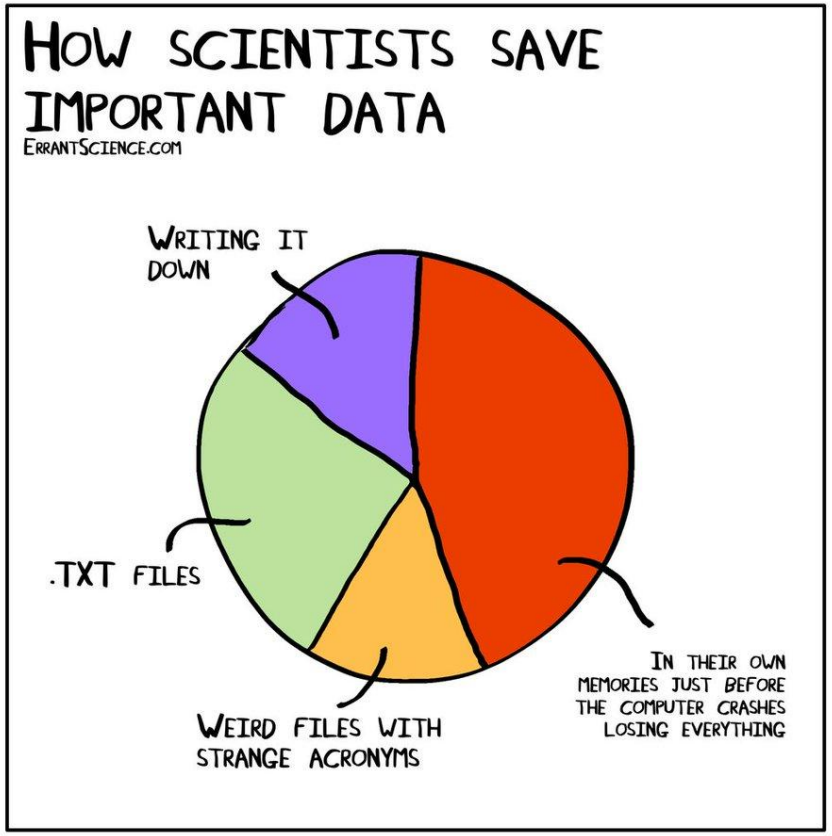
Data Management Planning tools

RDMkit

DSW

Laura Portell Silva
BSC, ELIXIR-ES

ELIXIR-Converge has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 871075.

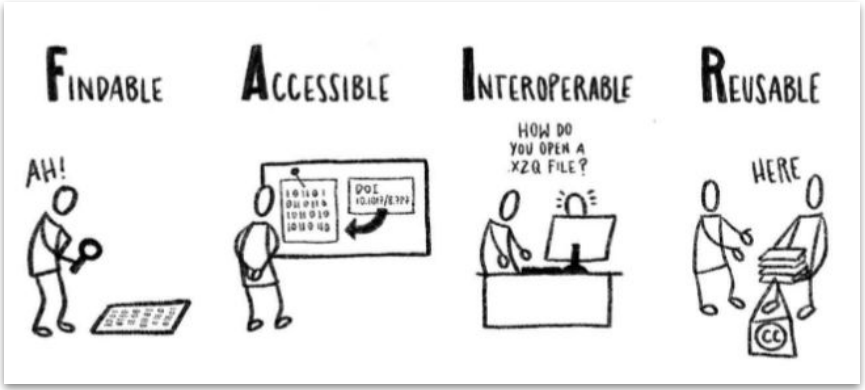

HOW SCIENTISTS SAVE IMPORTANT DATA
ERRANTSCIENCE.COM
WRITING IT DOWN
.TXT FILES
WEIRD FILES WITH STRANGE ACRONYMS
IN THEIR OWN MEMORIES JUST BEFORE THE COMPUTER CRASHES LOSING EVERYTHING

Research Data Management (RDM)

FAIR management of the research data throughout all the steps of the data lifecycle, from the first steps of data management planning to the final steps of depositing data to public archives.

Research Data Lifecycle

FAIR data



Goal

Support FAIR data management for the diversity of life sciences across Europe via national nodes

With an infrastructure that is

- Internationally **aligned**
- **Scalable** to 500 000 life scientists
- Long-term **sustainable**
- **FAIR** data at its heart

ELIXIR Supporting FAIR data in and across nodes and beyond

FAIR Services & Resources

Registries, standards, ontologies, identifiers, data management platforms, stewardship tools, templates

Specific Communities

Standards, Infrastructure, Pilots Human Data, Biodiversity

Trusted Repositories

Deposition Databases & Portals Scalable Curation
Sustainability

FAIR Data Techniques

Workflows, reproducible processing, transparent reporting & provenance
FAIR assessment & evaluation
FAIRification methods

FAIR Policy and Activism

FAIR principles, assessments
FAIR leadership & partnering at the global, European and national level.

FAIR Expertise and training

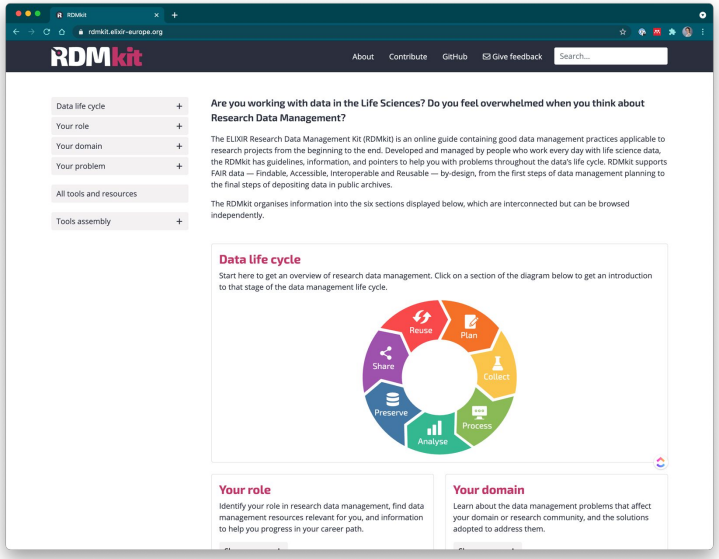
Capability frameworks, skills, data expert network, training portal

RDMkit goal

Provide guidance to the available support!

RDMkit supporting FAIR-by-design

Website focal point for guidance, information, best practice, examples



RDM through the **life cycle of projects** as outlined in **DMPs**

How to help projects with their RDM and FAIR data?

Researchers

One stop shop of information, signposting resources for developing & executing project Data Management Plans.

Data managers and data stewards

guides to complement their own expertise and resources.

Project consortium coordinators

For research data management, practice and potential partners.

Funding agencies and policy makers

Assess and evaluate Data Management Plans, develop open science policies and highlight open science requirements.

RDMkit components

Written by field workers in the language of the community

1. Data Lifecycle

3. Guidelines both generic & specific

5. Tools and resources

2. Roles

4. Example tool assemblies

6. Training resources

1

Data lifecycle

Planning

What is Data Management Planning?

Data Management Planning consists of defining the strategy that you plan to use for managing data and documentation generated within the project. It is about thinking upfront what's the best way to avoid problems or unexpected costs related to data management, and set the conditions for your research data to achieve the highest possible impact in science, even after the end of the project.

. . .

Why is Data Management Planning important?

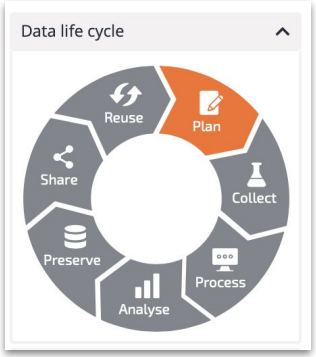
It is good research practice to take care of your research data and have a data management plan. It will make your work more efficient, facilitate team work and use of services and tools. Moreover, a detailed DMP would help in making your research data more FAIR. Advantages of making a data management plan:

. . .

What should be considered for Data Management Planning?

- Research organisation and funders often require a DMP as part of the application for grants or latest, when the project is funded. Therefore, consider guidelines, policies and tools for data management planning required by your funder.
- Data management should be planned in the early stages of a research project. Preferably, the DMP should be filled in before starting data collection. However, the DMP is a **living document** and should be updated as the research project progresses to match e.g. an update of the infrastructures, research softwares or a novel collaboration.

. . .



Related RDMkit pages in "Your tasks"

Compliance monitoring & measurement: measure compliance to data management regulations and standards.

Data management plan: how to write a Data Management Plan (DMP).

Data protection: how to make research data compliant to GDPR.

2 Roles

Data steward policy

Description

As a policy data steward, I focus on policy development and the implementation of research data management practices in my organisation. I work in close collaboration with directors, policy makers and funders, and I help establish suitable and sufficient data stewardship services and infrastructure.

It is my job to coordinate and align efforts on the quality, security and management of my organisation's data, thus effectively supporting the research process. I have a good knowledge of local, national and international procedures and regulations, such as my institution's research code, national codes of conduct or the EU privacy legislation. I translate these general concepts into practical guidelines for my organisation, and coordinate their implementation and monitoring.

. . .

Learning path

Institutes across Europe have started hiring professional data stewards. A policy oriented data steward is expected to be competent in the following areas:

- Develop, implement, monitor and evaluate policies regarding research data that are endorsed by researchers and aligned with internal and external stakeholders and effectuate change management
- Communicate about the FAIR data and Open Science principles to researchers, support staff and relevant stakeholders, thus creating awareness within institutes and organisations
- Give advice on RDM and formats for data management plans (DMP) to a broad audience

IT

Policy

Research

Related RDMkit pages in "Your tasks"

Compliance monitoring & measurement: measure compliance to data management regulations and standards.

Licensing: how to license research data.

Data management plan: how to write a Data Management Plan (DMP).

Data protection: how to make research data compliant to GDPR.

Sensitive data: how to identify different types.

Training

TeSS - ELIXIR's training portal

RDNL - Essentials for Data Support

Mantra - RDM training

FAIR guiding principles

Data Carpentry lessons

RDNL & DCC - Delivering RDM Services

NPOS/ELIXIR data steward competency framework

ELIXIR Data Management Network

Science Europe - Practical Guide to the International Alignment of RDM

3.1 Guidelines for life science RDM tasks

Data management plan

What template should you use to draft your Data Management Plan (DMP)?

Description

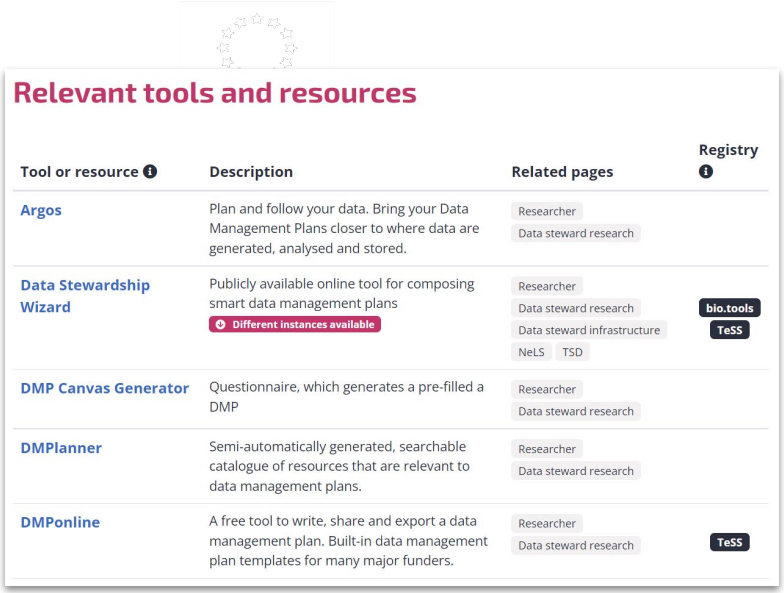
A number of DMP templates are currently available, originating from different funding agencies or institutions. Moreover, there are ongoing efforts to develop templates for machine-actionable DMPs.

Considerations

- Each funding agency could require or recommend a specific DMP template.
- Your institution could require and recommend a DMP template.
- Template could be presented as list of questions in text format or in a machine-actionable format.

Solutions

- Consult the documentation of your funding agency or institution, or contact them to figure out if they require or recommend a DMP template.
- A core DMP template has been provided by Science Europe.
- From the Horizon Europe Programme Guide and the Horizon Europe Annotated Model Grant Agreement you can read DMP guidelines and access the Horizon Europe DMP template.



Relevant tools and resources

Tool or resource	Description	Related pages	Registry
Argos	Plan and follow your data. Bring your Data Management Plans closer to where data are generated, analysed and stored.	Researcher; Data steward research	
Data Stewardship Wizard	Publicly available online tool for composing smart data management plans. Different instances available	Researcher; Data steward research; Data steward infrastructure; NeLS; TSD	bio.tools; TeSS
DMP Canvas Generator	Questionnaire, which generates a pre-filled a DMP	Researcher; Data steward research	
DMPlanner	Semi-automatically generated, searchable catalogue of resources that are relevant to data management plans.	Researcher; Data steward research	
DMPonline	A free tool to write, share and export a data management plan. Built-in data management plan templates for many major funders.	Researcher; Data steward research	TeSS

3.2 Guidelines for particular life science domains

- Plant sciences
- Marine metagenomics
- Human data
- Biomolecular simulation data
- Intrinsically disordered proteins
- Microbial biotechnology
- Epitranscriptome data
- Proteomics
- Toxicology data
- Bioimaging data

Human data

Introduction

When you do research on data derived from human individuals, there are additional aspects that must be considered during the data life cycle. Note, much of the topics discussed on this page will refer to the EU General Data Protection Regulation (GDPR) as it is a central piece of legislation that affects basically all research done on human subjects in the EU and on individuals residing in the EU. Much of the information on this page is of a general nature when it comes to working with human data, an additional focus is on human genomic data and the sharing of such information for research purposes.

Planning for, and collection of, human research data

Description

For research on human data, you must follow established research ethical guidelines and legislations. Preferably, planning for these aspects should be done before starting to handle personal data and in some cases such as in the case of the GDPR, it is an important requirement by laws and regulations.

Considerations

- Have you got an **ethical permit** for your research project?
  - To get an ethical permit, you have to apply for an **ethical review** by an **ethical review board**.

Solutions

- Tryggve ELSI Checklist is a list of Ethical, Legal, and Societal Implications (ELSI) to consider for research projects on human subjects.

Relevant tools and resources

Tool or resource	Description	Related p
BBMRI-ERIC's ELSI Knowledge Base	The ELSI Knowledge Base is an open-access resource platform that aims at providing practical know-how for responsible research.	Data prote Sensitive da Data stewa Data stewa
Beacon	The Beacon protocol defines an open standard for genomics data discovery.	Researcher Data stewa Data stewa
BIONDA	BIONDA is a free and open-access biomarker database, which employs various text mining methods to extract structured information on biomarkers from abstracts of scientific publications	Data storag Proteomics

Tool assemblies

TransMed Assembly

What is the TransMed data and computing tool assembly?

The TransMed data and computing tool assembly is an infrastructure provided by ELIXIR Luxembourg for clinical and translational projects. TransMed assembly provides the tools for managing ongoing projects that often require the management of cohort recruitment, and processing of samples, data and metadata. This entails GDPR-compliant and secure data collection, storage, curation, standardisation integration and analysis of clinical data and associated molecular, imaging and sensor/mobile data and metadata.

Who can use the TransMed data and computing tool assembly?

All researchers can use tools in the TransMed assembly individually or in combination depending on their project needs. Most of the tools in the TransMed assembly are open-source and can be re-used. ELIXIR Luxembourg provides know-how transfer and training on the tool assembly upon request from researchers and data steward organisations. To make a request please contact info@elixir-luxembourg.org.

For what purpose can the TransMed assembly be used?

Relevant tools and resources

Tool or resource	Description	Related pages	Registry
Ada	Ada is a performant and highly configurable system for secured integration, visualization, and collaborative analysis of heterogeneous data sets, primarily targeting clinical and experimental sources.	Data analysis	
Atlas	Free, publicly available web-based, open-source software application developed by the OHDSI community to support the design and execution of observational analyses to generate real world evidence from patient level observational data.	Data steward research, Researcher	bio.tools, TeSS
DAISY	Data Information System to keep sensitive data inventory and meet GDPR accountability requirement.	Data steward infrastructure, Data steward policy, Human data, Data protection	bio.tools
Data Catalog	Unique collection of project-level metadata from large research initiatives in a diverse range of fields, including clinical, molecular and observational studies. Its aim is to improve the findability of these projects following FAIR data principles.	Documentation and metadata	TeSS
ELIXIR-AAI	The ELIXIR Authentication and Authorisation Infrastructure (AAI)	Sensitive data, NeLS assembly, TSD for sensitive data - Norway	TeSS
FAIR Cookbook	FAIR Cookbook is an online resource for the Life Sciences with recipes that help you to make and keep data Findable, Accessible, Interoperable and Reusable (FAIR)	Compliance monitoring & measurement, Data steward research	

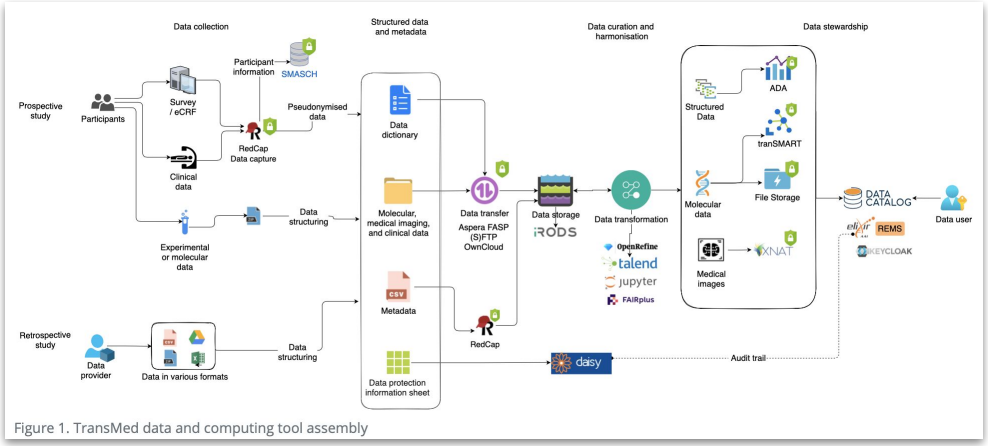


Figure 1. TransMed data and computing tool assembly

5

Searchable list of tools mentioned in guidelines

All tools and resources

Edit me

This is the main tool and resource list of our website. This is a curated list which means that not all tools or resources that exist for a certain topic are listed here. This is mainly because we do not intend to be a registry. In most cases you will only find back the tools or resources that are mentioned in the different pages. Most pages will show a filtered list of this main table at the end of the page. We link to tools and resources to related information in ELIXIR registries: related policies and standards in FAIRsharing, scientific and technical descriptions of the resource in bio.tools, and related training in TeSS. If you see a missing link with one of the registries or a mistake, please open an issue or check our how to add a tool or resource guide.

Show 25 entries Search:

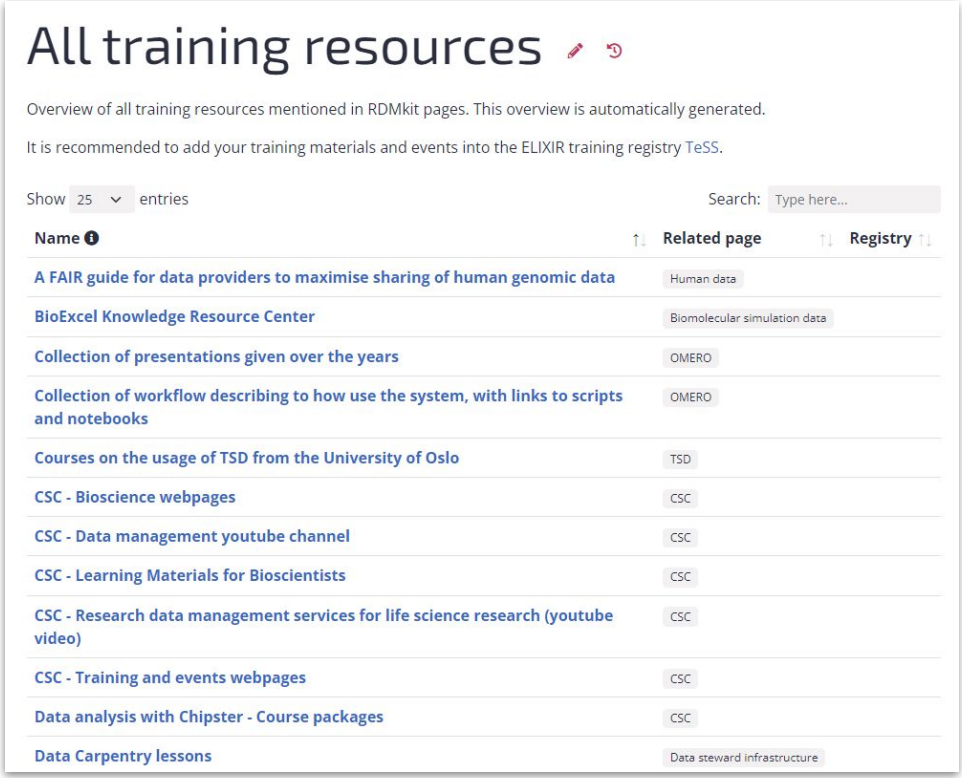
Tool or resource	Description	Related pages	Registry
Choose a license	Choose an open source license	Licensing, Researcher, Data steward research, Data steward policy	TeSS
Common Workflow Language (CWL)	An open standard for describing workflows that are build from command line tools	Data steward infrastructure, Data steward policy, Researcher, Data analysis	FAIRsharing, TeSS
COmputational Modeling in Biology NEtwork (COMBINE)	An initiative to bring together various formats and standard for computational models in biology	Microbial biotechnology	
Conda	Open source package management system	Data steward infrastructure, Data analysis	TeSS
Consent Clauses for Genomic Research	A resource for researchers when drafting consent forms so they can use language matching cutting-edge GA4GH international standards	Human data	
COPO	Portal for scientists to broker more easily rich metadata alongside data to public repos.	Documentation and metadata, Researcher, Plant sciences	bio.tools, FAIRsharing
COVID-19 Molecular Structure and Therapeutics Hub	COVID-19 Molecular Structure and Therapeutics Hub	Biomolecular simulation data	

Links to registries

elixir TeSS FAIRsharing elixir bio.tools

6

Searchable training resources



All training resources

Overview of all training resources mentioned in RDMkit pages. This overview is automatically generated.

It is recommended to add your training materials and events into the ELIXIR training registry TeSS.

Show 25 entries

Search: Type here...

Name	Related page	Registry
A FAIR guide for data providers to maximise sharing of human genomic data	Human data	
BioExcel Knowledge Resource Center	Biomolecular simulation data	
Collection of presentations given over the years	OMERO	
Collection of workflow describing to how use the system, with links to scripts and notebooks	OMERO	
Courses on the usage of TSD from the University of Oslo	TSD	
CSC - Bioscience webpages	CSC	
CSC - Data management youtube channel	CSC	
CSC - Learning Materials for Bioscientists	CSC	
CSC - Research data management services for life science research (youtube video)	CSC	
CSC - Training and events webpages	CSC	
Data analysis with Chipster - Course packages	CSC	
Data Carpentry lessons	Data steward infrastructure	

RDMkit by numbers

All 23 nodes in ELIXIR, national and European projects

78
Pages

10 Domains
15 Tasks
10 Tool Assemblies

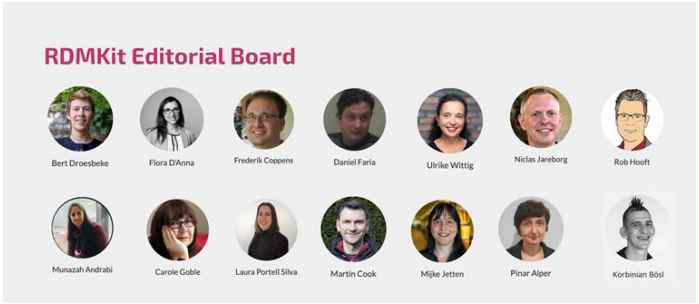
364
Tools and resources

115
Contributors

ELIXIR Data Managers Network

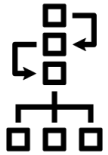
Trainers

Domain experts



RDMkit

A sustainable, open, ongoing community effort



Contribution & Editorial Processes

Simple, sustainable platform

Contentathons & Focus groups

RDMkit: a toolkit that scales and enables

- Support the data needs of Europe's estimated **500,000** life scientists
- A user-oriented guide to the FAIR RDM activities of ELIXIR
- Part of **Horizon Europe** and **ERC** Guidelines

Pool the expertise of the community for the community

Support researchers to know and utilise RDM services

Increase self-sufficiency, build capacity and skills in every institute

First step towards FAIR RDM:
Data Management Planning

Document describing the intended data management strategy for collection, process, store, archiving, publish research data.

It is necessary for:

- Sharing and reusing scientific data
- Guarantee quality of the results
- Justify and optimise the usage of public resources

A DMP:

- Covers all the steps of the research project
- Has to be updated periodically
- Is extended further than the durability of the project

Data Management Plan (DMP)

Content

1. Data summary
- Description,
- origin
- total size
- formats
- usefulness…

2. FAIR data
2.1 Making data findable (metadata)
2.2 Making data openly accessible
2.3 Making data interoperable
2.4 Increase data re-use (licences)

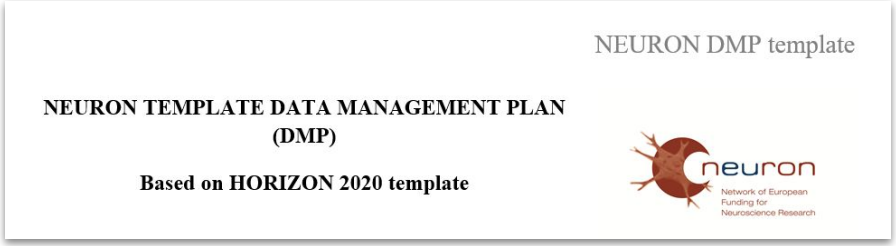
3. Allocation of resources

4. Data security

5. Ethical aspects

6. Other relevant aspects

Format



Online tools

- https://dmptool.org/
- https://dmponline.dcc.ac.uk/
- https://www.sigma2.no/easydmp-user-documentation
- https://ds-wizard.org
- https://dmp.csuc.cat/

Data Management Plan (DMP)

From a simple project, follow the steps to prepare a **DMP**:

1. **Understand** the topic of the project
2. **RDM lifecycle**: relevant datasets to collect, reuse, create
3. **Organization** and **documentation** of the data
4. **Publication** and **preservation** of the results
5. **Resources** needed for the RDM of the project
6. Possible **security** and **legal** issues (confidentiality, licenses, access permissions…)

Data Management Plan content

1. Summary of data

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the **types** and **formats** of data generated/collected
- Specify if existing data is being **reused** (if any)
- Specify the **origin of the data**
- State the **expected size of the data** (if known)
- Outline the data **utility**: to whom will it be useful

- What is the purpose of your project?
- What datasets could be part of this DMP?
- Who is the main responsible to fill in this section?
- Who is the content interesting for?

Data Management Plan content

2.1 FAIR: Finability

METADATA

- Do they describe the data and its characteristics?
- Do you allow automatic search?
- Do they contain relevant information, with meaningful keywords?
- Allows you to distinguish between similar datasets?

Allow people to find the data they are looking for

- No relevant search results
- Can't differentiate between similar datasets

Data Management Plan content

2.2 FAIR: Accessibility

- Where is data and metadata stored, so that they are accessible?
- Is data access labeled correctly?
- Is data available via direct download in a searchable resource?
- Is data available through open, free, and standard protocol?
- Does the access protocol allow for an authentication and authorization procedure if necessary?

Registry of research data repositories: re3data.org

REQUIREMENTS

- **Persistent identifiers** to allow easy and persistent referencing.
- **Metadata is accessible,** even when the data are no longer available.
- If certain **datasets** remain **closed,** the reasons for not giving access should be given.

Easy access to data

- Unclear how to access data
- No direct download option

Data Management Plan content

2.3 FAIR: Interoperability

EMBL-EBI Ontology Lookup Service

https://www.ebi.ac.uk/ols/index

VOCABULARY

- Is data described with standard vocabulary?
- Is the correct ontology being chosen?
- Are standard ontologies correctly linked?
- Is data standard labeled correctly?

- Unclear/ambiguous data description
- No link to referred standard

Easy to interpret data content, inside and outside the discipline

Data Management Plan content

2.4 FAIR: Reusability

Choose tools with built-in license wizards: B2SHARE

- Are data and products licensed?
- Are data formats provided in the metadata?
- Is all relevant online documentation provided?
- Are there links to relevant tools and services provided?
- Is there personal (GDPR) or confidential data that needs special treatment?

- Documentation is readily available
- Data has a license

- Unclear what data contains (documentation lacking or too technical)
- Unclear under what conditions you can reuse the data

Data Management Plan content

3. Allocation of resources
Expenses, resources and responsibilities

4. Data security
Backups, permissions

5. Ethical aspects
Privacy, personal or sensitive data….

6. Other relevant aspects
Case-specific DM issues

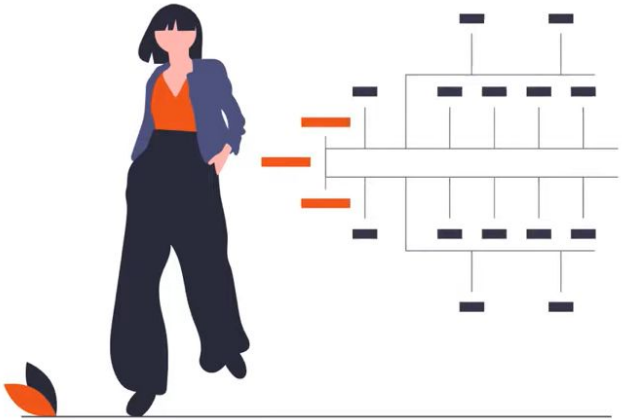
Data Stewardship Wizard

- A tool for generating data management plans in various formats.
- Using smart questionnaires to guide researchers.
- Customizable questionnaire templates and document export formats.
- Online collaboration.

Data Stewardship Wizard

Knowledge Model

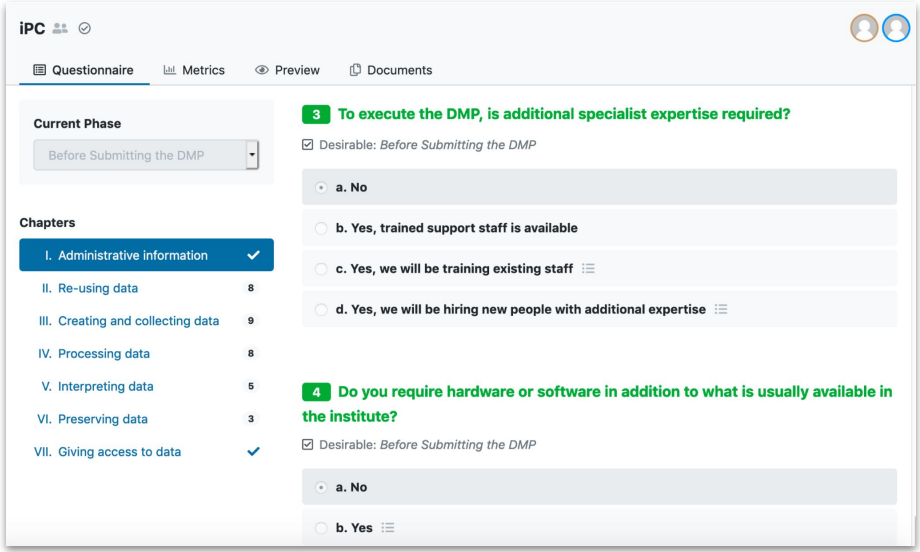
- Contains the **knowledge** about what should be asked and how.
- **Tree-like structure** of **Chapters, Questions, Answer** and additional resources.
- Use of **template** for the questionnaire.
- Provided by DSW or institution data stewards.



Data Stewardship Wizard

Questionnaire

- Interactive form how to get the answers from users based on the Knowledge Model.
- Contains report on the FAIR metrics.
- Can be exported to a human-readable (e.g., PDF) or machine-actionable (e.g., maDMP) document.



Data Stewardship Wizard

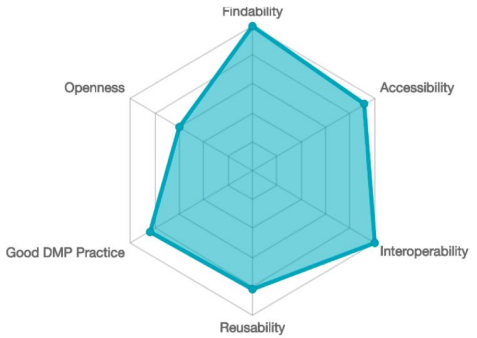
FAIR metrics report example

Summary Report

Answered (current phase): 130/163

Answered: 167/218

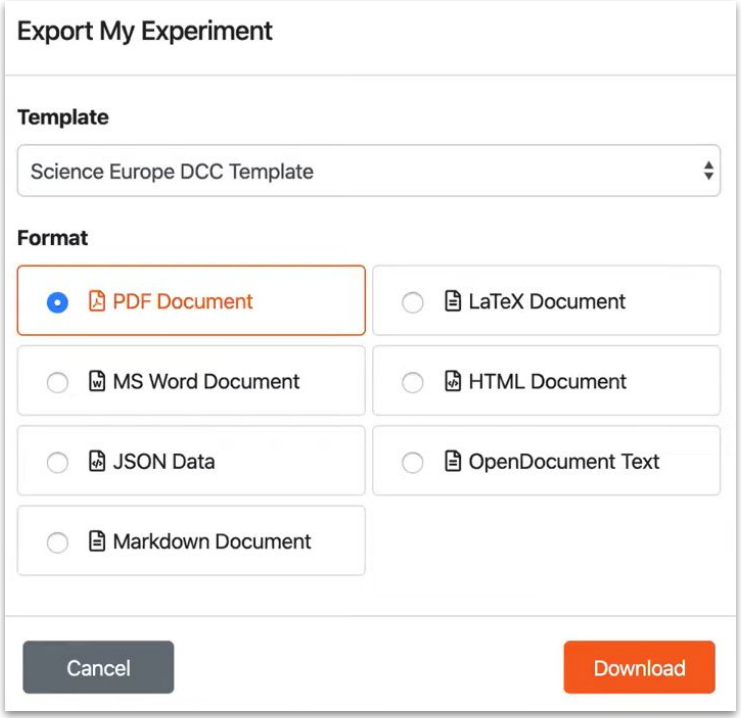
Metric	Measure
Findability	1.00
Accessibility	0.92
Interoperability	1.00
Reusability	0.82
Good DMP Practice	0.84
Openness	0.60



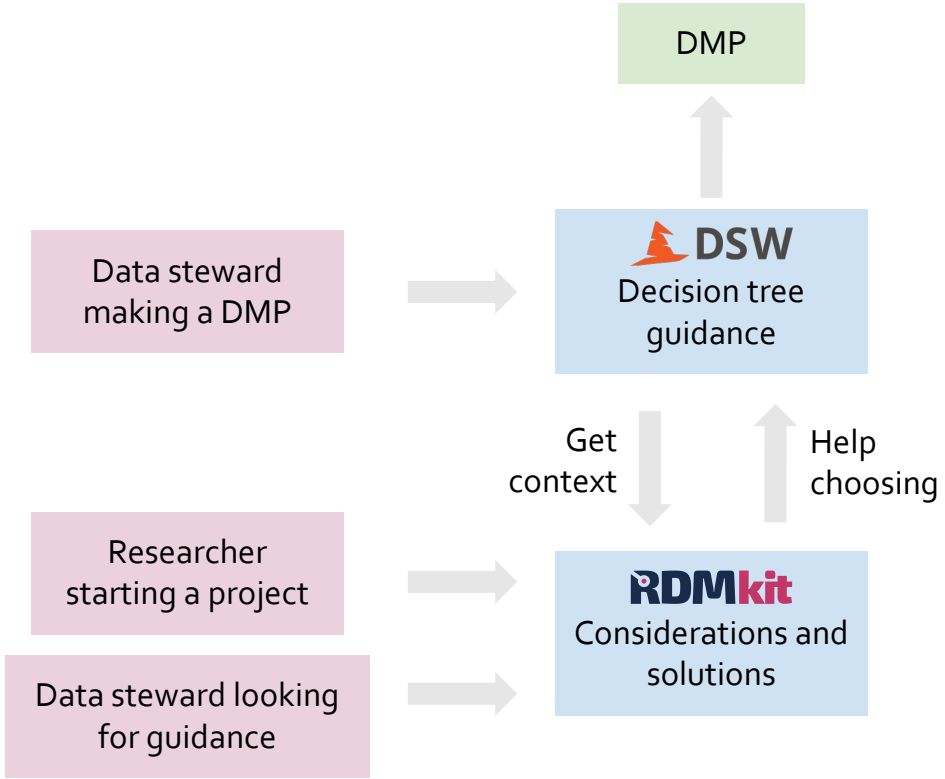
Data Stewardship Wizard

Exporting the DMP

- Documents can be generated from a filled questionnaire.
- Customizable document templates.
- Various document formats.
- DSW provided templates:
    - **Science Europe**
    - **Horizon 2020**
    - **maDMP**



Connection between the RDMkit and DSW



Thank you!

RDMkit

https://rdmkit.elixir-europe.org/

DSW

https://ds-wizard.org/get-started