

---

# Good Research Data Management: From Theory to Practice

Volker U. Schwartze

---

# Outline

1. Research data management – The What, Why and How?
  - Definitions and concepts
  - Requirements and challenges
  
2. Where can I get support regarding research data management?
  - Services provided by the FSU Jena
  - Additional information sources



---

# Research Data Management – The What, Why and How

---

# What is (Research) Data?

- No clear/consistent definition (discipline-specific definitions)
- Set of values/information/findings that are the result of observations, measurements, surveys etc.<sup>1</sup>

<sup>1</sup> Duden (<https://www.duden.de/woerterbuch>)

---

# What is (Research) Data?

- German Research Foundation (Deutsche Forschungsgemeinschaft, DFG):
  - Measured values
  - Laboratory findings
  - Survey data
  - Audiovisual information
  - Texts
  - Objects from a collection/samples
  - Test methods (e.g. questionnaires, software,...)

DFG (2015) „DFG Guidelines on the Handling of Research Data“

---

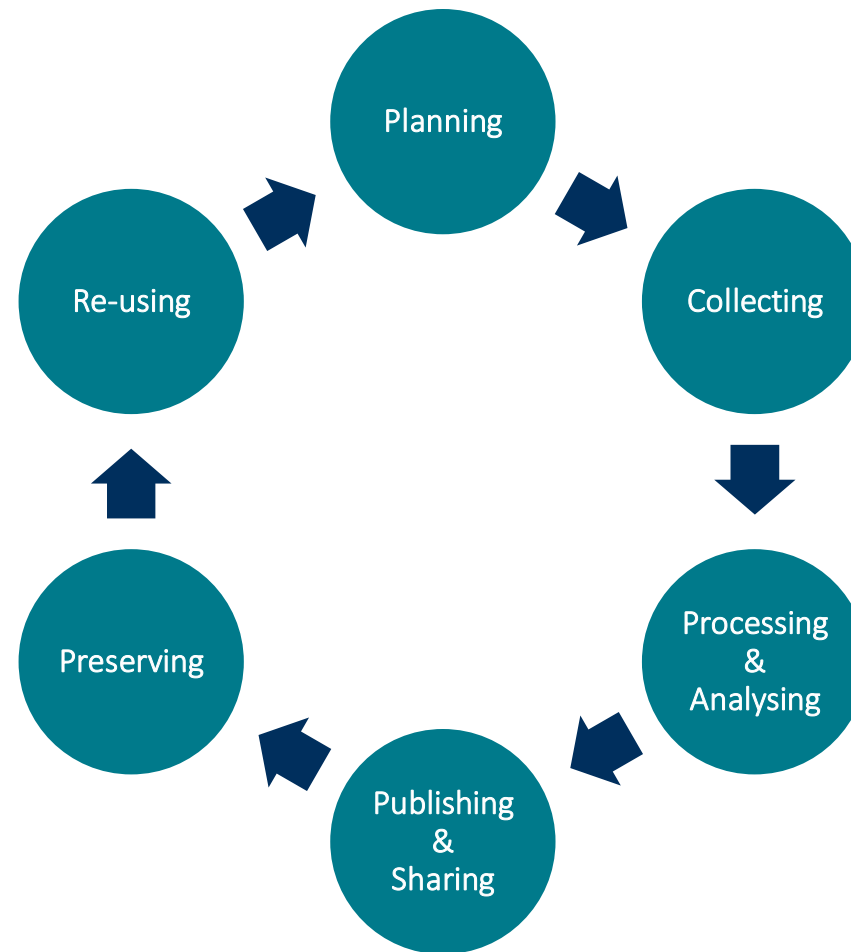
# What is Research Data Management?

Data management refers to all aspects of creating, housing, delivering, maintaining, and archiving and preserving data.  
It is one of the essential areas of responsible conduct of research.<sup>1</sup>

<sup>1</sup> Mantra 2017: 5

---

# The Research Data Life Cycle



---

# Why Research Data Management?

- Demanded by funding agencies and institutions
- Good scientific practice
- Verifiability and reproducibility
- Saves time and resources
- Data security and prevention of data loss
- Re-use of data (no need for unnecessary repetitions)





# Importance of Research Data Management Skills

Data Scientist



Industry (R/D)



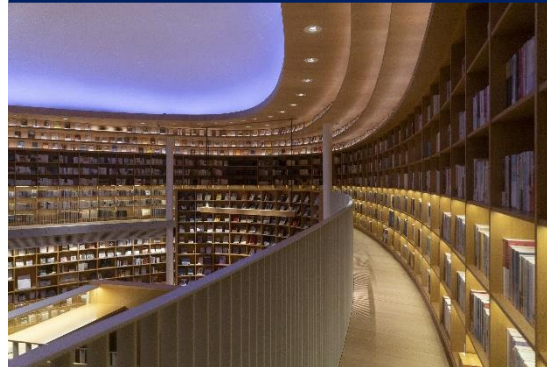
Academia



Consultant / Startup



Science-related positions



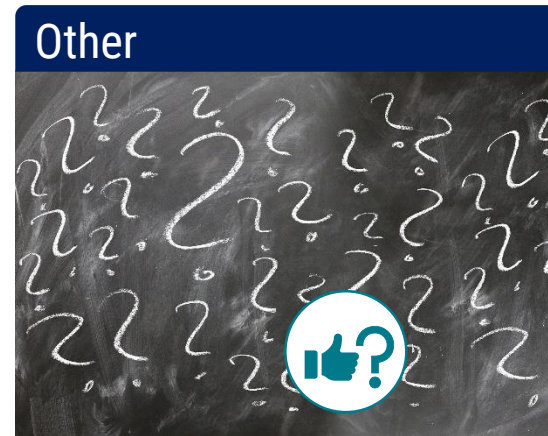
Other



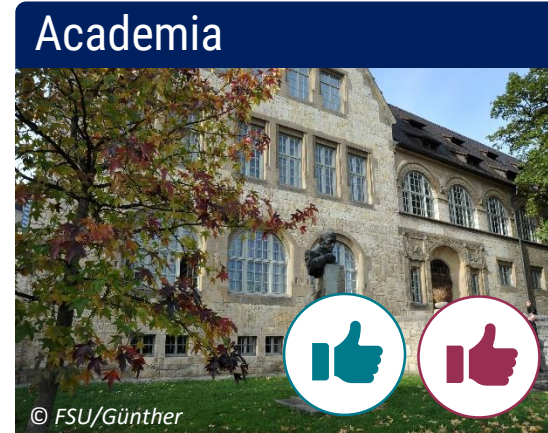
# Importance of Research Data Management Skills



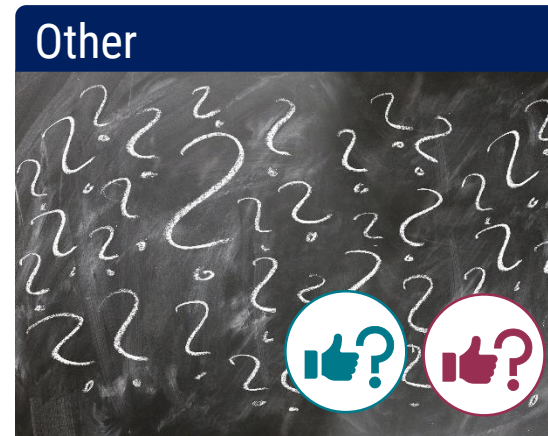
Data Management Skills



# Importance of Research Data Management Skills



Data Management Skills



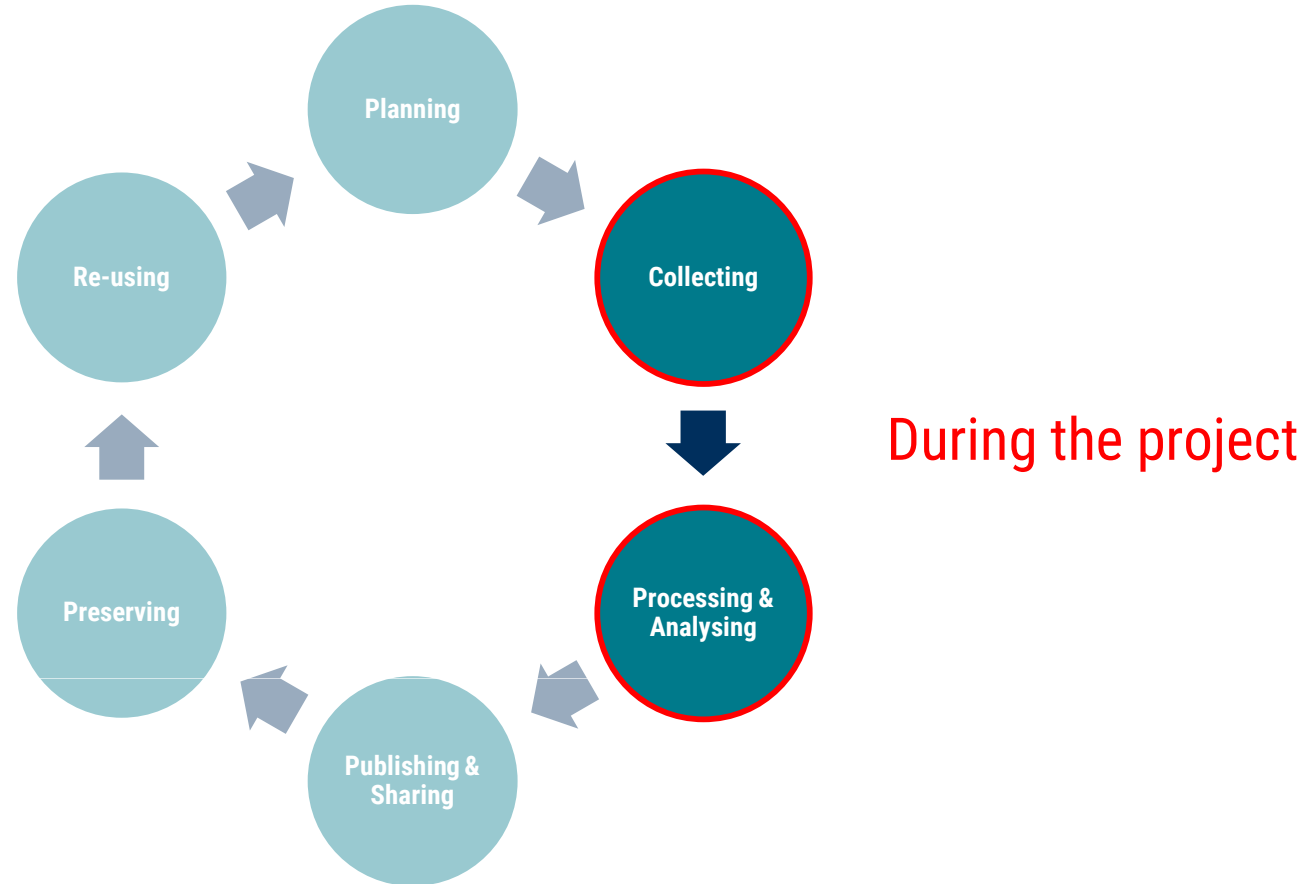
Open Data &  
Reproducible Research

---

# Research Data Management – When and How?



# Research Data Management – When and How?





---

## Storage and Backup

---

## Reasons for Data Loss

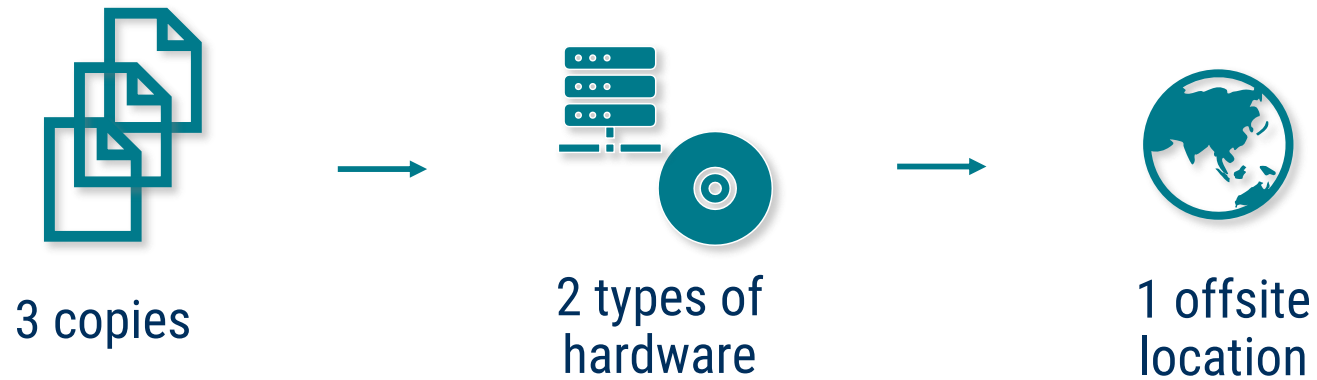
- Hardware defects
- Software errors
- Malware or hacking
- Mistakes by user (accidental deletion, change, renaming)
- Blackout, water damage, fire, ...
- Theft or burglary



---

# Storage and Backup

- 3-2-1 rule

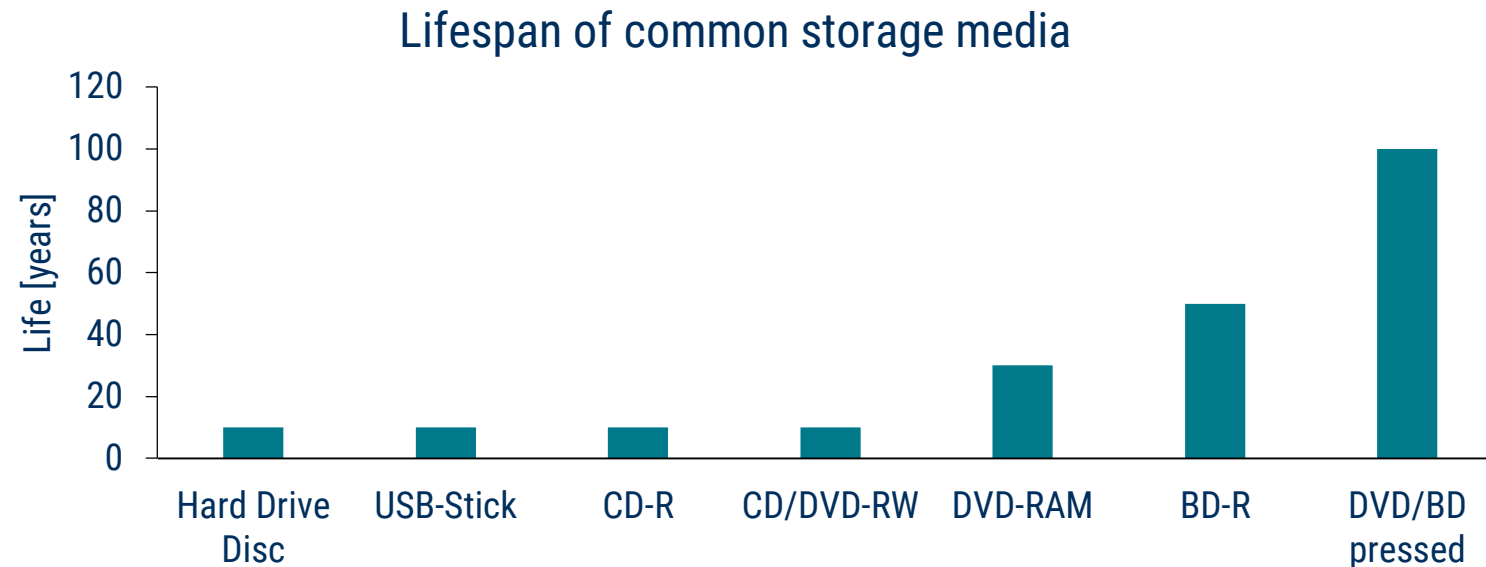


- Regular backups



# Storage and Backup

- Lifetime of storage restricted → ever 2-5 years new hardware



J. Rex (2019), DOI: 10.5281/zenodo.2579580

# Solutions offered by the URZ



**Home or project folder**



**Cloud**



**Server**

**Back-up frequency**

Daily

Daily

Dialy/weekly

**Hold-back time**

3 months

7 days

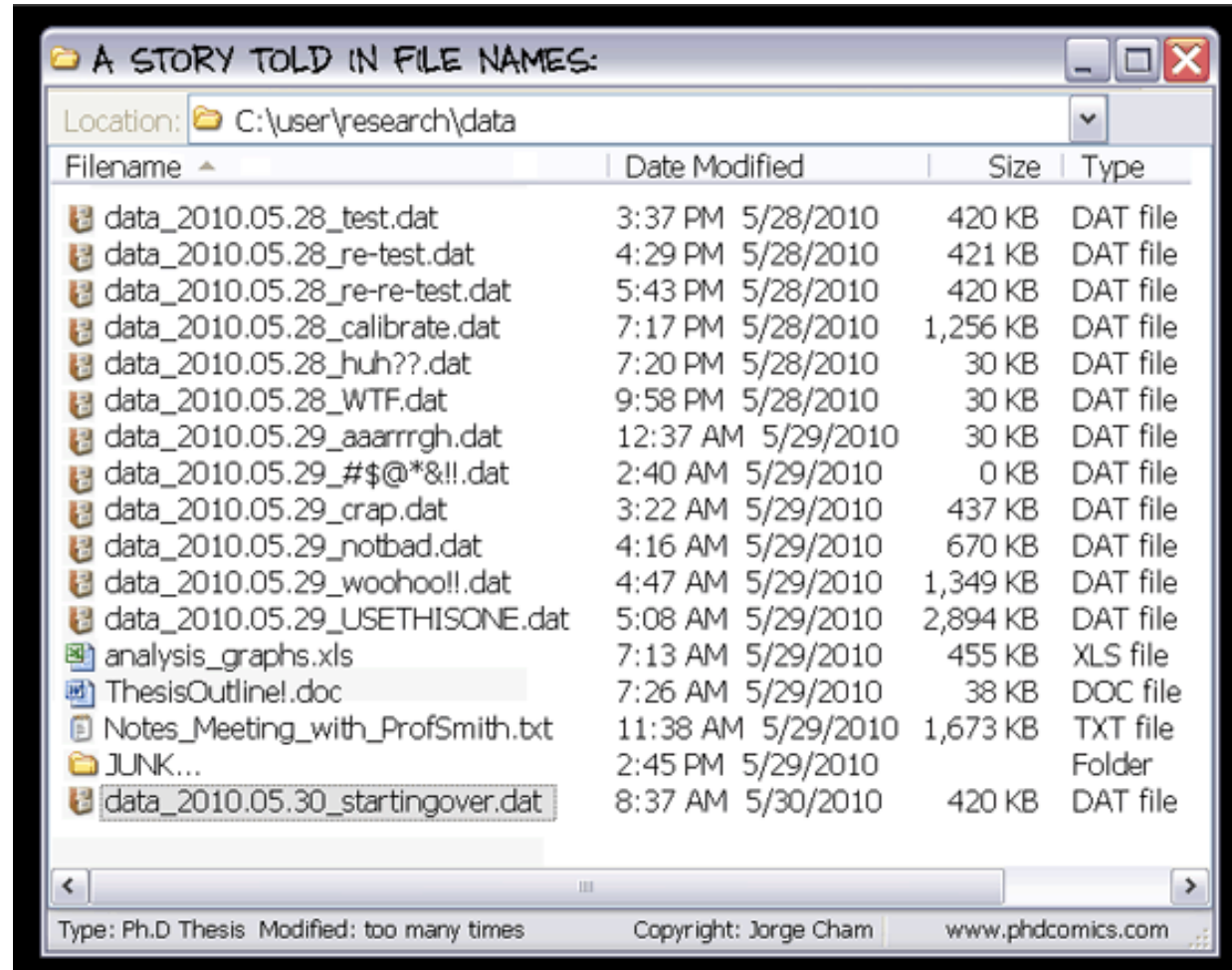
3 months

Slide from André König (URZ)



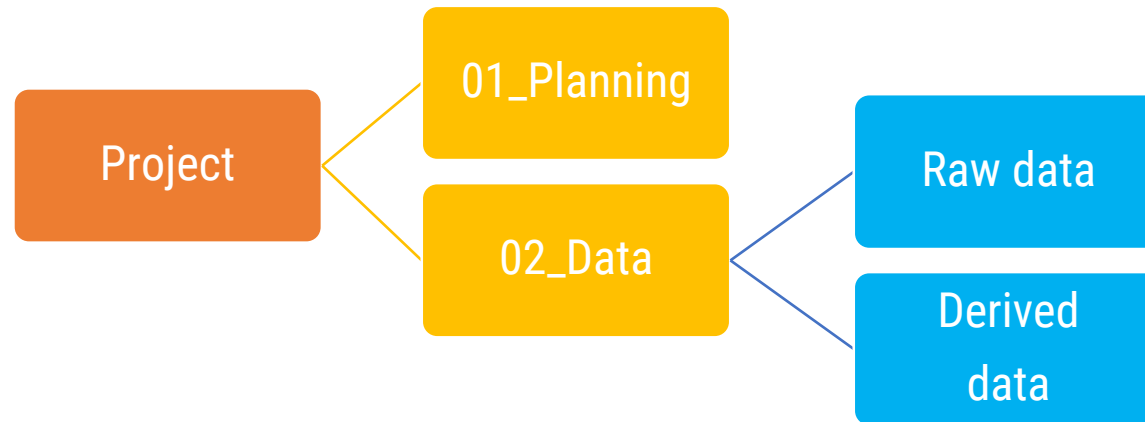
## — Data Organisation

# Data Organisation Gone Wrong!



---

# Data Organisation – Folder Structures

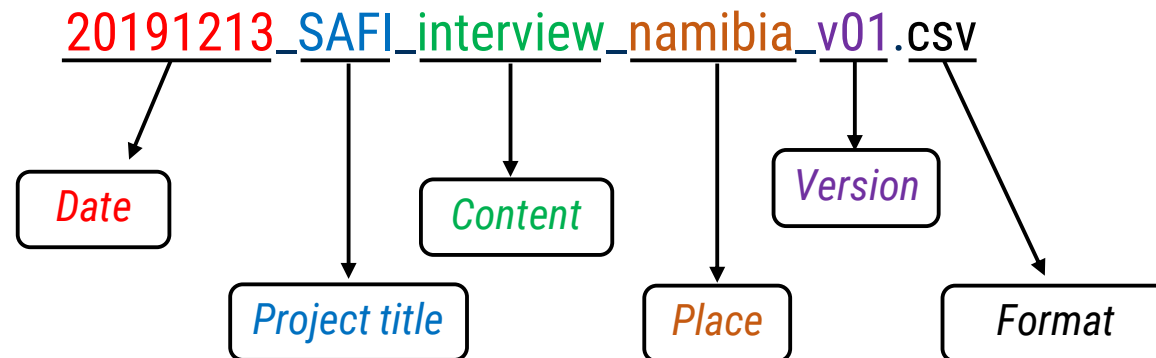


- Clear, hierarchical folder structure
- Short, telling names
- Numbering of folders (01, 02, ...)
- README file for explanations and abbreviations

# Data Organisation – Naming of Files and Folders

- Choose a convention, write it down and stick to it!
- As short as possible, as long as necessary
- Document versions (Version 01, 02,... Instead of 1, 2,...)
- README file may help (e.g. with abbreviations)

Example:



A close-up photograph of a person's hands writing in a notebook. The person is wearing a dark red, ribbed sweater. Their left hand is resting on the left page of the notebook, and their right hand is holding a yellow pencil, writing on the right page. The notebook has lined pages. The background is slightly blurred, showing a light-colored surface.

## — Data Documentation

---

# Why should you document your work?

- Essential element of good scientific practice
- Replicability (even after years)
- Easier collaboration
- Re-use of data by other persons
- Avoid unnecessary repetitions
- Efficient workflows now and in the future

„Without documentation research data is worthless.“





---

## Documentation of data

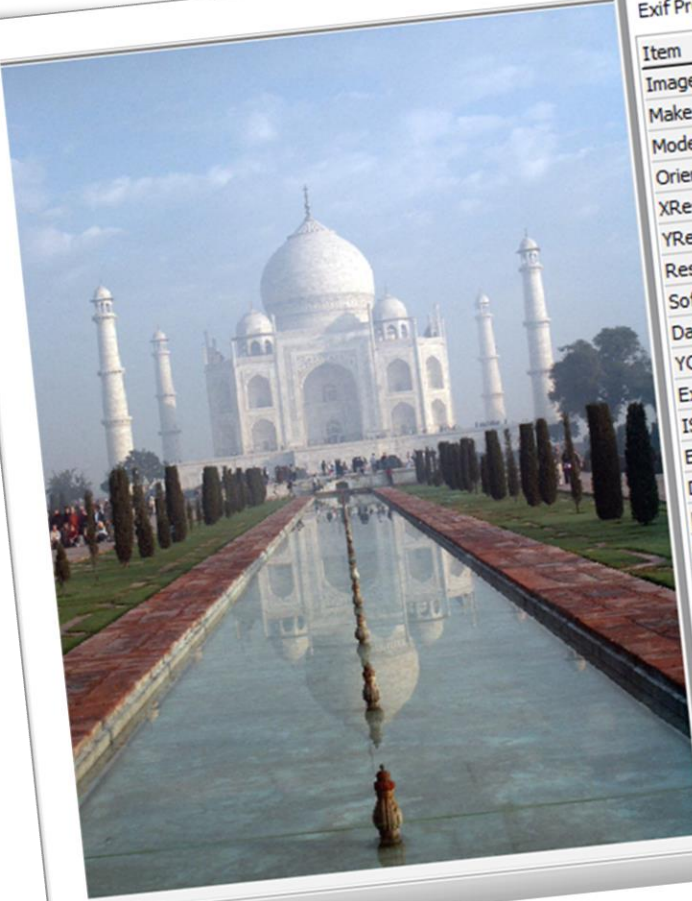
- WHO collected the data?
- WHAT is the content of the data?
- WHY were the data collected?
- WHEN were the data collected?
- WHERE were the data collected?
- HOW were the data produced and processed?



„Der, die, das,  
Wer, wie, was,  
Wieso, weshalb, warum...“  
[German intro of Sesame Street]

# What is metadata?

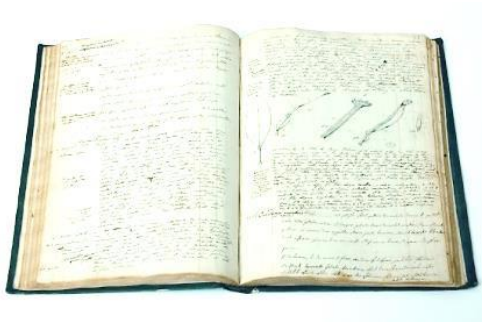
- „Data about data“
- Provide structured information about a certain aspects of data



Exif Properties		Maker Data	Summary
Item	Details		
Image Description	DCF 1.0		
Make	Minolta Co., Ltd.		
Model	DIMAGE S304		
Orientation	Normal		
XResolution	72.00		
YResolution	72.00		
Resolution Unit	Inch		
Software	Adobe Photoshop CS Win		
Date Time	2005:02:28 10:08:32		
YCb Cr Positioning	Centered		
Exposure Program	Normal		
ISO Speed Ratings	100		
Exif Version	"0210"		
Date Time Original	2001:01:02 15:43:30		
Date Time Digitized	2001:01:02 15:43:30		
Components Configuration	YCbCr		
Shutter Speed Value	0.0039 sec (1/256)		
Aperture Value	F6.0		
Exposure Bias Value	0/10		
Max Aperture Value	F3.7		
Metering Mode	MultiSegment		
Light Source	Unidentified		
Flash	Off		
Focal Length	11.81 mm		
Flash Pix Version	"0100"		

<http://www.alamy.com/photodata>

# How are metadata recorded?



**Lab book**



Human readable

```
<?xml version="1.0"?>
<dwr:DarwinRecordSet
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://rs.tdwg.org/dwc/dwcrecord/ http://rs.tdwg.org/dwc/xsd/tdwg_dwc_classes.xsd"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
  xmlns:dwr="http://rs.tdwg.org/dwc/dwcrecord/">
  <dcterms:Location>
    <dwc:locationID>http://guid.mvz.org/sites/arg/127</dwc:locationID>
    <dwc:country>Argentina</dwc:country>
    <dwc:countryCode>AR</dwc:countryCode>
    <dwc:stateProvince>Neuquén</dwc:stateProvince>
    <dwc:locality>25 km al NNE de Bariloche por Ruta 40 (=237)</dwc:locality>
  </dcterms:Location>
  <dwc:Occurrence>
    <dcterms:type>PhysicalObject</dcterms:type>
    <dcterms:modified>2009-02-12T12:43:31</dcterms:modified>
    <dcterms:rightsHolder>Museum of Vertebrate Zoology</dcterms:rightsHolder>
    <dcterms:rights>Creative Commons License</dcterms:rights>
  </dwc:Occurrence>
</dwr:DarwinRecordSet>
```

**Digital metadata**

(XML, README file, electronic lab notebook)



Human and machine-readable

- + searchable
- + structured
- + standardised

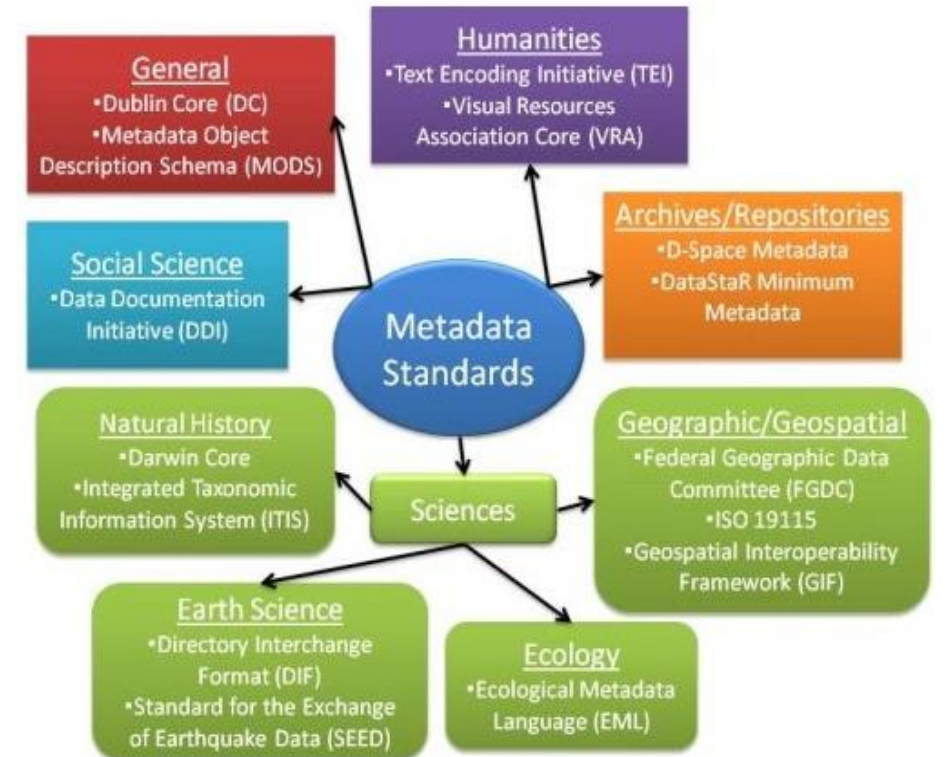
# Metadata standards

## Broadly applicable standards

- ISO 639 Code for the representation of the names of languages. e.g. eng – English; ger/deu – German
- ISO 8601 Codes for the representation of date and time. e.g. 2014-06-19T13:15:30Z

## Domain-specific standards

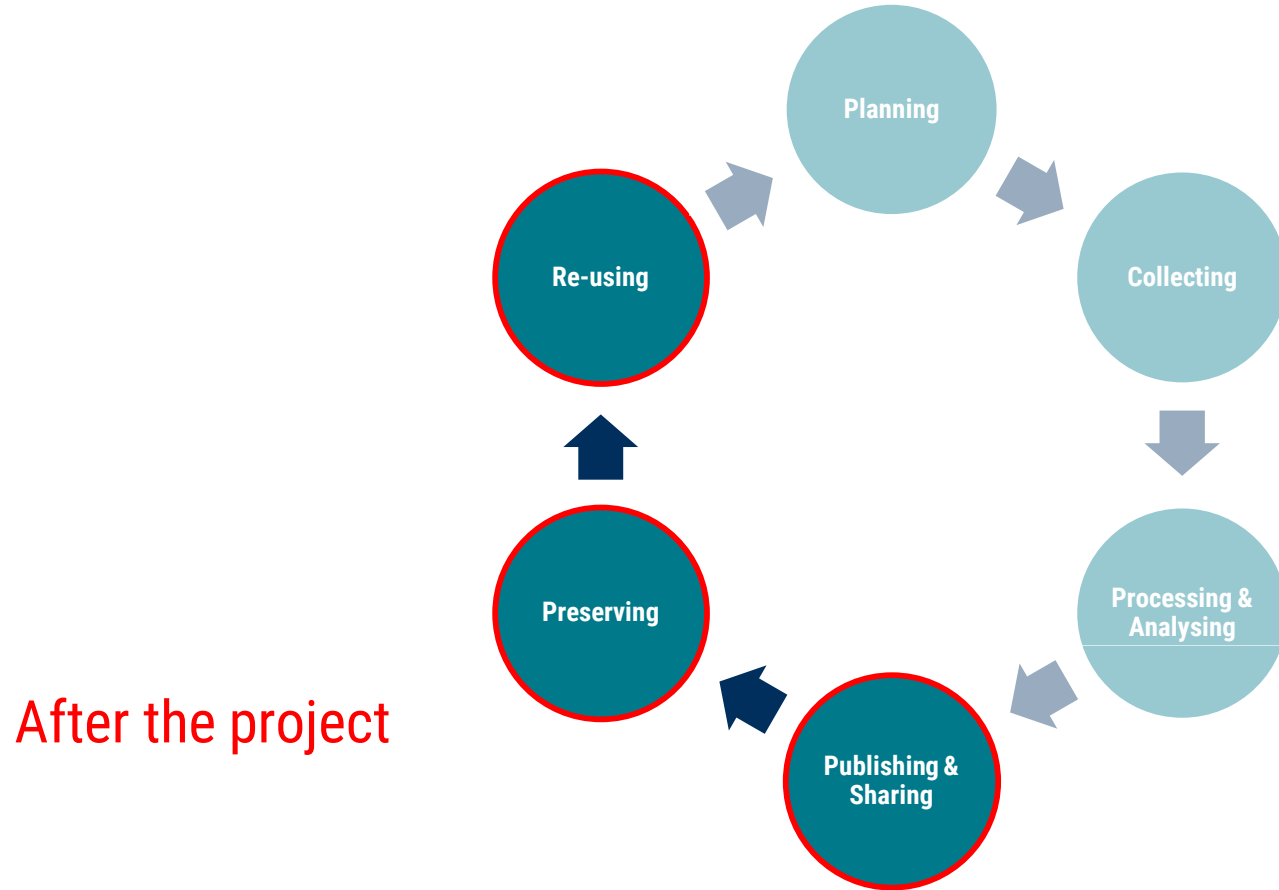
- Consistent terms/definitions/nomenclature
- Domain-specific vocabularies and ontologies



Metadata Concept Map by Amanda Tarbet is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

---

# Research Data Management – When and How?



---

# How will the data be managed **after** the project?

- ALL data have to be stored for at least 10 years (Good Scientific Practice)
- Data selections for long-term archiving (for eternity):
  - Data that can not be reproduced (unique measurements/observations in time and space)
  - Data that can only be reproduced with a lot of effort (e.g. cost/computer intensive analysis)
  - Data that are the basis for publications

---

# Data Preservation



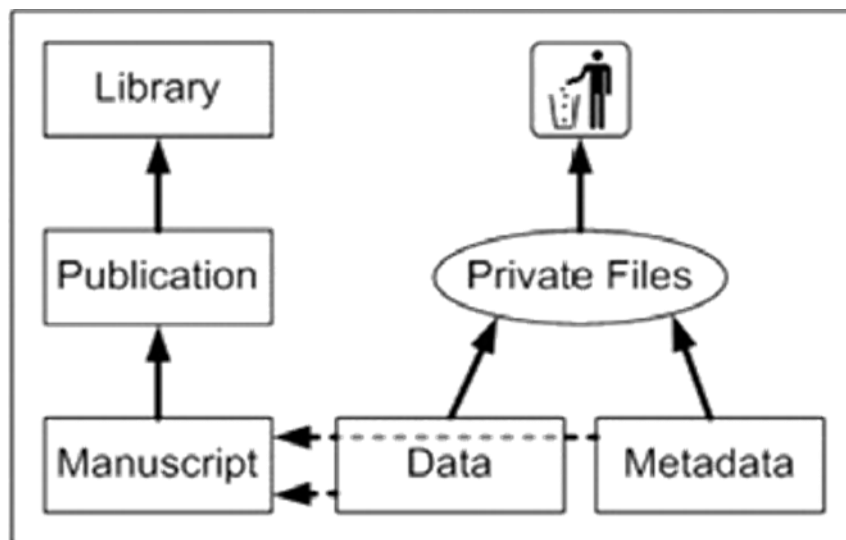
---

# How and where do I preserve my data?

- Data publication or data archiving?
- Maintenance of existing infrastructure?
- Migrating data into a (public) data repository?
- Legal considerations (sensible / personal data)?
- Access level of data (open, limited user group, closed)?
- Embargo periods?
- Guidelines by funding organisation (data publication encouraged / requested)?

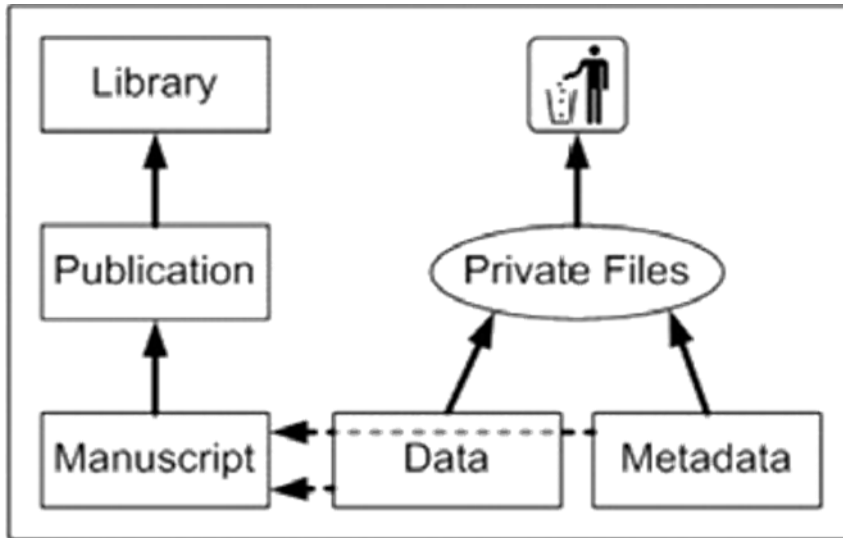


# Data Preservation / Publication: Ideal vs. Reality

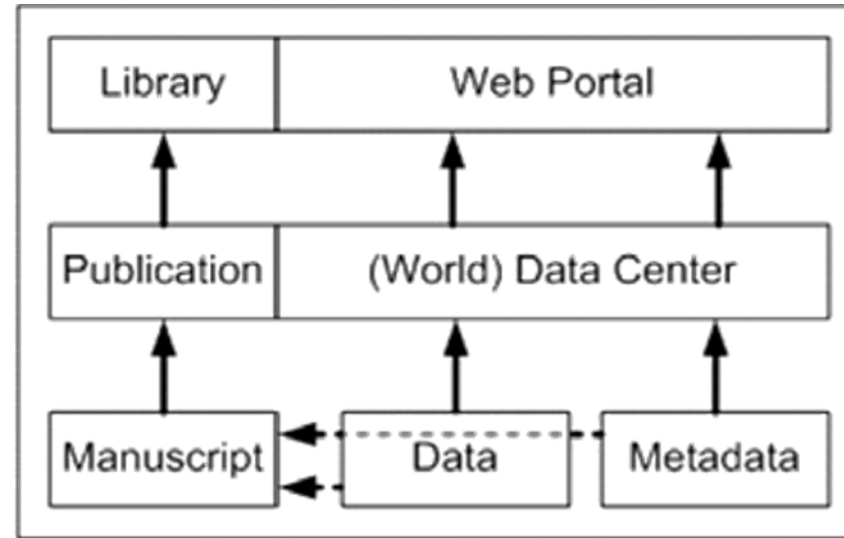


Reality

# Data Preservation / Publication: Ideal vs. Reality



Reality



Ideal



---

## Data Publication

# How can I publish my data?

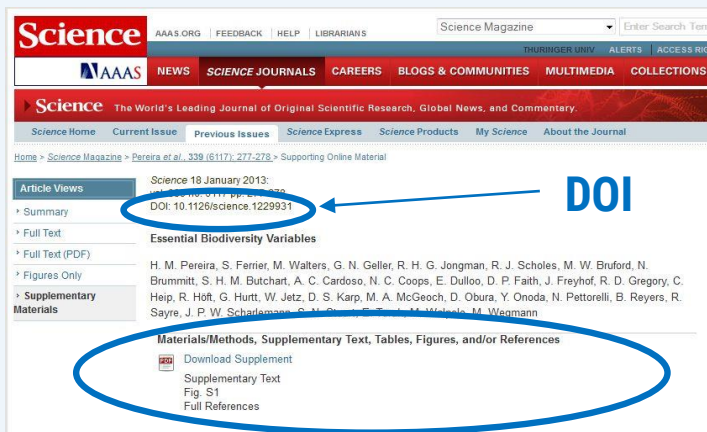
## Supplement of an article in scientific journal

The screenshot shows the Science journal website interface. At the top, there is a navigation bar with 'Science' logo, 'AAAS.ORG', 'FEEDBACK', 'HELP', 'LIBRARIANS', and a search box. Below this is a secondary navigation bar with 'NEWS', 'SCIENCE JOURNALS', 'CAREERS', 'BLOGS & COMMUNITIES', 'MULTIMEDIA', and 'COLLECTIONS'. The main content area displays the article title 'Essential Biodiversity Variables' and the authors 'H. M. Pereira, S. Ferrier, M. Walters, G. N. Geller, R. H. G. Jongman, R. J. Scholes, M. W. Bruford, N. Brummitt, S. H. M. Butchart, A. C. Cardoso, N. C. Coops, E. Dullo, D. P. Faith, J. Freyhof, R. D. Gregory, C. Heip, R. Hoff, G. Hurtt, W. Jetz, D. S. Karp, M. A. McGeoch, D. Obura, Y. Onoda, N. Pettorelli, B. Reyers, R. Sayre, J. P. W. Scharlemann, G. M. Stoeckli, S. Thorne, M. Walz, M. Wegmann'. A blue circle highlights the DOI '10.1126/science.1229931' with an arrow pointing to the text 'DOI'. Another blue circle highlights the 'Supplementary Materials' section, which includes 'Download Supplement', 'Supplementary Text', 'Fig. S1', and 'Full References', with an arrow pointing to the text 'Supplement'.

Supplement

# How can I publish my data?

## Supplement of an article in scientific journal



The screenshot shows the Science journal website. The article title is "Essential Biodiversity Variables" by H. M. Pereira et al. The DOI is 10.1126/science.1229931. Under the "Supplementary Materials" section, there is a link to "Download Supplement" and "Supplementary Text Fig. S1".

DOI

Supplement

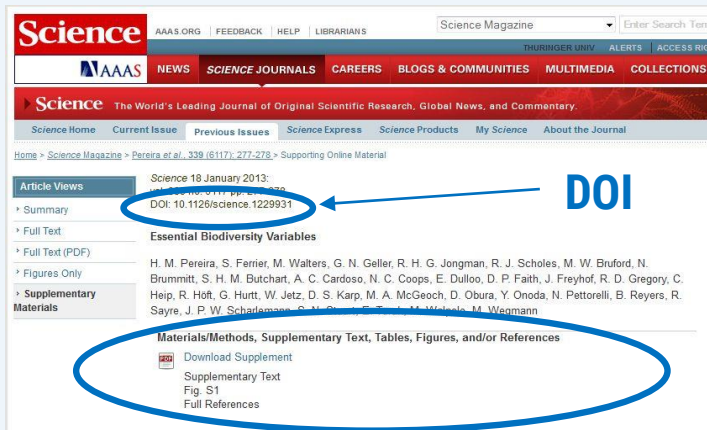
## Data Journal



A collage of logos for various data journals and publishers, including GENOMICS Data, (GIGA)<sup>n</sup> SCIENCE, RMetS Geoscience Data Journal, CODATA, Nuclear Data Sheets, DATA SCIENCE Journal, Biodiversity Data Journal, Journal of chemical & engineering data, npg nature publishing group, WILEY ONLINE LIBRARY, and SCIENTIFIC DATA.

# How can I publish my data?

## Supplement of an article in scientific journal

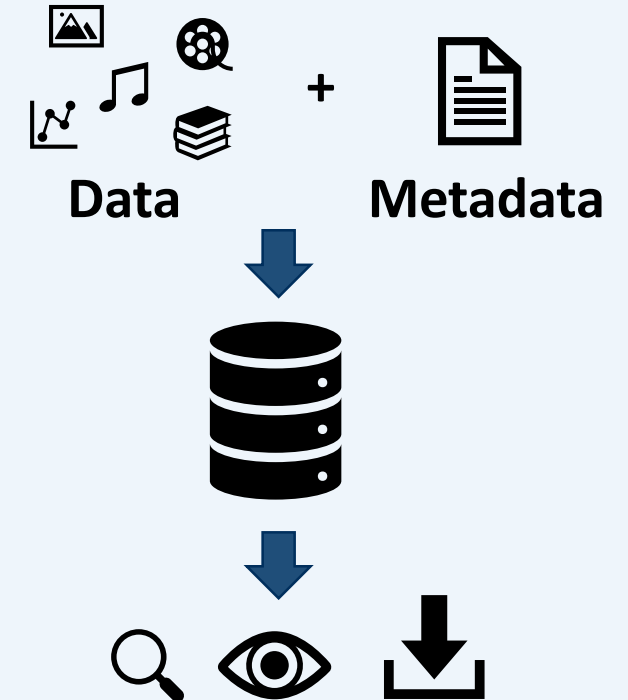


Supplement

## Data Journal



## Data Repository



---

# Which repository should I use?

- Search for suitable repositories (e.g. re3data.org)

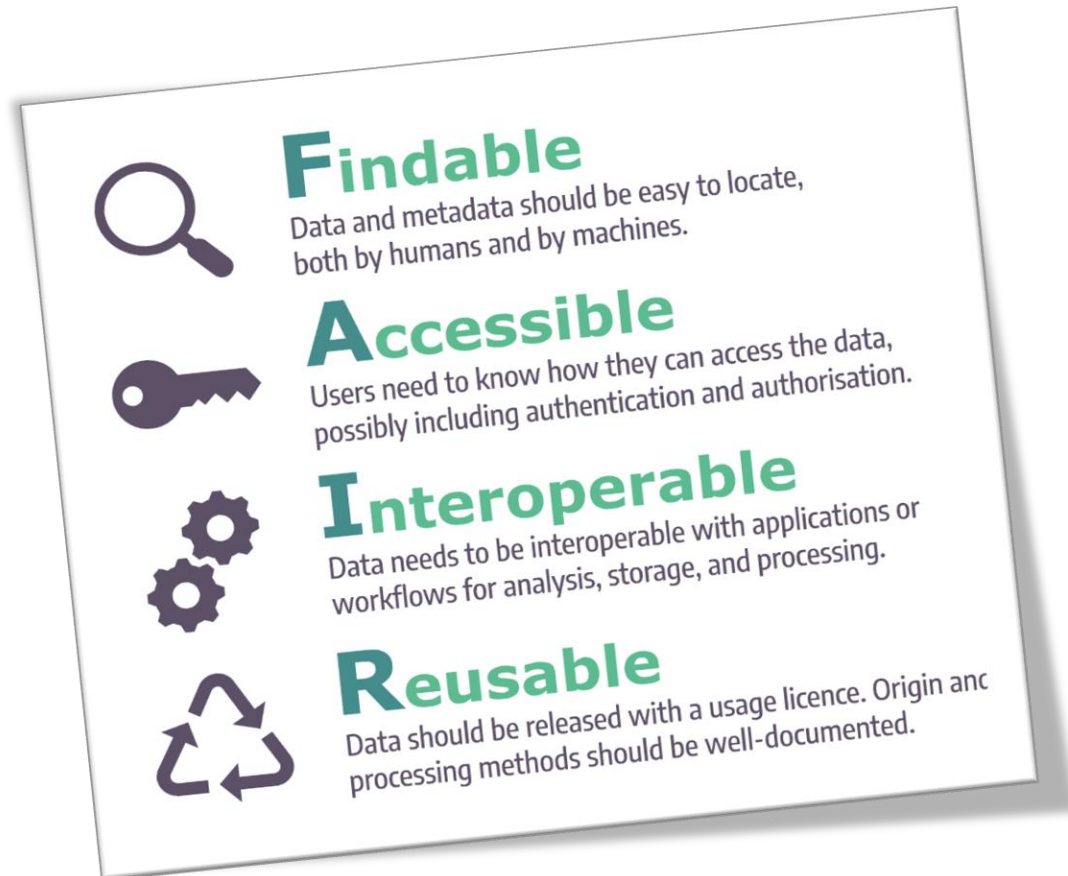


- Domain/topic-specific repositories (preferable)

- Generic or institutional repositories (if no specific repository is suitable)



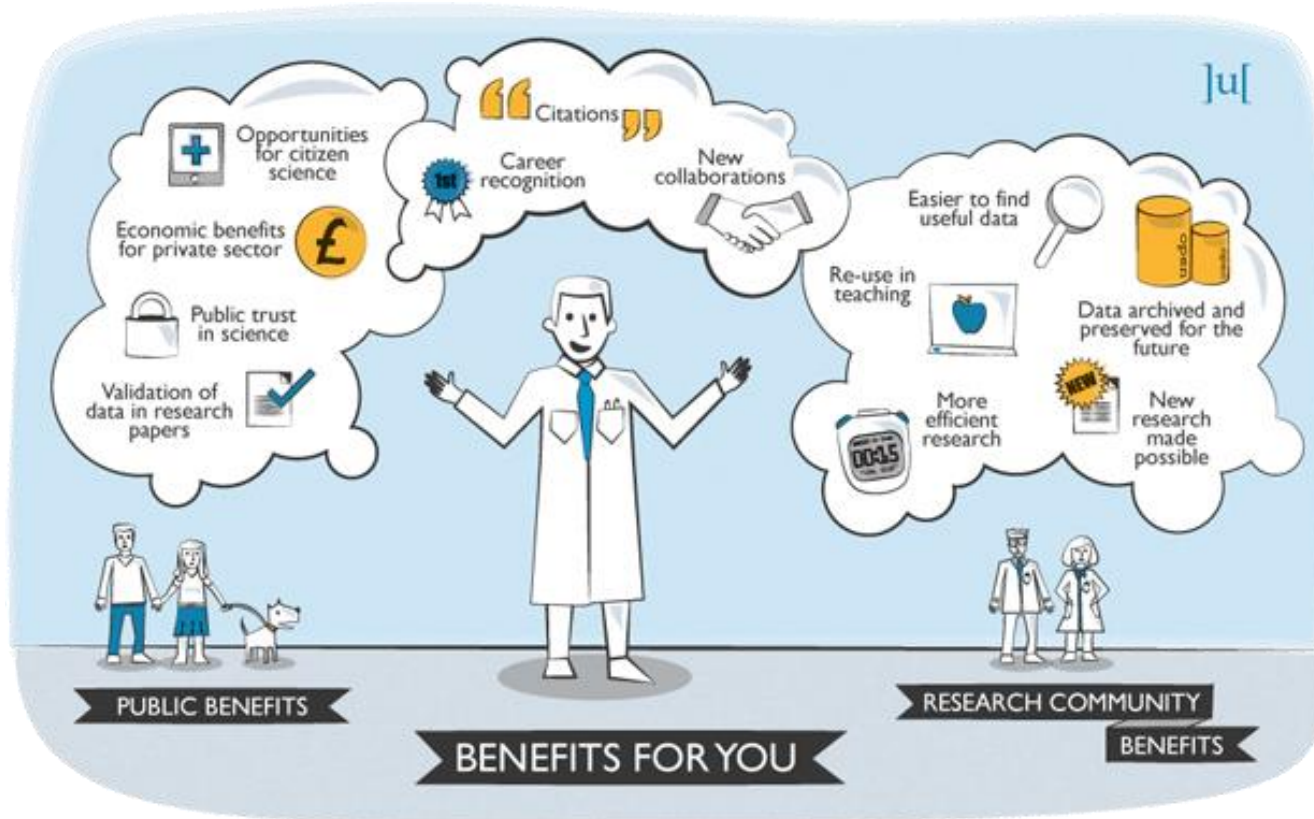
# What is FAIR Data?



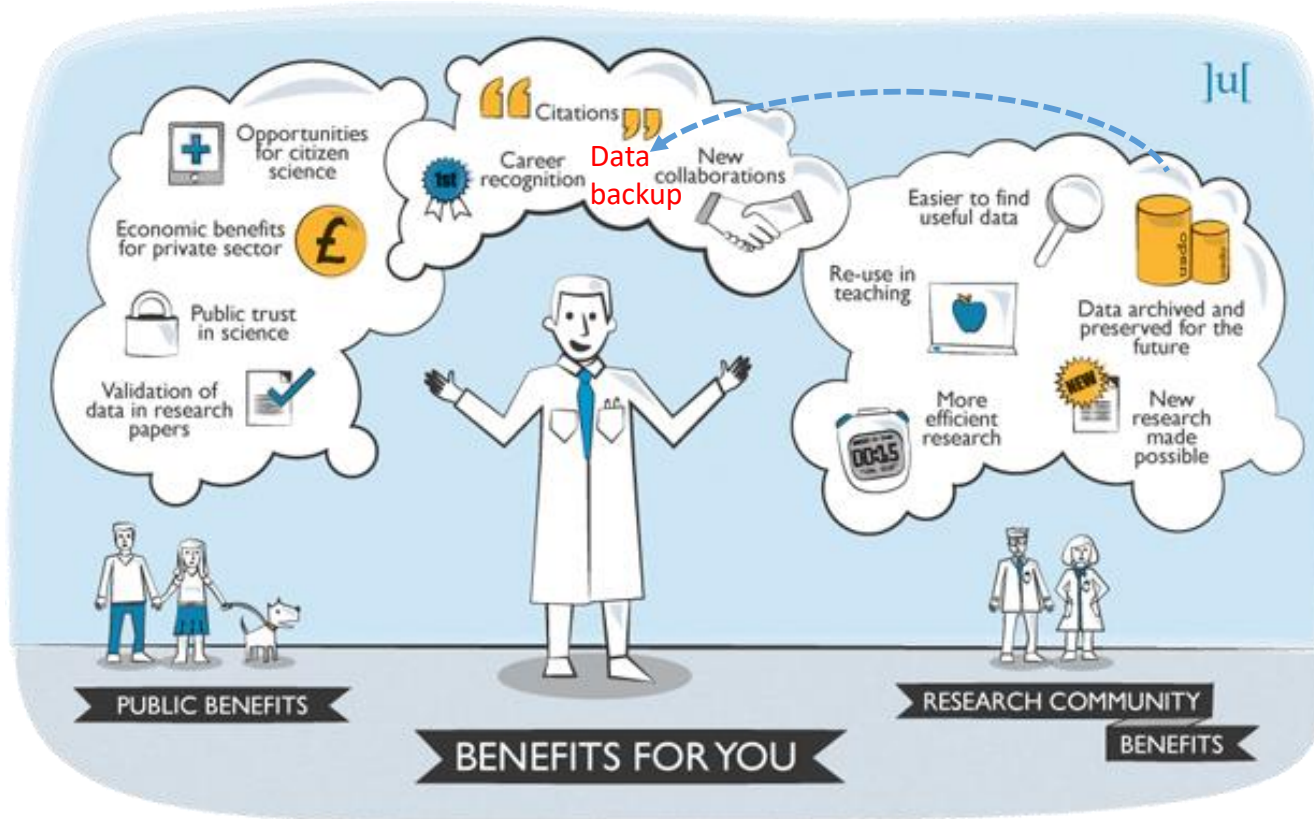
- Published in 2016
- Basic requirements for re-usable data defined
- Increasingly important in science
- Part of the DFG “Guidelines for Safeguarding Good Scientific Practice”



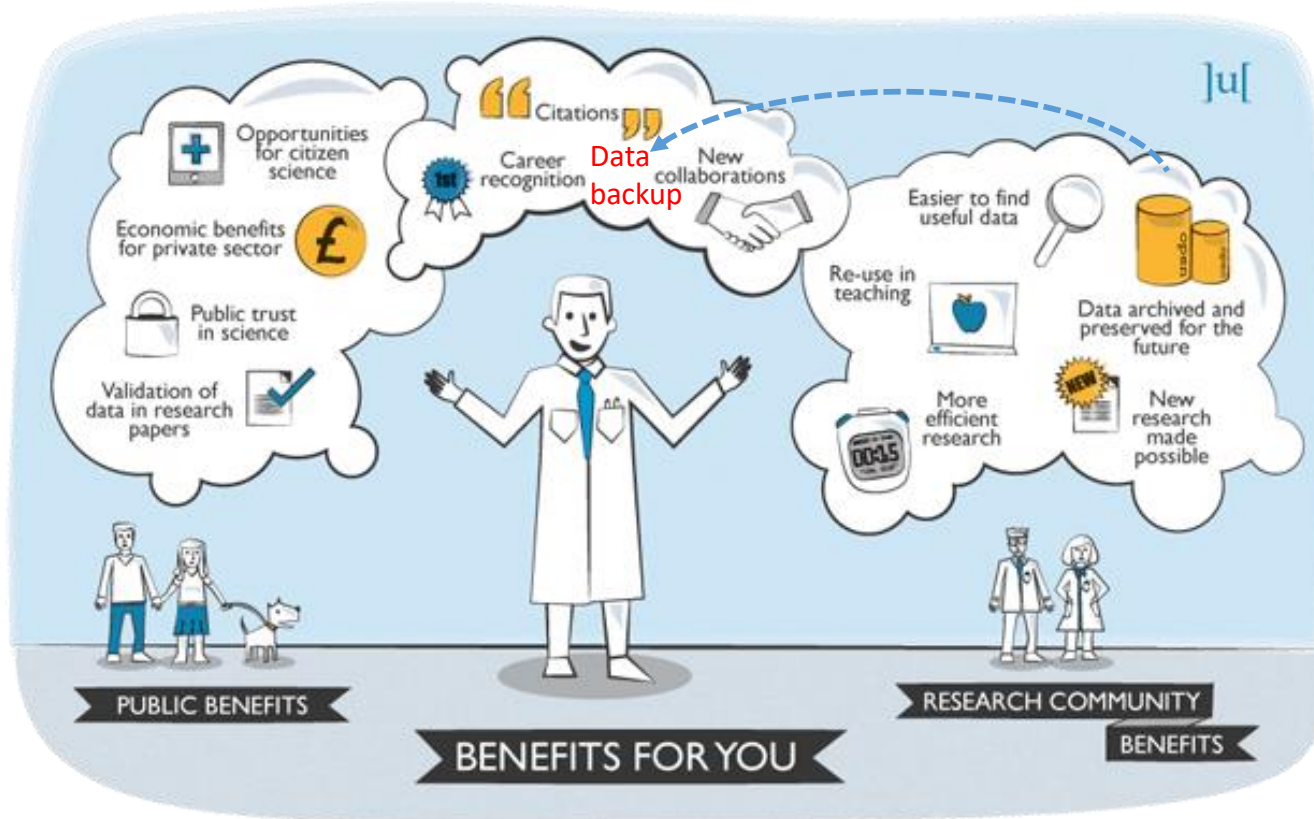
# Why should I publish my data?



# Why should I publish my data?



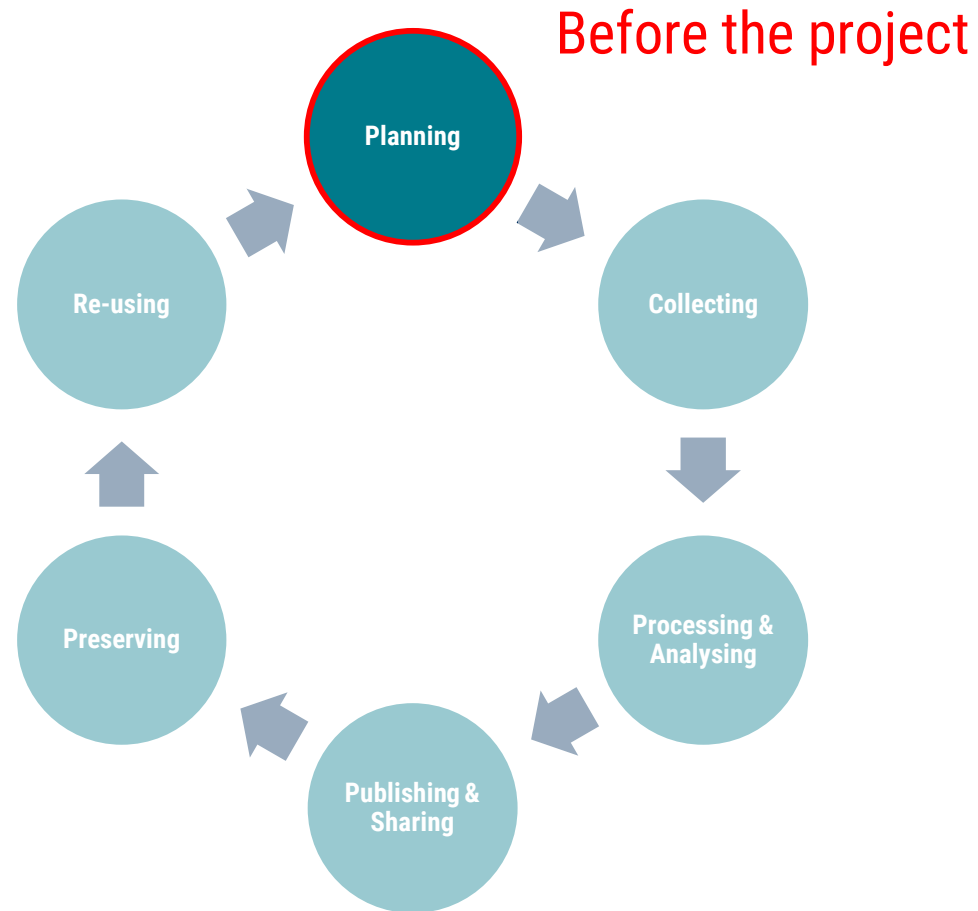
# Why should I publish my data?



& required by funding agencies, institutions and journals

---

# Research Data Management – When and how?



---

What needs to be planned **before** the start of the project?

To ensure that no aspect of research data management  
is forgotten during the project



Research Data Management Plan (DMP)

A hand with pink nail polish is writing 'My Plan:' in a notebook. The notebook is open on a wooden surface. The text is written in a cursive, handwritten style.

My Plan:

---

## Data Management Plans

# What is a DMP?

- (Formal) document that outlines how you will handle data during and after the project
- *Living document*
- Information about all stages of the data life cycle (incl. required resources)

## 1. Data Collection and Documentation



What kind of data are generated

How will data be generated

What metadata are needed

## 2. Ethics, legal and security Issues



How will ethical issues be handled

How are the data accessed

Are there copyright issues

Are there sensitive data

What about intellectual property rights

## 3. Data Storage and Preservation



How are the data stored?

Are there back up systems

How are data safely preserved

## 4. Data Sharing and reuse



How and where will the data be shared?

How are sensitive data protected

How can data be accessed

<https://www.uzh.ch/blog/hbz/2018/11/15/data-management-plan-in-a-nutshell/>

---

# Why is a DMP needed?

- Provides a guideline during the project
- Defines responsibilities
- Several funding agencies request it as part of a grant proposal (e.g. EU)





---

## Research Data Management Support

---

# The Research Data Management Helpdesk

- Founded in 2015
- Central contact point for all aspects of RDM



Roman  
Gerlach



Annett  
Schröter



Jeanin  
Jügler



Bettina  
Färber



Volker  
Schwartze

---

# The Research Data Management Helpdesk



## Individual consultation

- Advice on all topics of RDM
- Before, during and after projects
- e.g. data publication, DMP, funding proposals



## Teaching and Training

- Graduate academy courses
- Sessions in existing lectures
- Workshops for departments, projects, etc.
- Coachings incl. development of DMP



## Information material on our website

- Links to material on different RDM topics
- Development of new material
- e.g. online courses, information sheets, policies, templates



## Public relations / Awareness

- Raising awareness for RDM
- e.g. FDM-Tage, Coffee lectures

---

## Further support



### Universitätsrechenzentrum (computer center)

- Storage space and backup
- Cloud services
- Software



### Rechtsamt (legal office)

- Legal advice and forms
- e.g. protection of personal/patient data



### Ethics Committee

- Consultation for projects
- Templates for declarations of consent and information for volunteers



### Thüringer Universitäts- und Landesbibliothek (ThULB)

- Access to literature, information and data
- Publication of research data (DBT)

# Upcoming events

GA Course  
„Make your data count“  
(Life Sciences)

26. + 28.4. 2021

GA Course  
„Nachhaltig forschen“  
(Humanities)

17. + 19.5. 2021

Das Thüringer Kompetenznetzwerk  
Forschungsdatenmanagement präsentiert...

Forschungsdatenmanagement  
**Online Coffee Lectures 2021**

27.01.2021	5S Data: Setz dich auf deine 5 Buchstaben und organisiere deine Daten! Kevin Lang (Bauhaus-Universität Weimar)
24.02.2021	Datendokumentation: Die Basis hoher Datenqualität Roman Gerlach (Friedrich-Schiller-Universität Jena)
31.03.2021	Der Datenmanagementplan: Früh geplant, ist halb gewonnen Jessica Rex (Technische Universität Ilmenau)
28.04.2021	Rechte an Forschungsdaten Rechtsamt (Friedrich-Schiller-Universität Jena)
26.05.2021	Open Science in allen Facetten Annette Strauch (Stiftung Universität Hildesheim)
30.06.2021	Publikation von Forschungsdaten: Ein Gewinn für alle! Annett Schröter (Friedrich-Schiller-Universität Jena)

---

# Where can I get further information?



[German speaking RDM website](#)



[European Collaborative Data Infrastructure](#)



[Research Data Management Training](#)



[Data Management Short Course for Scientists](#)



Thank you for your attention!

I am happy to answer your questions.

---

## Pleas quote as follows

Gerlach, Roman, Färber, Bettina, König-Ries, Birgitta, Schröter, Annett, Rzymiski, Christoph, Schwartze, Volker, Steiner, Petra & Olena Tykhostup (2021) Good Research Data Management: From Theory to Practice. Presentation by Volker Schwartze, 04.02.2021, Promovierendentag der Friedrich-Schiller-Universität, Jena.



Copyright licence: (CC BY 3.0 DE)