Facilitating FAIRness with a Stockholm University machine-actionable DMP Online template

Joakim Philipson - Research Data Analyst
(Stockholm University, RDM team)

Focus on Open Science 2021-06-18
National requirement

- As from 2019, all grantees from Swedish Research Council must have a data management plan if the research generates research data.

- Swedish Research Council (VR) & Association of Swedish Higher Education Institutions (SUHF) collaborated on partially reworked version of Science Europe’s Core Requirements for Data Management Plans – now a template in the DMP Online tool – but only free text fields

- Stockholm University offers DMP Online tool as default to all our researchers at dmp.su.se via SSO local login

- Other DMP tools: ARGOS, DSW, DMPTool, easyDMP, RDMO - ... all more or less converging to RDA Common Standard to produce more machine-actionable, maDMPs
Multiple purpose of SU maDMP template

1. *Easing the administrative burden on researchers* to fill in the DMP by means of preset multiple-choice answer options, drop-down menus, radio buttons etc., including extensive local *Guidance*

2. *Facilitating* semi-automated *review and evaluation of FAIRness*, while retrieving needed information for RDM administration

3. *Conforming to the RDA DMP Common Standard* for machine-actionable DMPs by validation against *maDMP-schema-1.1*
Features of SU maDMP template

- *Preserves structure* (sections I-VI) and *original questions from Science Europe – Swedish Research Council (VR)* in headings, but with more specific questions and answers (checkboxes, drop-down menus, radio buttons: yes/no/unknown)

- Using DMP Online (Roadmap) API (v0) to get output (as JSON) of entire DMPs (questions and answers) for further processing

- **Workflow:**
  1. APIv0 curl GET https://dmp.su.se/plans/72835 ->
     ```
     $ curl -G -k "https://dmponline.dcc.ac.uk/api/v0/plans?plan=72835" -H "Authorization:Token token=XYZ123456789sTuV012345" > DMP72835.json
     ```

  2. Oxygen: *dmp72385.json* to *dmp72385.xml* conversion ->

  3. *dmp72385.xml* to *dmp72385transform.xml* with XSLT
     [allowing FAIR-evaluation] ->

  4. *dmp72385transform.xml* to *dmp72385transform.json* ->

  5. *dmp72385transform.json* validated with RDA maDMP
Test example DMP of SU-VR template

Testing SU maDMP model for Climate Data: Sharing the seas.

Project Details

<table>
<thead>
<tr>
<th>Project title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing SU maDMP model for Climate Data: Sharing the seas.</td>
</tr>
</tbody>
</table>

mock project for testing, practice, or educational purposes

Project abstract

Using this published dataset as basis example:
Crone, Beatrice; Wassénius, Emmy; Lillepold, Kate, 2020,
"Dataset: Sharing the seas: A review and analysis of ocean sector interactions in the Anthropocene",
https://doi.org/10.7910/DVN/S6GTUS, Harvard Dataverse, V2

Project Start 09/01/2020

Project End 06/30/2021

ID SU-DMP72835

Select Guidance
There is no additional guidance for this template.
Testing SU maDMP model for Climate Data: Sharing the seas.

<table>
<thead>
<tr>
<th>Project Details</th>
<th>Contributors</th>
<th>Plan overview</th>
<th>Initial DMP</th>
<th>Full DMP (undeveloped)</th>
<th>Share</th>
<th>Download</th>
</tr>
</thead>
</table>

expand all | collapse all

0: Note on personal data! (1 / 1)

I: Description of data – reuse of existing data and/or production of new data (7 / 7)

II: Documentation and data quality (3 / 3)

III: Storage and backup (3 / 3)

IV: Legal and ethical aspects (3 / 3)

V: Accessibility and long-term storage (8 / 8)

VI: Responsibility and resources (3 / 3)

VII: Funding requirement fulfilled for initial version (1 / 1)

VIII: DMP administrative information (4 / 4)
Demo DMP example of SU-VR template: section 0

Testing SU maDMP model for Climate Data: Sharing the seas.

0: Note on personal data! (1 / 1)

Please sign the following declaration before proceeding with your DMP!

"As a SU user of DMP Online I am aware that, except for information about project members such as PI and contact person, no personal data - i.e. any information relating to an identified or identifiable natural person (a 'data subject', that is a living person), should be registered in this DMP. If necessary, to avoid this, any actual data registered here (even only as examples) will be completely anonymised before registration."

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
- No

Save

Answered 3 months ago by joakim.philipson@su.se
Testing SU maDMP model for Climate Data: Sharing the seas.

Demo DMP example of SU-VR template: I-Q1
Testing SU maDMP model for Climate Data: Sharing the seas.

Demo DMP example of SU-VR template: II-Q1

Original (VR) questions:

- How will the material be documented and described, with associated metadata relating to structure, standards and format for descriptions of the content, collection method, etc.?
- How will data quality be safeguarded and documented (for example repeated measurements, validation of data input, etc.)?

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 (URI).

1. Dataverse/StockholmUniversityLibrary
2. su.Fishshare.com
3. SND
4. Zemodo/StockholmUniversityLibrary
5. GitHub
6. README-file
7. Bolin Climate Research DB
8. Other: https://...
9. Manually (not recommended)

Additional Information

https://osdb.library.gu.se/OceanInteractions/

Guidance

The first four options of repositories are described here on SU research data web pages, where you also will find direct links to them. These repositories are also curated by members of the Research Data Management Team at SU, which means you can get further local support when creating metadata for your project using these services.

The Bolin Centre for Climate Research at SU have their own domain specific repository (database), with particularly good visualization of data. It is curated and administered by staff at the Bolin Centre.

If other, please specify URI (link).

Please note, datasets containing personal data, i.e., any information relating to an identified or identifiable natural person (a "data subject", that is a living person), should as a rule not be uploaded to cloud-based repositories with servers subject to jurisdiction outside EU-EEA, e.g., Dataverse (presently at Harvard.edu may change to local SU-server later), su.Fishshare.com. (SU instructions regarding the use of cloud services are here.)
Data repositories curated by SU RDM-team

- Bolin Centre DB
- Dataverse
- SND - https://snd.gu.se

- Create standard metadata (for DMP q & a)
- Pre-order DOI for your dataset (or private link)
- Curation and counseling from SU
- Harvest and transform for archival (su.figshare, so far)
- One storage / backup option during project (embargo)?
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.)

☐ 1. Data clean(s)ing
☐ 2. File format and software description
☐ 3. Integrity check of data files
☐ 4. Pre-registration
☐ 5. Repeated measurements
☐ 6. Suppl. documentation (e.g. codebook / variable description, ELN, questionnaire, stimuli)
☐ 7. Validation of data input
☐ 8. Registered reports
☐ 9. Other (please specify!)

Additional Information

https://ros.fi/o/, https://doi.org/10.1016/j.poacfin.2018.11.003
Testing SU maDMP model for Climate Data: Sharing the seas.

## Demo DMP example of SU-VR template IV-Q1

### IV: Legal and ethical aspects (3 / 3)

**Original (VR) questions:**
- How is data handling according to legal requirements safeguarded, e.g. in terms of handling of personal data, confidentiality and intellectual property rights?
- How is correct data handling according to ethical aspects safeguarded?

**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)?

- [ ] yes
- [ ] no
- [ ] unknown

**Save**

Answered 1 minute ago by joakim.philipson@su.se

This answer removes 2 questions from your plan.

---

The General Data Protection Regulation (GDPR) provides the following definition (article 4.1):

>'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

If you need help to find out whether you will be processing personal data in your project, answer "unknown" and the SU-RDM team will be automatically notified and will contact you about your potential need for support.
Q3: 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. dianrnr.) in the text field below.

Yes
No
Unknown
Yes, but not (a–e)

Additional Information

B / I / E / 

Demon DMP example of SU-VR template IV-Q3

First list (a–e): Answering yes to any of the points a–e above means that you may need to apply for ethical review for research taking place in Sweden. See information here. For research taking place in other countries, other rules may apply.

For support concerning ethical review of research involving humans, please contact etik@fs.su.se.

Second list (F–I):

F. Animal research, and the use of data from animal research, may require ethical review or other approvals. See information concerning animal research here (information in Swedish). Some information in English is provided here.

For guidance concerning what applies to animal research, please contact etik@fs.su.se.

G. The use of genetic resources and/or traditional knowledge associated with genetic resources is regulated in the Nagoya Protocol, implemented through the EU Access and Benefit-Sharing (ABS) Regulation. Information from the Swedish Environmental Protection Agency is found here. For guidance concerning matters related to the Nagoya Protocol and The ABS Regulation, please contact info@fs.su.se.
DMP example of SU-VR template: API output & transform

1. Raw.json

```
1. { 
2.   "id": 72835, 
3.   "title": "Testing SU maDMP model for Climate Data: Sharing the seas.", 
4.   "grant_number": null, 
5.   "last_updated": "2021-06-16 14:05:01 UTC", 
6.   "creation_date": "2021-03-04 10:42:54 UTC", 
7.   "test_plan": true, 
8.   "template": { 
9.     "title": "SU-VR template", 
10.    "id": 1456260114 
11.   }, 
12. }
```

2. Convert to .xml

```
<xml version="1.0" encoding="UTF-8">
  <dmp xmlns:xlink="http://www.w3.org/1999/xlink"
      xmlns:xsii="http://www.w3.org/2001/XMLSchema-instance"
      xmlns:local="local-functions"
      OBJID="72835">
    <LABEL> "Testing SU maDMP model for Climate Data: Sharing the seas." 
    <title> "Testing SU maDMP model for Climate Data: Sharing the seas." 
    </title>
  </dmp>
```

3. Transform & FAIR eval.

```
<FAIRScore>
  <section1>
    <DataDescription>
      <FileFormatName>
        <option7. File names with only character set [0-9A-Za-z-_.], no spaces. <option>
      </FileFormatName>
      <F-value1. F-value>
        <Option6. Descriptive file names (e.g. "20210101StockholmMeanTemperatures.csv") <Option>
      </F-value1.>
      <R-value1. R-value>
        <Option4. Proprietary file formats (.doc; .xlsx) <Option>
      </R-value1.>
      <R-value1. R-value>
        <Option2. Non-proprietary file formats (.csv, .txt, .json) <Option>
      </R-value1.>
      <R-value1. R-value>
        <Option1. File formats (.pdf; .csv; .txt) <Option>
      </R-value1.>
      <FileFormatName>
        <option7. File names with only character set [0-9A-Za-z-_.], no spaces. <option>
      </FileFormatName>
    </DataDescription>
  </section1>
</FAIRScore>
```

4. Valid RDA maDMP-schema-1.1.json
Thanks for your attention!

Contact

opendata@su.se