acceptability and may request modifications prior to award as appropriate.

Plans must be approved by the funding institute prior to award.

NIH Link

Principles and open research data

FAIR

– Introduced under the umbrella of Open Science →
  – Main objective to ensure availability and utility of projects research data
  – Measures taken to maximize access and re-use
  →Data sharing

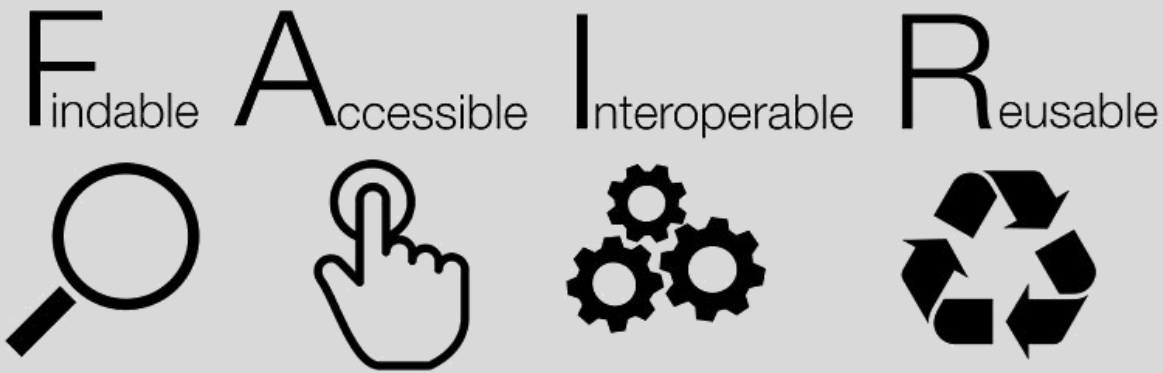


Image CC-BY-SA by SangyaPundir

FAIR guiding principles for scientific data

Wilkinson MD et al. 2016

- To be **Findable**
  - F1. (meta)data are assigned a globally unique and persistent identifier (e.g., Digital Object Identifier)
  - F2. data are described with rich metadata (defined by R1 below)
  - F3. metadata clearly and explicitly include the identifier of the data it describes
  - F4. (meta)data are registered or indexed in a searchable resource

FAIR guiding principles for scientific data

Wilkinson MD et al. 2016

- To be **Accessible**
  - A1. (meta)data are retrievable by their identifier using a standardized communications protocol
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
  - A2. metadata are accessible, even when the data are no longer available

FAIR guiding principles for scientific data

Wilkinson MD et al. 2016

UNIVERSITÄT BERN

- To be **Interoperable**
  - I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
  - I2. (meta)data use vocabularies that follow FAIR principles
  - I3. (meta)data include qualified references to other (meta)data

FAIR guiding principles for scientific data

Wilkinson MD et al. 2016

- To be **Reusable**
  - R1. meta(data) are richly described with a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible data usage license
  - R1.2. (meta)data are associated with detailed provenance
  - R1.3. (meta)data meet domain-relevant community standards

Data Management Plan

Basics

UNIVERSITÄT BERN

- Key element of good data management
- Describes (active) management of research data over the full life cycle of a project/study

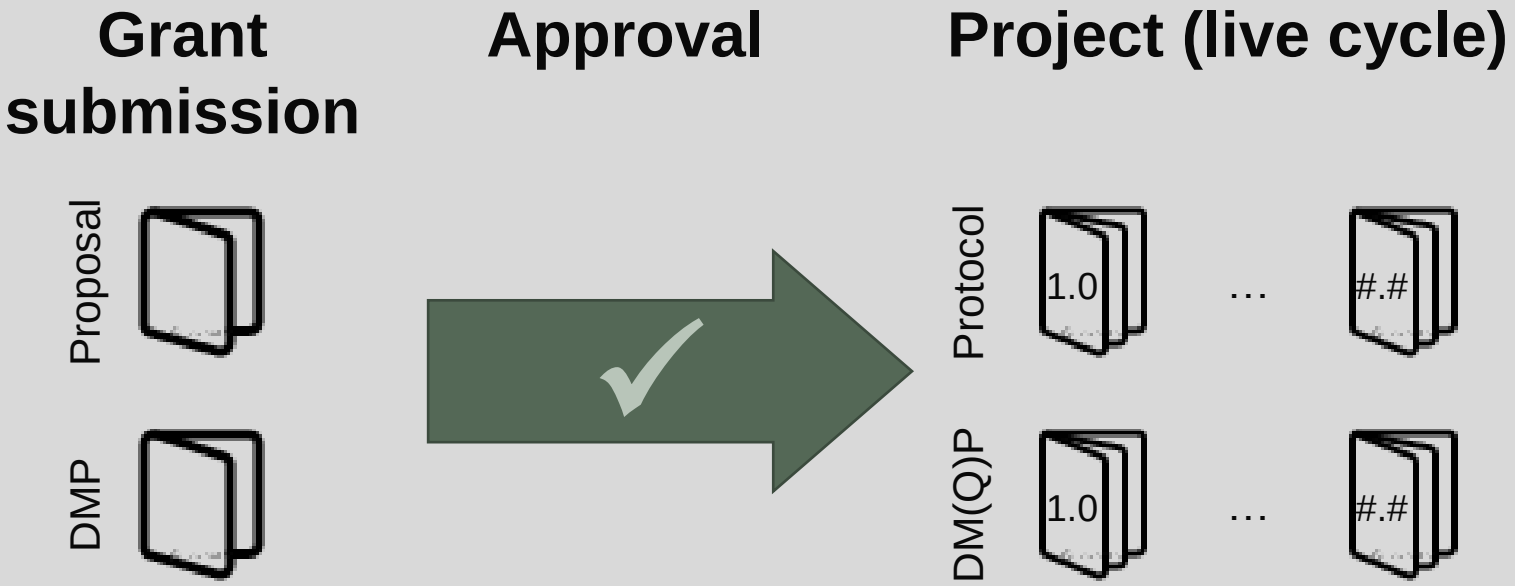
- A living but controlled document i.e. expected to change but formalized release process (versioning and signatures)

Data Management Plan

Overview

- A controlled document describing
  - the handling of research data during & after the end of the project
  - what data will be collected, processed and/or generated
  - which methodology & standards will be applied
  - whether data will be shared/made open access and
  - how data will be curated & preserved (including after the end of the project).
- For most funding bodies mainly a (too generic for clinical trials) description to ensure FAIR principles (Findable, Accessible, Interoperable, Reusable)

From grant proposal to project



Data Management Plan

Swiss National Science Foundation

UNIVERSITÄT BERN

- Data collection and documentation
  - What data will you collect, observe, generate, or re-use?
  - How will the data be collected, observed or generated?
  - What documentation and metadata will you provide with the data?
- Ethics, legal, and security issues
  - How will ethical issues be addressed and handled?
  - How will data access and security be managed?
  - How will you handle copyright and Intellectual Property Rights issues?
- Data storage and preservation
  - How will your data be stored and backed-up during the research?
  - What is your data preservation plan?
- Data sharing and reuse
  - How and where will the data be shared?
  - Are there any necessary limitations to protect sensitive data?
  - I will choose digital repositories that are conform to the FAIR Data Principles (confirmation).
  - I will choose digital repositories maintained by a non-profit organization (confirmation).

SNSF DMP

Example for standardized text

UNIVERSITÄT BERN

- If CTU Bern data management full service
- Example: 2.2 How will data access and security be managed?
  - *Data entered into the CRFs are transferred to the database using Secure Sockets Layer (SSL) encryption. Data will be stored in dedicated servers, which are located in dedicated, locked server rooms with restricted access. All servers are regularly backed-up using a multi-level system. Access to the eCRF and database is restricted to trained users and is logged. A role concept regulates permission for each user.*

SNSF DMP

Example for standardized text

- If CTU Bern data management full service and statistics
- Example: 1.1 What data will you collect, observe, generate or re- use?
  - *The data collected in this project is mainly determined by the pre-specified outcomes. It will be collected prospectively either from patient charts/clinical documentation/artifacts or directly from the patient. All collected data will have a dedicated source. A source data location log at each study site enables the exact location of all source documents. Data will be collected in a structured and coded manner. The format and coding of each variable will be defined in the codebook/data dictionary. For some analyses, variables will need to be generated from the collected data. This process is programmed in statistical software and therefore completely reproducible. In exceptional cases, freetext will be collected which will be coded in a second step. Data in this project is collected longitudinally. For each participant, baseline and follow-up data will be collected. Follow-up visits will be done at the following time points: add visit schedule/timepoints here. At baseline, data to characterize the participant population will be collected whereas outcome data will be collected at each follow-up visit.*

Data Management Plan

EU Horizon (2020 and Europe)

- Manual with guidance and template available
- Data summary
  - Types and formats of data
  - Origin of data
  - Expected size of data
  - …
- FAIR data
  - Making data findable including provision of metadata
  - Making data openly accessible
  - …
- Allocation of resources
- Data security
  - During project and afterwards
- Ethical aspects
- Other issues

(Clinical Trial) Data Management Plan

Details

- Roles & Responsibilities (related to data)
- Description of collected data
- Case Report Forms (development and description)
- Clinical Data Management System
  - Infrastructure (server, back-up, access control)
  - Study-specific implementation (Codebook, coding, imports, deployment, change management)
- Data collection and entry (including training)
- Quality control and assurance (monitoring, validation, queries, …)
- Database closure
- Data transfer and export
- Archiving
- Data preservation
- Data sharing (repository, shared artifacts, data request process including governance, legal aspects)

EU Horizon Europe

Data Guidelines

UNIVERSITÄT BERN

- Open Data policy
- FAIR Data Principles
- All datasets have been deidentified in accordance with the Safe Harbor method before submission
- Spreadsheet data
- Data should be deposited in a stable and recognised open repository under a CC BY or CC0 license prior to article submission
- Software source code – should be made available

Link

Data sharing

Clinical trials data

- Critically think about governance
  - As a researcher in an academic institution, you work as part of an organization (e.g., Uni Bern or Inselspital) and not as private person → data belongs to the institution
  - You represent the institution and act on its behalf (usually) for a project but what happens when you leave (time horizon: at least 10 years after project end)?
  - No open access to data but on request → after signing a Data Sharing Agreement
- Repository
  - BORIS Portal – Research Data
  - Re3data: use this registry of research data repositories to find a repository for your data
- Metadata
- What type of data sets: whole database or report-/publication-based

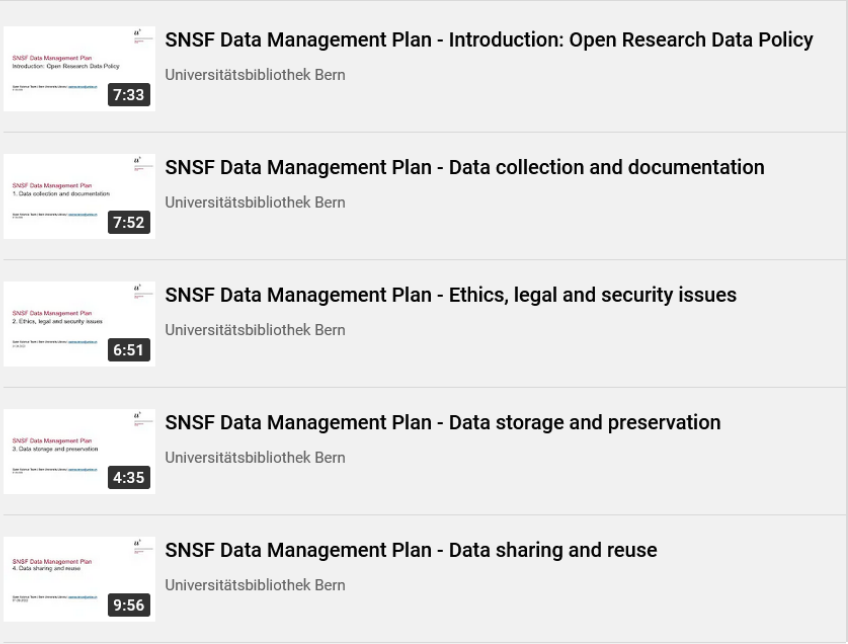
Metadata etc.

Repository

- Metadata
  - Clinicaltrial.gov number and all other identifiers (e.g. KEK number)
  - Protocol data elements (WHO/clinicaltrials.gov)
  - Use controlled vocabulary/ontologies
    - Medical Subject Headings
    - SNOMED
- Documents (latest versions!)
  - Protocol
  - Case Report Forms
  - Informed Consent Form
  - Codebook (& dummy dataset?)
  - Data Management (Quality) Plan
  - Statistical Analysis Plan
  - Request process including requirements/criteria, conditions, and Data Sharing Agreement

Open Science Team

Data Management Plan Support

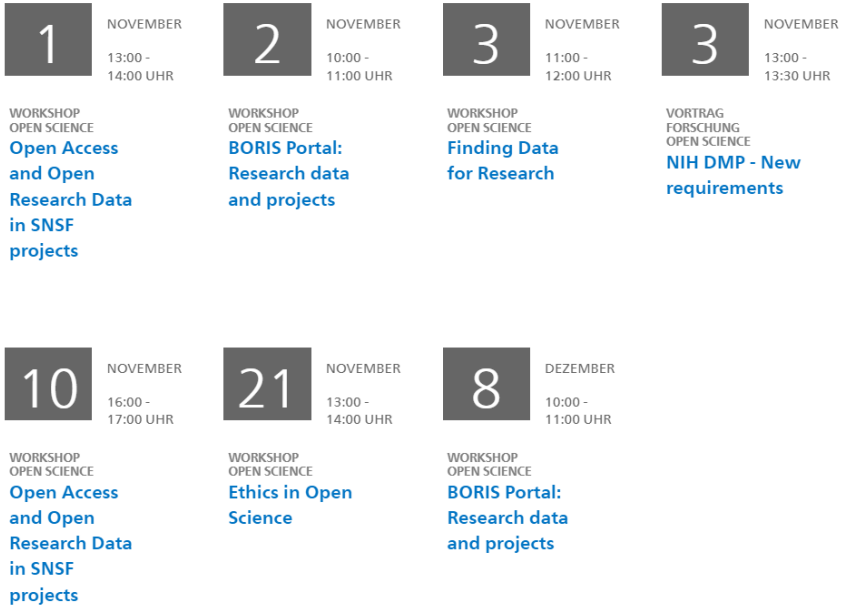


How to write a DMP? Watch video modules via YouTube

Data Management Plan review is free of charge and can be submitted online or via openscience@unibe.ch

Open Science Team

Courses & Workshops

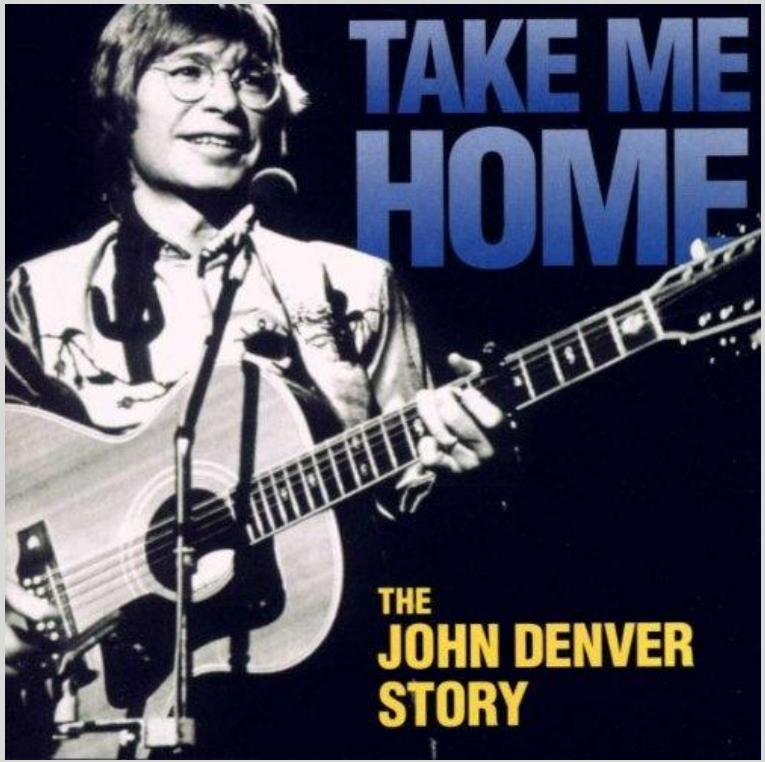


Workshops & courses in research data management: Link

@ openscience@unibe.ch

The take home messages

– DMP ≠ DMP
  → Funder: abbreviated, more high-level, ‘standardized’ description
  → Project: extensive and detailed description keeping independent user/reader 5-10 years after project end in mind
– If in doubt
  → Contact CTU Bern
  → Open Science Team (openscience@unibe.ch)



Thank you

for your attention!

Sven Trelle, CTU Bern

Olga Churakova, Open Science Team

UNIVERSITÄT BERN