

Data Management Plans

Tomasz Miksa

tomasz.miksa@tuwien.ac.at

Agenda

Why do we need to manage data properly?

What are Data Management Plans (DMPs)?

How to create a DMP?

Researchers trying to reuse data...

Conversation of two researchers

- Can I see your data?
- It's on my USB stick
- Can I have it?
- I have in a box and I have moved recently
- Can I have it?
- I forgot to label the boxes...
- (half a year later)
- Thanks, for the USB. However, I cannot read the hexadecimal file on it. How do I open it?
- You need a special program
- What program?
- ...



Hanson, Karen; Surkis, Alisa; Yacobucci, Karen: Data Sharing and Management Snafu in 3 Short Acts.
<https://doi.org/10.5446/31036>

Variety of solutions

In response to these needs many solutions were proposed and are being implemented

- **FAIR principles**
- **open access** to scientific publications and data
- research **data repositories** to host the data
- **persistent identifiers** to locate the data
- **data management plans**
- ...

WHAT IS A DATA MANAGEMENT PLAN (DMP)?

Data Management Plan

DMP is a formal document

It outlines what you will do with your data **during** and **after** you complete your research

It ensures your data is safe for the **present** and the **future**

[from University of Virginia Library]



DMP is an awareness tool!

DMP makes you think

- what data you will use and where you get it from
- what infrastructure, software, licenses are needed
- what will be the output of your research
- how you will share your research outputs

DMP helps you organise yourself better

DMP can reveal how solid your methodology is

- is it a 'fishing expedition'?




DMPs are used worldwide

- Required by
 - research funders
 - institutions, e.g. universities

Research Data Management

General Information

Research data management is an integral part of good research practice (see » [Research Integrity & Research Ethics](#)). The FWF therefore requires a data management plan (DMP) for all projects approved as of 1 January 2019. A DMP describes how data and their metadata are collected, organised, stored, published, shared, and archived for a specific project. Furthermore, the DMP outlines how the data will be made  FAIR, which means Findable, Accessible, Interoperable and Reusable. The » [FWF's Open Access Policy to Research Data](#) must be taken into account when drafting the DMP.

<https://www.fwf.ac.at/en/research-funding/open-access-policy/research-data-management>



Example: Projects funded by European Commission

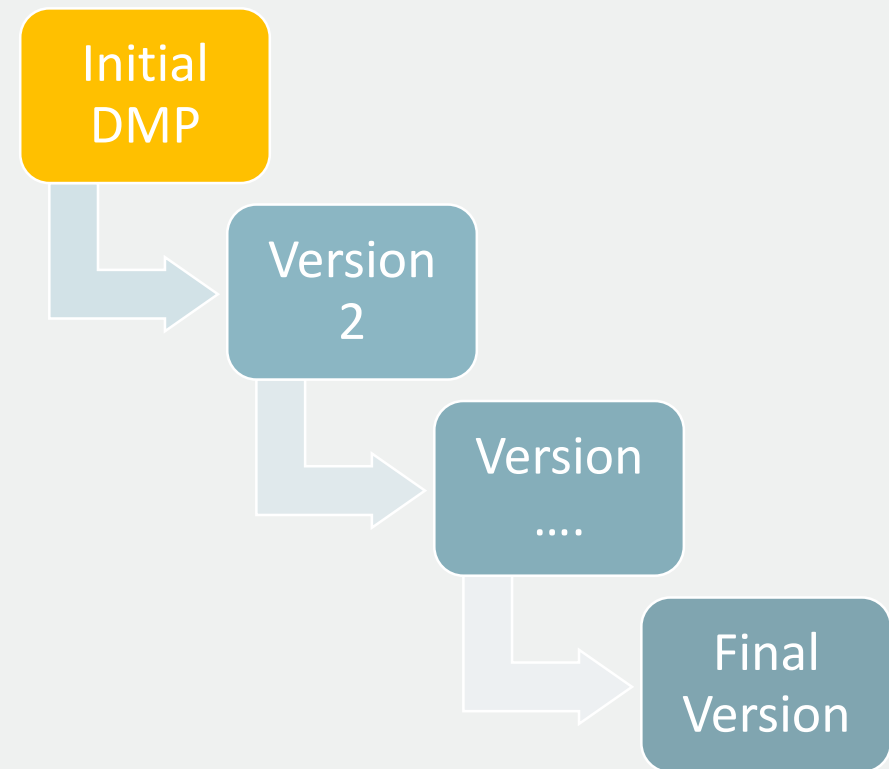
DMP is a living document

First version

- within the first 6 months

Updated versions

- when significant changes occur
 - new datasets
 - changes in policies
- periodic reporting
 - project reviews
- end of project



HOW TO CREATE A DMP?

How to create a DMP?

Most cases by

- filling out a template
- answering questions from a checklist

Using software tools

- users choose appropriate funders template
- only relevant questions and guidance is presented

Science Europe Guidelines

Basis for many funder templates



Table of Contents

Foreword by Dr Thierry Damerval	2
Introduction	4

GUIDANCE FOR ORGANISATIONS: CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS	7
---	---

GUIDANCE FOR ORGANISATIONS: CRITERIA FOR THE SELECTION OF TRUSTWORTHY REPOSITORIES	11
--	----

GUIDANCE FOR RESEARCHERS: Translating the Core Requirements into a DMP template Guiding the Selection of Trustworthy Repositories	15
---	----

GUIDANCE FOR REVIEWERS: Evaluation Rubric for Data Management Plans	31
--	----

Notes and References	51
Annex: Compatibility with the FAIR Data Principles	52

4 For procedural elements of implementing DMPs, see the RDA DMP Common Standards Working Group: <https://www.rd-alliance.org/groups/dmp-common-standards-wg>



FWF Example

Based on the SE requirements

I General Information		
I.1 Administrative information	Provide information such as name of principal investigator, FWF project number, and version of DMP	<ul style="list-style-type: none"> - Provide the relevant grant information. - Consider regular updates of the DMP.
I.2 Data management responsibilities and resources	<p>Who (for example, role, position, and institution) will be responsible for data management?</p> <p>What resources will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?</p>	<ul style="list-style-type: none"> - Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised. - For collaborative projects, explain the co-ordination of data management responsibilities across partners. - Explain how the necessary resources (for example, time) to prepare the data for sharing/preservation have been costed in. Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, and repository charges.
II Data Characteristics		
II.1 Data description and collection or re-use of existing data	<p>How will new data be collected or produced and/or how will existing data be re-used?</p> <p>What data (types, formats, and volumes) will be collected or produced?</p>	<ul style="list-style-type: none"> - Explain which methodologies or software will be used if new data are collected or produced. - State any constraints on re-use of existing data if there are any. - Explain how data provenance will be documented. - Give details on the kind of data: for example, numeric (databases), textual (documents), image, audio, or video. - Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example, pdf, xls, doc, txt, or rdf).

DMP tools



DMP Online

- <https://dmponline.dcc.ac.uk/>

Data Stewardship Wizard

- <https://ds-wizard.org>

Argos

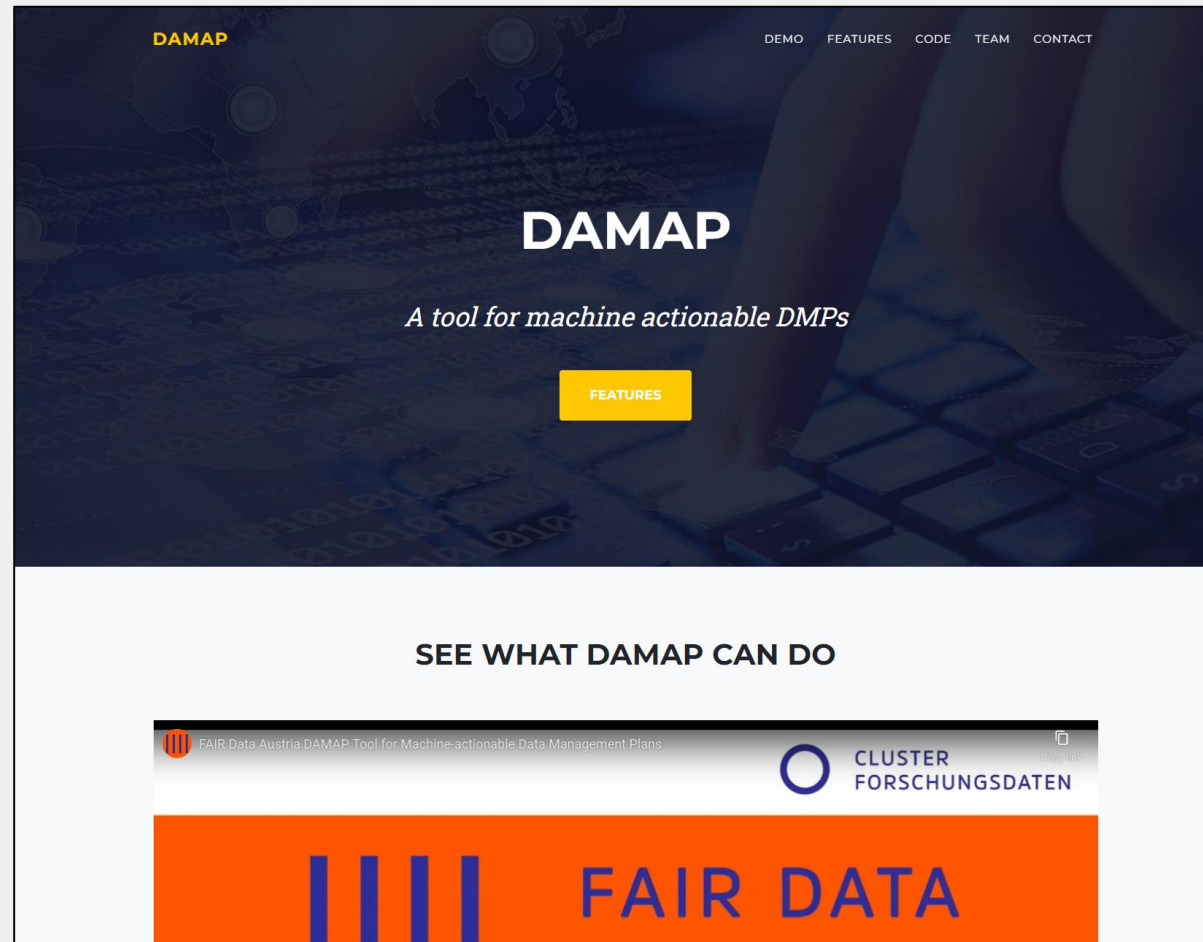
- <https://argos.openaire.eu/splash/>

RDMO

- <https://rdmorganiser.github.io/en/>

A screenshot of the DMP ONLINE web application. The interface is primarily orange and grey. At the top right, it says 'Signed in as miksa@ifs.tuwien.ac.at'. The main header includes the 'DMP ONLINE' logo and navigation links: 'View plans', 'Create plan', 'About', 'Future plans', 'Help', and 'Change language'. The current page is titled 'FFG Webinar Horizon 2020 Example' and shows '0/71 questions answered'. Below the title are tabs for 'Plan details', 'Initial DMP', 'Detailed DMP', 'Final review DMP', 'Share', and 'Export'. The 'Initial DMP' tab is active, showing a progress bar with two sections: '1. Data summary (1 question, 0 answered)' and '2. FAIR data (4 questions, 0 answered)'. The '2. FAIR data' section is expanded, displaying the text: 'In general terms, your research data should be 'FAIR' that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.' Below this is a sub-section '2.1 Making data findable, including provisions for metadata:' with a bulleted list of requirements. A text editor with a toolbar (bold, italic, list, link, grid) is visible below the list. On the right side, there is a 'Guidance' panel with a 'Share note' button and a box containing 'EC Guidance' text: 'The Research Data Alliance provides a Metadata Standards Directory that can be searched for discipline-specific standards and associated tools.' A 'Save' button is located at the bottom left of the text editor area.

DMP tool at TUW



(and what I should also do!)

WHAT SHOULD I WRITE IN FACT?

FWF Template – running example

I. General information

II. Data Characteristics

III. Documentation and Data Quality

IV. Data Storage, Sharing and Long-Term Preservation

V. Legal and Ethical Aspects

This lecture: not an exhaustive walk-through! Only interesting/relevant aspects!

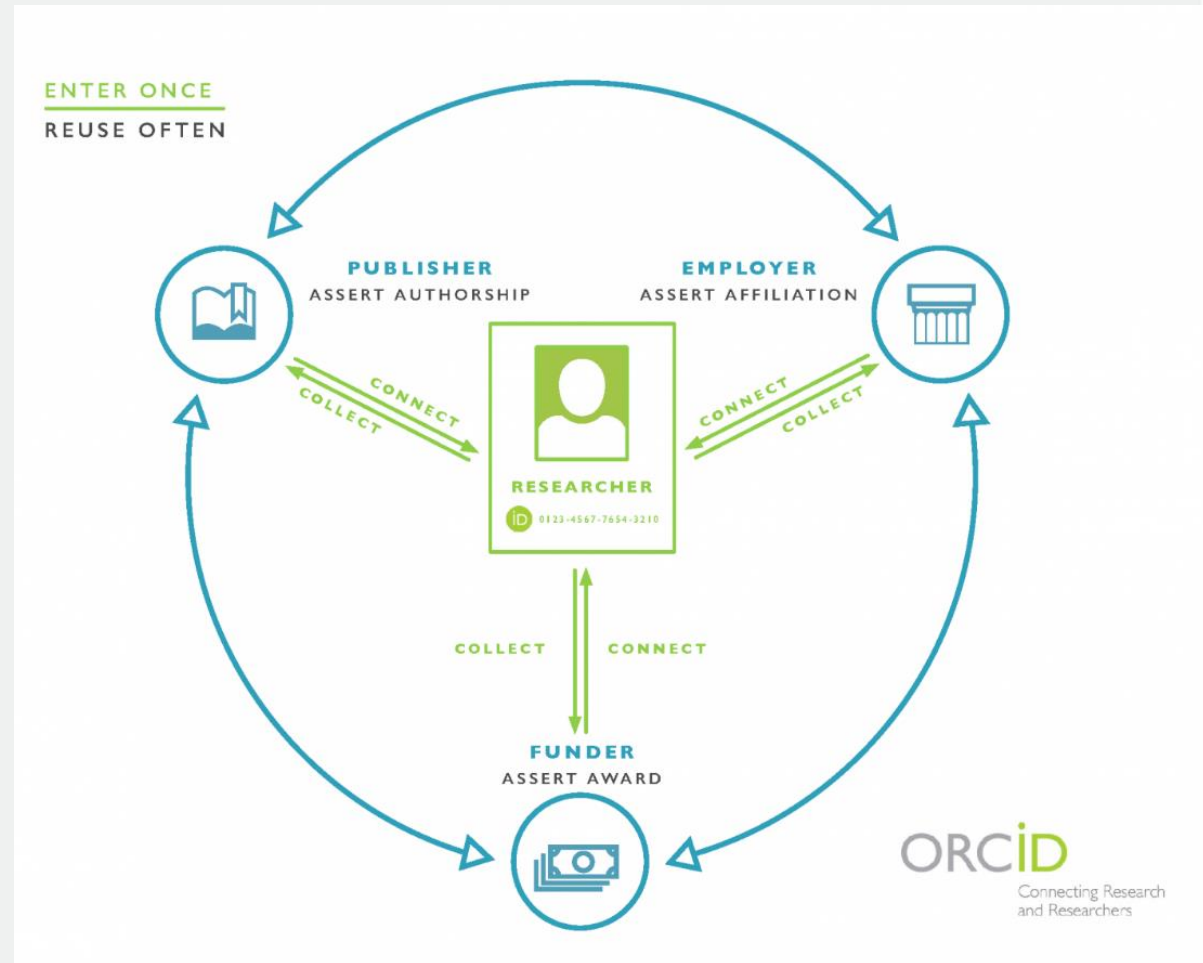
I General Information	
I.1 Administrative information	
I.2 Data management responsibilities and resources	
II Data Characteristics	
II.1 Data description and collection or re-use of existing data	
III Documentation and Data Quality	
III.1 Metadata and documentation	
III.2 Data quality control	
IV Data Storage, Sharing, and Long-Term Preservation	
IV.1 Data storage and backup during the research process	
IV.2 Data sharing and long-term preservation	
V Legal and Ethical Aspects	
V.1 Legal aspects	
V.2 Ethical aspects	

I. GENERAL INFORMATION

ORCID – persistent identifier for people

ORCID ID

- Unique person ID
- ORCID assigned once
- Person can change affiliations (jobs)
- Example: 0000-0002-4929-7875



Search English

ORCID EDIT YOUR RECORD ABOUT ORCID CONTACT US HELP

Connecting Research and Researchers

4,115,029 ORCID iDs and counting. [See more...](#)

Daniel Mietchen

ORCID ID
<https://orcid.org/0000-0001-9488-1870>

[Print view](#)

Also known as
 D. Mietchen, Mietchen, Daniel, Mietchen, D., EvoMRI, D Mietchen, Mietchen D, Mietchen-D

Country
 Germany

Keywords
 open science, open data, open access, magnetic resonance microscopy, evolution, biodiversity, social machines, vocal learning

Websites
[Twitter](#)
[Wikidata](#), [Wikipedia et al.](#)
[GitHub](#)
[Open Science Q & A](#)
[Scholia](#)

Other IDs
[Scopus Author ID: 7801384320](#)
[ResearcherID: A-7748-2009](#)

Employment (2) Sort

National Center for Biotechnology Information: Bethesda, MD, United States
 2015-03-01 to present | Intramural researcher (Computational Biology Branch)
 Source: Daniel Mietchen

Museum für Naturkunde - Leibniz-Institut für Evolutions- und Biodiversitätsforschung: Berlin, Berlin, Germany
 2013-08-16 to 2015-02-28 | Researcher (Digital World)
 Source: Daniel Mietchen

Works (64) Sort

Machine-actionable data management plans (maDMPs)
 Research Ideas and Outcomes
 2017-04-05 | journal-article
 DOI: [10.3897/rio.3.e13086](#)
 Source: CrossRef Metadata Search Preferred source

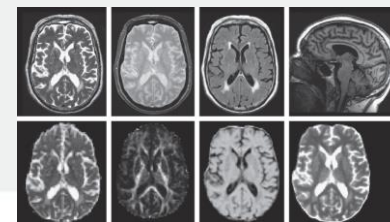
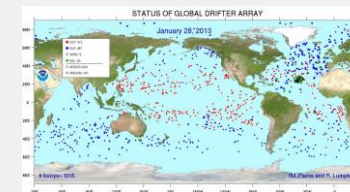
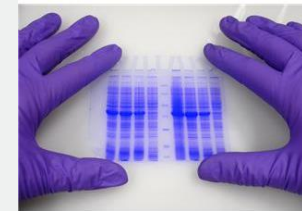
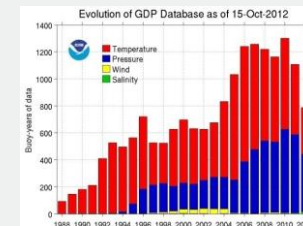
Progress in promoting data sharing in public health emergencies
 Bulletin of the World Health Organization
 2017-04-01 | journal-article
 DOI: [10.2471/blt.17.192096](#)
 Source: CrossRef Metadata Search Preferred source

Strategies and guidelines for scholarly publishing of biodiversity data

II. DATA CHARACTERISTICS

What is data?

- Instrument measurements
- Experimental observations
- Still images, video and audio
- Text documents, spreadsheets, databases
- Quantitative data (e.g. survey data)
- Survey results & interview transcripts
- Simulation data, models & software
- Slides, artefacts, specimens, samples
- Questionnaires
- Sketches, diaries, lab notebooks ...



Data Summary

Type

- text, spreadsheets, software, models, images, movies, audio, patient records, etc.

Source

- human observation, laboratory, field instruments, experiments, simulations, compilations, etc.

Volume

- total volume of data, number of files, etc.

Data and file formats

- non-proprietary formats
- used within community

III. DOCUMENTATION AND DATA QUALITY

What is in the picture?



Metadata – Atlas Of Living


Atlas Of Living Australia ALA Apps ALA Info Search the Atlas Search

NatureShare - 2380_Gymnorhina_tibicen
HumanObservation of *Cracticus tibicen* | Australian Magpie recorded on 2011-04-17T12:32:00+1000


Flag an issue Contact curator

Dataset
Event
Taxonomy
Geospatial
Images
Data quality tests (1 4 21 13 48)
Additional political boundaries information
Environmental sampling for this location

Location of record



Images



Photographer: Russell Best

Dataset

Data resource	NatureShare
Catalogue number	2380_Gymnorhina_tibicen
Basis of record	Human observation
Observer	Best, R. Russell <i>Supplied as 'Russell Best'</i>
Rights	CC BY 2.5 AU
More details	http://natureshare.org.au/observation/2380/
Photographer	Russell Best
Rights holder	Russell Best via NatureShare
Occurrence remarks	Tags: Female
Occurrence status	present
Abcd identification qualifier	Not provided

Event

Record date	[date not supplied] <i>Supplied date '2011-04-17T12:32:00+1000'</i>
Event remarks	Photo date/time used.

Taxonomy

Scientific name	<i>Cracticus tibicen</i> <i>Supplied scientific name 'Gymnorhina tibicen'</i>
Taxon rank	Species
Common name	Australian Magpie
Kingdom	Animalia
Phylum	Chordata
Class	Aves
Order	Passeriformes
Family	Artamidae
Genus	<i>Cracticus</i>
Species	<i>Cracticus tibicen</i>

Metadata – Atlas Of Living Australia

Dataset

Data resource	NatureShare
Catalogue number	2380_Gymnorhina_tibicen
Basis of record	Human observation
Observer	Best, R. Russell <i>Supplied as "Russell Best"</i>
Rights	CC BY 2.5 AU
More details	http://natureshare.org.au/observation/2380/
Photographer	Russell Best
Rightsholder	Russell Best via NatureShare
Occurrence remarks	Tags: Female
Occurrence status	present
Abcd identification qualifier	Not provided

Metadata – Atlas Of Living Australia

Event

Record date	[date not supplied] <i>Supplied date "2011-04-17T12:32:00+1000"</i>
Event remarks	Photo date/time used.

Taxonomy

Scientific name	<i>Cracticus tibicen</i> <i>Supplied scientific name "Gymnorhina tibicen"</i>
Taxon rank	Species
Common name	Australian Magpie
Kingdom	Animalia
Phylum	Chordata
Class	Aves
Order	Passeriformes
Family	Artamidae
Genus	<i>Cracticus</i>
Species	<i>Cracticus tibicen</i>
Taxonomic issues	No issues
Name match metric	Exact match The supplied name matched the name exactly.

Metadata – Atlas Of Living Australia

Geospatial

Country	Australia
State or territory	Victoria
Local government area	Macedon Ranges (S)
Latitude	-37.421078
Longitude	144.61954
Geodetic datum	EPSG:4326
Biome	Terrestrial
Verbatim longitude	144.619541
Verbatim latitude	-37.421077

Location of record



Standards and metadata

Metadata

- helps to understand and interpret data
- provides details about experiment setup
 - who, when, in which conditions, tools, versions, etc.
- helps identify and discover new data

Use community standards to enable interoperability

<http://www.dcc.ac.uk/resources/metadata-standards>

IV. DATA STORAGE, SHARING AND LONG-TERM PRESERVATION

Managing data during research

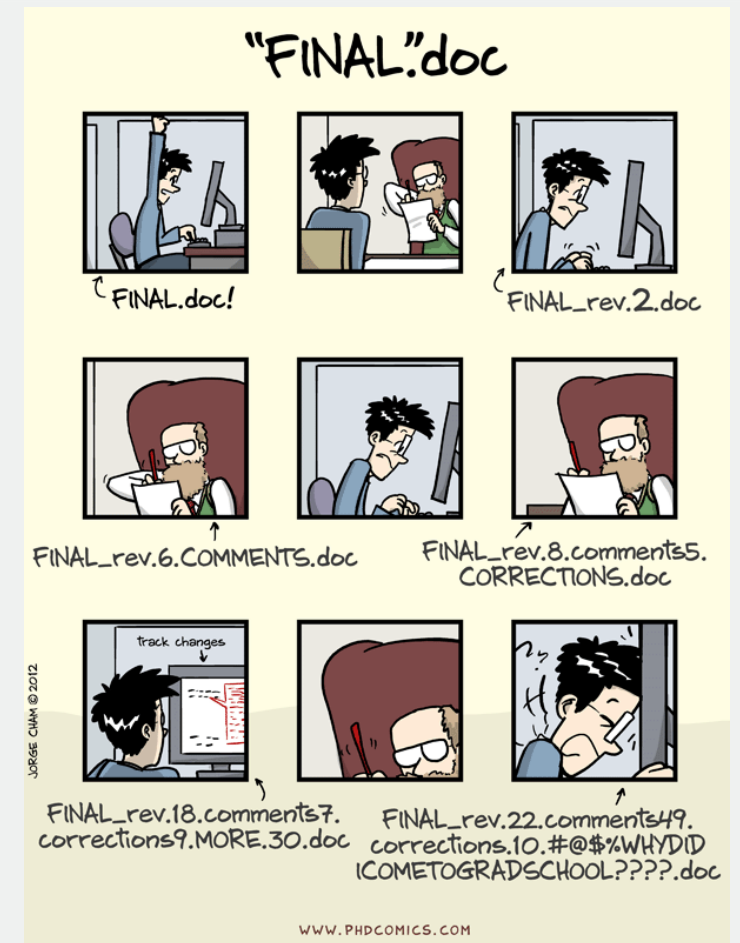
If you loose your data there will be nothing to share!

Recreating or recollecting data can be

- impossible
 - e.g. observational data
- too expensive
 - e.g. cost of computational power

How do you manage data during the project?

- file naming convention
- versioning
- backups
- should the access be restricted?
- who is responsible?

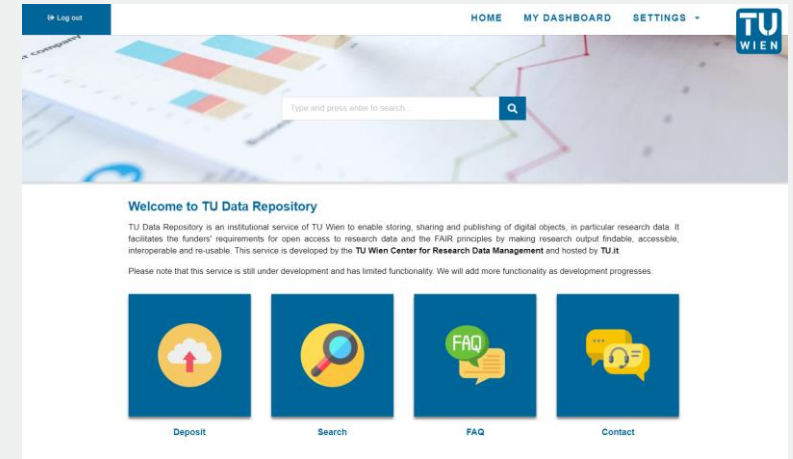


Backup vs archiving and preservation (traditional view)



Data managed during the project

- Changed/deleted
- Backup



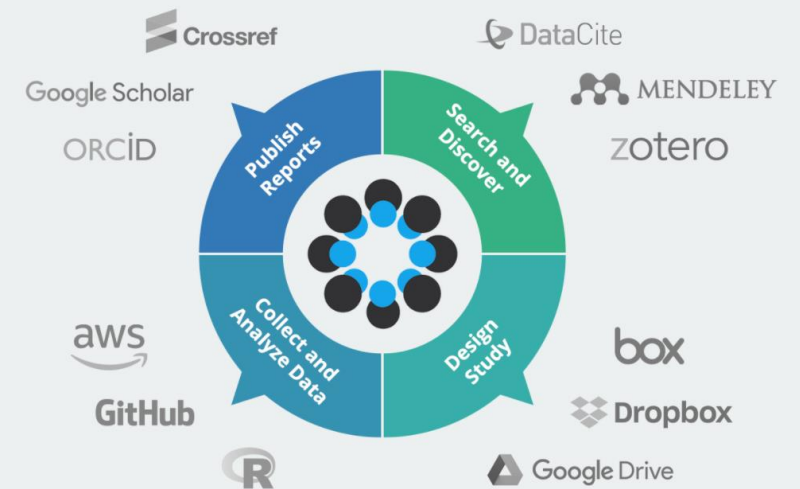
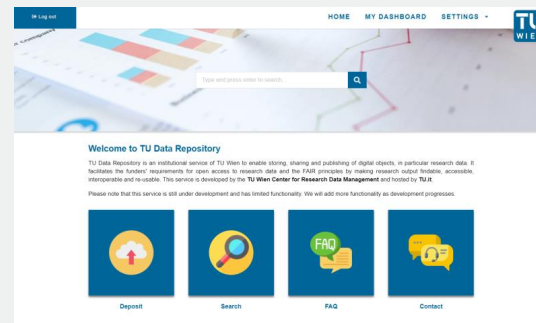
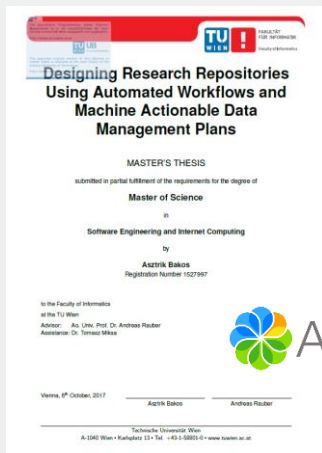
Stable snapshot of data

- Moved into a repository
- Enriched with metadata and licensing details
- Not only backups

What makes a system a repository?

Backup vs archiving and preservation (new approaches)

- No need to differentiate between project and post-project phases
- One system can be used for managing and preserving data



<https://www.cos.io/products/osf>

Archiving and preservation

Which data will be shared?

- What has to be kept?
- What can't be recreated?
- What is potentially useful to others?
- What legally must be destroyed?

Where will the data be deposited?

- not all of the data must be shared in the same way

Are there any embargo periods?

For how long?

What is the cost and who will pay for it?

Which license to use?

Where to find a repository?

1. Use Domain specific repository

- e.g. chEMBL (if you work with molecules)

2. Use Institutional repository

- e.g. researchdata.tuwien.ac.at (if you work at TU Wien)

3. Search registry to find a relevant one

- e.g. re3data.org

4. Use *catch-all* repository

- e.g. zenodo.org

re3data.org


Repository details


TU Data

General Institutions Terms Standards

Name of repository	TU Data
Repository URL	https://researchdata.tuwien.ac.at/
Description	TU Data is an institutional repository of TU Wien to enable storing, sharing and publishing of digital objects, in particular research data. It facilitates the funders' requirements for open access to research data and the FAIR principles by making research output findable, accessible, interoperable and re-usable. This service is developed by the TU Wien Center for Research Data Management and hosted by TU.it.
Content type(s)	Standard office documents Archived data
Keyword(s)	FAIR interdisciplinary
Repository type(s)	institutional
Mission statement for designated community	https://researchdata.tuwien.ac.at/
Research data repository language(s)	English
Data and/or service provider	service provider data provider

[Back to search](#) [Submit a change request](#) [Get a badge](#)

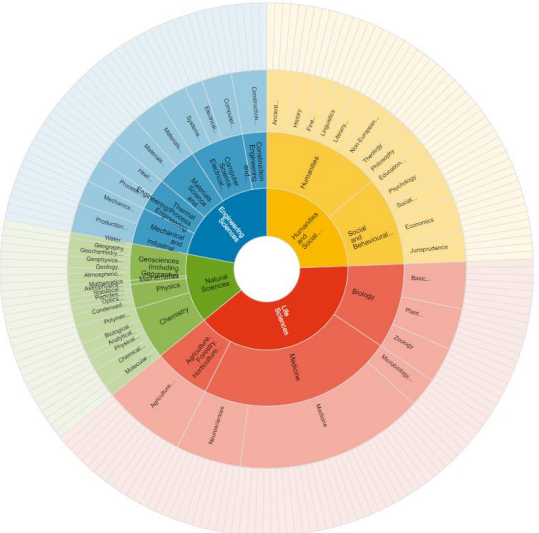
 Cite this re3data.org record:
re3data.org: TU Data; editing status 2021-06-02; re3data.org - Registry of Research Data Repositories.
<http://doi.org/10.17616/R31NJMYD> last accessed: 2022-03-25

re3data.org Search Browse Suggest Resources Contact 

Browse by subject

Graphical Text

click to zoom into subjects or to select a bottommost subject in the hierarchy as filter for the re3data search page
ctrl + click on a top subject to select it as filter



V. LEGAL AND ETHICAL ASPECTS

Creative Commons

CC-BY (Attribution)

- allows anyone to use, re-use, and remix a work without restriction, also commercially
- You must give appropriate credit, provide a link to the license, and indicate if changes were made.



CC BY-SA (Attribution – ShareAlike)

- all new works must carry the same license



CC BY-ND (Attribution- NoDerivs)

- reuse, but no changes



CC BY-NC

- no commercial use

CC BY-NC-SA

CC BY-NC-ND



Software Licenses

Choose correct license for your software

- Apache, MIT, GNU, BSD, ...

Check licenses of libraries you reuse in your software

- Example: GNU GPL vs GNU LGPL
 - GPL enforces the reusing software to be GPL (also public)
 - LGPL code must be clearly marked, rest of the software can have different license (can be private)

Software licenses can also be used for data

Choose an open source license

Which of the following best describes your situation?

- I want it simple and permissive.**
The MIT License is a permissive license that is short and to the point. It lets people do anything they want with your code as long as they provide attribution back to you and don't hold you liable.
Babel, .NET Core, and Rails use the MIT License.
- I'm concerned about patents.**
The Apache License 2.0 is a permissive license similar to the MIT License, but also provides an express grant of patent rights from contributors to users.
Elasticsearch, Kubernetes, and Swift use the Apache License 2.0.
- I care about sharing improvements.**
The GNU GPLv3 is a copyleft license that requires anyone who distributes your code or a derivative work to make the source available under the same terms, and also provides an express grant of patent rights from contributors to users.
Ansible, Bash, and GIMP use the GNU GPLv3.

What if none of these work for me?

- My project isn't software.**
There are licenses for that.
- I want more choices.**
More licenses are available.
- I don't want to choose a license.**
You don't have to.

The content of this site is licensed under the Creative Commons Attribution 3.0 Unported License. About Terms of Service Curated with ❤️ by GitHub, Inc. and You!

<https://choosealicense.com>

SUMMARY

Tips for writing DMPs

DMP can reveal how solid your work is

Seek advice - consult and collaborate

When answering questions from checklists write coherent text

Be specific when referring to tools and standards

Assign responsibilities and name responsible personnel

Welcome to the Center for Research Data Management

We are a team of TU Wien employees with different professional backgrounds and support you as a researcher and project driver at TU Wien in handling research data along the entire data life cycle.

Many issues can be solved easily by phone or email, but we also offer personal talks and information events tailored to your needs.

For a quick overview of RDM at TU Wien, we recommend our Quick Start Guide.



Consulting



Training



Data practice



Team



News



RDM infos & tips

Contact us

Write to us or give us a call:
research.data@tuwien.ac.at
Phone: +43 1 58801 44002

Twitter: [@RDMTUWien](https://twitter.com/RDMTUWien)

Center for Research Data Management
Favoritenstraße 16 (top floor), 1040 Vienna



Technical RDM services & tools

of TU Wien: data repository, DMP-Tool, GitLab instance