Data Management Plan: a required data travelogue

The GUIDE for DMP

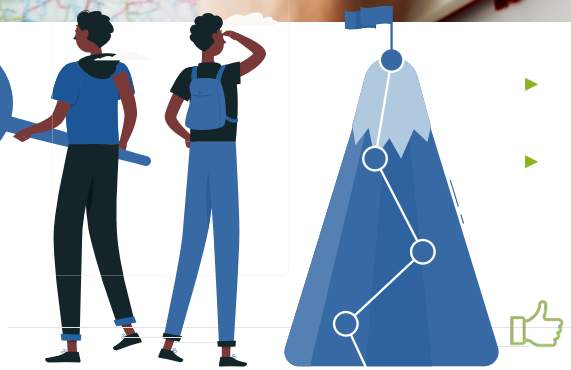
▸ OPEN SCIENCE ▸ FAIR ▸ 2021

Cecilia Mascia[a], *Vittorio Meloni*[a], *Alessandro Sulis*[a],
Caterina Giorgia Carboni[b], *Franco Cappai*[b] *and Francesca Frexia*[a]

[a]*Center for Advanced Studies, Research and Development in Sardinia (CRS4)* - [b]*Sardegna Ricerche*

DMP: a resource for the ages

Flower photo created by rawpixel.com - www.freepik.com



Illustrations created by stories - www.freepik.com

- Data reuse
- Result reproducibility
- EU Commission requirement

Aided DMP compilation

Picture adapted from clouds vector created by vectorjuice - www.freepik.com

DSW DATA STEWARDSHIP WIZARD

argos

DMPTool Build your Data Management Plan

DMPONLINE

https://ds-wizard.org/
https://argos.openaire.eu/
https://dmptool.org/
https://dmponline.dcc.ac.uk/

DMP v1.0

- Questionnaire-like
- Partially completed

further elaboration

See also:

Perception of researchers in the preparation of data management plans, Viviane Veiga – OS FAIR lightening talk (22 sept 2021)

Feedback from H2020 beneficiaries on writing a DMP - D. Spichtinger: https://doi.org/10.12688/openreseurope.13342.1

The guide

- Analysis of several DMP templates between the most used ones – e.g., those from the **Swedish Research Council**, the **Digital Curation Centre** and the **European Commission**.
- Substantial overlap in content and structure.

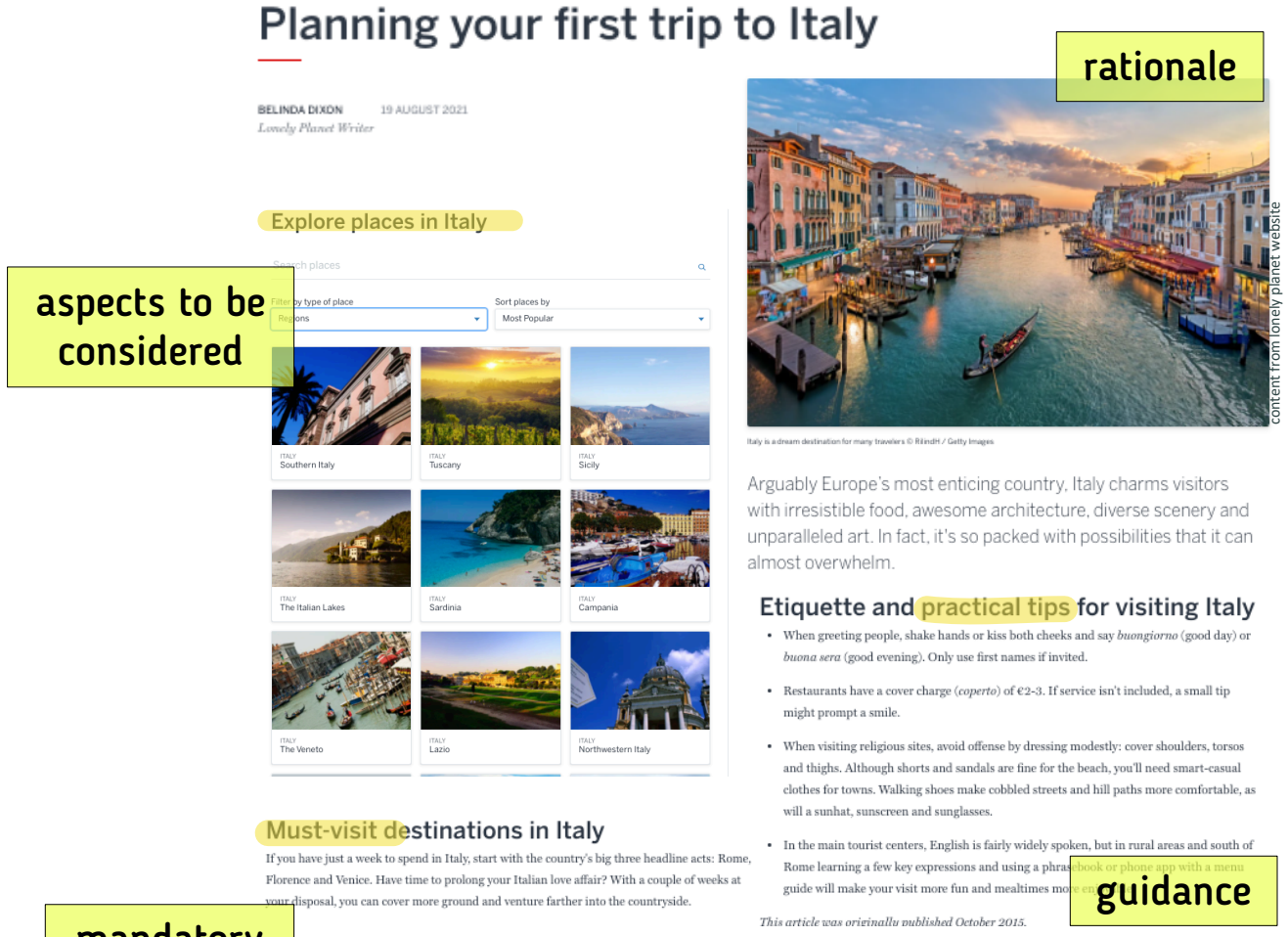
- Available on Zenodo at: https://doi.org/10.5281/zenodo.5140990
- Testing scenario: 18 independent clinical studies within the **I FAIR program**

F. Cappai, C. G. Carboni, E. D'Aloja, G. Fotia, F. Frexia, G. Serra, G. Sotgiu, P. Uva, G. Zanetti (2019), I FAIR Program: the Sardinian way to support and fund independent clinical studies that want to be Findable Accessible Interoperable Reusable, in: The Ecosystem of Evidence Conference – Abstract book, GIMBE, Evidence for Health, p. 15. URL: https://www.ebhcconference.org/Abstract_book_2019.pdf

The structure of the guide

Five main dimensions:

1. General information;
2. Data management:
   a. Collection and generation;
   b. Documentation and metadata;
   c. Shareability, publication and reuse;
   d. FAIRness;
3. Resources;
4. Security, storage and backup;
5. Ethical, legal, administrative and privacy aspects.



* specialised for the I FAIR programme

Write down as much information as possible

Five main dimensions:

1. General information;
2. Data management:
   a. Collection and generation;
   b. Documentation and metadata;
   c. Shareability, publication and reuse;
   d. FAIRness;
3. Resources;
4. Security, storage and backup;
5. Ethical, legal, administrative and privacy aspects.

3.3.2. Documentation and metadata

Rationale: Describe all types of documentation (README files, metadata, etc.) that will be provided to support data understanding and reuse. Metadata should at least include basic details allowing other users (computer or human) to find the data (minimally: the file name, a persistent identifier, collection date, access conditions, etc.). Furthermore, the documentation may include details on: the methodology used, the performed processing and analytical steps, variable definitions, references to vocabularies used, as well as units of measurement, possibly following existing community standards and guidelines. This section should explain how this information will be prepared and shared **[SNSF].**

Aspects to be considered:

Documentation content and purpose (data description, definitions of variables, units of measurement, contextual information, procedures, processing, data quality measures, etc.); documentation availability (where recorded and how is accessible) and formats (e.g., a 'readme' text file, file headers, code books, lab notebook , etc.); metadata (e.g., purpose, types, formats and standard scheme, standards, machine-readability, generation or collection); data searchability (unique and persistent identifiers, workflows for data identification, etc.); software requirements and necessary knowledge for (meta)data processing **[SNSF], [DCC], [NSF], [UniBi], [HRB], [NWO], [SRC]**.

Guidance:

- Where these are in place, researchers are advised to use community metadata standards. The Research Data Alliance maintains a Directory of Metadata Standards **[NWO].**
- Note the standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies) **[NSF]**.

Mandatory Elements (for I FAIR)

Documentation content and purpose (data description, definitions of variables, units of measurement, contextual information, procedures, processing, data quality measures, etc.); documentation availability (where recorded and how is accessible) and formats (e.g., a 'readme' text file, file headers, code books, lab notebook, etc.); software requirements and necessary knowledge for (meta)data processing.

- http://rd-alliance.github.io/metadata-directory/

Remembering the “exceptionalism”

3.6. Ethical, legal, administrative and privacy aspects

Rationale: Ethical issues in research projects require specific measures in addition to ordinary data management procedures, such as data anonymization, ethics committees approval, formal consent agreements, etc. In this section, all the relevant ethical issues in the project and their respective countermeasure should be outlined. Note that some of the content listed here may overlap with what is defined in the previous section. If additional legal or administrative issues are identified, such as Intellectual Property Rights (IPRs) and ownership, they should be treated here as well **[SNSF],[HRB].**

Aspects to be considered:

Security provisions for sensitive/personal data (data protection, recovery, secure storage and transfer); permissions for data handling (collection, processing, storing, etc.); relevant legislation and standards (e.g., GDPR); confidentiality agreement; informed consent (formal agreement, gaining, etc.); ethical issues; ethical review (IRB, Ethic Committee, protocols, approval dates, etc.); specific risks/measures for data security; support staff; data anonymization/pseudonymization and encryption; code of conduct; any national/institutional

3.5. Security, storage and backup

Rationale: Data backup, recovery and security measures are of an utmost importance, especially when working with personal or other sensitive data. This section should outline which measures are adopted, how they are put into practice, eventual standards or regulations considered (e.g., ISO 27001-Information security management) and the main procedures or facilities for storage, processing or transfer of personal or other sensitive data. **[NWO],[SNSF],[H2020].**

Aspects to be considered:

Data and metadata security provisions (e.g., data recovery, automatic data backup, data access rights, technological solutions, responsible staff, etc.); storage and retention (e.g., certified repository, institutional/third parties storage, online/offline storage, locations, multiple copies, formats, etc.); security standards and legislation (e.g., regulations, confidentiality agreement, etc.); security risk analysis (e.g., main concerns, estimation, how they can be managed, etc.); personal and sensitive data protection measures (e.g., secure storage, transfer, anonymisation, pseudony

[H2020],[SNSF],[DCC],[SRC], [UniBi],

3.3.3. Shareability, publication and reuse

Rationale: This paragraph should describe the main aspects of data sharing, including methods, repositories, how the reuse is regulated, timing for release and restrictions **[SNSF].**

Aspects to be considered:

Data repositories (e.g., trustworthy data repository, indexed in a catalogue, direct handling of data requests, proprietary/not-commercial, generic/domain-specific platforms, FAIR compliance); data discoverability; sharing conditions and policies (data selection, licenses, Data Access Committee, chargeable access, timing of data release, reason for delay if applicable, justification in case of no shareability, acknowledgement for data reuse, access procedures) and limitations motivations (legal, ethical, copyright, confidentiality or other clauses); obligation for data release (if existing); long-term preservation plan (long-term value

Italy

Ethical questions raised in search for Sardinian centenarians' secrets

Samples from residents of Sardinia's 'Blue Zone' famed for longevity have been sold to for-profit research firm Tiziana

▲ For every 100,000 people in Sardinia there are 24 centenarians; in the Blu
Photograph: DEA/L Romano/De Agostini/Getty Images

Stephanie Kirchgaessner *in the Blue Zone, Sardinia*

@skirchy

Fri 12 Aug 2016 16.53 BST

There is something like gold flowing through the ve
Maria Tegas, and everyone wants a piece of the trea

The centenarian, who lives in a poor and remote ar
Sardinia - in one of 14 villages known to geneticists
the Blue Zone - has not had an easy life. Orphaned

Parliamentary questions

103k 17k

30 June 2016

P-005318-16

Question for written answer P-005318-16
to the Commission
Rule 130
Giulia Moi (EFDD)

Subject: Protection of human DNA databases

Answer in writing

SharDNA, a (mostly publicly-owned) research company specialising in genomics, over the years has tested some 13 000 DNA samples donated by the inhabitants of ten towns and villages in the Ogliastra area, thus building up an extremely valuable genetic database. The events that resulted in the bankruptcy of SharDNA in 2012 have changed the purpose of the research, transforming it into a commercial activity. The importance of this database is due to the fact that the samples relate to one of the longest-living populations in the world, living in an isolation that has enabled ancient genes to be preserved unchanged over the centuries. Given that, for genetic data, a high level of protection of the right to the protection of personal data is always required, taking into account the Unesco Declaration on Human Genetic Data, whereby the collection of genetic material must be based on the principle of 'prior, free, informed and express consent, without inducement by financial or other personal gain', can the Commission answer the following questions:

— In its view, should legislation not be drawn up in order to prevent the commercial use of such sensitive information?
— What tools are currently available to the Commission to prevent the marketing of such an important heritage?
— Will it encourage the adoption of instruments with a view to protecting such personal rights, a task that is currently left to the discretion of individual Member States?

Original language of question: IT

Legal notice - Privacy policy

Something about the SharDNA case:
https://doi.org/10.1007/s12687-017-0328-2
https://doi.org/10.1007/978-3-030-49388-2_17

Remembering the “exceptionalism”

3.6. Ethical, legal, administrative and privacy aspects

Rationale: Ethical issues in research projects require specific measures in addition to ordinary data management procedures, such as data anonymization, ethics committees approval, formal consent agreements, etc. In this section, all the relevant ethical issues in the project and their respective countermeasure should be outlined. Note that some of the content listed here may overlap with what is defined in the previous section. If additional legal or administrative issues are identified, such as Intellectual Property Rights (IPRs) and ownership, they should be treated here as well **[SNSF],[HRB].**

Aspects to be considered:
Security provisions for sensitive/personal data (data protection, recovery, secure storage and transfer); permissions for data handling (collection, processing, storing, etc.); relevant legislation and standards (e.g., GDPR); confidentiality agreement; informed consent (formal agreement, gaining, etc.); ethical issues; ethical review (IRB, Ethic Committee, protocols, approval dates, etc.); specific risks/measures for data security; support staff; data anonymization/pseudonymization and encryption; code of conduct; any national/institutional

3.5. Security, storage and backup

Rationale: Data backup, recovery and security measures are of an utmost importance, especially when working with personal or other sensitive data. This section should outline which measures are adopted, how they are put into practice, eventual standards or regulations considered (e.g., ISO 27001-Information security management) and the main procedures or facilities for storage, processing or transfer of personal or other sensitive data. **[NWO],[SNSF],[H2020].**

Aspects to be considered:
Data and metadata security provisions (e.g., data recovery, automatic data backup, data access rights, technological solutions, responsible staff, etc.); storage and retention (e.g., certified repository, institutional/third parties storage, online/offline storage, locations, multiple copies, formats, etc.); security standards and legislation (e.g., regulations, confidentiality agreement, etc.); security risk analysis (e.g., main concerns, estimation, how they can be managed, etc.); personal and sensitive data protection measures (e.g., secure storage, transfer, anonymisation, pseudony
[H2020],[SNSF],[DCC],[SRC], [UniBi],

3.3.3. Shareability, publication and reuse

Rationale: This paragraph should describe the main aspects of data sharing, including methods, repositories, how the reuse is regulated, timing for release and restrictions **[SNSF]**.

Aspects to be considered:
Data repositories (e.g., trustworthy data repository, indexed in a catalogue, direct handling of data requests, proprietary/not-commercial, generic/domain-specific platforms, FAIR compliance); data discoverability; sharing conditions and policies (data selection, licenses, Data Access Committee, chargeable access, timing of data release, reason for delay if applicable, justification in case of no shareability, acknowledgement for data reuse, access procedures) and limitations motivations (legal, ethical, copyright, confidentiality or other clauses); obligation for data release (if existing); long-term preservation plan (long-term value

Guidance:
- Consider using a generic FAIR-compliant platform (see the example) for data deposit if there are no suitable specific repositories for the considered research field **[SNSF];**
- Repository Finder can help to find an appropriate repository to deposit research data; data licenses can be expressed through the commonly used Creative Commons licenses and a wizard can help to choose the correct one. Check also the Five Recommendations for FAIR Software. **[NWO]**.

Guidance:
“Accessible data” doesn’t imply “open data”, the approach should be “as open as possible, as closed as necessary”[4]. The Registry of Research Data Repositories provides a useful listing of repositories that you can search to find a place of deposit.

Guidance:
The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Practical TIPS from the guide:

- https://repositoryfinder.datacite.org/
- https://ufal.github.io/public-license-selector/
- https://fair-software.nl/
- https://www.re3data.org/
- https://b2share.eudat.eu/

A living guide for a living document

- A **complementary approach** to existing tools and good practices.
- Good initial feedback from the **I FAIR Program**.
- Plans: **systematic evaluations** and refinement.

THANK YOU

contact ▸ cecilia.mascia@crs4.it
the guide ▸ https://doi.org/10.5281/zenodo.5140990