Plan Overview

A Data Management Plan created using DMPonline

Title: SU-EOSC Nordic 5.3.2 maDMP project v38

Creator: Joakim Philipson

Affiliation: Stockholm University

Template: SU-VR template

ORCID iD: 0000-0001-5699-994X

Project abstract:

Continuation and development of SU-VR template, transformation file (*SUDMP2maDMP1-1.xsl*) and FAIRness evaluation schema (Schematron *SUDMP-FAIReval.sch*). From v36, substantial changes have been made to the transformation file (.xsl), now version 0.97 and the Schematron schema (v0.3) for FAIRness evaluation, following a change in conversion algorithm JSON to XML to JSON in Oxygen XML Editor 24.1, with addition of . The transformation file *SUDMP2maDMP1-1.xsl* v0.97 also uses a direct download of the DMP export.json, converted to xml, as a parameter document for the provision of some information elements lacking in the APIv0 output, notably start and end-date of project. To both .xsl and .sch files have been added validation measures to check consistency between identifiers and id-type.

Further, *SUDMP2maDMP1-1.xsl* v0.97 has changed from relative to absolute xpaths in the transformation, necessitated by a change in output order of keys/elements from APIv0, now seemingly random.

The SU-VR template itself has only minor changes with update of links in Guidance and cosmetic edition of question formulations. Restored conditional question *IV:Q2* (about sensitive data) with a 'no' answer removing questions *IV:Q3-4* (on data protection and ethical review).

ID: 100659

Start date: 01-10-2021

End date: 30-10-2022

Last modified: 23-05-2022

Grant number / URL: EOSC-Nordic-857652

SU-EOSC Nordic 5.3.2 maDMP project v38

0: Note on personal data!

Q1: I have read and understood the above declaration and hereby certify that this DMP contains no personal data except for information about project members such as PI and contact person.

- Yes

I: Description of data – reuse of existing data and/or production of new data

Q1: Data Quality Assurance measures (FAIR data): please check multiple options that will apply to assure quality and integrity of data collected, created or reused.

- 7. File names with only character set [0-9A-Za-z-_.], no spaces.
- 6. Descriptive file names (e.g. '20200910SanFranciscoDaylight1pm.mp4')
- 2. Sustainable file formats (e.g. .pdf; .csv; .txt)
- 1. Non-proprietary file formats (e.g. .csv, .txt, .json, netCDF)

Q2: Dataset ID: at this initial planning stage, please find one main identifier (e.g. a DOI, Handle, URL, ...) for the entire dataset(s) in the project where possible, even if it comprises several data files of different types.

https://doi.org/10.7910/DVN/MGZBAL

Q3: Dataset Identifier Type of your Dataset ID above in Q2, please check the corresponding option in the list below!

- doi

Q4: Dataset Description (Abstract) - please describe the dataset(s) in the project! The description can be at a rather simple conceptual level, which does not have to point to individual data files.

Stockholm University DMP template (pdf) in DMP Online, SU-VR template, is based on Swedish Research Council (VR) and Science Europe model (sections and original questions), but with more specified question and answer options by means of multiple choice checkboxes , dropdown menus, radio buttons for increased machine actionability of output. Includes XSLT for transformation of output (present) and Schematron schema for evaluation of prospective FAIRness of project RDM, as described by transformed API v0 output of DMPs. Also includes DMP Online instances DMP using SU-VR template with raw API (v0) JSON output, converted to XML, then transformed using XSLT-file and converted back to JSON to check compliance with maDMP-schema-1.1 (included as reused in dataset) and direct dmp export file (json).

Q5: Title of dataset

SU-EOSC Nordic 5.3.2 maDMP project

Q6: Are you re-using datasets that already have a definite distribution (identifier, access point or location, title ...)

- yes

RDA DMP Common Standard current schema madmp-schema-1.1.json

Q7: Access-point(s): url(s) that gives access to the re-used dataset(s) or resource(s). (This/these could be simple URLs or identifier-URLs such as a DOI, separated by commas)

https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/df3aada9359cca14ea4a02374512a5c165c818ff/examples/JSON/JSON-schema/1.1/maDMP-schema-1.1.json

Q8: Data access: please indicate the access conditions for the re-used dataset!

- open

Q9: License(s) of *re-used data* or *software* - please select the license(s) from the menu. (Multiple choice possible. If *Other*, please give a URI or other file location of the given license.)

- 12. Other

Unlicense: https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/LICENSE.md

Q10: Type of dataset(s) / resource type of the main dataset(s) of the project described by answers to Q2 / Q5 (thus, not primarily of re-used datasets).

- software

Q11: Issue date (YYYY-MM-DD) of dataset in Q2 / Q5.

2022-05-23

Q12: Format(s) of *re-used* datasets. Please give file name(s) of re-used data files with *file extensions / suffixes*, separated by commas

maDMP-schema-1.1.json

II: Documentation and data quality

Q1: How will metadata be created for your dataset? If by *use of a repository* (*recommended*), please specify which, either from the given options, or - if Other - by giving a link(s) / URL(s) [if multiple separated by commas] as Additional Information below. Please*, do not write whole texts here* with line or paragraph breaks*, as this prevents automatic processing and evaluation of the DMP!*

- 8. Other: https://...
- 6. README-file
- 1. Dataverse/StockholmUniversityLibrary

https://gitlab.com/JoakimPhilipson/SUmaDMProjectEOSC_Nord5-3-2

Q2: Which metadata standards and vocabularies will you employ for general and domain specific metadata?
(Multiple options possible. Some of them may overlap, then it is unnecessary to check all that hold a particular vocabulary specified in the text field.)

- 11. Other - please specify!
- 7. LOV - Linked Open Vocabularies
- 3. DDI
- 1. Metadata from II:Q1

- **FaBiO, the FRBR-aligned Bibliographic Ontology**: https://sparontologies.github.io/fabio/current/fabio.html
- **CASRAI Research Data Management Glossary**: https://casrai.org/rdm-glossary/

Q3: Which of the following data quality documentation and safeguard measures, if any, would you consider applying to your dataset?
(Multiple options possible. For options 4. *Pre-registration,* 6. *Supplementary documentation* or 9. *Other*, please specify to the extent possible in the comment area, e.g. by giving the URLs of particular services you intend to use for certain data quality measures.)

- 7. Validation of data input
- 3. Integrity check of data files
- 2. File format and software description

III: Storage and backup

Q1: Where will you store and backup your data during the project? (Multiple options possible.)

- 6. Repository in II:Q1
- 3. Own harddrive
- 1. SU Sunet Drive

Q2: What volume (X) of data will you need to store and backup?

- < 1 GB

Q3: What security measures will you need to employ to protect your data during the research process? (Multiple options can be selected).

- 1. Password protection

IV: Legal and ethical aspects

Q1: Will the creation, collection or reuse of dataset(s) in your project entail processing of *personal data*, i.e. any information relating to an identified or identifiable natural person

(a '*data subject*', that is a *living* person)?

- no

Q2: Will the creation, collection or reuse of dataset(s) in your project entail any of the following:

- **(a) processing of *special categories of personal data* according to the General Data Protection Regulation (EU 2016/679), i.e. personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation**
- **(b) processing of personal data regarding violations of law that include crimes, judgments in criminal cases, penal law sanctions, or administrative deprivation of liberty**
- **(c) physical interventions on research subjects or deceased persons**
- **(d) methods with the purpose of affecting a research person physically or mentally, or which includes an apparent risk of injuring the research subject either physically or mentally**
- **(e) studies of biological material that has been taken from a living or deceased person, and can be traced to that person**

Or, further, will the creation, collection or reuse of dataset(s) in your project include:

- **(F) Data from animal research**
- **(G) Data on genetic resources and/or traditional knowledge associated with genetic resources**
- **(H) Data that can be used for military purposes or concerning products that can be used for military purposes**
- **(I) Data that are sensitive in some other respect**

(Possible ethical review documentation applying to any of points (a-e) will be asked for separately.) If you answer yes to any of the points (F-I) and there is already relevant documentation or applications, please provide (a) reference(s) to any application(s)/ approval(s)/decision(s)/document(s), if possible by URL(s) / PID(s) such as DOI(s) giving direct access, or registration no. (Swe. *diarienr.*) in the text field below.

- no

Q5: Intellectual Property Rights - License(s) of data. Please select the usage license(s) for dataset(s) and/ or software *produced in your project*. (Multiple options possible. If *Other*, please specify by a URI or other file location for each of the additional license(s), separated by commas, in the comment area.)
[For *previously existing datasets* that you will be *re-using* , corresponding license(s) are entered in section *I:Q9 Description of data - re-use ...*]
If you choose the option '*No license*', being aware that it *might* make your dataset less FAIR, please state the reason for this choice. [There are *fully legitimate reasons*,

concerning e.g. personal data and sensitive data, that cannot be shared.]

- 09. MIT-license
- 06. CC0-1.0

V: Accessibility and long-term storage

Q1: Where will datasets, documentation and/or metadata be made accessible? (Means or location of *distribution*).

- 3. SU Archive
- 1. Repository in answer to II:Q1

Q2: What will be made directly accessible (e.g. via repository in Q1, or as supplement to online journal)?

- Software scripts
- Metadata AND all data files

Q3: When will data files and/ or metadata and documentation be made accessible?

- On completion of research project
- Continuously, as they are produced and updated

Q4: How will you ensure that all data files, documentation and metadata are transferred to SU digital archive for long-term preservation?

- Sunet Drive transfer
- Direct transfer to temp. archive (MADI) on request

Q5: Will specific systems, software, source code or other types of services be necessary in order to understand, partake of or re-use / analyse data in the long term?

- no

Q6: Will the software you will use to collect, create, handle, transform, refine or analyse data also be needed to replicate or rerun experiments, partake of your datasets or open datafiles?

- yes

Q7: Will the software / code you will use to collect, create, handle, transform, refine or analyse data be ... (multiple options possible)

- 1. Non-proprietary/Open Source (e.g. Python, R, XSLT)

Q8: Will you be using Software in the "cloud" / Software-as-a-Service (SaaS) to create, handle, transform, refine or analyse data ?

- No

VI: Responsibility and resources

Q1: Who is responsible for data management and (possibly) supports the work with this while the research project is in progress?

- 2. Myself
- 1. PI

Q2: What resources will be required for data management to ensure that data fulfil the FAIR principles? (Multiple options possible.)

- 1. RDM-team support
- 3. Repository account(s)

Q3: Please estimate total extra costs (C) for data management, that is not covered by

grant funding (or regular SU services, such as RDM-team support).

- < 10000 SEK

VII: Funding requirement fulfilled for initial version

Q1: I hereby certify that the prefect / responsible head of department or institute has (re)viewed this initial DMP as fulfilling the requirements for funding. I am aware that answering Yes will send this Initial version of the DMP to Archive for long-term preservation, and that future editing will then be in Phase 2, the final version.

- No

VIII: DMP administrative information

Q1: Please give an Identifier *of the Contact Person* designated in Project details (even if same as PI), - *not* the *name* of the contact person, but only the identifier-string (that is within the " " in the examples below).

0000-0001-5699-994X

Q2: Please select Type of Identifier given as answer to Q1 above.

- orcid

Q3: Affiliation (Department / Institution) of Contact Person. Please select main Department / Institution affiliation *from drop-down menu* (ordered after faculties as in this list), or else choose "Other" and specify in comment area below!

- Stockholm University Library

Q4: Language used for this DMP. Please select!

- eng: English

Q5: Funder(s). Multiple choice possible. If Other, please specify funder name(s) in the Additional Information text field, if more than one separated by commas.

- Other

Stockholm University, EOSC Nordic

Q6: Grant ID. Please specify, if possible as a URL. (The Grant ID can often be the same as the PROJECT-ID in SweCris, e.g. https://www.vr.se/swecris#/project/2010-00383_VR)

EOSC-Nordic-857652

Q7: Funding status. Please choose one from the dropdown menu.

- granted

IX: Full DMP - additional Datasets and identifiers, Reference list and Project end

Q1: Additional dataset(s)
Please fill in the *Default answer* table below in accordance with the given *Example answer* by replacing *None* in *Title, Identifier and Type* with *real values* for your dataset(s) after the *T1:, Id1: , Type1:* etc. You can add / delete rows if needed, but make sure the *new entries* are still in *italics* and leaving the last row without real values with *None* (as this will help us process your DMP data output for review.)
Identifier type: select from the same list as in section I-Q3: *ark, doi, handle, url, other.*
If *other*, please specify the type of dataset ID below the table as e.g. *"Local filename"* or *"Project-ID".*
The Description, Type of dataset (software, images, text, spreadsheets, sound, video, other) and Issue date for these additional datasets will as default be the same as for the main dataset described in section I: Q4, Q10 and Q11, so you might have to adjust these answers to fit for all datasets, or specify these new values for each additional dataset (below the table).

Short Title of dataset	Identifier of dataset	Identifier Type
T1: SU-EOSC Nordic 5.3.2 maDMP project v36	*Id1: https://dmponline.dcc.ac.uk/plans/94605*	*Type1:url*
T2: SU-EOSC Nordic 5.3.2 maDMP project v35	*Id2: https://dmponline.dcc.ac.uk/plans/90875*	*Type2: url*
T3: SU-EOSC Nordic 5.3.2 maDMP project v34	*Id3: https://dmponline.dcc.ac.uk/plans/85409*	*Type3: url*
T4: None	*Id4: None*	*Type4: None*
T5: None	*Id5: None*	*Type5: None*

Q2: List of References / Sources / Publications (other than reused datasets, in sec. I:Q6-9).
To be updated during all research project.

References: authors (year): *titles*	**Identifiers / Locations**
Hasan, Adil et al. (2021): *D5.5 Actionable Modular Data Management for brokering distributed resources.* (Working paper)	10.5281/ZENODO.5713787
Kim, S. (2020): *Machine-actionable Data Management Plans Model Analysis and Improvement Direction.*	https://doi.org/10.1633/JISTAP.2020.8.4.2
Tomasz Miksa, Simon Oblasser, Andreas Rauber (2021): *Automating Research Data Management Using Machine-Actionable Data Management Plans*. (ACM Trans. Manage. Inform. Syst. 13, 2, Article 18)	https://doi.org/10.1145/3490396

Q3: Research project ended?
Please indicate if the research project described by this DMP is completed, so the full and final version of this DMP can be sent to long-term archive.

- No