eResearch Alliance

Love your data!

Data Management Basics for study and research

12.08.2020

Timo Gnadt

gnadt@sub.uni-goettingen.de

eResearch Alliance

Survey: Who are you? (1)

A: I am a bachelor student

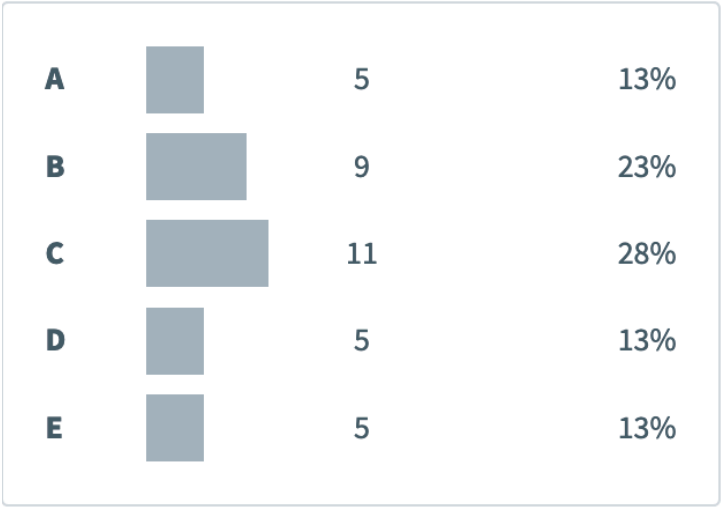
B: I am a master student

C: I am a doctoral/PhD student or scientific assistant

D: I am a PostDoc

E: None of the above

Answers from the participants:



Survey: Who are you? (2)

I am from the …

A: natural/structural sciences

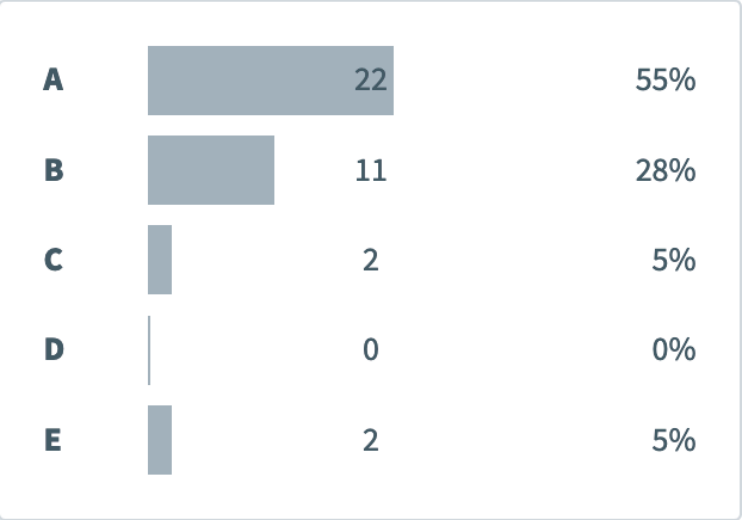
B: arts and humanities

C: social sciences

D: Jurisprudence

E: None of the above

Answers from the participants:



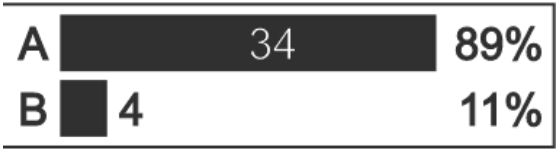
Survey: Who are you? (3)

I am studying / doing a PhD / working

A: in Göttingen

B: not in Göttingen

Answers from the participants:



Structure

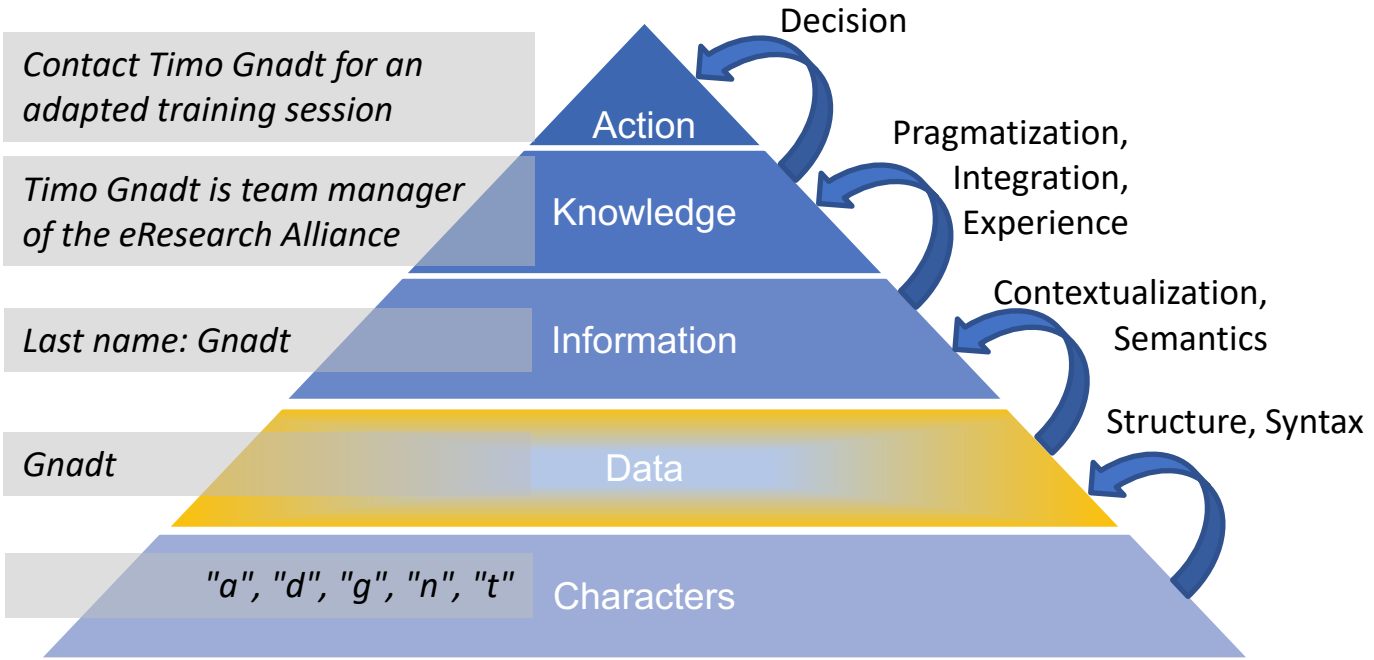
- Introduction to Research Data Management
- Storing data
- Organizing data
- Describing data
- Sharing data
- Further links and information

Comments, questions and suggestions for this training can be entered here:

https://pad.gwdg.de/LoveYourData_20200812?both

Introduction to Research Data Management

What is 'data'?


Contact Timo Gnadt for an adapted training session
Timo Gnadt is team manager of the eResearch Alliance
Last name: Gnadt
Gnadt
"a", "d", "g", "n", "t"
Action
Knowledge
Information
Data
Characters
Decision
Pragmatization, Integration, Experience
Contextualization, Semantics
Structure, Syntax

What is 'data'?

"A re-interpretable representation of information in a formalised way, suitable for communication, interpretation or processing."

translated into Digital Curation Centre

Data are representations of observations, objects or other units used as evidence of phenomena for the purpose of research or scholarship.

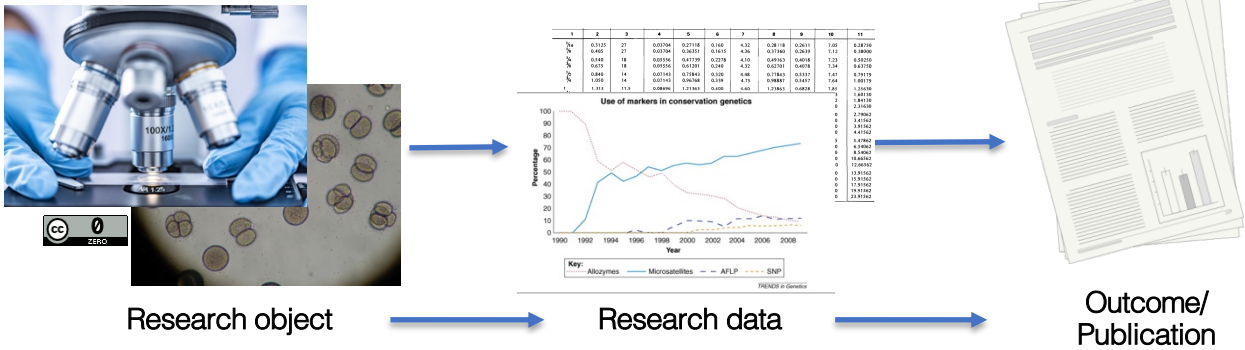
(*Christine Borgmann, UCLA,2014*)

What is research data?

All representations of information that you use in your research:

Types: Statistics, interviews, simulations, measurement data from experiments, observation data from instruments, text with semantic annotations, 3D scans, model drawings, numerical representations, ...

Shapes: Video, audio, images, spreadsheets, e-mails, paper documents, binary data, software, text files, lab notebooks, ...



Graph from: Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick PW (2010). Conservation genetics in transition to conservation genomics. Trends Genet 26: 177-187.

What is *your* research data?
or: What will *your* research data be?

Answers from the participants:

- genomic data
- metric data derived from GIS grids
- Archival materials
- electrophysilogical recordings
- Video data
- neuroimaging (MRI)
- observation from field testing
- Genomic data
- observational data from biodiversity surveys
- material and geological data
- Observations from greenhouse plants
- Fluorescence, electrophysiological data
- Oral History data and Archival materials
- Manuscripts
- field survey

Types of research data

Type	Characteristics	Example
Observations	Data is collected in real time Mostly irreplaceable	Sensor data Survey data
Experiments	Mostly created in the laboratory Reproducible but expensive	Gene sequences Chromatogram
Simulations	Generated from test models Model and metadata more important than output	Climate models Economic models
Derived data	Derived or compiled from other data, reproducible	Text Mining 3D models
References	Collection of smaller data sets Mostly published	Gene sequence database Primary text sources
Digital copies	Digital version of an analog object, reproducible as long as the original exists	Manuscripts

Source: Dominique Ritze, Kai Eckert and Magnus Pfeffer. Research data.
In: PatrickDanowski, (Open) Linked Data in Libraries; 122-138, DeGruyter Saur, Berlin, 2013

Research data - a valuable investment

Source: European Space Agency: Rosetta and Philae at comet, on flickr.

Rosetta Mission, 2004-16

Duration:

- >10 years preparation
- 10 years from the beginning to the data

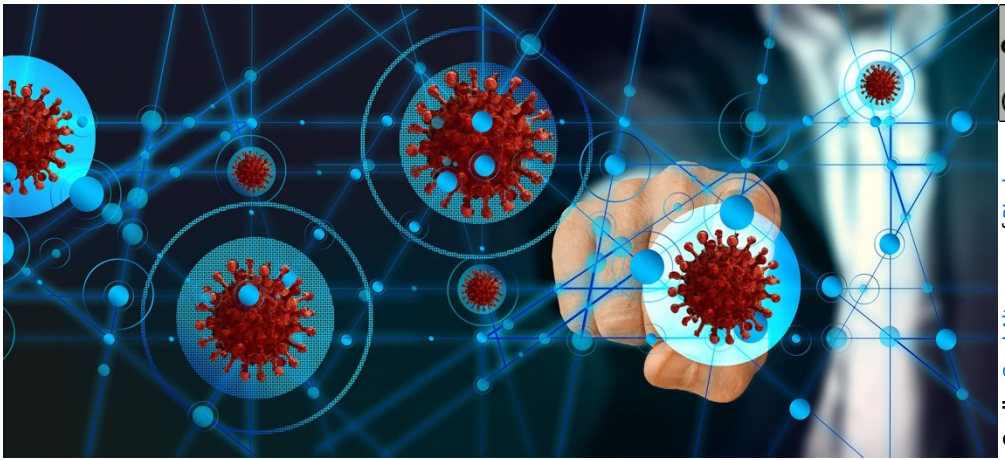
costs:

- over € 1.000.000.000

Result:

- some cool photos
- many data
- *a radically new theory about the origin of the universe?*

Research data - a resource in great demand

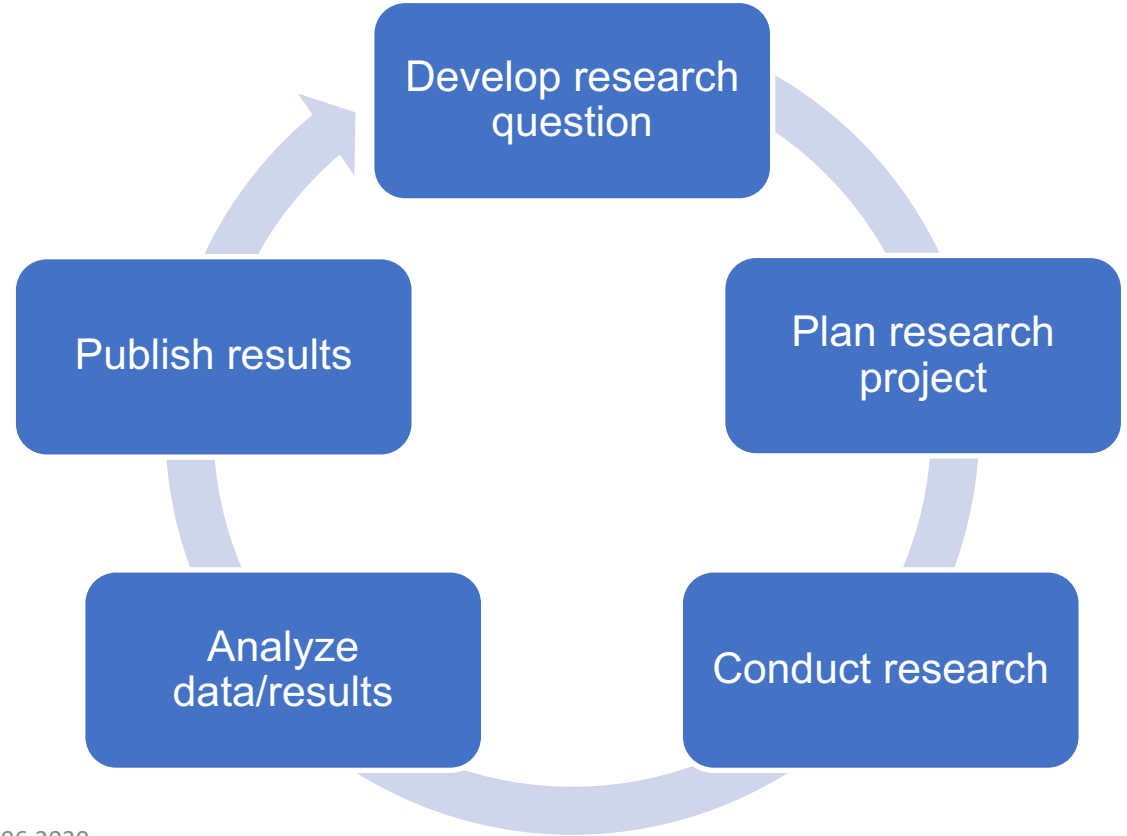


Quelle: Gerd Altmann auf Pixabay, CC0

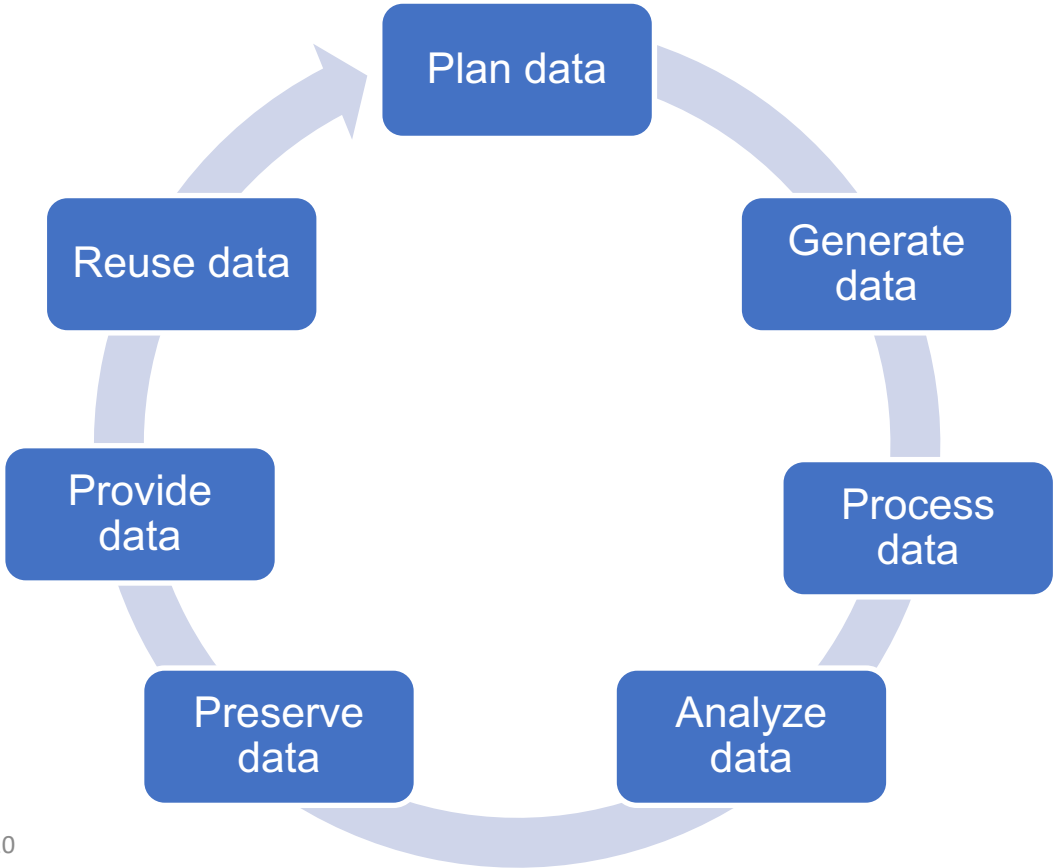
Data with reference to COVID-19

- medical, molecular biological, epidemiological...
- mathematical, sociological, geographical, psychological, …

Research cycle



Research data cycle

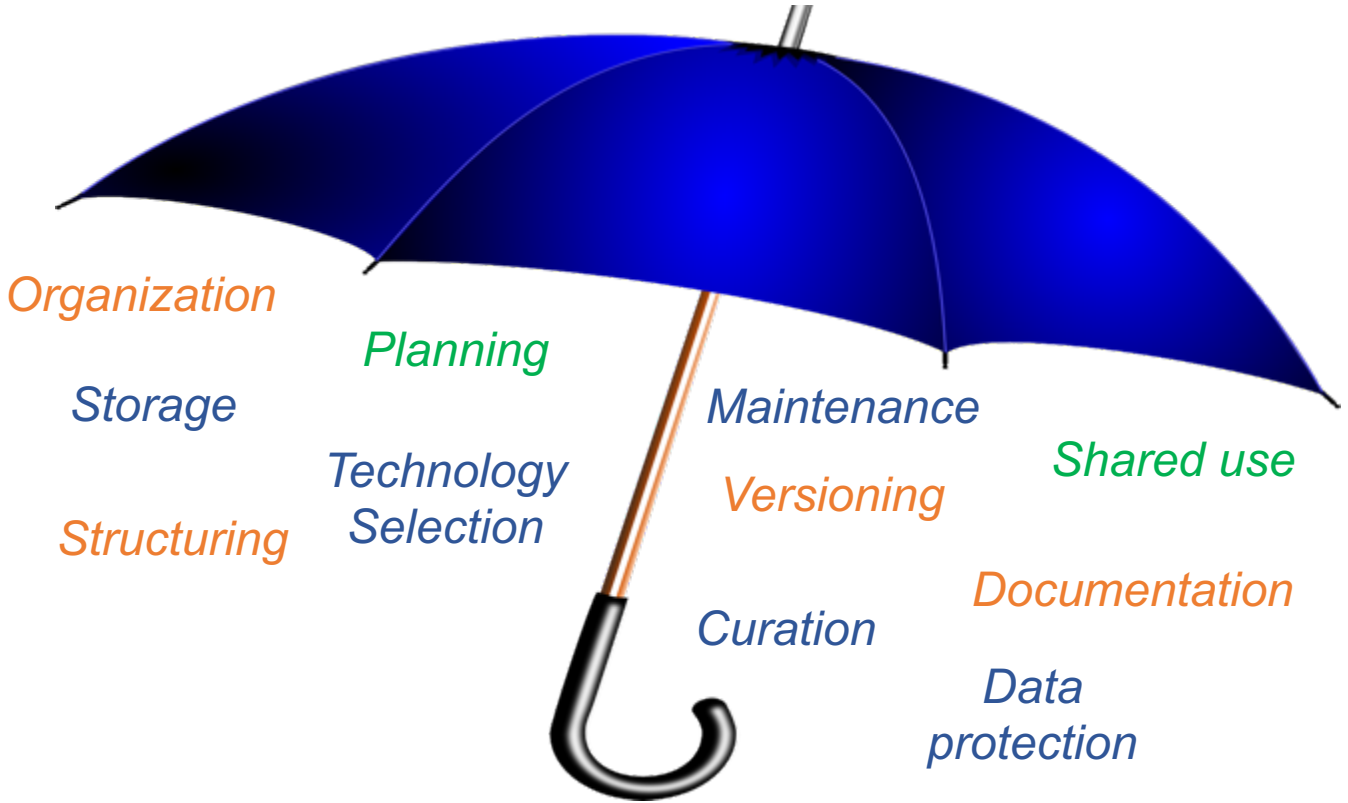


What is research data management?

Answers from the participants:

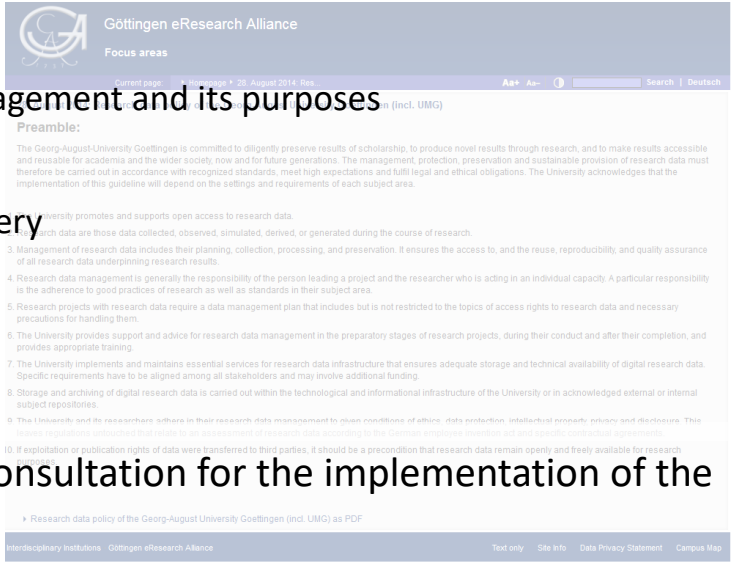
- propper documentation and sustainable storage
- data collection and classification
- sorting out the methods for saving data in the way to then easily find it. Also systematise the data according to the content or technical characteristic..
- saving and making data accessible to others
- finding the digital space to store it

What is research data management?



Research data policy of the Georg-August-University of Göttingen

- Officially published on August 28, 2014
- One of the first German universities with such a guideline
- Topics covered:
  - Research data, research data management and its purposes
  - Data management plans
  - Support, training and service delivery
  - Storage solutions
  - Ethical and legal standards
  - Open Access
- **eResearch Alliance:** Support and consultation for the implementation of the policy for the Göttingen Campus



Source: http://www.uni-goettingen.de/en/488918.html

Why research data management?

Answers from the participants:

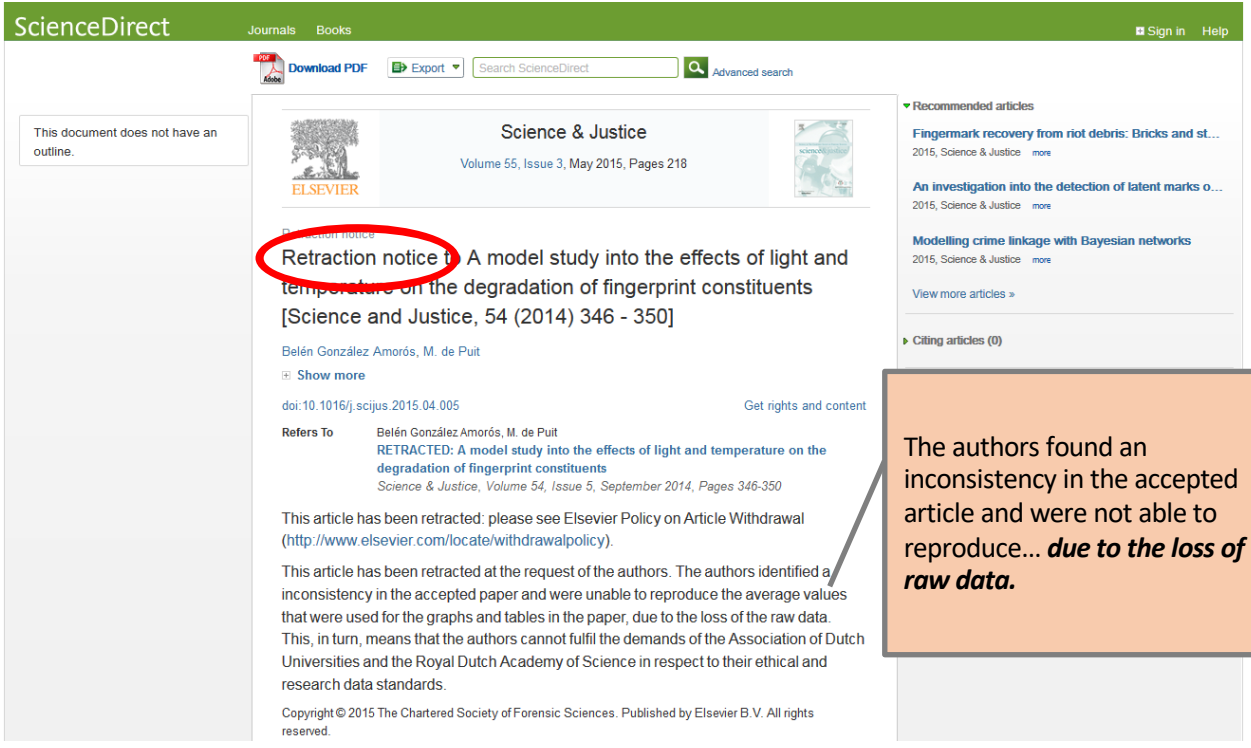
- possibility of re-analysis
- collect data for a larger context than own data

Why research data management?

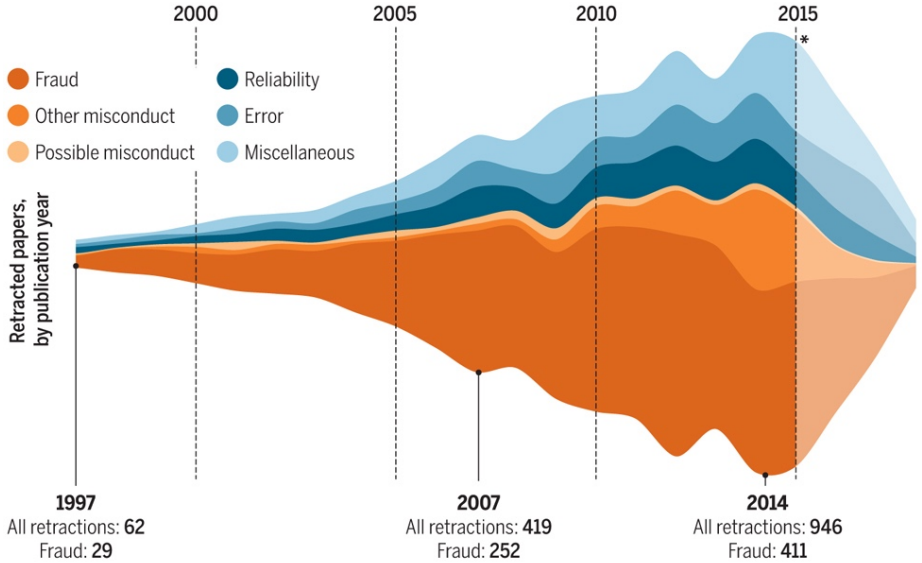
- **Increase your efficiency and the value of your work / research**
  - Prevent data loss
  - Avoid unnecessary work
  - Improve data quality
- **Good scientific practice**
  - Reproducibility, accountability and compliance
  - "Primary data as a basis for publications must be stored securely in the institution where they were created for ten years in a permanent form". *(DFG, Proposals to safeguard good scientific practice, 1998)*
  - Request from funding organisations (e.g. EU Horizon 2020)
- **Data exchange with colleagues**
  - Research can be *very* expensive, and the only result of long research projects and trips is often data.
  - Compared to the costs of data collection, the costs of data management are low.
  - Productive data exchange is simply a matter of efficiency.

Why research data management?

Why research data management?



Why research data management?



Sources:

Jeffrey Brainard et al., **Rethinking retractions,** *Science* 26 Oct 2018: Vol. 362, Issue 6413, pp. 390-393
DOI: 10.1126/science.362.6413.390
GRAPHIC: J. YOU/SCIENCE; DATA: RETRACTION WATCH

Why research data management?

- **Improve your research**
- **Good scientific practice**
- **Data exchange with colleagues**
- **Data Publication**
  - Required by increasing number of journals
  - Avoid retractions
  - Get cited for your data!
- **Enabling new types of research**
  - Feedback loops between empirical and modelling approaches
  - Initiation of research questions in completely different areas

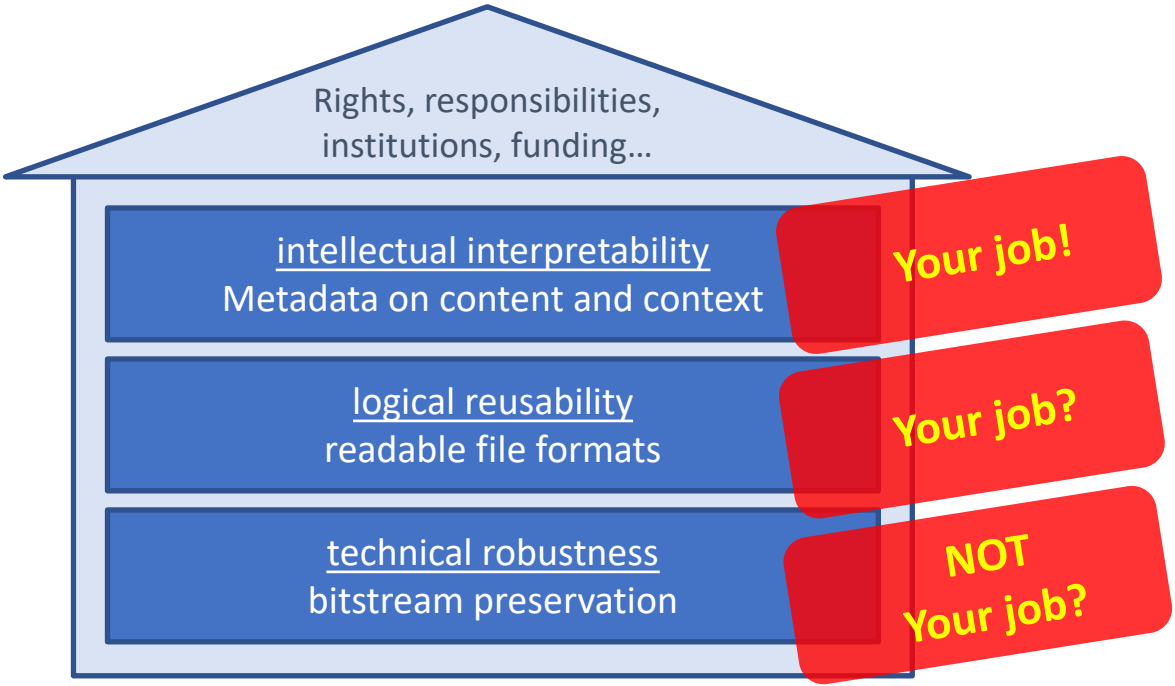
Publications are arguments of authors, and data are the evidence to support these arguments.

(Christine Borgmann, UCLA, 2014)

The deeper meaning of Research Data Management

Source: cmhughes on pgfplots

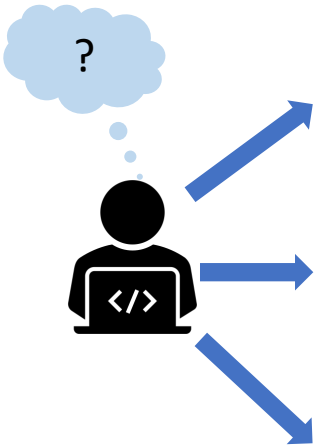
Levels of data curation



Illustrative video: Data management SNAFU in 3 short acts

Storing data

Why do I want to store my data?



Purpose	Solution
short-term storage, backup copy	• institutional backup • individual backup
long-term storage, archiving	• Data archive, • institutional archive solution or • individual archiving
making data available to others	Data repository: • institutional, • generic or • subject-specific

Yes, we store – what for?

	Backup	Archiving	Sharing
Purpose	Ability to recover data in case of data loss or error propagation	Enabling validation through persistent storage of data used in publications	Enable review, citation and reuse of data sets (data exchange)
Characteristics	Duplication of current work data & interim results	Archive format (e.g. zip) with all associated & relevant data / files (ideally including metadata)	Format defined by the repository; discipline-specific metadata standards
Regularity	Regularly during the work phase or project duration	Once for each relevant data set, usually at the end or after the work phase	Once for each selected data set, either during or after the work phase
Expenses	Varying, e.g.: set up once, check regularly	Establish predefined procedure with data archive (e.g. data/computing center)	Documented process, sometimes accompanied by repository support

Survey: Backup 1

Do you back up your research data?

If so, how?

Answers from the participants:

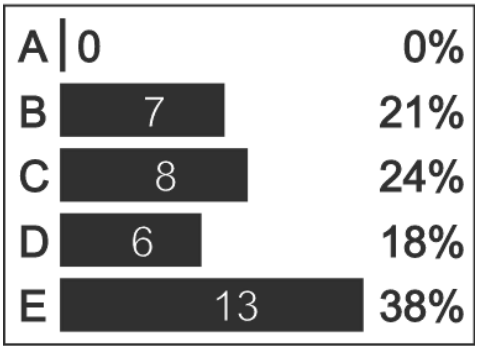
A: No, never.

B: Yes, on a USB stick.

C: Yes, on an external hard disk or CD/DVD.

D: Yes, in the cloud.

E: Yes, through a service of my institute/university.



Survey: Backup 2

How often do you back up your research data?

A: Not at all

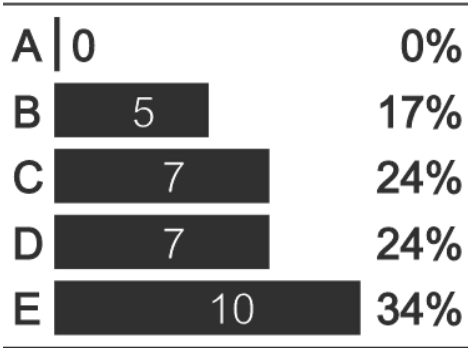
B: 1-2 times a year

C: 3-12 times a year

D: Several times a month

E: (Almost) daily

Answers from the participants:



Survey: Backup 3

Have you ever tried to recover a deleted file?

Answers from the participants:

A: No, never.

B: Yes, but not successfully

C: Yes, successfully but with high effort

D: Yes, that's no problem for me

Survey: Backup 4

Can you revert to an earlier version of a file?

Answers from the participants:

A: No, I don‘t know how.

B: Yes, I think so.

C: Yes, I've done that before.

D: Yes, that's no problem for me.

Survey: Backup 5

Who is responsible for backup and storage services at your institute?

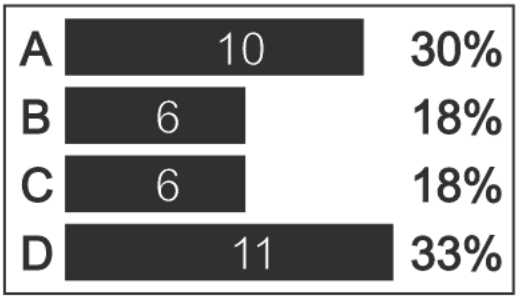
Answers from the participants:

A: I don't know.

B: I think I could find out.

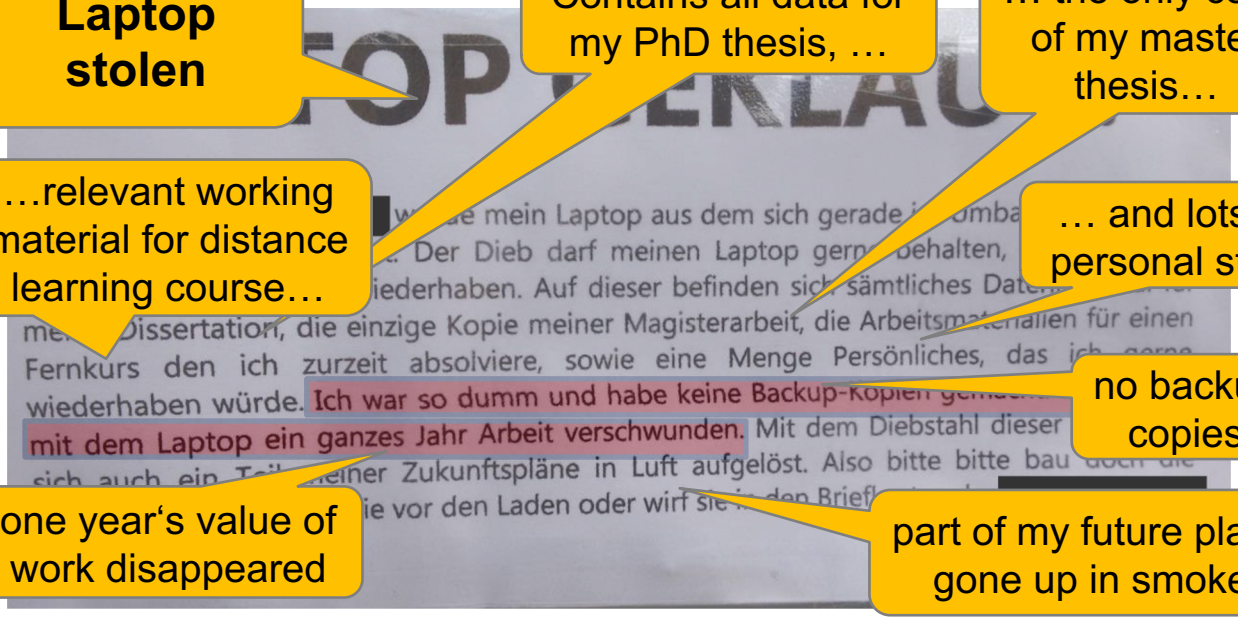
C: I know who I can ask about that.

D: I have already had contact with them.



Why backup?

Notice at the bus stop Jüdenstraße/Göttingen



Why backup?

Source: Gino on flickr

Source: steviep187 on flickr

Source: Kjell Eson on flickr

Why backup?

Source: University of Southampton, Department of Electronics and Computer Science, 2005

Source: Frankfurt University of Applied Sciences,

March 2020
© Frankfurt Fire Department

Causes for data loss

Answers from the participants:

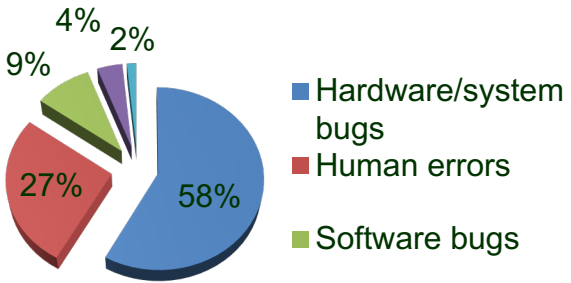
- virus attack
- errorness deleting
- messy storing
- laptop crash

Causes for data loss

- Malware / theft / destruction
- Software failures
  - Program errors / bugs / software updates
  - Features

    (e.g.: overwriting in Dropbox during synchronization)
- Hardware failures
  - Bad design / cheap parts / defects
  - Age
  - Dropped laptops / HDDs
  - Liquids (water, coffee, coke)
  - Lightning strikes / electrical impulses
- Human error
  - Accidental deletion
  - Lack of knowledge

Source: a man working at home while eating breakfast
by Socialeurope via flickr



Source: Kroll Ontrack, 2007, Robin Harris

Further reading : disasters and tales of data loss, statistics on how data gets lost

Backup Principles

- **3 copies**
- **2 different media**
- **1 remote storage location**

- Create multiple backups
- Expect human errors (keep older versions)
- Do not use backup drives for file sharing
- Store backups physically separate from your PC/laptop
- Check your backups regularly
- Practice the worst case and do a full system recovery
- Discuss the topic with friends to learn about their best practices
- Include your mobile devices in your planning

Backup software

eResearch Alliance

Operating system	Integrated Backup-SW	Comment
Windows 7	File Recovery	• Requires configuration to copy not only local libraries • Can create a bootable image
Windows 8/10	File History („File version history“)	• Saves only local libraries • Can be configured for individual libraries and folder exclusion • Cannot create a bootable image
Mac OS	Time Machine	• Saves everything except excluded folders • Can use encryption • Can be used to restore a nonbooting Mac
Ubuntu	File Recovery	• Uses encryption and compression • Can use cloud storage

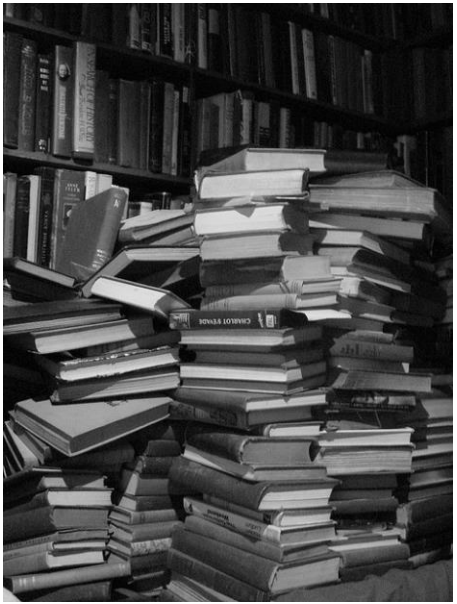
Operating system	Free other backup software
Windows	Personal Backup, PureSync, Paragon Backup&Recovery, Robocopy, …
Mac OS	Carbon Copy Cloner, SuperDuper, ...
Ubuntu	Rsync, Back in Time

GWDG storage services

Name	Backup	Shared Usage	Comment
Fileservice / Active Directory	Yes	Possible	• Network drives, e.g. P: , possibly more • are saved automatically
IBM Tivoli Storage Manager (TSM)	Yes	No	Offer for institutes to centrally secure all local workstations
CrashPlanProE	Yes	No	• Individual backup solution • GWDG license: €26,- per year
ownCloud	Yes	Yes	Free storage space: 50 GB
GRO.data (Dataverse)	Partially	Yes	Primarily for data exchange and data storage for later publication
HSM	No	No	For archiving data from completed projects
GitLab	No	Yes	Versioning; not for large data volumes

Organizing data

Why organize?

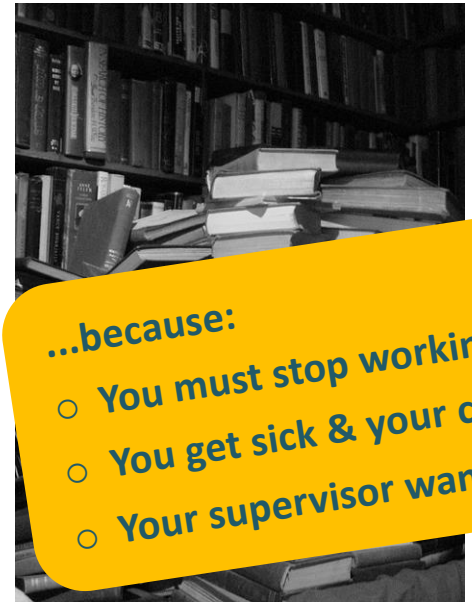


Source: austinevan on flickr

Organize your files so that you and your colleagues can find and access things when you need them.

Source: twechy on flickr

Why organize?



Source: austinevan on flickr

Organize your files so that you

...because:

- You must stop working on A and work on B for 2 weeks.
- You get sick & your colleagues have to finish your joint publication
- Your supervisor wants your results from 4 months ago - *in 4 minutes!*

...ings when you need them.

Source: twechy on flickr

File naming conventions

To stay organized, you should define the following:

- A self-describing folder structure or a tagging scheme
- What information should be included in file names?
- How should file names be structured?
- How should files be referenced and how should they be exchanged?

... especially when working in a team!

Self-explanatory file name:

```
20200812_Presentation_eRA_v42.pptx
```

vs. short file name:

```
eRA_final.pptx
```

Avoid special characters

```
" „ " " ‘ ’ ` {} < > :
; / \ ? ! $ & ~ *
```

File naming conventions

To stay organized, you should define the following:

- A self-describing folder structure or a tagging s
- What information should be included in file nam
- How should file names be structured?
- How should files be referenced and e exchanged?

Use what works for YOU

AND STICK TO IT !

... especially when working in a team!

Self-explanatory file name:

```
20200812_Presentation_eRA_v42.pptx
```

vs. short file name:

```
eRA_final.pptx
```

Avoid special characters

```
" „ " " ‚ ‘ ` { } < > :
; / \ ? ! $ & ~ *
```

Versioning

```
20200812_Presentation_eRA_V13.pptx
20200812_Presentation_eRA_V13final.pptx
20200812_Presentation_eRA_V13new-final.pptx
20200812_Presentation_eRA_V13final-finalv1.pptx
20200812_Presentation_eRA_revised_v01a.pptx
```

Best practice:

- Use consecutive **version numbers** and **initials of the author**
  - No unreliable descriptors in file names like „final“
  - **Rather use folders** for marking/sorting various purposes and for structuring
- When you work with others, agree on a common naming system
- AND/OR: Use a file versioning system
  - e.g. gitlab or ownCloud

Version control with git

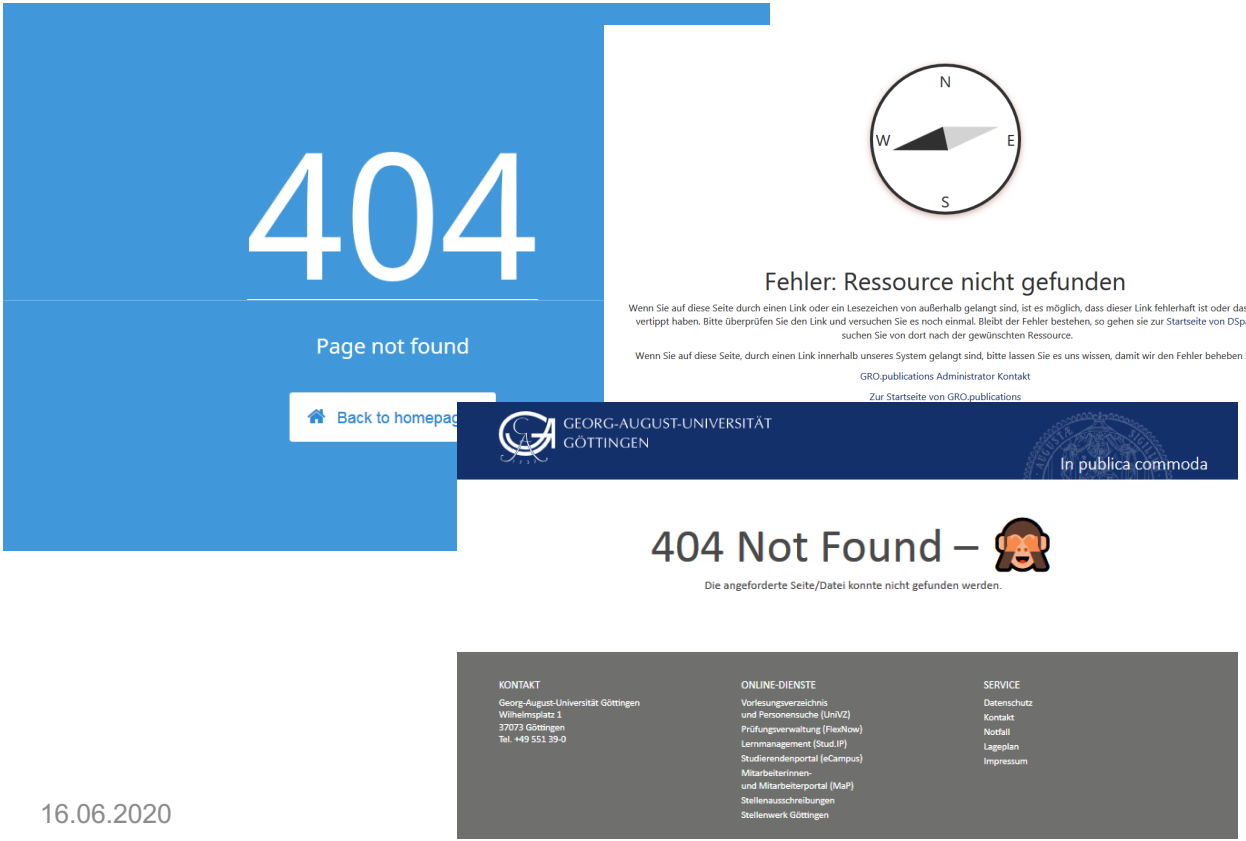
- Mainly used in software development
  - many functionalities for the support of development processes
- Also usable for versioning of documents

Source: Darby on gitlab.com

GWDG gitlab

- Web based versioning system
- Advice and support for the establishment of your projects by GWDG
- Connection to the GWDG user administration
- Central monitoring, system stability and backup by the GWDG
- Carpentries workshops on gitlab by SUB

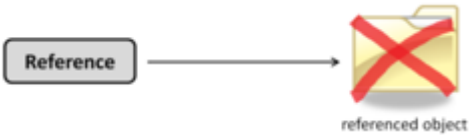
Persistent identifiers (PIDs)



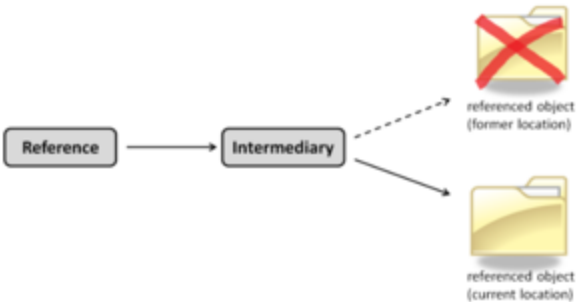
What are persistent identifiers (PIDs)?

Common references like URLs point directly to the location of an object:

If the location of the object changes, the reference points to nothing. The referenced object cannot be found:



The basic idea behind the concept of PIDs is to introduce an **intermediary** between the reference and the referenced object:



This agent monitors all movements or changes made to the object and always forwards requests to the current location.

Adapted from: Kálmán, Tibor: Sustainable referencing of digital objects using persistent identifiers, nestor/DigCurV School 2012, 22-24 October 2012.

What are persistent identifiers (PIDs)?

- Prevention of dead links
- Unique naming (referencing) of a digital resource (e.g. journal article or research data)
- Assignment of a permanent and uniquely referenceable code to be resolved on the internet
- Examples:
  - DOI 10.17192/bfdm.20181.7816
  - Handle hdl:11304/6eacaa76-c275-11e4-ac7e-860aa0063d1f
  - EPIC 21.11101/0000-0000-9D43-4
  - URN urn:isbn:0451450523
  - PURL http://purl.abcd.org/ABC/DEF/200
  - ORCID 0000-0001-2345-6789

How do I get a PID?

- Publications: From the publishing journal or repository
- Data: Depositing in a data repository
- At the Göttingen campus:
  - GRO.data (Campus repository)
  - SUB Göttingen: GOEDOC, GoeScholar, University Press
  - GWDG: ePIC PID Service

Describing data

Explain these data

1000025	5	1	1	1	2	1	3	1	1	2
1002945	5	4	4	5	7	10	3	2	1	2
1015425	3	1	1	1	2	2	3	1	1	2
1016277	6	8	8	1	3	4	3	7	1	2
1017023	4	1	1	3	2	1	3	1	1	2
1017122	8	10	10	8	7	10	9	7	1	4
1018099	1	1	1	1	2	10	3	1	1	2
1018561	2	1	2	1	2	1	3	1	1	2
1033078	2	1	1	1	2	1	1	1	5	2
1033078	4	2	1	1	2	1	2	1	1	2
1035283	1	1	1	1	1	1	3	1	1	2
1036172	2	1	1	1	2	1	2	1	1	2
1041801	5	3	3	3	2	3	4	4	1	4
1043999	1	1	1	1	2	3	3	1	1	2
1044572	8	7	5	10	7	9	5	5	4	4
1047630	7	4	6	4	6	1	4	3	1	4

Explain these data

Sample code number	Clump Thickness	Uniformity of Cell Size	Uniformity of Cell Shape	Marginal Adhesion	Single Epithelial Cell Size	Bare Nuclei	Bland Chromatin	Normal Nucleoli	Mitoses	Class
id number	1-10	1-10	1-10	1-10	1-10	1-10	1-10	1-10	1-10	(2 for benign, 4 for malignant)
1000025	5	1	1	1	2	1	3	1	1	2
1002945	5	4	4	5	7	10	3	2	1	2
1015425	3	1	1	1	2	2	3	1	1	2
1016277	6	8	8	1	3	4	3	7	1	2
1017023	4	1	1	3	2	1	3	1	1	2
1017122	8	10	10	8	7	10	9	7	1	4
1018099	1	1	1	1	2	10	3	1	1	2
1018561	2	1	2	1	2	1	3	1	1	2
1033078	2	1	1	1	2	1	1	1	5	2
1033078	4	2	1	1	2	1	2	1	1	2
1035283	1	1	1	1	1	1	3	1	1	2
1036172	2	1	1	1	2	1	2	1	1	2
1041801	5	3	3	3	2	3	4	4	1	4
1043999	1	1	1	1	2	3	3	1	1	2
1044572	8	7	5	10	7	9	5	5	4	4
1047630	7	4	6	4	6	1	4	3	1	4

Source: **Breast Cancer Wisconsin (Diagnostic) Data Set**, Dua, D. and Graff, C. (2019). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science.

What is metadata?

Metadata are data "descriptions"

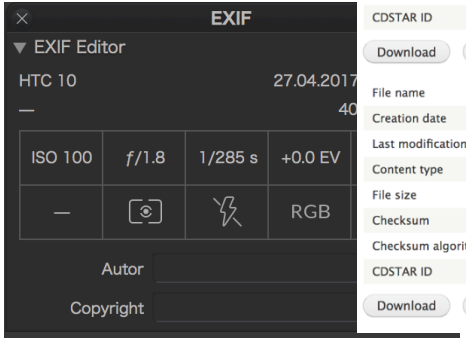
- **WHO** recorded the data?
- **WHAT** is the content of the data?
- **WHEN** were the data recorded?
- **WHERE** (geographically) were the data collected?
- **HOW** were the data recorded?
- **WHY** were the data recorded?

Source: USFWS - Pacific Region on flickr

What is metadata?

- Different definitions, depending on the perspective
- Practical approach: Metadata...
  - describe objects in a *structured and standardized* way
  - can help in the *selection and identification of* resources
  - can describe how to *use* data *correctly* or how to *reproduce* it
  - can describe everything: literature, a painting, places, a set of data…
  - can be digitally linked to objects (embedded) or added separately

Metadata is everywhere



Show record

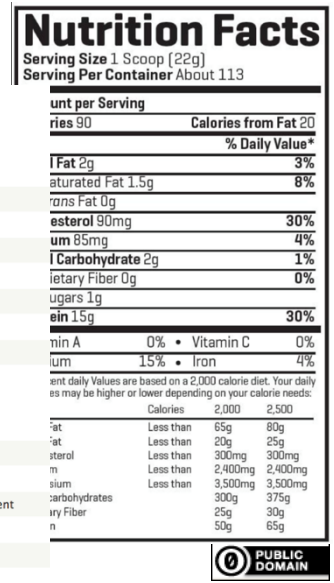
2 Files

File name	RE-DATA-Season1_2017-03-02.xlsx
Creation date	2017-04-24 21:09:41
Last modification date	2017-04-24 21:09:41
Content type	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
File size	625848
Checksum	48928b915f2e06eafa7d9105c3a7132b
Checksum algorithm	MD5
CDSTAR ID	EAEA0-0FEC-9E6F-7147-0/RE-DATA-Season1_2017-03-02.xlsx

Download Delete file

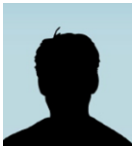
File name	RE-season1-evaluation_2017-03-02.docx
Creation date	2017-04-24 21:09:40
Last modification date	2017-04-24 21:09:40
Content type	application/vnd.openxmlformats-officedocument.wordprocessingml.document
File size	2824324
Checksum	f9cf1a2aa9ab55865b71aea4f7300e59
Checksum algorithm	MD5
CDSTAR ID	EAEA0-0FEC-9E6F-7147-0/RE-season1-evaluation_2017-03-02.docx

Download Delete file



Metadata: What needs to be in there?

Who created **what**, **how**, **when**, **where** and **why**?



Timo Gnadt
gnadt@sub.uni-goettingen.de

r	x	y	abs
35	0.4	34	36
535	0.5	2	777
63		2.6	67
4	1.3	61	5

Excel table with test data for training purposes

Random number generator for changing field data

July 26, 2016

On my Office Windows PC

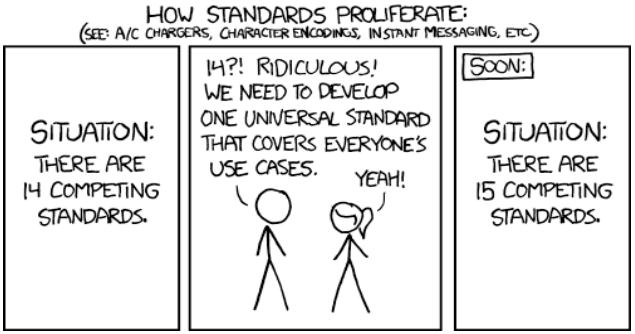
For use in a workshop

Contents:

- **Description of** the object
- **Methodology** and **instruments**
- **Units** of measurement
- **References** to related data
- **Definitions of** jargon, acronyms, code
- **Technical information** about the file

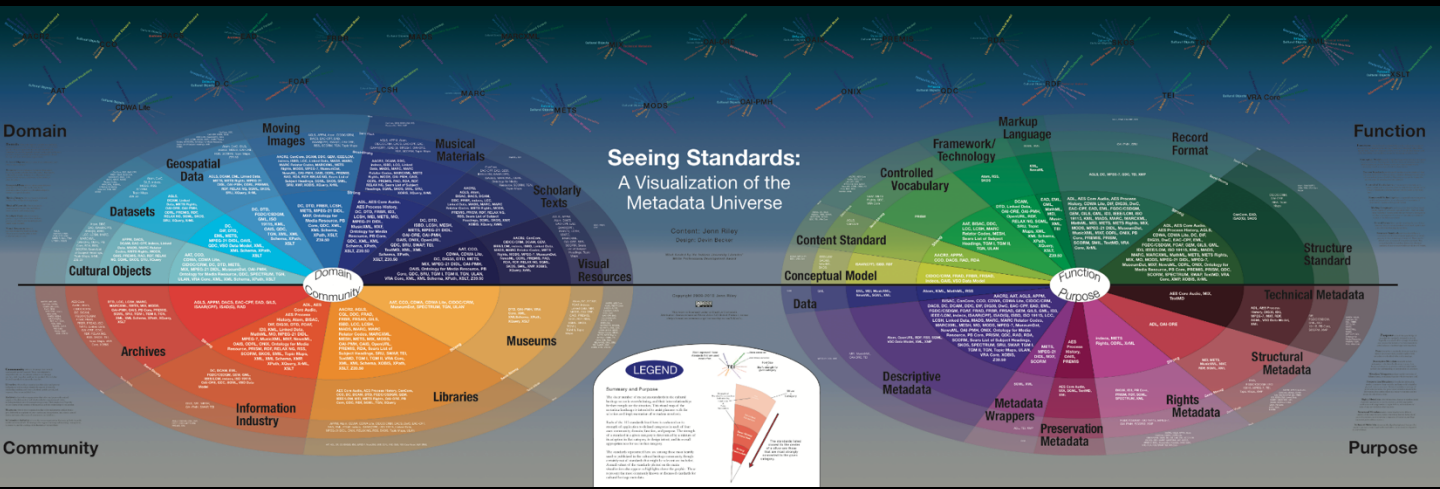
Why Metadata Standards?

- A standard provides a structure with which data can be described:
  - Common terms to ensure consistency
  - Common definitions for easier interpretation
  - Common language to facilitate communication
  - Common structure for quick information retrieval
- For search and retrieval, standards offer:
  - a documentation structure in a reliable and predictable format for computer interpretation
  - a uniform summary description of the data set



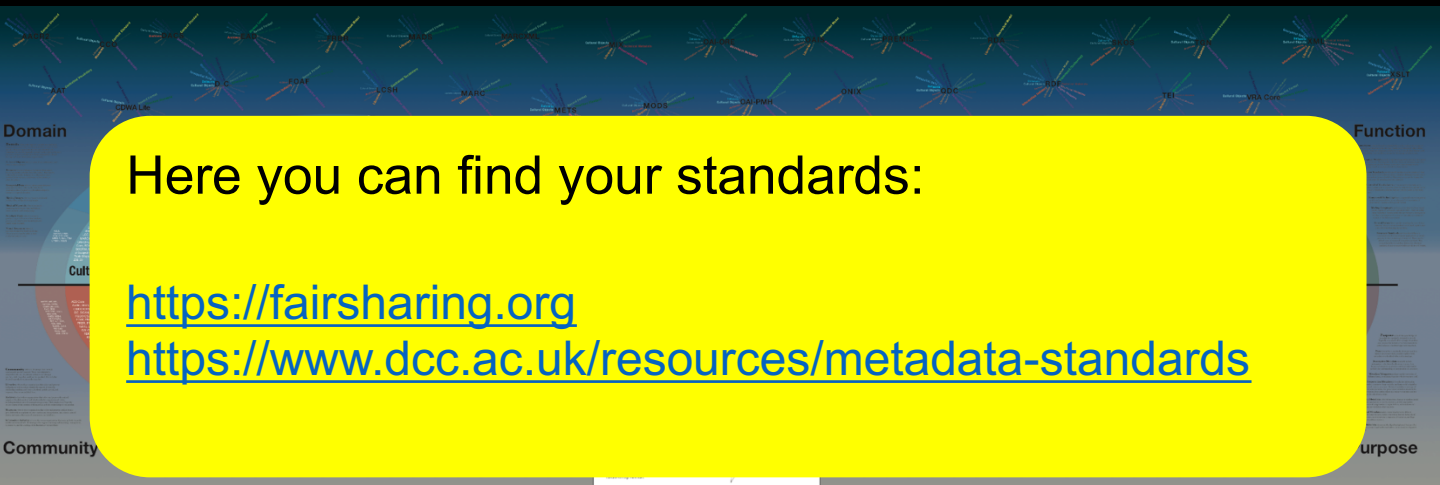
Source:
XKCD: Standards
Randall Munroe

Metadata standards



Seeing Standards: A Visualization of the Metadata Universe
Jenn Riley, Devin Becker
http://jennriley.com/metadatamap/

Metadata standards



Here you can find your standards:

https://fairsharing.org
https://www.dcc.ac.uk/resources/metadata-standards

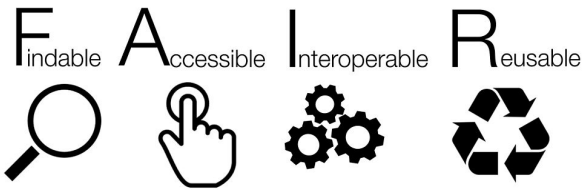
Seeing Standards: A Visualization of the Metadata Universe
Jenn Riley, Devin Becker
http://jennriley.com/metadatamap/

FAIR Data Principles

FAIR Data Principles

Compilation of guiding principles for research data

- Objective: To make data findable, accessible, interoperable and reusable*.
- Address data producers and data publishers to promote the maximum use of research data
- FAIR refers to both people and machines



Source: SangyaPundir on Wikimedia Commons

- Published in 2016:

  Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.*
  The FAIR Guiding Principles for scientific data management and stewardship.
  Sci Data **3,** 160018 (2016). https://doi.org/10.1038/sdata.2016.18

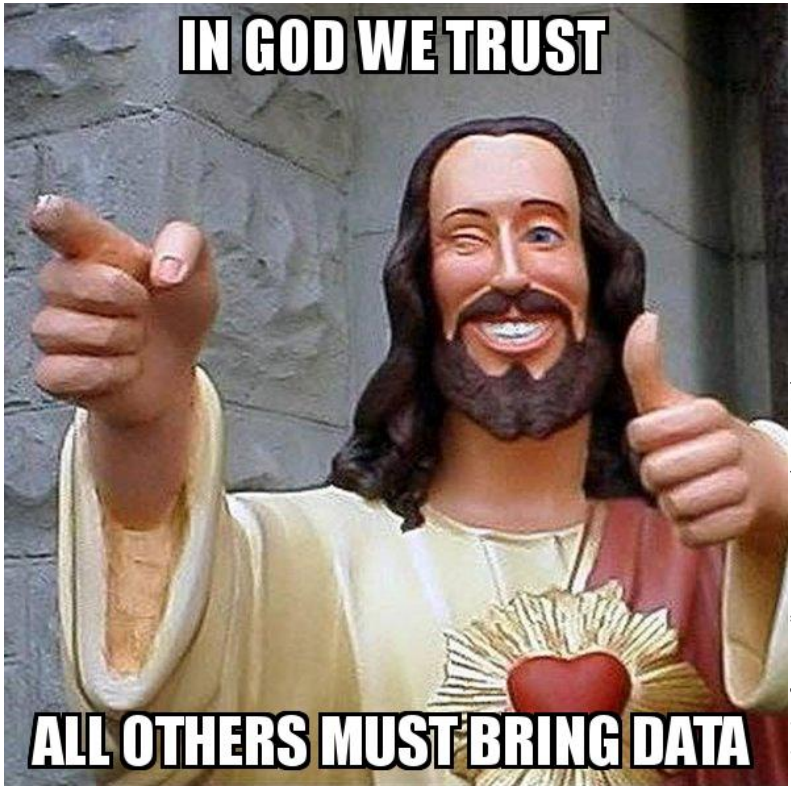
- Available here:

  https://www.force11.org/group/fairgroup/fairprinciples

 *) Detailed listing of the principles on slides at the end of the presentation

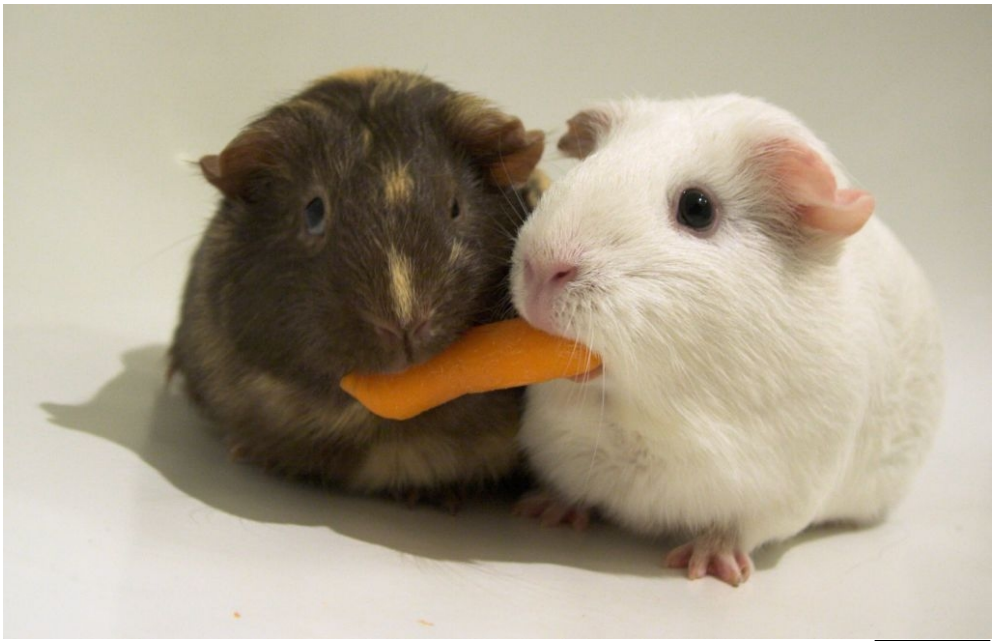
Sharing data

Data sharing - Motivation



Citation from: William E. Demming (1900-1993)

Why share?



Source: Sharing by ryancr via flickr

Why share?


Source: Karen Arnold on Pixabay

Reputation

- Receive recognition for high quality research
- Better understanding of your methods
- Enables others to review the work
- Recognition for contributions to the research community
- Expand research beyond your discipline


Source: mohamed Hassan on Pixab

Funding

- The provision of data and/or publications may be a requirement of your funding agency.
- It can make your project application more attractive, even if sharing data is not required.

Why share?

Source: Richard Matthews on flickr: dart (2011)

Impact

- Sharing makes your data:
  - Easier to find
  - More easily accessible
- Open data/publications lead to increased citations

Reuse

- Starting point for a supplementary study
- Test data for new software and algorithms
- Data for teaching purposes
- Contexts that cannot be assessed at present

Data sharing - concerns

Source: All he does is eat eat eat, Jannes Pockele via flickr

- Stockpiling for bad times
- Nobody likes to polish
- Dirt behind the scenes
- Atmosphere of fear
- Small fishes & unicorns

Self use

No documentation

Work in progress

Theft and abuse

Insignificance

Value over time?

Embargo!

Do it for yourself!

"Working set"

Trust in science

The future is unpredictable

Data sharing - Real barriers

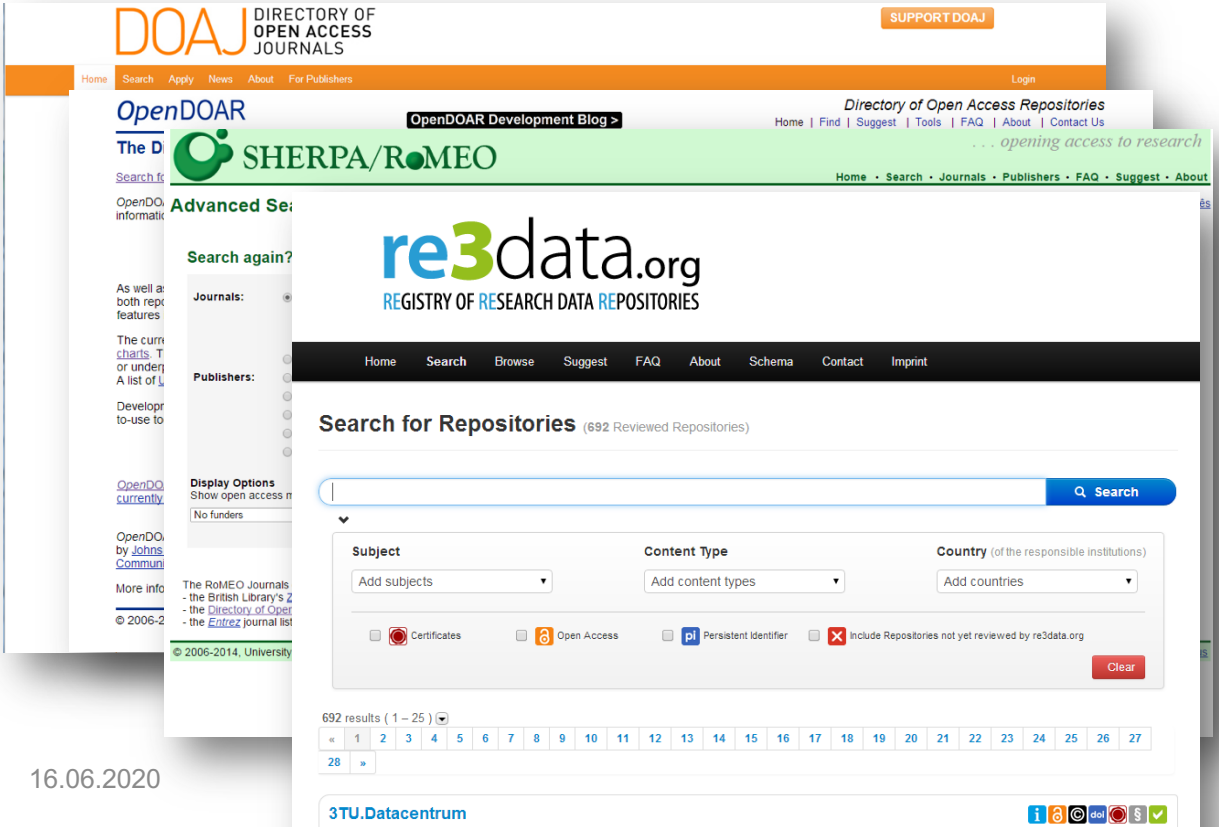
- Place
  - No common tradition
  - No repository
  - No expertise
- Funds
  - no funds
- Rights
  - No carte blanche

Source: Simatai_Great_Wall by Arian Zwegers on Wikimedia Commons

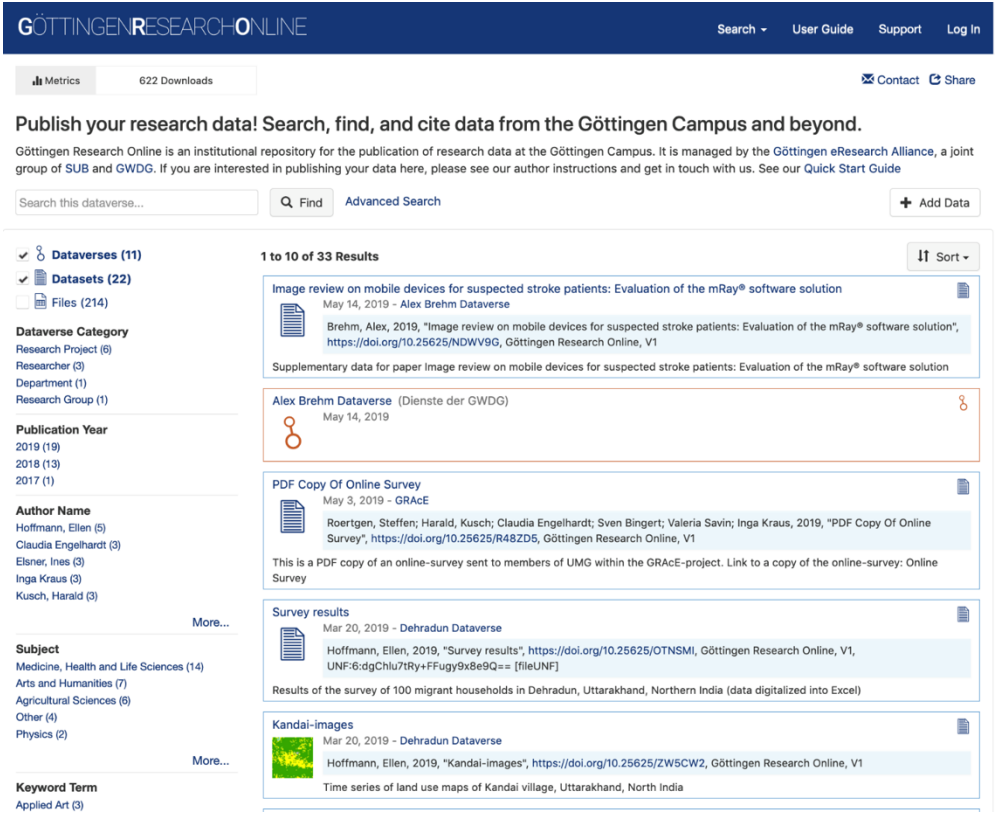
Modes of sharing

Way of transfer	Access conditions	Usage condition
Peer-to-peer	restricted	none
Webspace	upon request	Agreement
repository	Embargo	License
	open	

Finding Open Access journals and repositories



GRO.data: Research Data Repository



Recent uploads

September 1, 2017 (v20) Software Open Access

View

matplotlib/matplotlib v2.1.0rc1

Michael Droettboom; Thomas A Caswell; John Hunter; Eric Firing; Jens Hedegaard Nielsen; Nelle Varoquaux; Benjamin Root; Elliott Sales de Andrade; Phil Elson; Darren Dale; Jae-Joon Lee; Jouni K. Seppänen; Antony Lee; Ryan May; Damon McDougall; David Stansby; Andrew Straw; Paul Hobson; Tony S Yu; Eric Ma; Christoph Gohlke; Steven Silvester; Charlie Moad; Adrien F. Vincent; Jan Schulz; Peter Würtz; Federico Ariza; Cimarron; Thomas Hisch; Nikita Kniazev

matplotlib: plotting with Python

Uploaded on September 1, 2017

19 more version(s) exist for this record

August 30, 2017 (v1) Working paper Open Access

View

Introducing Parsl: A Python Parallel Scripting Library

Babuji, Yadu; Brizius, Alison; Chard, Kyle; Foster, Ian; Katz, Daniel S.; Wilde, Michael; Wozniak, Justin

Researchers frequently rely on large-scale and domain-specific workflows to conduct their science. These workflows may integrate a variety of independent software functions and external applications. However, developing and executing such workflows can be difficult, requiring complex...

Uploaded on August 30, 2017

August 24, 2017 (v2) Dataset Open Access

View

Aligned ISNI and Ringgold identifiers for institutions

Delpeuch, Antonin

This dataset provides a correspondence between ISNI and Ringgold identifiers, by combining two datasets: Open ISNI for Institutions, available at http://isni.ringgold.com/ , which provides metadata for institutions identified by ISNI. The dataset of institutions used by ORCID for disambiguation,...

Uploaded on August 24, 2017

1 more version(s) exist for this record

August 22, 2017 (v2) Dataset Open Access

View

Supplementary Data: Status of the scalar singlet dark matter model (arXiv:1705.07931)

Other services on campus

eResearch Alliance

Name	available through	Purpose/comments
Jupyter notebooks	GWDG	Live editing and execution of text, diagrams, equations and code in a web browser
CodiMD pad	GWDG	Collaborative text editing
Electronic laboratory notebook	UMG	(Re)organisable, searchable and storable research documentation
Self-study online courses on digital competencies	SUB	Courses on literature search, IT basics, data security, data visualization, OER *https://www.uni-goettingen.de/en/565228.html*
Open Access Publication Fund	SUB	full coverage for up to € 2.000,- for publication in OA journals
Video Conferencing	GWDG	Various solutions

GWDG services

SERVICES

Storage Services
- Data Archiving
- Backup
- GWDG CrashPlan PROe
- File Service
- GWDG Cloud Share
- GWDG ownCloud
- Cryptshare

Email & Collaboration
- E-Mail-Service (MS Exchange 2010)
- Spam and Virus Filtering
- Mailing Lists
- MS Sharepoint
- Managed Services
- Project Management Service
- Etherpad
- ShareLaTeX
- Rocket.Chat
- GitLab

Server Services
- Virtual Server
- Hosting/Housing of Servers
- Web Hosting
- GWDG Cloud Server
- FTP-Server
- Puppetserver

Network Servies
- IP Address Management System
- Cable and Route Management System
- System Monitoring
- Setting up eduroam
- Integration into the Active Directory
- User Management with OpenLDAP
- Client Management
- Client Management for macOS and iOS

Application Services
- Persistent Identifier (PID)
- High Performance Computing
- Library Service Aleph
- Database Service Oracle
- Database Service MySQL
- Application and Registration Services
- Plagiarism Prevention
- Online Surveys
- Bioinformatics Programs
- Statistics Programs
- Jupyter

IT Security Services
- Vulnerability Scans on Network-attached Equipment
- Public-Key-Infrastructure (PKI)
- Authentication and Authorization Infrastructure (AAI)
- Virus Protection (Sophos Update Service)

General Services
- Identity Management
- Courses
- Software and Licence Management
- Videoconferencing
- Computer Lending Pool
- Print & Scan Services

IT Consulting
- Scientific Data Management
- IT Security
- Hardware Purchase
- Apple Support Centre
- Establishing Directory Services (AD, LDAP)
- Planning of Data Transmission Networks

https://www.gwdg.de/services

Further reading

- DataOne
  - https://www.dataone.org/education-modules
- MANTRA
  - http://mantra.edina.ac.uk/
- "Reproducible research with jupyter notebooks"
  - https://reproducible-science-curriculum.github.io/rr-jupyter-workshop/
- „IT in a nutshell“ – a self-learning course by SUB Göttingen on IT basics
  - https://www.uni-goettingen.de/en/613871.html

eResearch Alliance

Göttingen eResearch Alliance

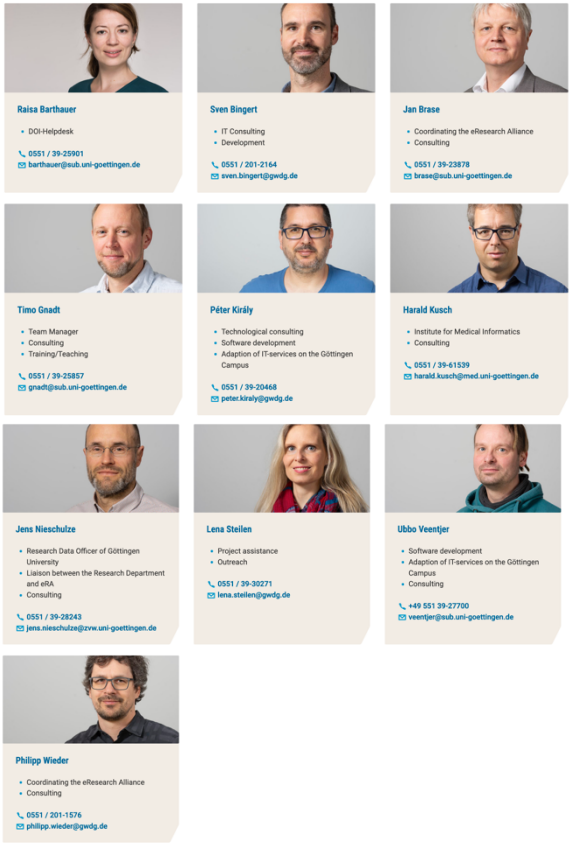
Team

- Various discipline-specific expertise
  - mainly in the natural sciences, humanities, computer science
- Jointly operated by

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

- Partner: Department of Research, University Medicine (UMG)
- Comprehensive expertise on eResearch topics



What the eRA can do for you

- Consulting / Support / Services
  - Research Data Management
  - Publication strategies
  - Digital methods, software and technologies to improve research projects
  - Central point of contact for experts & expertise throughout the campus
- Training
  - Discipline- or project-specific or general
  - Information material / knowledge database
- Cooperation
  - Project partnership
  - Project as a service

www.eresearch.uni-goettingen.de

Thank you for your participation!

The comments, questions and suggestions of the participants for this training can be found here:

https://pad.gwdg.de/LoveYourData_20200812?both

CONTACT:

info@eresearch.uni-goettingen.de

www.eresearch.uni-goettingen.de

FAIR Data Principles

Traceable:

F1. (Meta)data is assigned a globally unique and persistent identifier.

F2.data are described with extensive metadata.

F3 (Meta) data are registered or indexed in a searchable resource.

F4.metadata specify the data identifier.

Source: *https://www.force11.org/group/fairgroup/fairprinciples*

FAIR Data Principles

Accessible:

A1. (Meta-)data can be retrieved by means of their identifier using a standardised communication protocol.A1.1 the protocol is open, free and universally implementable .A1.2 the protocol allows an authentication and authorisation procedure if required.

A2.metadata are accessible even if the data are no longer available.

Source: *https://www.force11.org/group/fairgroup/fairprinciples*

FAIR Data Principles

Interoperable:

I1. (Meta-)data use a formal, accessible, shared and generally applicable language to represent knowledge

I2. (Meta-)data use vocabularies that follow the FAIR principles.

I3 (Meta) data contain qualified references to other (meta) data.

Source: *https://www.force11.org/group/fairgroup/fairprinciples*

FAIR Data Principles

Reusable:

R1.meta(data) have a variety of accurate and relevant attributes

.R1.1. (Meta)data are released with a clear and accessible data use licence.R1.2. (Meta)data are associated with their provenance.R1.3. (Meta)data meet domain-relevant Community standards.

Source: *https://www.force11.org/group/fairgroup/fairprinciples*