Intro to Research Data Management at MPI-EVA

Bret Beheim
IMPRS Core Seminar, 7 Nov. 2024

Intro to Research Data Management at MPI-EVA

1. Overview of RDM resources @ MPI
2. Discussion about principles of RDM
3. Writing a RDM plan
4. Applied example: wrangling raw data files in R

Intro to Research Data Management at MPI-EVA

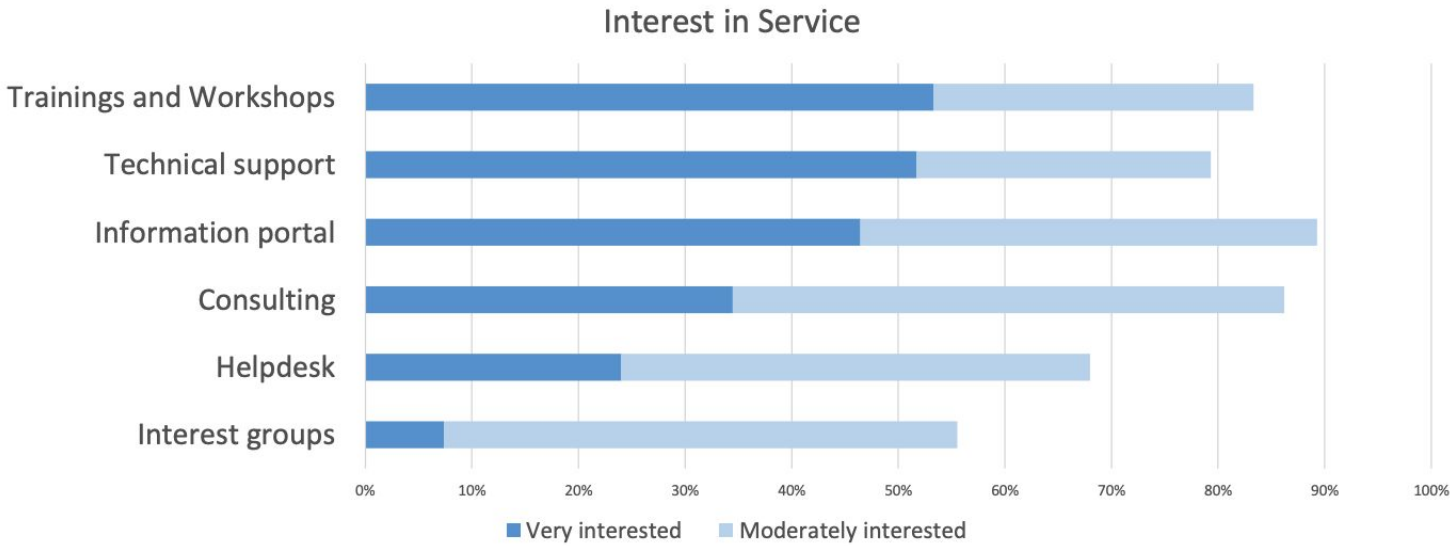
1. Overview of RDM resources @ MPI
2. Discussion about principles of RDM
3. Writing a RDM plan
4. Applied example: wrangling raw data files in R

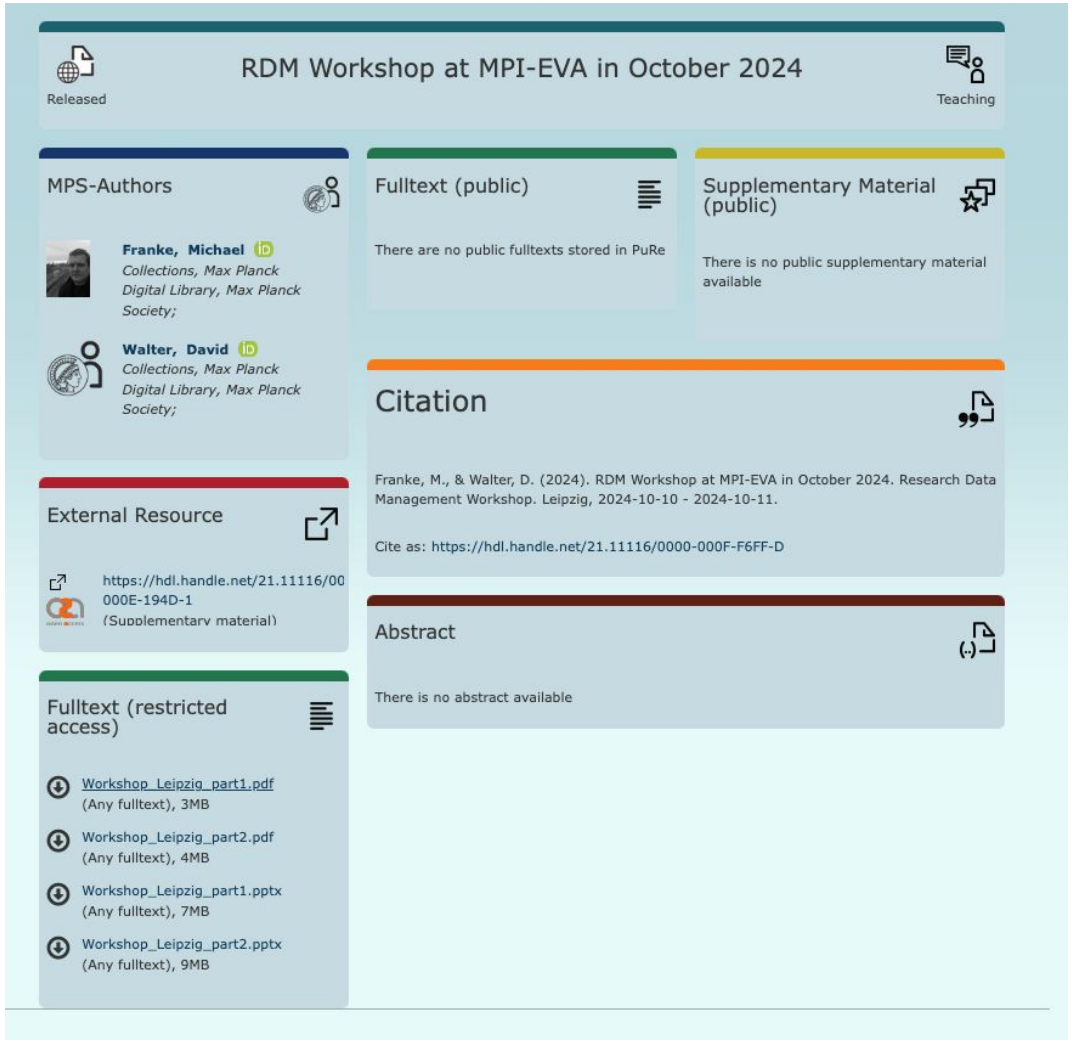
RDM Working Group @ MPI-EVA









Interest in Service
Trainings and Workshops
Technical support
Information portal
Consulting
Helpdesk
Interest groups
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Very interested
Moderately interested



https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_3559200

Genevieve Housman October 28

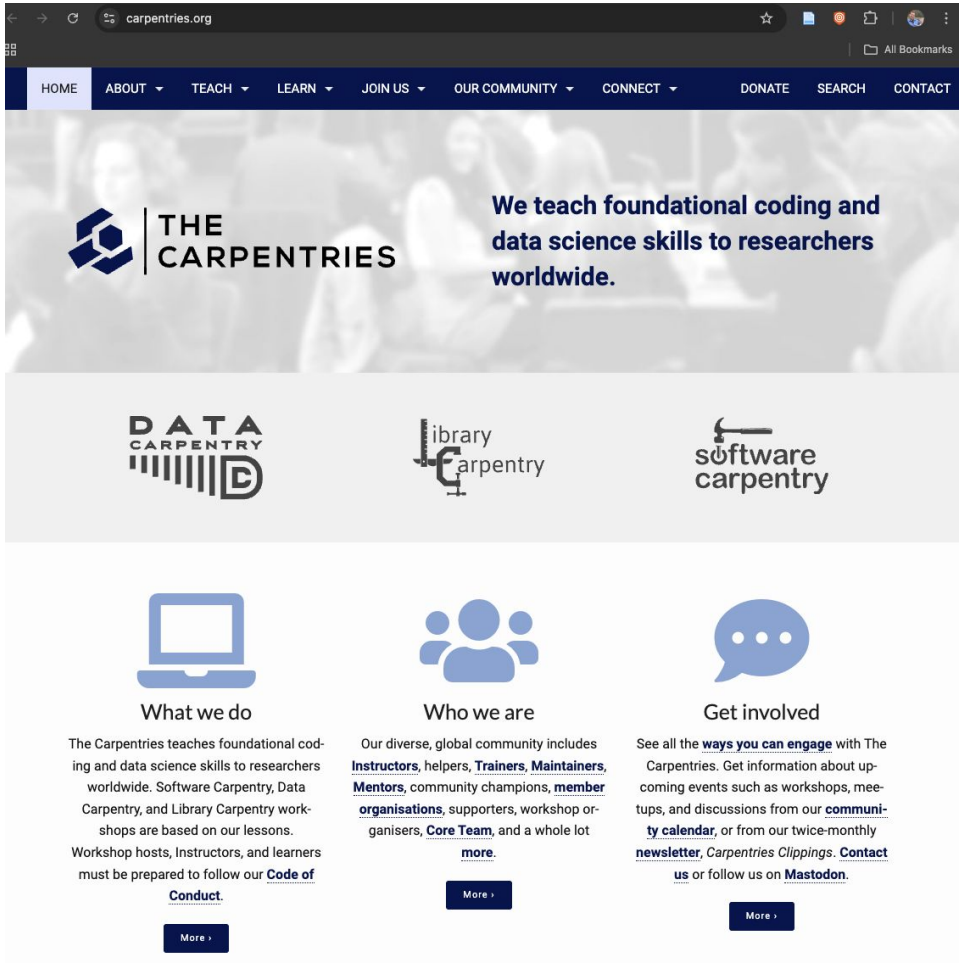
Announcing the start of 🤠 **Data Roundup Office Hours** 🤠 on October 30th in H4.10!

There's more to managing research data than wrangling. If you have questions about organizing, storing, maintaining, publishing, citing, or archiving data, come join us to round up some dogies... er, data!

Date: Wednesdays, every two weeks
Time: 12 - 1 pm (bring your lunch 🥪)
Location: H4.10
Hosts: Research Data Management (RDM) Working Group

Show less

Only visible to users in ~Town Square


carpentries.org
All Bookmarks
HOME
ABOUT
TEACH
LEARN
JOIN US
OUR COMMUNITY
CONNECT
DONATE
SEARCH
CONTACT
THE CARPENTRIES
We teach foundational coding and data science skills to researchers worldwide.
DATA CARPENTRY
library Carpentry
software carpentry
What we do
The Carpentries teaches foundational coding and data science skills to researchers worldwide. Software Carpentry, Data Carpentry, and Library Carpentry workshops are based on our lessons. Workshop hosts, Instructors, and learners must be prepared to follow our Code of Conduct.
More ›
Who we are
Our diverse, global community includes Instructors, helpers, Trainers, Maintainers, Mentors, community champions, member organisations, supporters, workshop organisers, Core Team, and a whole lot more.
More ›
Get involved
See all the ways you can engage with The Carpentries. Get information about upcoming events such as workshops, meetups, and discussions from our community calendar, or from our twice-monthly newsletter, Carpentries Clippings. Contact us or follow us on Mastodon.
More ›

Max Planck Digital Library Resources

- https://rdm.mpdl.mpg.de - general RDM links
- https://edmond.mpdl.mpg.de - permanent public data archive
- https://rdmo.mpdl.mpg.de - RMD planning templates
- https://keeper.mpdl.mpg.de - 1TB cloud storage and private archiving
- https://labfolder.mpdl.mpg.de - electronic lab notebook

Identifiers

Digital Object Identifier (DOI)

- For digital objects
- https://www.doi.org

Public domain

Open Researcher and Contributor ID (ORCID)

- For persons
- https://orcid.org

Public domain, CC0

Research Organization Registry (ROR)

- For organisations
- https://ror.org

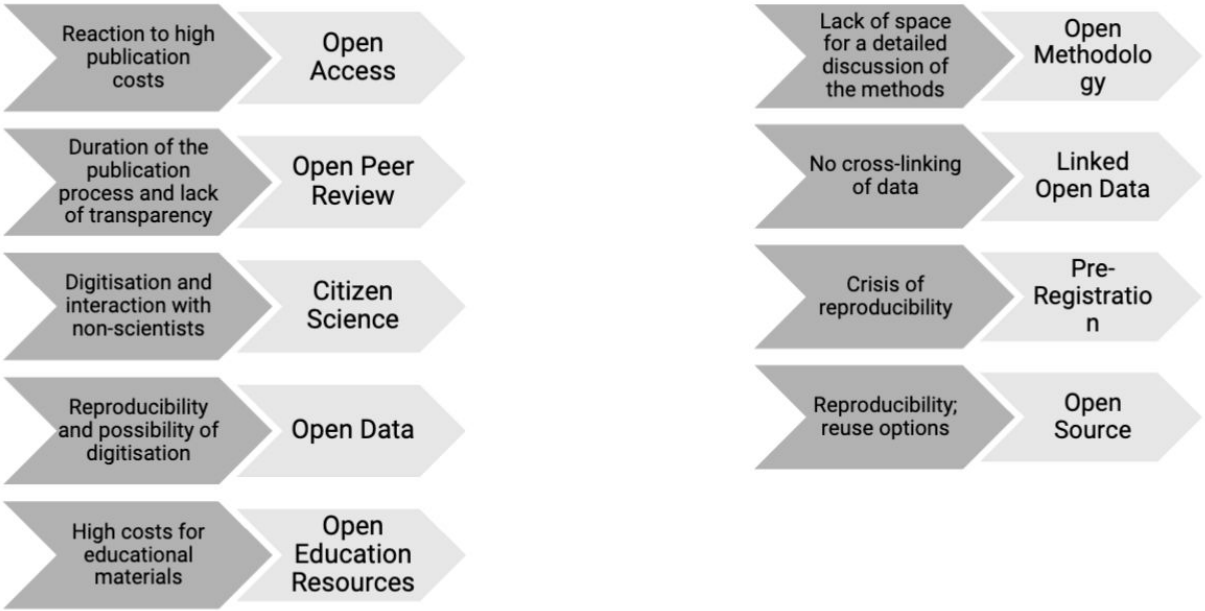
Research Organization Registry, CC BY 4.0, https://doi.org/10.5281/zenodo.4701802

Franke, M., & Walter, D. (2024). RDM Workshop at MPI-EVA in October 2024. Research Data Management Workshop. Leipzig, 2024-10-10 - 2024-10-11.

Intro to Research Data Management at MPI-EVA

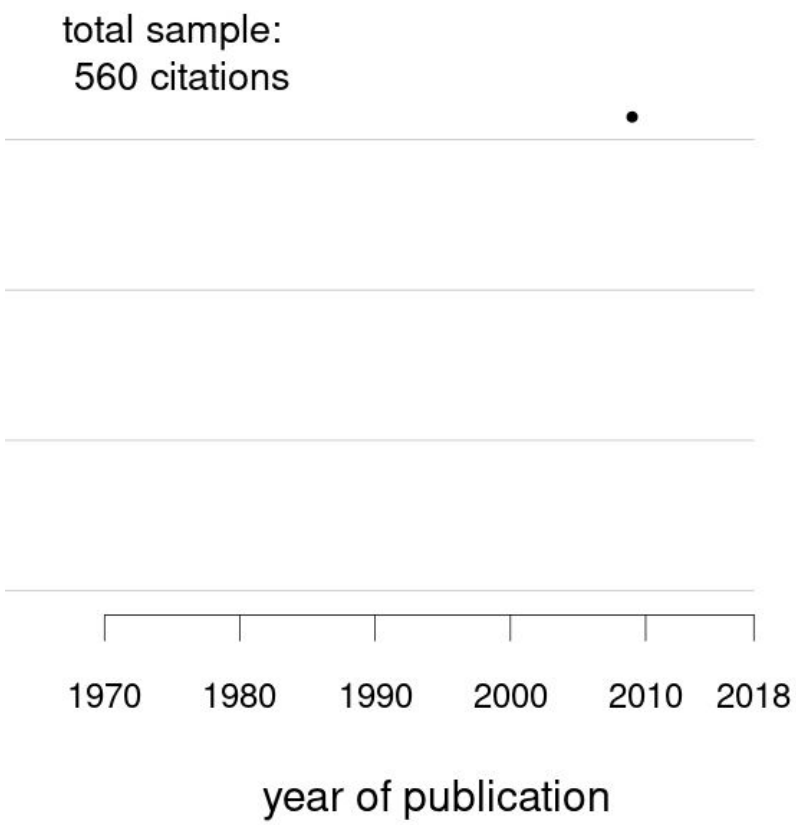
1. Overview of RDM resources @ MPI
2. Discussion about principles of RDM
3. Writing a RDM plan
4. Applied example: wrangling raw data files in R

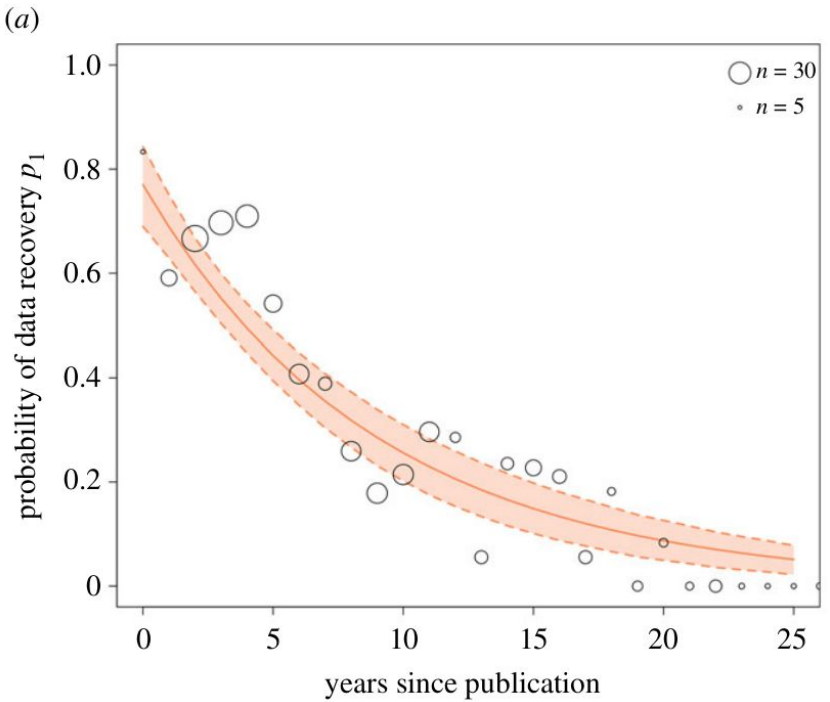
Hypothesis: Open Science is a Reaction



Franke, M., & Walter, D. (2024). RDM Workshop at MPI-EVA in October 2024. Research Data Management Workshop. Leipzig, 2024-10-10 - 2024-10-11.

Evolutionary Anthropology Reproducibility Study (EARS)





Minocher, et al. (2021) RSOS

MPG “Good Scientific Practice”

“**2.4 Securing and storing primary data – Documentation and archiving**

[..] The **Institute Management** is expected to **provide** the usual **storage media** for the field concerned and to guarantee that information stored both digitally and in analogue format is **secured** and **remains accessible**. The framework conditions must be such that protection from unauthorized access, loss, destruction, theft, and manipulation can be guaranteed. [..]

Research **Group Leaders** and **individual** researchers are obligated to make use of the protection options provided by Institute Management and retain and **store** both research **data and research results**. It does not matter in this context whether the research results are published or not.“

(https://www.mpg.de/197494/rulesScientificPractice.pdf, p. 41)

Franke, M., & Walter, D. (2024). RDM Workshop at MPI-EVA in October 2024. Research Data Management Workshop. Leipzig, 2024-10-10 - 2024-10-11.

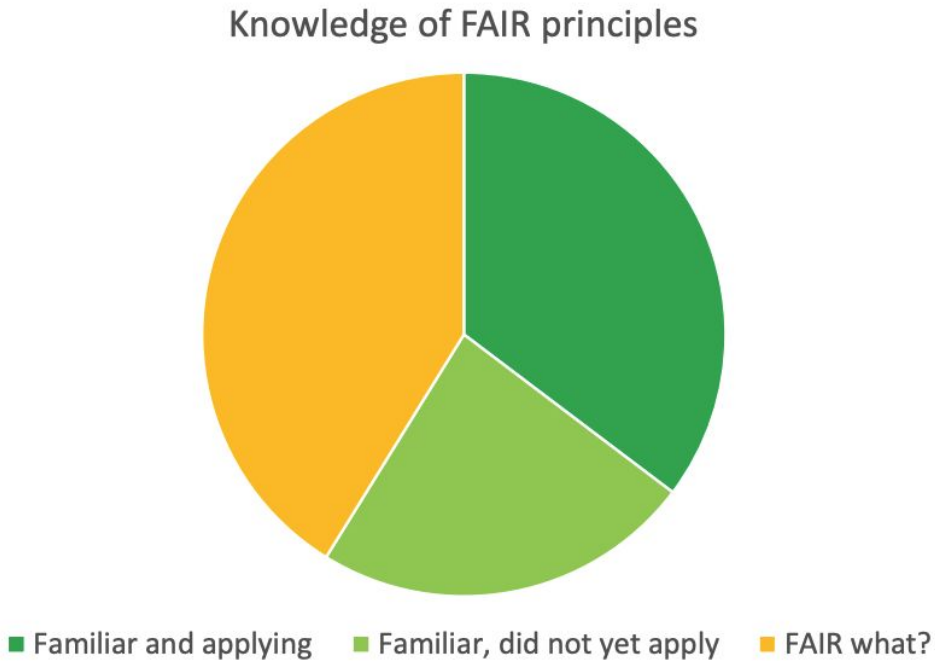
Legal Aspects of Data Publishing

- Who owns the data?
  - The scientist, the supervisor, the Max Planck Institute, the Max Planck Society?
  - There is no universally valid statement on this!
  - Talk to your supervisor. Have a look at your employment contract and local research guidelines. The publication approval process must be clarified.
  - Find out more! It's better to clarify things earlier rather than after the data publication (= too late).

https://doi.org/10.5281/zenodo.3674561

53

Franke, M., & Walter, D. (2024). RDM Workshop at MPI-EVA in October 2024. Research Data Management Workshop. Leipzig, 2024-10-10 - 2024-10-11.


Knowledge of FAIR principles
Familiar and applying
Familiar, did not yet apply
FAIR what?

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

https://blog.ukdataservice.ac.uk/fair-data-assessment-tool/

CARE Principles for Indigenous Data Governance

Collective Benefit.

Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.

C1. For inclusive development and innovation
C2. For improved governance and citizen engagement
C3. For equitable outcomes

Authority to Control.

Indigenous Peoples' rights and interests in Indigenous data must be recognized and their authority to control such data respected.

A1. Recognizing rights and interests
A2. Data for governance
A3. Governance of data

Responsibility.

Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self determination and collective benefit.

R1. For positive relationships
R2. For expanding capability and capacity
R3. For Indigenous languages and worldviews

Ethics.

Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

E1. For minimizing harm and maximizing benefit
E2. For justice
E3. For future use

I

(Legislative acts)

REGULATIONS

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 27 April 2016

on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

(Text with EEA relevance)

rijksuniversiteit groningen

GDPR is the enemy?!

› GDPR strongly emphasizes ***transparency***

- Justify how much data you will collect → ***power analysis!***
- Justify what you will do with the data → ***analysis plan!***
- Justify who will have access to the data → ***RDMP!***
- Prevent data loss → ***carefully archive your data!***

› **GDPR is fully compatible with the goals of open, transparent science!**

Jacob Jolij, “GDPR vs Open Science”, Oct 2024 talk

Intro to Research Data Management at MPI-EVA

1. Overview of RDM resources @ MPI
2. Discussion about principles of RDM
3. Writing a RDM plan
4. Applied example: wrangling raw data files in R

Intro to Research Data Management at MPI-EVA

1. Overview of RDM resources @ MPI
2. Discussion about principles of RDM
3. Writing a RDM plan
4. Applied example: wrangling raw data files in R

Please get "*all-uploads.zip*" at
https://github.com/babeheim/rdm-workshop-2024