Stockholm University: SU-VR template

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
- No

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.

- 1. Non-proprietary file formats (e.g. .csv, .txt, .json, netCDF)
- 4. Proprietary file formats (.doc ; .xls)
- 8. Other measures (please specify!)
- 2. Sustainable file formats (e.g. .pdf; .csv; .txt)
- 3. Software specific file formats (e.g. Matlab - .mat; Stata - .dta)
- 5. Datestamped file names (e.g. '20201011StockholmMeanTemperatures.csv')
- 7. File names with only character set [0-9A-Za-z-_.], no spaces.
- 6. Descriptive file names (e.g. '20200910SanFranciscoDaylight1pm.mp4')

Guidance:

FAIR data stands for the principles that data should be **F**indable, **A**ccessible, **I**nteroperable and **R**e-usable. Please note, these principles can apply even if data cannot for various reasons be published openly. FAIR data is *not* the same as open data. But it can be used as *one* measure of data quality.
When answering this question, options 1 and 2 for *file formats* will make your datasets more **I**nteroperable and **R**e-usable, and so will score higher on FA**IR** data. (For sustainable and preferred file formats, please consult further: Sustainable digital file formats from Library of Congress, and / or the SND file format guide .)
Options 5, 6 and 7 for *file names* will make your data files more **F**indable (and thus also contribute to your **F**AIR data score). Option 7, using only is also a strong requirement for archiving in compliance with Swedish National Archives (*Riksarkivet*) rules, and also facilitates software processing of data files.

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.

Example Answer:

- https://hdl.handle.net/11353/56789
- 2020-99999_VR

Guidance:

At this level in the initial phase, since a dataset is defined here only at a conceptual level, so if detailed identifiers of individual datasets or files are still lacking, there is the possibility of regarding the entire project as one dataset, and use e.g. the Project-ID or Funder / Grant Number as the Dataset ID. *Please avoid whitespace* in the Dataset ID, so preferably use underscore or camelCase instead if necessary.

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

- doi
- handle
- other
- ark
- url

Guidance:

Please note that the Type option selected here should describe the type of the Dataset ID entered in Q2 above. If **other**, please specify the type of dataset ID used in the **Additional Information** as e.g. *"Local filename"* or *"Project-ID",* correponding to your entry of Dataset ID in Q2.

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.

Example Answer:

"Images of *Gadus morhua* from Helgeland in Norway 2014"

Q5: **Title** of **dataset**

Example Answer:

"Fish catch in a Norwegian fiord"

Guidance:

The title here can be quite general at this initial stage. If you already have a dataset record (even unpublished) in a repository like Dataverse, Figshare or Zenodo, you can use the Title of that dataset record.

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

- no
- unknown
- yes

Guidance:

Questions **Q6-Q9** concern only the sources or access points of **Reused datasets** that have already definite identifiers, locations and formats ("distribution" in the vocabulary of DCAT). An answer **"no"** or **"unknown"** to **Q6**, will **remove** the following questions **Q7-Q9**, **Q12** but will allow a comment in the text area. Remember that all answers can be edited later, if things change.

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)

Example Answer:

"http://some.repo.org/records/1234", "https://doi.org/10.17045/sthlmuni.12248576.v1"

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

- closed
- open
- shared

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.)

- 10. MPL-2.0
- 11. OSL-3.0
- 12. Other
- 13. No license
- 02. CC-BY-SA-4.0
- 01. CC-BY-4.0
- 03. CC-BY-ND-4.0
- 05. CC-BY-NC-ND-4.0
- 04. CC-BY-NC-4.0
- 06. CC0-1.0
- 07. GPL-2.0
- 08. GPL-3.0-or-later
- 09. MIT-license

Example Answer:

Other:
https://opendefinition.org/licenses/odc-by/ , https://opensource.org/licenses/CDDL-1.0

No license:
Personal data (video/audio interviews)

Guidance:

01. **CC-BY-4.0**
02. **CC-BY-SA-4.0**
03. **CC-BY-ND-4.0**
04. **CC-BY-NC-4.0**
05. **CC-BY-NC-ND-4.0**
06. **CC0-1.0** **Public domain**

Software Licenses:

07. **GPL-2.0**
08. **GPL-3.0-or-later**
09. **MIT-license**
10. **MPL-2.0**
11. **OSL-3.0**

For **Other** licenses, a comprehensive list is found at:
https://spdx.org/licenses/
You could also find an appropriate license here:
https://ufal.github.io/public-license-selector/
or here: https://opendefinition.org/licenses/

Then please click the **See full text** button (shown when hovering over license title), and / or copy the corresponding URL of the license selected from the menu bar and paste it in the text field (with multiple license URLs separated by comma).

If you choose the option '*No license*', being aware that it *might* make your dataset *less* **FAIR*** and **Re-usable**, please state the reason for this choice!
[There are *fully legitimate reasons*, concerning e.g. personal data and sensitive data, that cannot be shared publicly.]

*TO BE RE-USABLE:
...
R1.1. (meta)data are released with a clear and accessible data usage license.

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
- spreadsheets
- sound
- video
- other
- text
- images

Guidance:

Generic type of dataset, as indicated by main file types. If **other** please specify in comment area, e.g. "presentation".

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

Guidance:

If the conceptual dataset is still in process, and has not been shared or published, the issue date might be given e.g. as the first/latest date that files and or metadata were posted, uploaded to or modified in a data repository, described below (in answer to section II:Q1). If nothing else is applicable, select current date when this question in your DMP is first answered.

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

Example Answer:

1920-2019StockholmMeanTemperatures.**csv**, 20200910SanFranciscoDaylight1pm.**mp4**

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!*

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

Example Answer:

If other, please specify URI (link):
https://www.pangaea.de/

Guidance:

The ***first four*** options of ***data repositories*** are briefly described on our **su.se/researchdata** pages, where you will also find direct links to them. These repositories are *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 that are subject to jurisdiction outside EU-EES, e.g. *Dataverse* (presently at Harvard.edu) and *su.figshare.com*. (SU instructions regarding the use of cloud services are here.)

However, these repositories can still be used, without uploading datafiles,* to create standard *metadata records* only (provided the *metadata* itself does not hold any personal data, that is.) Such formal metadata records can describe datasets that are safely stored elsewhere, with possible links or references to access points for authorized users.

Please also note that a README-file, even when created through a research organisation template, can never fully replace formal metadata according to a standard used by a data repository and created through a web form. Formal metadata, following a common metadata standard by means of a repository web form may allow SU-RDM team to harvest and transform dataset deposits for automated mandatory archival, as a service to researchers. This will help alleviate the administrative burden on you to send in the same metadata and datafiles once again for archiving.

*[*Zenodo.org* actually demands at least *one* "datafile" uploaded, to allow creation of a metadata record, but this could be a simple README.txt-file, that just reproduces some of the descriptions in the formal metadata record.]

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.)

- 7. LOV - Linked Open Vocabularies
- 3. DDI
- 1. Metadata from II:Q1
- 2. DataCite
- 4. BARTOC Vocabularies
- 5. GEMET Vocabularies
- 6. Getty Vocabularies
- 8. OLS - Ontology Search
- 9. Ontobee
- 10. RDA metadata standards
- 11. Other - please specify!

Example Answer:

For *options 4, 6-9*, or if *Other* - please specify in the text area, preferably by name of metadata standard / ontology / vocabulary with corresponding IRI / URI link, e.g.:

- Cell Line Ontology: http://bartoc.org/en/node/1021
- Art & Architecture Thesaurus (AAT): https://www.getty.edu/research/tools/vocabularies/aat

- Darwin Core: https://rd-alliance.github.io/metadata-directory/standards/darwin-core
- Envo - The Environment Ontology: https://www.ebi.ac.uk/ols/ontologies/envo
- GOLD - General Ontology for Linguistic Description: http://purl.org/linguistics/gold
- MeSH: https://bioportal.bioontology.org/ontologies/MESH

Guidance:

Using common metadata standards and domain specific vocabularies / ontologies with links (IRIs / URIs) will make your data more **FAIR** - **F**indable, **A**ccessible, **I**nteroperable and **R**e-usable.

1. **Metadata from II:Q1** - if your answer to Q1 includes some reference to a specific repository (in particular the first four options in Q1), possible metadata standards can be inferred from this by the RDM-team.
2. DataCite is a general metadata standard which many repositories (including the first four options in Q1) provide either as default or as export format.
3. DDI - **Data Documentation Initiative** is a particularly rich metadata standard (used by *Dataverse* and *SND* in Q1)
4. BARTOC - **Ba**sic **R**egister of **T**hesauri, **O**ntologies & **C**lassifications, by type, subject, language, license...
5. GEMET - **Ge**neral **M**ultilingual ***E****nvironmental* **T**hesaurus (EEA - European Environment Agency), terms in 36 languages (incl. Swedish).
6. Getty Vocabularies - within arts & humanities mainly and Thesaurus of Geographic Names (TGN).
7. LOV - Linked Open Vocabularies holds > 725 different vocabularies for all types of research fields.
8. OLS - Ontology Lookup Service > 260 ontologies, mainly biomedical, from EMBL-EBI / Elixir.
9. Ontobee - holds >225 ontologies, mostly for life sciences (overlap with 8.), but also e.g. mathematics, software.
10. RDA Metadata Directory - general and domain specific metadata standards within several research areas.

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.)

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

Guidance:

1. ***Data clean(s)ing*** and consolidation will contribute to secure quality of data overall. (Here are some possible tools to use for this: frictionlessdata.io ; goodtables.io ; openrefine.org)

2. A detailed ***description*** (preferably as part of the general metadata entered in a repository web form) of ***file formats*** *and* ***software*** that was used to produce the dataset and possibly (in the case of software specific formats) also needed to open or understand/interpret the

data will help make your dataset more Re-usable - and thus more FAI**R**!

3. ***Integrity checks*** of datafiles are facilitated in data repositories by means of *checksum* algorithms such as MD5, SHA-256 etc. [https://en.wikipedia.org/wiki/Checksum, https://en.wikipedia.org/wiki/SHA-2] or *UNF* - Universal Numeric Fingerprint (used in *Dataverse*). These may be particularly relevant to apply when data files are moved from one location to another, to check that nothing was lost on the way or that data has not been "corrupted" since a previous upload. Automatic integrity checks, producing a hash check sum, are built-in to certain applications, but there are also tools to download freely for this, such as DROID from National Archives, UK.

4. ***Preregistration*** "is the practice of depositing a research question and study design with a registration service or journal before conducting a scientific investigation. ... Sharing your research design also increases the credibility and reproducibility of your results and helps address publication bias." [https://plos.org/open-science/preregistration/]

"Preregistration separates *hypothesis-generating* (exploratory) from *hypothesis-testing* (confirmatory) research. Both are important. But the same data cannot be used to generate *and* test a hypothesis, which can happen unintentionally and reduce the credibility of your results. Addressing this problem through planning improves the quality and transparency of your research. This helps you clearly report your study and helps others who may wish to build on it." [https://www.cos.io/initiatives/prereg]
There are several other sites offering preregistration, e.g. https://osf.io/ and https://aspredicted.org/, but to be meaningful the preregistration should be "locked" during research, so that it could not be changed and adapted to results post factum.
For this option, please give the URL of the service / site you intend to use in the comment area, if possible!

6. ***Supplementary documentation***: any information that helps others - and yourself(!) - understand, interpret and evaluate the data, even in the distant future. Examples may be *codebooks* / *variable descriptions*, *ELN* - Electronic Lab Notebooks,(anonymous) questionnaires / interview questions, stimuli used to solicit responses ...

7. ***Validation*** schemas or references to online services / software scripts (including *regular expressions*) used for data validation. (Validation could also be part of *1. Data clean(s)ing*.)

8. ***Registered reports*** "Registered Reports is a publishing format used by over 250 journals that emphasizes the importance of the research question and the quality of methodology by conducting peer review prior to data collection. High quality protocols are then provisionally accepted for publication if the authors follow through with the registered methodology."

III: Storage and backup

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

- 7. Other repository (please specify)
- 6. Repository in II:Q1
- 2. SU Box
- 3. Own harddrive
- 4. External harddrive
- 8. USB sticks
- X. Other option (please specify)
- 5. "Cloud server" (please specify)
- 1. SU Sunet Drive
- 9. SNIC-facility (HPC, PDC ...)

Example Answer:

Other repository: https://osf.io

Guidance:

Ideally, data files should be backuped at 2-3 different locations to be secure. Option 3 "*Own harddrive*" here means harddrive on your own work computer.

Please note that *all researchers and PhD-students affiliated with SU* will get **200 GB** in **SU Sunet Drive**, which is a secure local storage at SU. In the near future it may also be used for upload and publishing datasets to be accessible via the SND metadata catalog. Here you can also safely store *personal data.* More information about data storage at Stockholm university: su.se/store (på svenska: su.se/lagra). You will be able to login again to your alloted storage space at Sunet Drive here: https://drive.sunet.se/. (Use "Sign in with your organization or institution account (SWAMID/eduGAIN)", choose "Stockholm University" and then your SU credentials.)

If you need more space (> 200 GB) for a project you can order it from here (needs approval from your SU head of department / prefect): arkivportal.it.su.se/order_extra_space-research_data

Please note further that *if* your dataset contains ***personal data***, i.e. any *information relating to an identified or identifiable natural person* (a 'data subject', that is a living person)*, several* of the *options* above (e.g. 2, 5-7) may **not** be used. Consider again the guidance to Section II:Q1 - datasets containing *personal data* should as a rule *not* be uploaded to cloud-based repositories with servers subject to jurisdiction outside EU-EES, e.g. *Dataverse* (presently at harvard.edu) or *su.figshare.com*. (SU instructions regarding the use of cloud services are found here.)

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

- < 1 GB
- 501 GB - 1 TB
- 1 - 10 GB
- 11 - 200 GB
- 201 GB - 500 GB
- 1 TB < X < 10 TB

Guidance:

Please note that all researchers and PhD-students affiliated with SU will get 200 GB in **SUNET Drive** (covering the three smallest storage size options above), which is a secure local storage at SU, and can also be used for upload and publishing datasets to be accessible via the SND metadata catalog.
For larger storage needs, you will need to get approval from your institution (prefect) to cover the costs for that. If you want to order extra storage space, please fill in this order form. Note that you must be connected via VPN to access the form outside of SU's intranet.

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

- 3. Encrypted partition/files on own harddisk/Sunet Drive
- 4. Private links
- 1. Password protection
- 5. Two-factor authentication
- 2. Encrypted USB-stick
- 6. Other (please specify)

Guidance:

1. Passwords should be unique (not re-used from other services that you use) and sufficiently long (the more characters, the better protection).
For options 2. Encrypted USB-stick and 3. Encrypted partition/files on own harddisk or in Sunet Drive/Next Cloud, you might get help from SU IT-department, or download easy-to-use free encryption software like 7-zip or VeraCrypt.
4. Private links for sharing data only with a restricted selection of individuals are offered by some repositories, such as Dataverse and su.figshare.com.
5. Two-factor authentication means using two methods of authentication together to give access, such as e.g. *password* + a numeric code (for one time use) sent to your mobile phone.

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
- yes
- unknown

Guidance:

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.

Does the processing of personal data raise other ethical concerns? I.e. could the processing result in individuals feeling a loss of control of their personal data, disclosure resulting in a sense of loss of privacy or concern of negative bias in the future?

Here is a basic Instruction for the processing of personal data in research.

If you will be processing *personal data* for any purpose in your project, you might have to perform a **PIA** - a Privacy Impact Assessment or a **DPIA** - Data Protection Impact Assessment, as it is called in **GDPR**.

Here is a software tool to help perform a **DPIA** if needed: Carry-out-Data-Protection-Impact-Assesment

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.

- yes, but not (a-e)
- yes
- unknown
- no

Guidance:

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*. For research taking place in other countries, other rules may apply.
You will then also need to inform research subjects and ask for their consent.

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**.
For guidance concerning matters related to the Nagoya Protocol and The ABS Regulation, please contact info@fs.su.se.

H. Products that may be used for both civil and military purposes are called **dual-use products**, and may include chemicals, devices and software. Also knowledge about production and use of such products may in itself be a dual-use product. There are rules concerning export and transfer of dual-use products, see information here. For support concerning dual-use products, please contact the export control administrator at Stockholm University: lars.resman@su.se

I. Data may be **sensitive** in several ways. *Personal data* may be sensitive *even if* they do *not* belong to the special categories of personal data listed in the General Data Protection Regulation (article 9. 1), *listed under point **a**)* above. Such sensitive personal data may include data related to an individual's sense of privacy or that may lead to negative bias in the future, etc. Also data that are *not personal data may risk* harming people, for example by

stigmatizing certain groups. Such data must be handled with caution. Please contact etik@fs.su.se for further support.

Data may also be sensitive and require special protection if it concerns matters relating to *national security*, *industrial cooperation* projects, etc. Such data might need review by legal counsel.

Q3: What measures will you take to prevent illegitimate access to personal data or other sensitive data in your dataset(s)?

- 5. Encryption of data
- 4. User access control
- 3. Aggregation of data subjects groups (to prevent indirect identification)
- 2. Full anonymization (deleting key for re-identification)
- 1. Pseudonymization (with retention of key for possible re-identification in other location).
- 7. Other (please specify below!)
- 6. Confidentiality check (Sekretessprövning enligt OSL)

Guidance:

Multiple options can be selected. If *Other* option selected, please specify in comment area.

1. Pseudonymization means that personal identifiers are replaced in a way that allows for identification only with the help of additional information that is stored separately from the data. Please note that pseudonymized data are *still* considered *personal data* according to GDPR.

2. Anonymization means that all identifying links to the research subjects are irreversibly destroyed (and "backwards identification" is not possible). Successful anonymization means that GDPR no longer applies to this data, which might then be shared more freely, e.g. through a data repository (as in section II:Q1).

[*However*, you may still be obliged to submit the original, non-anonymized data to the SU-Archive for long-time preservation to be handled according to the Swedish Act on Public access to information and secrecy (pdf 2 MB) (SFS 2009:400) and the *Act containing supplementary provisions to the GDPR* (SFS 2018:218; pdf 208 kB).
This in principle would give the shared data status as only pseudonymized, but it has in other cases been judged sufficient as long as it is unlikely that the key for re-identification will ever be accessible to those taking part of the pseudonymized data.]

3. Aggregation of data can mean "*to group data in such a way that individual records no longer exist and cannot be distinguished from other records in the same grouping*". [Protiviti] *However*, aggregation does not automatically guarantee complete anonymization in the sense above.

Read more about anonymization and aggregation here:
Anonymization-and-GDPR
GDPR.eu/recital-162-processing-for-statistical-purposes
Iapp.org: aggregated-data-provides-a-false-sense-of-security
Protiviti: aggregation-and-anonymization

4. User access control may be by authorization mechanisms, password, personal

permission etc. (or a combination of such measures).

5. Encryption of data may be implememnted by downloading easy-to-use free encryption software like
7-zip or VeraCrypt , or by support from the IT-department.

6. Confidentiality check is performed together with the SU-Archive in accordance with the Swedish Act on Public access to information and secrecy. Choosing this option will send a notification about this to the SU RDM-team.

Q4: Has ethical review been performed for your project?
If *yes*, or there is a pending *application,* please provide (a) reference(s) to the relevant ethical review application(s)/approval(s)/decision(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.

- yes
- not yet; will apply / have applied for it
- no

Guidance:

Please give any additional information or links to such information about a possible application for ethical review that you might have in the comment area.

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.]

- 13. No license
- 11. OSL-3.0
- 10. MPL-2.0
- 09. MIT-license
- 03. CC-BY-ND-4.0
- 02. CC-BY-SA-4.0
- 01. CC-BY-4.0
- 06. CC0-1.0
- 05. CC-BY-NC-ND-4.0
- 04. CC-BY-NC-4.0
- 08. GPL-3.0-or-later
- 07. GPL-2.0
- 12. Other

Guidance:

01. **CC-BY-4.0**
02. **CC-BY-SA-4.0**
03. **CC-BY-ND-4.0**

04. **CC-BY-NC-4.0**
05. **CC-BY-NC-ND-4.0**
06. **CC0-1.0 Public domain**

Software Licenses:

07. **GPL-2.0**
08. **GPL-3.0-or-later**
09. **MIT-license**
10. **MPL-2.0**
11. **OSL-3.0**

For **Other** licenses, a comprehensive list is found at:
https://spdx.org/licenses/
You could also find an appropriate license here:
https://ufal.github.io/public-license-selector/
or here: https://opendefinition.org/licenses/
Then please click the **See full text** button (shown when hovering over license title), and / or copy the corresponding URL of the license selected from the menu bar and paste it in the text field (with multiple license URLs separated by comma).

If you choose the option '*No license*', being aware that it *might* make your dataset *less* **FAIR*** and **Re-usable**, please state the reason for this choice!
[There are *fully legitimate reasons*, concerning e.g. personal data and sensitive data, that cannot be shared publicly.]

*TO BE RE-USABLE:
...
R1.1. (meta)data are released with a clear and accessible data usage license.

Example Answer:

Other:
https://opendefinition.org/licenses/odc-by/ , https://opensource.org/licenses/CDDL-1.0

No license:
Personal data (video/audio interviews)

V: Accessibility and long-term storage

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

- 5. Other (please specify)
- 1. Repository in answer to II:Q1
- 4. On direct request from authorized users
- 3. SU Archive - default
- 2. Supplement to journal article / publication

Guidance:

Multiple options possible for different datasets / files. This question regards all research documentation that is not subject to confidentiality measures according to Swedish legislation on Public access to information and secrecy. Long-term archiving in local SU Archive is mandatory, even for research information to which confidentiality applies.

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

- Other documentation
- Software scripts
- Metadata AND all data files
- Metadata and some datafiles
- Only metadata
- Code book / variable descriptions

Guidance:

Multiple options possible.

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

- Continuously, as they are produced and updated
- After embargo expires
- Other (please specify!)
- On completion of research project
- Only after publication of journal article / paper

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
- Automatic harvest & transfer from repository
- Other (please specify!)
- E-mail attachment
- Direct transfer to temp. archive (MADI) on request

Guidance:

For long-term storage and preservation of data files, metadata and documentation, archiving in SU digital archive is mandatory. *Presently* we offer *Automatic harvesting* and transformation to archival format for all metadata and data files uploaded to *su.figshare.com* and *zenodo.org* (of the repositories named under Documentation and data quality, Q1), which means the transfer to SU archive is automatic for these files so that researchers need not do anything more to have them delivered. In the near *future* we hope to *extend* this service *to* more repositories, such as *Dataverse* and *SND*.

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?

- yes
- no
- unknown

Guidance:

This question mainly concerns **archiving** and prospects for long-term **preservation / sustainability** of your dataset(s), and - if the answer is yes - the **software** that was used to produce the data.
This is a resource to evaluate how sustainable that software may be:
https://www.software.ac.uk/resources/online-sustainability-evaluation

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
- no
- unknown

Guidance:

This question mainly concerns how **accessible** your datasets will be for evaluation, **re-use** and **replication** of results - that is more F**A**I**R**.
If your answer is *yes***,** please describe in the text area what software you will use (references with versionnr, URLs) or how it can be accessed.
If the **software** / code used is created within the research project itself, to ensure that it is also more **FAIR**, you might consult these recommendations: https://fair-software.eu/ .

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

- 3. Other (please specify!)
- 2. Proprietary/Commercial (e.g. Matlab, Stata)
- 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
- Yes (please specify!)

Guidance:

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 or processed by cloud-services unless SU has a clear *Personal Data Processing Agreement (* Swe. *"Personuppgiftsbiträdesavtal")* with the service provider of the SaaS / "cloud" software server agent. (SU instructions regarding the use of cloud services are here.)

Please also note that if you intend to use commercial online services or need special software licenses you might be subject to *Swedish Law on public procurement* (**LOU** - *Lagen om Offentlig Upphandling*), and should consult with *upphandling@su.se* .

Example Answer:

Yes: NVivo Transcription
https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/about/nvivo/modules/transcription

Q9: Does SU have a *Personal Data Processing Agreement* (Swe. "Personuppgiftsbiträdesavtal") with the service provider of the SaaS / "cloud" software server agent?

- no
- unknown
- yes

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
- 3. Contributors with Role: Data Manager
- 5. Others (please specify!)
- 4. RDM-team at SU
- 1. PI - default

Guidance:

Multiple options possible. RDM team can help with data management mainly through curation of research data deposited in su.figshare.com, zenodo/stockholmuniversitylibrary, Dataverse and SND (above), and with counselling regarding archiving and preservation, ethical and legal questions.
Option 3. **Contributors with Role: Data Manager**, designated in previous screen (*dmp.su.se/plans/[DMP-Id No.]/contributors*).

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

- 2. Internal secure storage
- 7. Other (please specify)
- 6. External secure storage
- 5. Extra costs for staff
- 4. Special software
- 3. Repository account(s)
- 1. RDM-team support

Guidance:

Options 1. and most often even 3. will not incur any extra costs for your project.

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).

- > 500000 SEK
- 100000 < C < 500000
- 10000 < C < 50000 SEK
- 50000 < C < 100000 SEK
- < 10000 SEK

Example Answer:

Estimated cost for secure data storage (4 years): 100000 SEK
Data steward (consultant): 500000 SEK

Guidance:

Possible specification of estimated costs in commentary field.

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.

- Yes
- 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).

Guidance:

Please give here only the identifier-string (that is within the " " in the examples above)!
For personal integrity reasons, do *not* use a *Swedish personnummer* (civic registration no.) as

identifier-string. Better then to use an e-mail address.

The ***type*** of identifier is selected in next question. Naturally, it should correspond to the identifier entry here.

Example Answer:

[**orcid**:]
"0000-0001-5699-994X"
[**isni**:]
"0000000078417464"
[**other**:]
"https://viaf.org/viaf/19904335/"
"name.nameson@swemail.se"

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

- isni
- other
- openid
- orcid

Guidance:

If **other,** please specify in comment area.

Example Answer:

other: URI
other: ScopusID
other: e-mail

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!

- Asian and Middle Eastern Studies
- Other
- Teaching and Learning (merged from 2022-01-01)
- Sociology
- Social Anthropology
- Astronomy
- ----------
- Swedish Institute for Social Research (SOFI)
- Stockholm Centre for Organizational Research (SCORE)
- Stockholm Business School
- Institute for International Economic Studies (IIES)
- Statistics
- Special Education
- Ethnology, History of Religions and Gender Studies
- English
- Culture and Aesthetics
- Archaeology and Classical Studies
- Stockholm Resilience Centre
- Nordic Institute for Theoretical Physics (NORDITA)
- Baltic Sea Centre
- Bergius Botanical Garden
- Materials and Environmental Chemistry (MMK)
- ----------
- Stockholm University Library

- Biology Education (BIG)
- Molecular Biosciences, The Wenner-Gren Institute (MBW)
- Ecology, Environment and Plant Sciences (DEEP)
- Zoology
- ----------
- Environmental Science
- Geological Sciences
- Physical Geography
- ----------
- Social Work
- Mathematics and Science Education (MND)
- Mathematics
- Meteorology (MISU)
- Physics
- ----------
- Biochemistry and Biophysics (DBB)
- Organic Chemistry
- Public Health Sciences
- Psychology
- Political Science
- Human Geography
- Education
- Economics
- Economic History and International Relations
- Criminology
- Child and Youth Studies
- Computer and Systems Sciences
- ----------
- Law
- ----------
- Swedish Language and Multilingualism
- Slavic and Baltic Studies, Finnish, Dutch and German
- Romance Studies and Classics
- Philosophy
- Media Studies
- Linguistics
- Language Education
- Humanities and Social Sciences Education
- History

Example Answer:

Other:
SciLife Lab

Q4: **Language** used for this DMP. Please select!

- und: Other
- spa: Español
- deu: Deutsch
- fra: Français
- eng: English - default
- swe: Svenska

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
- Östersjöstiftelsen
- Statens Geotekniska Institut
- Ragnar Söderbergs stiftelse
- Riksbankens jubileumsfond
- IFAU
- FORMAS
- FORTE
- Vinnova
- VR - Swedish Research Council

Example Answer:

If Other:
Wellcome Trust, Rymdstyrelsen

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)

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

- rejected
- granted
- applied
- planned

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).

Guidance:

Identifier type should be selected 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".*

Example Answer:

Short Title of dataset	**Identifier of dataset**	**Identifier Type**
T1: Public version 2.0 of figs2FGS-CSP	*Id1: https://doi.org/10.5281/zenodo.3778653*	*Type1: doi*
T2: RDMtoolkit4re-use	*Id2: https://github.com/RDMtoolkit4re-use/*	*Type2: url*
T3: None	*Id3: None*	*Type3: None*

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

Example Answer:

References: authors (year): *titles*	**Identifiers / URLs / Locations**
Kim, S. (2020): *Machine-actionable Data Management Plans Model Analysis and Improvement Direction.*	https://doi.org/10.1633/JISTAP.2020.8.4.2
Loeb, Jacques (1906): *Vorlesungen über die Dynamik der Lebenserscheinungen.*	Leipzig: Verlag von Johann Ambrosius Barth, 1906.
Smale, N. et al. (2020): *A Review of the History, Advocacy and Efficacy of Data Management Plans*	https://doi.org/10.2218/ijdc.v15i1.525

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.

- Unknown
- No
- Yes