



L-Università
ta' Malta



Research Data Management, drUM, & Responsible Data Sharing

Pavol Macejovsky & Steve Micallef
University of Malta Library
Open Science

L-Università ta' Malta



Roles & Responsibilities Research Data Management (RDM)

Researchers & research communities

- Incorporate best practices for research data
- Deposit data in **trusted repositories**
- Adhere to relevant policies and standards

Research institutions

- Develop policies & promote standards
- **Provide infrastructure, training, & support**
- Ethics & data protection governance
- Compliance & assurance
- Reward researchers for sharing their data

Research funders & publishers

- Set requirements
- Monitor compliance
- Encourage reuse/impact
- Reward researchers for sharing their data

Outline

- **Why Research data management (RDM) matters**
- **Real-world cases**
- **Fair data principles**
- **Research data life cycle:**
 - **Describing relevant stages**
- **Best practises for organising and documenting research data**
- **Data storage and security**
- **Tools for data anonymization and encryption**
- **The importance of archiving research data**
- **UM archiving of data**
- **Understand the drUM workflows**
- **AI in RDM**
- **Depositing data in drUM**



Why RDM matters

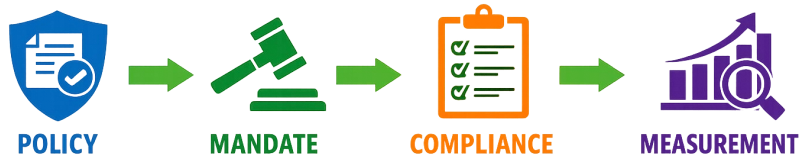


L-Università
ta' Malta

POLICY

Research Data Management Policy

POL-002-01





Today's Data, Tomorrow's Discoveries

Increasing Access to the Results of Research Funded by the
National Science Foundation

National Science Foundation



European Research Council
Established by the European Commission



**UK Research
and Innovation**



**SCIENCE
EUROPE**
Shaping the future of research





Proper Research Data Management (RDM) is mandatory for any Horizon Europe project generating or reusing research data. It is a key part of Horizon Europe's open science requirements.

The FAIR principles

Developing a Data Management Plan (DMP)

Providing access to research data in trusted repositories

Validation (and re-use) requirements

Costs of Research Data Management

In Horizon Europe, *beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the **FAIR principles***, and should at least do the following:

- Prepare a Data Management Plan (DMP) and keep it updated throughout the course of the project
- Deposit data in a trusted repository and provide open access to it ('as open as possible, as closed as necessary')
- Provide information (via the same repository) about any research output or any other tools and instruments needed to re-use or validate the data

Keep in mind that 'research data' is a very broad concept and certainly not limited to numerical/tabular data.



Data availability statements

Data availability statements are required for all articles published with MDPI. During the peer review and editorial decision process, authors can be asked to share existing datasets or raw data that have been analyzed in the manuscript, and whether they will be made available to other researchers following publication. Authors will also be asked for the details of any existing datasets that have been analyzed in the manuscript.

Below are the recommended Data Availability Statements:

Data availability status

Data available in a publicly accessible repository

Recommended Data Availability Statement

The original data presented in the study are openly available in [repository name, e.g., FigShare] at [DOI/URL] or [reference/accession number].

[Editorial policies](#)[Authorship](#)[Competing interests](#)[Research Ethics](#)

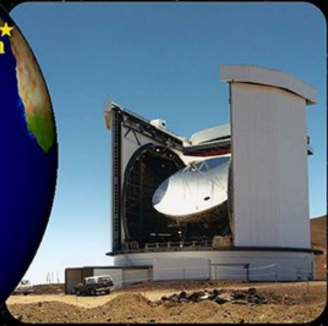
Reporting standards and availability of data, materials, code and protocols

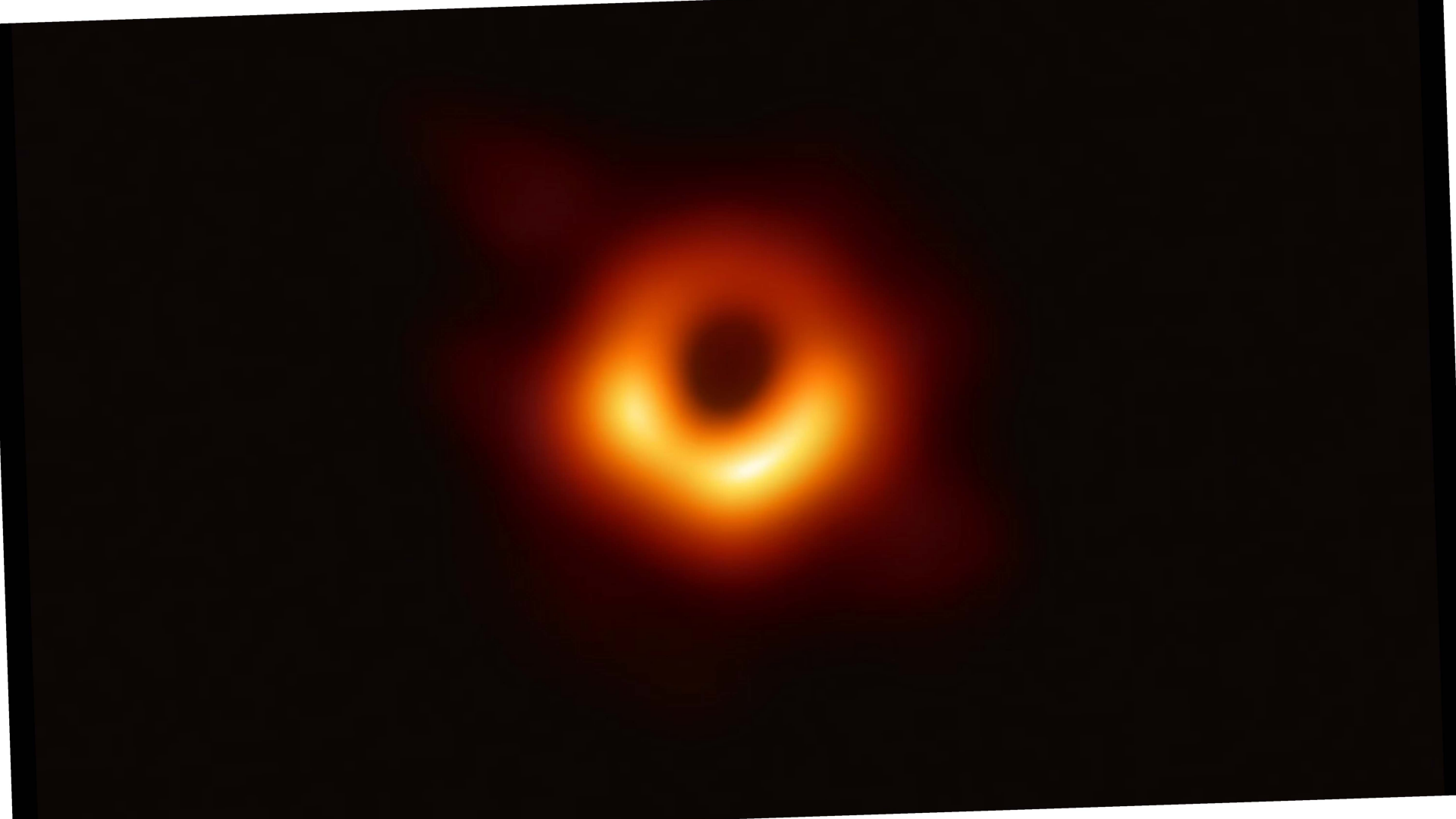
[Image integrity and standards](#)[Plagiarism and duplicate publication](#)[Corrections, Retractions and Matters Arising](#)[Peer Review](#)[Confidentiality](#)[Acknowledgements](#)[Preprints & Conference Proceedings](#)[Press and embargo policies](#)

Reporting standards and availability of data, materials, code and protocols

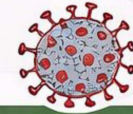
An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature Portfolio journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications.** Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must also be disclosed in the submitted manuscript.

After publication, readers who encounter refusal by the authors to comply with these policies should contact the chief editor of the journal. In cases where editors are unable to resolve a complaint, the journal may refer the matter to the authors' funding institution and/or publish a formal statement of correction, attached online to the publication, stating that readers have been unable to obtain necessary materials to replicate the findings.





COVID-19 Research and the Speed of Vaccine Development



Global Data Sharing



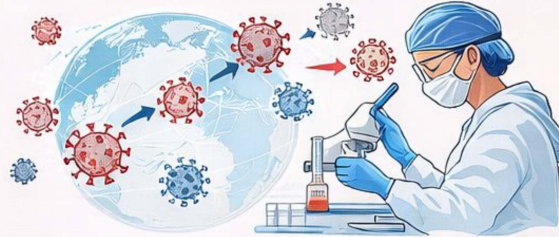
- Genome Sequences of the SARS-CoV-2 Virus, Clinical Trial Data, Epidemic Models



GenBank **GISAID**



Data Reuse for Vaccine Development



- Track Mutations
- Identify Variants
- Understand Transmission

moderna

Pfizer **BioNTech**

Reuse of Knowledge and Technology

Rapid Vaccine Development



- › Vaccines Developed

Impact of Open Data



- › Quick Scientific Response
- › Identification of COVID-19 Variants
- › Accelerated Global Health Crisis Response

BENEFITS & DEMAND



Citation visibility

Recent large-scale evidence finds a small but measurable **citation** association (~4%) for sharing data via repositories.

(Colavizza et al., 2024)



Researcher demand

>85% report they would share/use others' data if access were easy (survey).

(Tenopir et al., 2020)



COSTS & RISKS



"Not FAIR" data cost

EU analysis estimates **≥€10.2bn/year** cost for "Not FAIR" data.

(European Commission / PwC EU Services, 2019)



Lack of support

More than **60%** researchers receive **no help** to share data openly.

(State of Open Data 2023)



Data decay

Data availability decays over time (~**2.6%/year** drop in odds of availability).

(Briney, 2024)



F



A



I



R



FAIR “cut-ups” matching game

Make (Meta)data findable



Rich Descriptions

Use **clear titles**, meaningful **keywords**, and **funder/grant info**. Include methods, variables, and instruments.



Add Linkages

Connect to related **publication DOI(s)**, project pages, and code/software repositories.



Identity & Versioning

Include **versioning** details along with complete contact and affiliation information.



Harvestability

Metadata is harvestable via **OAI-PMH**, which significantly improves global discoverability.

Make (Meta)data accessible



Define Access Levels

Classify data as **Open, Restricted, or Embargoed**. Metadata should be visible automatically, even if files are restricted.



Access Procedures

Clearly state **how to access** restricted data, including necessary conditions, approval processes, and requirements.



Open Metadata Discovery

Ensure **metadata is openly accessible** even for restricted files to support global discovery and compliance.

Make data interoperable



Open Formats & Standards

Prefer **open / non-proprietary formats** and **widely used standards** (e.g., CSV, XML, GeoTIFF, NifTI, geoJSON).



Standard Schemas & Vocabularies

Use **metadata standards**, consistent field names, and controlled terms to ensure disciplinary alignment.



Consistent Encoding

Maintain consistent **encoding and conventions**: ISO dates, standard units, and GIS reference systems.

Make data reusable



Documentation Basics

Add a **README + codebook/data dictionary** (variables, units, etc.).



Data Provenance

Document **provenance** (how raw → processed; scripts/workflows; software versions).



Community Standards

Use community standards (standard variable names, ontologies where relevant).



Licensing & Reuse

Choose a **reuse-friendly licence** when possible (or clearly justify restrictions).



Limitations & Citation

State **limitations** (bias, data quality issues) and **how to cite**.

The original records

- Often not published openly
- Restricted/controlled access
- Secure internal storage or controlled environments
- Important to keep & hard to preserve



Cleaned, transformed, anonymised (if needed)

- Best candidate for data sharing
- Potential to be published openly
- A high value “preservation target”
- Archived with documentation (README + data dictionary/codebook).

Streets as at 1977	Streets added between 1977 and 2022	No of Addresses in Street	No of Addresses with a House Name	List of House Names (in alphabetical order)
Anglu Gatt Street		27	20	Ajax, Carmen, Casa Maria, El Shaddai, Iowa, Iria, Jasiane, Joe Rose, Le Petite Fleur, Lincoln, Melwood, Sorijiet Agostinjani, St Nicholas, Tanti, The Daffodils, The Willows, Tulip, Valverde, Vct Bldgs, Woodstone
Autumn Street		41	30	Agape, Aloha, Camelia, Cardijin, Carmen, Cyrean, Divine Mercy, Doris, Eflin Ville, Franco, Frankmar, Gardenia, Grognet, Harmony, Ioanthe, Ivy, Klarinett, Maple Leaf, Maria, Marija Bambina, Morning Glory, Obos, Reno, Royal Oak, Season Blooms, Serenity, Shalom, St Mary, Trinity
Bezbizija, Limits of		5	5	Bint, Il-Girna, Tal-Bid, Warda, Zefir
Biccerin Street		5	3	Belfiore, Blossom, St Joseph
Bidnija, Limits of		14	14	
Bridge Street		30	23	Avocet, Bridge Apts, Bridgewell, Casa Vinita, Chateau Blanche, Coral, Dew Drop, Ġilju, Ġnien il-Ghasel, Heart of Mary, Marpal, Our House, Rose Marie, Rozinjol, Santa Maria, Sapphire, Seven Hills, Shema, St Mary, The Elms, Theotokos, Walnut, White Rose
Brittania Street		15	4	Bethlehem, Crossroad, Marigold. Sempre Viva
Brittania Square		8	2	Madonna ta' Pompei, Santa Maria
Brittania Street Alley		3	2	L-Gharix , L-Ghorfa
Cali Street		15	12	Aries, Bethania, Cupid, Grudge, Jasmine, Mirage, Our Dream, Rivendell, Rose, St Joseph, Tulip, York

House names in ten Maltese localities

> <https://doi.org/10.60809/drum.24331825>



Cite

Download all (331.52 kB)

Share

Embed

+ Collect



Edit as administrator

	A	B	C	D	E	F	G	H	I	J	K	L
1	age	gender	bcom	economics	olevel	intermediate	advanced	selfworth	prime1	prime2	dicdon	ultdon
2	18	Female	1	0	0	0	0	2	0	0	1	5
3	19	Female	1	0	0	0	0	3	0	0	5	0
4	18	Female	1	0	3	4	0	3	0	0	5	5
5	20	Male	1	1	3	0	4	3	0	0	5	10
6	19	Male	1	1	3	0	4	4	2	2	0	5
7	21	Male	1	1	3	0	4	4	2	1	0	0.5
8	19	Female	1	1	3	0	0	3	0	0	5	2
9	21	Female	1	1	0	0	4	4	2	2	3	5
10	20	Female	1	1	3	4	0	4	0	0	5	5

[Guide](#)
[1 - Survey](#)
[2 - Clean Data Set](#)
[3 - Original Data Set](#)
[4 - Metadata](#)
[5 - Data for Figures](#)

Economics students : selfish by nature, altruistic by nurture

<https://doi.org/10.60809/drum.24137757>

Cite

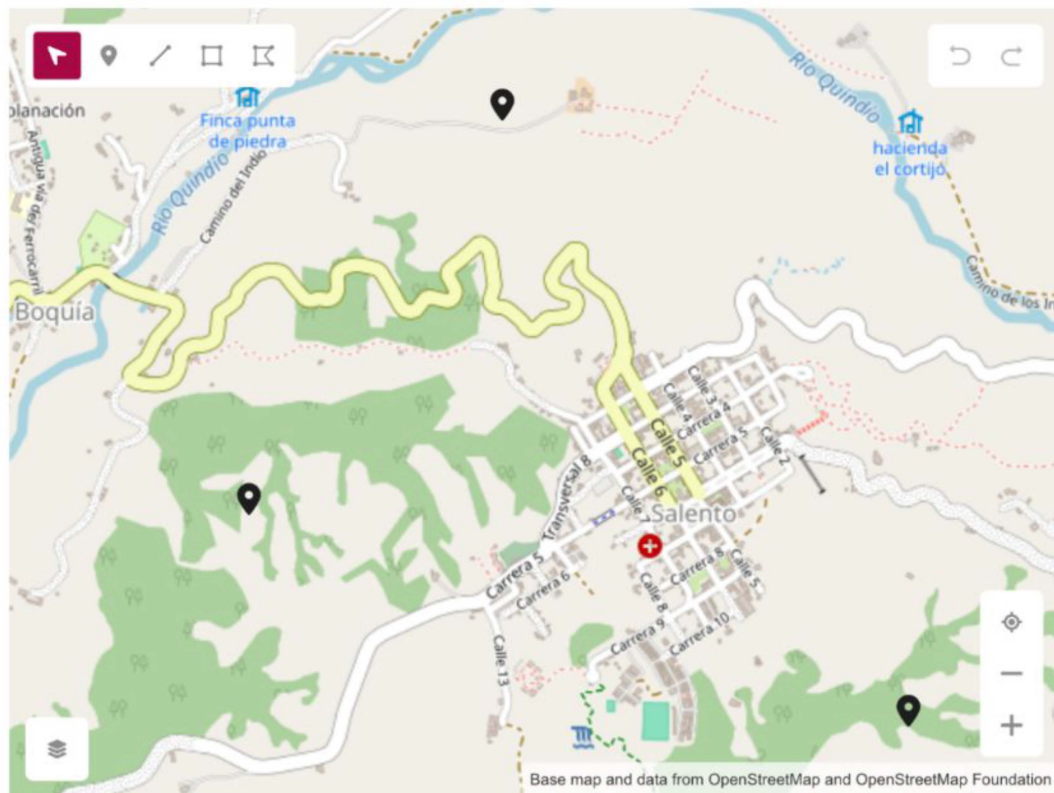
Download all (309.7 kB)

Share

Embed

+ Collect

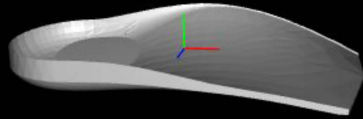
...



Features		GeoJSON			
#	Type	Description	Related file	Action	
> 1	Point	First collection point. 15 April 2025	soil_sample_metadata.csv	...	
✓ 2	Point	Second collection point. 16 April 2025.	soil_analysis_results.csv	...	
Point path		Latitude	Longitude		
1		4.646788696550...	-75.57449813322...		
> 3	Point	No description	None	...	
> 4	Point	No description	None	...	

Cancel

Done



MD5: 4a7c977c9363688586e304c1f9812f

_ESM.stl (634.33 kB)  

Additional file 2: of Custom-made foot orthoses: an analysis of prescription characteristics from an Australian commercial orthotic laboratory

The old way of doing research (science)

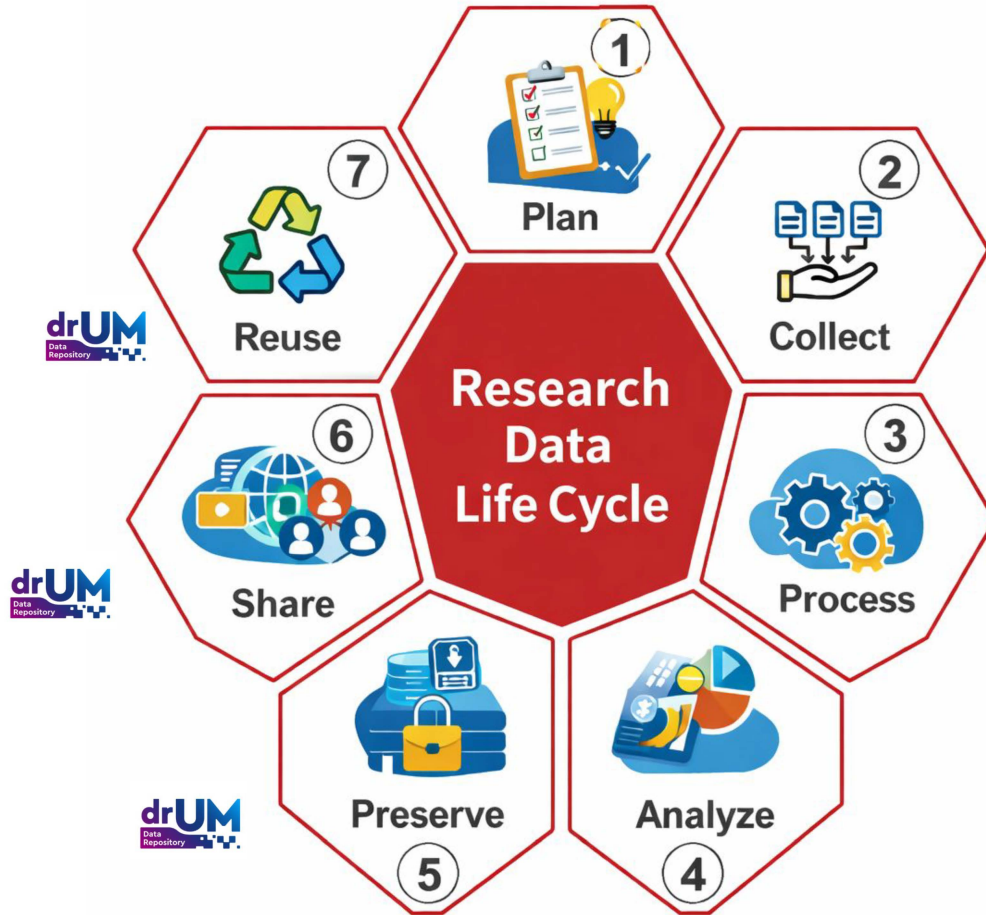
1. Researcher collects data (information)

2. Researcher interprets/synthesises data

3. Researcher writes paper based on data

4. Paper is published (and preserved)

5. Data is left to benign neglect, and eventually ceases to be accessible



- Citation & linking

- Scope & data needs
- Roles & responsibilities
- Plan for compliance & sharing

- Collection methods
- Quality control
- Legal and ethical considerations



- What will be archived/published
- **Repository & timing**
- Access & licence




- Cleaning & validation
- Data transformation
- Metadata/README updates

- Storage & organisation
- Access & permissions
- Backup & security controls

- Methods & alignment
- Tools & visualization
- Reproducibility

References

- Jiang, J., Maurici-Pollock, D., & Tang, R. (2025). Research Data Lifecycle (RDLC): An investigation into the disciplinary focus, use cases, creator backgrounds, stages and shapes of RDLC models. *International Journal of Digital Curation*, 19(1), 1-29.
- ELIXIR. (2024). *Data life cycle*. Retrieved March 17, 2026, from https://rdmkit.elixir-europe.org/data_life_cycle
- Bossaller, J., & Million, A. J. (2023). The research data life cycle, legacy data, and dilemmas in research data management. *Journal of the Association for Information Science and Technology*, 74, 701–706.
- Alper, P., D'anna, F., Droesbeke, B., Andrabi, M., Buono, R. A., Bianchini, F.,... Goble, C. (2025). RDMkit: A research data management toolkit for life sciences. *Patterns*, 6(9), 1–14.
- Li, S. (2024). *Information management*. Springer.

UM solution	Storage size	Security measures	Back-up	Internal collab	External collab	Sensitive info	Remote access
<p>Google Drive (UM Google Workspace)</p> 	Students: 5 GB; Staff: 25	ISO 27001 certified;	Sync/backup workflows (Drive for desktop)	Share folders/files, but continuity risk if the owner leaves UM.	Possible; manage external access	For sensitive files: use strong access control + encryption where appropriate (UM recommends AES-256 via 7-Zip/Keka).	Web/mobile/desktop
<p>University network drive (departmental/internal)</p> 	TBC (IT/department/project)	Not clearly stated, so confirm specifics with IT Services .	UM guidance stresses ensuring regular backups for chosen storage.	Recommended for collaboration within the UM community.	Typically limited vs cloud tools (best confirmed with IT).	Use access controls + encrypt restricted files; UM guidance highlights encryption for restricted materials.	Via UM VPN
	Based on requested quota	ISO 27001 certified; (high-level assurance practises).	Hosted on AWS; AWS S3 provides multiple backups for resilience	Workflow designed for archiving/publishing datasets	Projects (collaborative spaces with “contributors” and “viewers”)	Use restricted/embargo access; encryption in transit/at rest supports security.	Online repository (web access).

Best practices for organising and documenting research data



- Your filenames and folder structure has to make sense:

to you

to you - after a holiday

to you - after graduation/project/promotion



- Your research data will be 'shared' and used by:

- Colleagues within a project
- Project leaders/supervisors when contracts ends
- Etc.



File naming (Wilson, G., at all 2017); (Demerdash, 2025)



Machine readable

Avoid special characters (&%\$#/), spaces, and long names; stick to one convention `Research_data_management_drUM_Responsible_Data_Sharing_workshop`



Human readable

Reflects content, independent of folder structure, with 3-5 elements (e.g. project / experiment name, data type, creator initials, creation data, etc.);

`Pre_RDMworkshop_v5_2026-11-03`



Plays well with default ordering

Dates in YYYYMMDD format, use hyphens/underscores, and leading zeros; avoid 'final' or 'draft'

GNOME Commander



Bulk Rename Utility

Bulk Rename Utility

C:\Users\khh197\OneDrive - University of Texas at San Antonio\Documents\001_EXAMPLE

Name	New Name	Size	Created
file_example_XLS_10 (1).xls	file_example_XLS_10 (1).xls	8.50 KB	6/10/2024 10:29:48
file_example_XLS_10 (2).xls	file_example_XLS_10 (2).xls	8.50 KB	6/10/2024 10:29:57
file_example_XLS_10 (3).xls	file_example_XLS_10 (3).xls	8.50 KB	6/10/2024 10:30:35
file_example_XLS_10 (4).xls	file_example_XLS_10 (4).xls	8.50 KB	6/10/2024 10:31:05
file_example_XLS_10 (5).xls	file_example_XLS_10 (5).xls	8.50 KB	6/10/2024 10:32:27
file_example_XLS_10 (6).xls	file_example_XLS_10 (6).xls	8.50 KB	6/10/2024 10:32:57

Name	New Name	Size	Created
file_example_XLS_10 (1).xls	240610_RuizStudy_001.xls	8.50 KB	6/10/2024 10:29:48
file_example_XLS_10 (2).xls	240610_RuizStudy_002.xls	8.50 KB	6/10/2024 10:29:57
file_example_XLS_10 (3).xls	240610_RuizStudy_003.xls	8.50 KB	6/10/2024 10:30:35
file_example_XLS_10 (4).xls	240610_RuizStudy_004.xls	8.50 KB	6/10/2024 10:31:05
file_example_XLS_10 (5).xls	240610_RuizStudy_005.xls	8.50 KB	6/10/2024 10:32:27
file_example_XLS_10 (6).xls	240610_RuizStudy_006.xls	8.50 KB	6/10/2024 10:32:57

RegEx (1) R

Match

Replace

Inc.Ext. Simple v2

Replace (3) R

Replace

With

Match Case First

Remove (5) R

First n Last n

From to

Chars Words

Crop

Digits High Trim

D/S Accents Chars

Sym. Lead Dots

Add (7) R

Prefix

Insert

at pos.

Suffix

Word Space

Auto Date (8) R

Mode

Type

Fmt

Sep. Seg.

Custom

Cent. Off.

Numbering (10) R

Mode at

Start Incr.

Pad Sep.

Break Folder

Type

Case



```

./your-project-name/
├── 1_admin
│   ├── 1_proposals    <-- project background, plans, proposals and applications...
│   ├── 2_finance      <-- funding applications, receipts, project budgets...
│   └── 3_reports
├── 2_ethics
│   ├── 1_approval     <-- ethics application, approval docs...
│   └── 2_consent      <-- consent forms, questionnaires...
├── 3_experiment
│   ├── 1_collection   <-- data collection scripts, experimental stimuli and code...
│   ├── 2_data
│   │   ├── 1_raw      <-- original (read only) data files, e.g. raw reaction time data, etc...
│   │   ├── 2_interim  <-- intermediary files, pre-processed or transformed data...
│   │   └── 3_processed <-- final canonical datasets for modeling and stats...
│   ├── 3_analysis(*)
│   │   ├── funcs      <-- source functions for processing, modeling, visualisation...
│   │   ├── models     <-- trained models, model predictions, coefficients...
│   │   └── notebooks  <-- Jupyter notebooks for data exploration, plotting...
│   └── 4_outputs
│       ├── 1_figures  <-- generated figures and graphics for reporting...
│       └── 2_reports  <-- project reports, HTML, PDF...
└── 4_dissemination
    ├── 1_presentations <-- Keynote/PowerPoint slides, conference posters...
    ├── 2_publications  <-- journal articles, published conference proceedings...
    └── 3_publicity     <-- media coverage and publicity, newspaper articles, outreach...
  
```

* to facilitate version control of scripts in the analysis directory, boilerplate files are added: an MIT LICENSE.txt, a README.md, and a `__init__.py` to make the `funcs` directory act as a Python module.

Access Directory Hierarchies:

qr.codes/ta8uZC



Version control

▼ <https://doi.org/10.60809/drum.26056957>



<https://doi.org/10.60809/drum.26056957.v1>

Version 1



Operating on AWS

- Runs on Amazon Web Services (AWS)
- 99.9% uptime transparency

aws



Robust Security Measures



ISO 27001 Certified



Dedicated Security Team



24/7 System Monitoring

Scalable & Secure Infrastructure

- Deployed globally across AWS Availability Zones
- Regular vulnerability scans and penetration testing
- HECVAT audits completed and shared

Data Backup and Preservation

- Data encrypted in transit and at rest
- Regular backups across AWS multiple regions
- Integration with preservation systems



Security levels (CEESDA, 2022) (UK Data Service, 2025)




Security of research data

-  Encryption
-  Anonymization
-  Restricted access
-  Secure transport
-  Complete deletion

Security of computer system

-  Firewall
-  Antivirus software
-  Installing updates
-  Using secured WIFI
-  Password-protect
-  Device encryption

Physical data security

-  Lock & Lock
-  Don't leave unattended
-  Safe transport

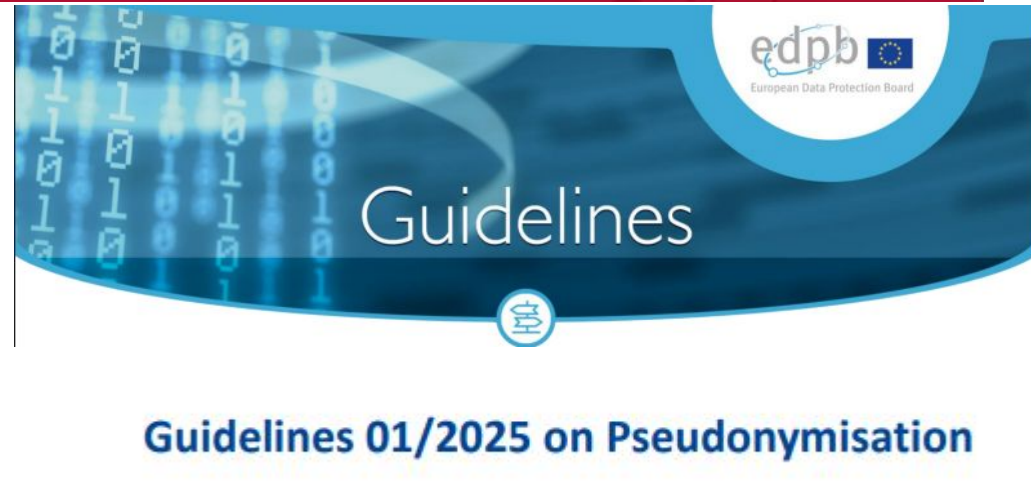
Anonymization

- **(GDPR Recital 26)** anonymous info is info that does not relate to an identified/identifiable person, and GDPR doesn't cover it (including for research/statistics).



Pseudonymization & Separation

Russell Fenech	P001
russel.fenech@...	P001e
6'2	6'2
95kg	95kg
381001A	P001





MRI Scanner

Secure **VPN Tunnel**
Restricted by **Firewall**



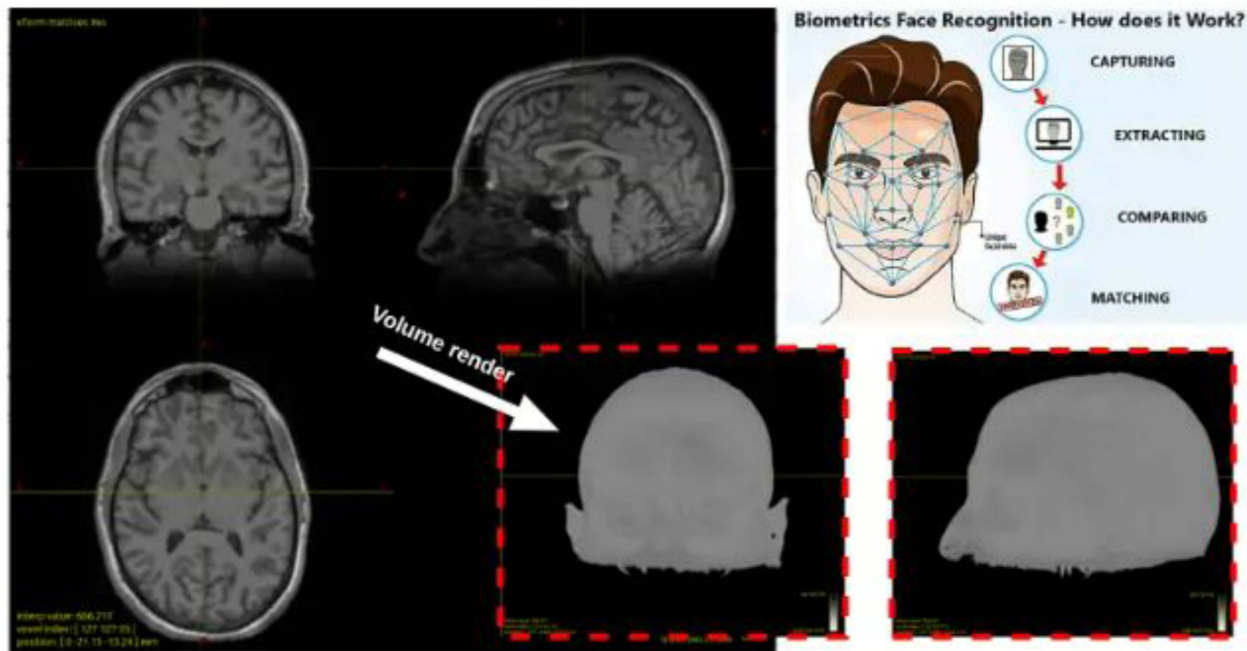
ORT(H)ANC



MySQL



drUM
Data
Repository



Biometric Risk

- MRI anatomic T1 images can reconstruct faces
- Therefore metadata removal is insufficient
- We implement defacing

mri_reface (Christopher G. Schwarz *et al.*) replaces identifiable face information in MRI/PET/CT scans to help prevent potential re-identification via face recognition. Faces are replaced with an **average face**, rather than removed, to better resemble a natural image and reduce effects on downstream brain measurement software

Lawful & Transparent



Explicit informed consent
(incl. **future research use**)

Data recognised as relating to
and **controlled by the subject**
--> Participants retain access,
withdrawal, and erasure rights

Privacy, Security & Governance Framework



No storage of identified data

Secure **transfer** (VPN + firewall
restrictions)

Cryptographic pseudonymisation
of DICOM metadata

Separation of identity mapping from
research datasets

MRI images **refaced to remove**
biometric identifiers

Participant Care: **Incidental**
findings protocol with appointed
radiologist

Ethical Open Science



BIDS-structured datasets
(interoperable & reusable)

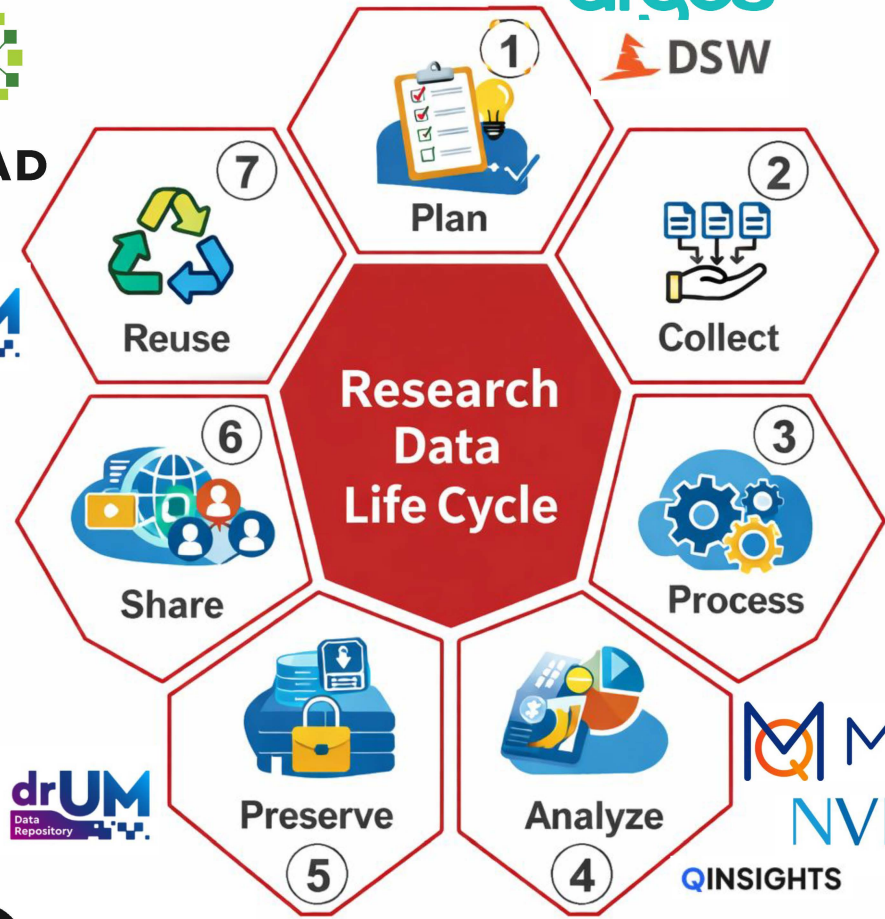
FAIR-aligned preparation for
controlled sharing

Only appropriately processed
data considered for archiving

Privacy protection as
foundation for reproducibility



L-Università ta' Malta



Common archiving & publishing principles across Europe

- Retain research data for a defined period to **verify results** (often ≥ 10 years)
- Use trusted repositories + publish **rich metadata** + assign **DOIs**
- Make data **FAIR** (Findable, Accessible, Interoperable, Reusable)
- Share at publication where possible; restrict when necessary, but keep metadata public

UM archiving research data



Secure, Long-Term Preservation

Ensuring a copy of datasets is kept securely by the host institution (UM) through its institutional data repository (drUM).



Global Visibility & Indexing



most open

share
remix
commercial

PD

CC 0
CC BY
CC BY SA

share
remix

CC BY NC SA

share

CC BY ND
CC BY NC ND

all rights reserved

least open

CC BY 4.0

CC0

MIT

GPL

GPL 2.0+

GPL 3.0+

Apache 2.0

All Rights Reserved - um-mt



RDM **kit**



What data should go to drUM, and when?



Processed /
curated



Electronic/Digital
file type



UM
research-related

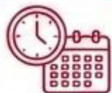


Adherence to FAIR
data principles



Ethically/legal
safe

When to deposit:



During the
project

Freeze a dataset
(embargo OK)



At publication
(recommended)

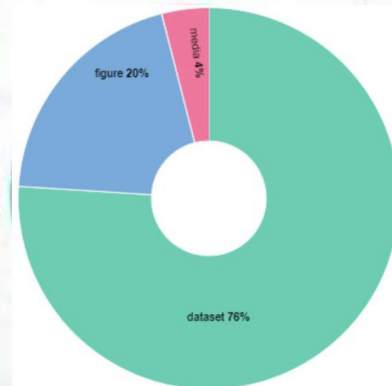
Publish processed data
+ code/workflow



Project
end

Version of record
for preservation

drUM Data Workflow & Repository Statistics



Supporting Research Data Management with Generative AI

(Kim, S. Y., Görz, M., & Geisler, S. (2025), (University of Turku, 2024), (RDA Guidance on AI Tools Usage, 2025)

Opportunities

- Data cleaning & Organization
- Metadata annotation
- Data Visualization (i.e knowledge graphs)
- Generation of synthetic datasets
- DMP support


Challenges & Risks


- Bias in AI models
- "Black Boxes" phenomenon
- Data quality
- Ethical and legal aspects

**Depositing
data in
drUM**




Thank you for your attention Questions?

 Open Science

 (+356) 2340 1234/5

 /um.libraryservices

 oar.lib@um.edu.mt

References

- Alper, P., D'Anna, F., Droesbeke, B., Andrabi, M., Buono, R. A., Bianchini, F., ... Goble, C. (2025). RDMkit: A research data management toolkit for life sciences. *Patterns*, 6, 101345. doi:10.1016/j.patter.2025.101345
- Bossaller, J., & Million, A. J. (2023). The research data life cycle, legacy data, and dilemmas in research data management. *Journal of the Association for Information Science and Technology*, 74, 701–706.
- Briney, K. A. (2024). Measuring data rot: An analysis of the continued availability of shared data from a single university. *PLOS ONE*, 19(6), e0304781. doi:10.1371/journal.pone.0304781
- CESSDA Training Team. (2017–2022). *File naming and folder structure*. In *CESSDA Data Management Expert Guide*. Bergen, Norway: CESSDA ERIC. Retrieved from <https://dmeq.CESSDA.eu/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure>
- CESSDA Training Team. (2017–2022). *Security*. In *CESSDA Data Management Expert Guide*. Bergen, Norway: CESSDA ERIC. Retrieved from <https://dmeq.CESSDA.eu/Data-Management-Expert-Guide/4.-Store/Security>
- Colavizza, G., Cadwallader, L., LaFlamme, M., Dozot, G., Lecorney, S., Rappo, D., ... Hrynaszkiewicz, I. (2024). An analysis of the effects of sharing research data, code, and preprints on citations. *PLOS ONE*, 19(10), e0311493. doi:10.1371/journal.pone.0311493
- Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLOS ONE*, 15(4), e0230416.
- Creative Commons. (n.d.). *Creative Commons licenses*. Retrieved from <https://creativecommons.org/share-your-work/use-remix/cc-licenses/>
- Demerdash, Y., Dockhorn, R., & Wilbrandt, J. (2025). Data organization made easy: Comprehensive folder structure template for early career life/natural science researchers. *Data Science Journal*, 24, 35. doi:10.5334/dsj-2025-035
- Digital Science, Hahnel, M., Smith, G., Schoenenberger, H., Scaplehorn, N., & Day, L. (2023). *The State of Open Data 2023*. figshare. doi:10.6084/m9.figshare.24428194

- ELIXIR. (n.d.). *RDMkit: Research data management toolkit for life sciences*. Retrieved April 30, 2026, from <https://rdmkit.elixir-europe.org/>
- ELIXIR. (2024). *Data life cycle*. Retrieved March 17, 2026, from https://rdmkit.elixir-europe.org/data_life_cycle
- Engineering and Physical Sciences Research Council. (2022). *Policy framework on research data: Expectations*. Retrieved from <https://www.ukri.org/who-we-are/epsrc/our-policies-and-standards/policy-framework-on-research-data/expectations/>
- European Commission. (2025). *Horizon Europe programme guide* (Version 5.1). Retrieved from https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf
- European Commission, Directorate-General for Research and Innovation, & PwC EU Services. (2018). *Cost-benefit analysis for FAIR research data: Cost of not having FAIR research data*. Luxembourg: Publications Office of the European Union. doi:10.2777/02999
- European Data Protection Board. (2025). *Guidelines 01/2025 on pseudonymisation*. Retrieved from https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_en
- European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679: General Data Protection Regulation. *Official Journal of the European Union*, L119, 1–88. Retrieved from <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- Figshare. (n.d.). *Security, stability, and ISO27001 certification*. Retrieved from <https://info.figshare.com/user-guide/security-stability/>
- Geisler, S. (2026). *Supporting research data management with generative AI: A modern FAIRytale?* [Lecture]. TIB AV-Portal. doi:10.5446/72494
- Hahnel, M. (2024, December 2). *Are we ready to start rewarding researchers for Open data?* OpenResearch.wtf. Retrieved from <https://www.openresearch.wtf/are-we-ready-to-start-rewarding-researchers-for-open-data/>
- Jiang, J., Maurici-Pollock, D., & Tang, R. (2025). Research Data Lifecycle (RDLC): An investigation into the disciplinary focus, intended use cases, creator backgrounds, stages and shapes of RDLC models. *International Journal of Digital Curation*, 19(1), 1–29. doi:10.2218/ijdc.v19i1.860

- Khare, S., Gurry, C., Freitas, L., Schultz, M. B., Bach, G., Diallo, A., ... Maurer-Stroh, S. (2021). GISAID's role in pandemic response. *China CDC Weekly*, 3(49), 1049–1051.
- Kim, S. Y., Görz, M., & Geisler, S. (2025). KONDA: An LLM-based tool for semantic annotation and knowledge graph creation using ontologies for research data. In *CEUR Workshop Proceedings* (Vol. 4065).
- KU Leuven. (n.d.). *Organising your data*. Retrieved April 19, 2026, from <https://www.kuleuven.be/rdm/en/guidance/data-standards/how-organise-data>
- Li, S. (2024). *Information management*. Springer.
- Macejovsky, P. (2020). *Research data management in higher education* [Master's thesis, Comenius University]. Retrieved from <https://opac.crzp.sk/?fn=detailBiblioFormChildE3FVS&sid=C67292DEA65D7A08E624F5CAC3F9&seo=CRZP-detail-kniha>
- Malato, G. (2022, May 12). *Image of Sgr A*, the black hole at the center of our galaxy*. U.S. National Science Foundation. Retrieved from <https://www.nsf.gov/science-matters/image-sgr-black-hole-center-our-galaxy>
- Malta Council for Science and Technology. (2021). *National open access policy: Malta*. Retrieved from <https://eosc.eu/wp-content/uploads/2022/11/National-Open-Access-Policy-2021.pdf>
- MDPI. (n.d.). *Instructions for authors: Data availability statements*. Retrieved from <https://www.mdpi.com/journal/data/instructions>
- Miotto, A., Toohey, J., Stockwell, L., Stapleton, S., Jennings, I., & Weaver, B. (2022). *6 - Folder structures*. In *Ten reproducible research things*. Griffith University. Retrieved from <https://gueslib.github.io/ten-reproducible-research-things/content/06-folders.html>
- National Science Foundation. (2015). *Today's data, tomorrow's discoveries: Increasing access to the results of research funded by the National Science Foundation* (NSF 15-052). Retrieved from <https://www.nsf.gov/reports/performance/public-access-plan-todays-data-tomorrows-discoveries>
- Nature Portfolio. (n.d.). *Reporting standards and availability of data, materials, code and protocols*. Retrieved from <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards>
- Open Source Initiative. (n.d.). *OSI approved licenses*. Retrieved from <https://opensource.org/licenses>

- OpenAIRE. (n.d.). *Amnesia*. Retrieved April 30, 2026, from <https://amnesia.openaire.eu/>
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175.
- Research Data Alliance. (2025). *RDA guidance on AI tools usage*. Retrieved from <https://www.rd-alliance.org/about/code-of-conduct/rda-guidance-on-ai-tools-usage/>
- Singh, A., Singh, A., Imbiriba, T., Erdogmus, D., & Borsoi, R. (2024). *KODA: A data-driven recursive model for time series forecasting and data assimilation using Koopman operators*. arXiv. doi:10.48550/arXiv.2409.19518
- Tedersoo, L., et al. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8, 192.
- Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., ... Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE*, 15(3), e0229003. doi:10.1371/journal.pone.0229003
- UK Data Service. (2025). *Research data handling and security guide for users*. Retrieved from <https://ukdataservice.ac.uk/app/uploads/cd171-researchdatahandling.pdf>
- University of British Columbia Library Research Commons. (n.d.). *Create a README file*. In *Research data management*. Retrieved April 20, 2026, from https://ubc-library-rc.github.io/rdm/content/03_create_readme.html
- University of Malta Library. (n.d.). *Research data lifecycle*. Retrieved April 30, 2026, from https://www.um.edu.mt/library/openscience/rdm/rdm_toolkit/rd_lifecycle/
- University of Malta Library. (n.d.). *Research data management FAQs*. Retrieved April 30, 2026, from https://www.um.edu.mt/library/openscience/rdm/rdm_faqs/
- University of Malta Library. (2025). *Research Data Management Policy* (Policy Code POL-002-01). Msida: University of Malta Library. Retrieved from <https://www.um.edu.mt/library/oar/handle/123456789/140264>
- University of Turku Library. (2026). *AI in research and research data management*. In *Research data management*. Retrieved April 30, 2026, from <https://utuguides.fi/c.php?g=719236&p=5319326>

- Vukovic, N. (n.d.). *Directory structure for science projects* [Computer software]. Retrieved April 19, 2026, from <https://github.com/vukovicnikola/templateproject>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. doi:10.1038/sdata.2016.18
- Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLOS Computational Biology*, 13(6), e1005510. doi:10.1371/journal.pcbi.1005510