Research data management

CRC 1073 / CRC 1633 / FOR 5522 / RTG 2455

13.06.2024

Timo Henne

henne@sub.uni-goettingen.de

CRC 1073
ATOMIC SCALE CONTROL OF ENERGY CONVERSION

DFG FOR 5522

eResearch Alliance

SFB 1633

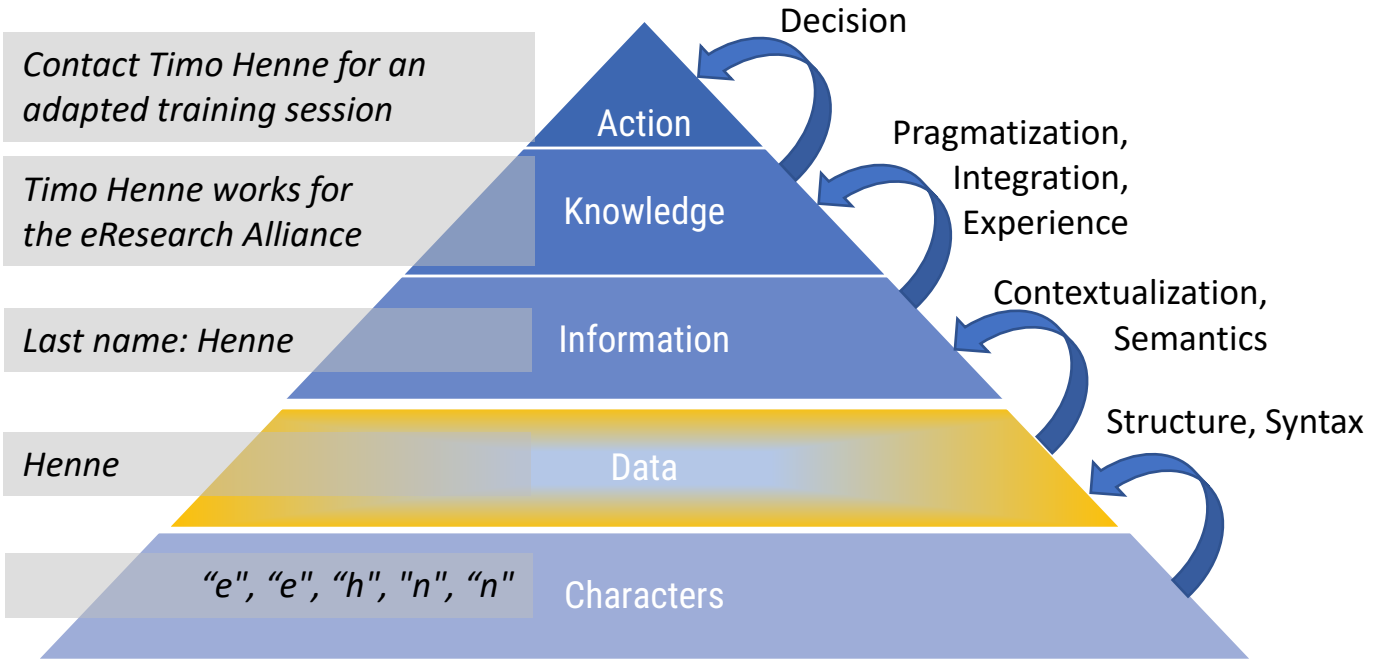
BENCh RTG2455

Outline

- Introduction to Research Data Management
- Storing Data
- Organization of Data
- Metadata
- Data Sharing
- Publishing research software
- Further Infos

Introduction to Research Data Management

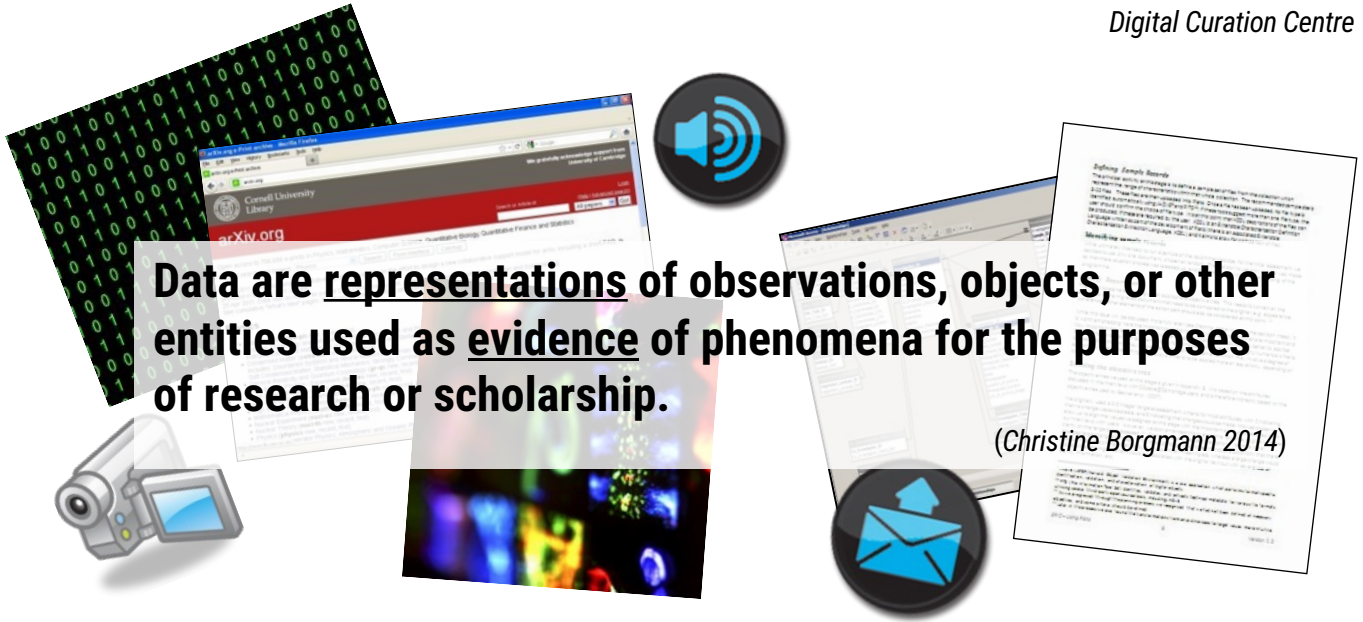
What is 'data'?


Contact Timo Henne for an adapted training session
Timo Henne works for the eResearch Alliance
Last name: Henne
Henne
“e", “e", “h", "n", “n"
Action
Knowledge
Information
Data
Characters
Decision
Pragmatization, Integration, Experience
Contextualization, Semantics
Structure, Syntax

What are ‘data’?

“A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing.”

Digital Curation Centre



Data are <u>representations</u> of observations, objects, or other entities used as <u>evidence</u> of phenomena for the purposes of research or scholarship.

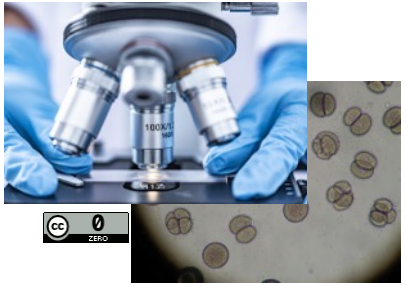
(*Christine Borgmann 2014*)

What are research data?

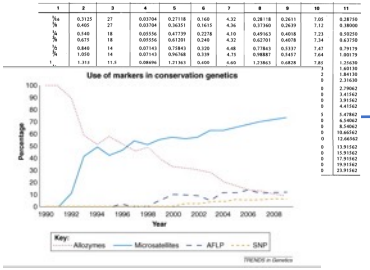
Any representation of information you use in your research:

Forms: Statistics, interviews, simulations, measurement data from experiments, observational data from instruments, text with semantic annotations, 3D scans, model drawings, scripts,…

Types: Video, audio, images, spreadsheets, emails, paper documents, binary data, software, text files, lab notebooks, ...



research object



research data

result/ publication

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. http://dx.doi.org/10.1016/j.tig.2010.01.001

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 often 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 2004-2016:

Duration:

- >10 years preparation
- 10 years from start to data

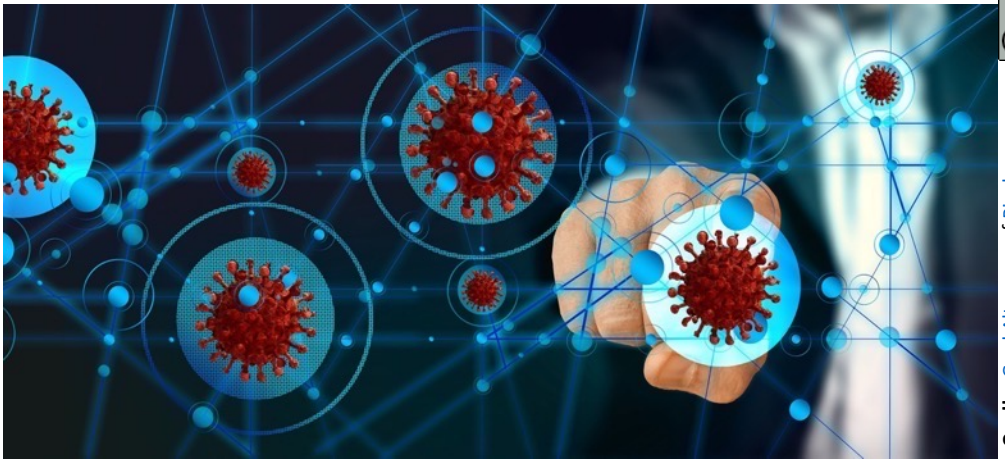
Costs:

- over € 1.000.000.000

Outcome:

- some cool photos
- lots of data
- *a radically new theory on the origin of the universe?*

Research data - a resource in great demand



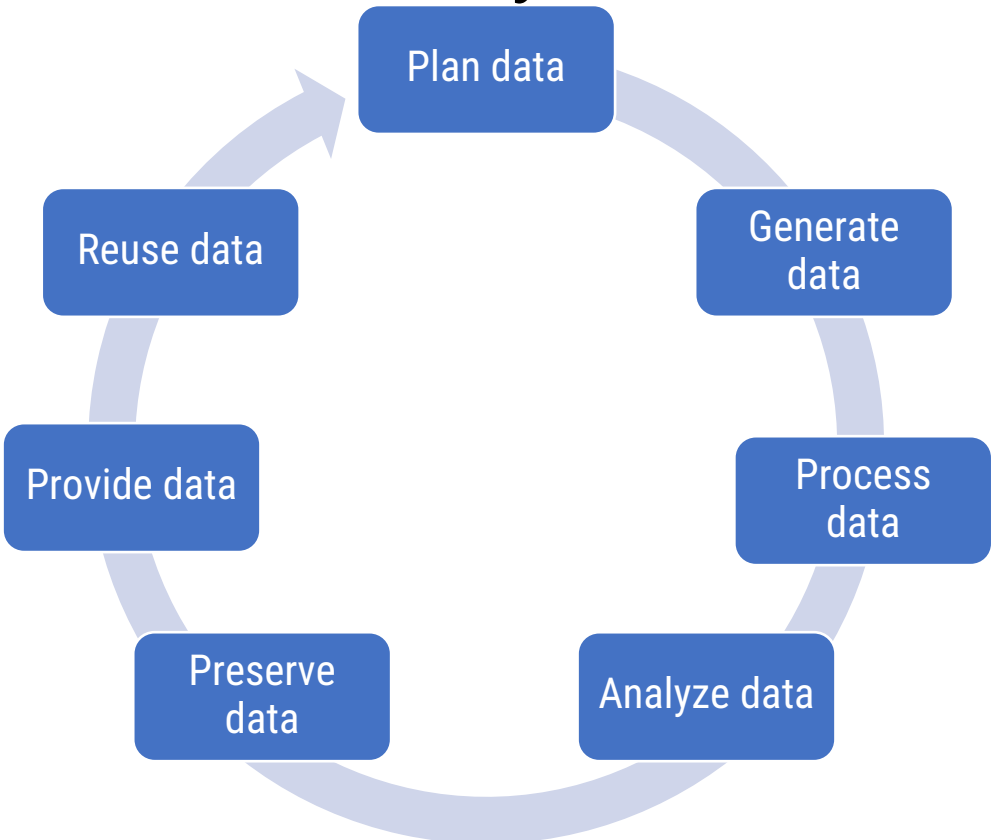
Quelle: Gerd Altmann auf Pixabay, CC0

Data with reference to COVID-19

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

Research cycle

Research data cycle

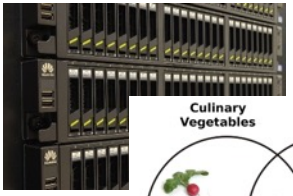


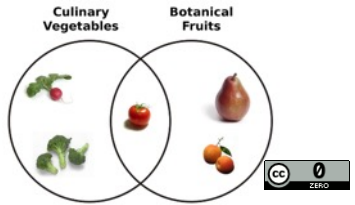
What is Research Data Management?

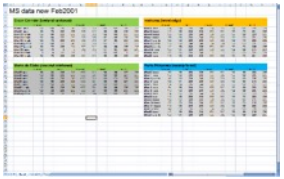
What is Research Data Management?

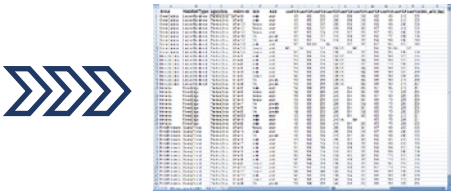
What is Research Data Management?

- Storage, Backup & Archiving
- Metadata and Documentation
- Data Quality
- File Names, Identifier and Versions
- Ethics, Rights and Licenses







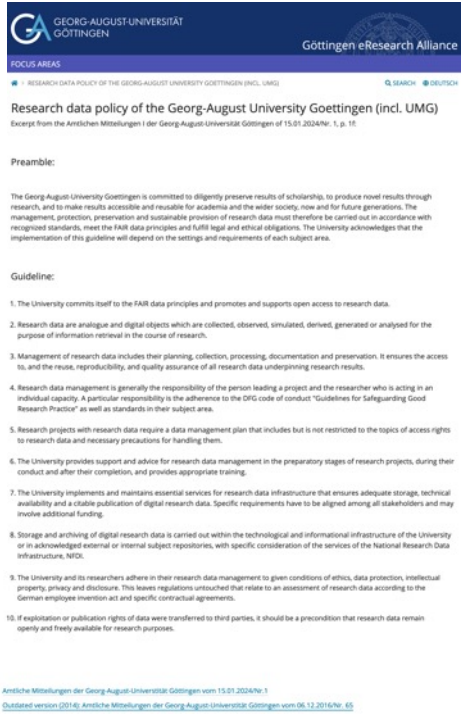


doi:10.10.1038/nphys1170

Thesis_final_v13b_revised.docx

Research Data Policy of the Georg-August Universität Göttingen

- Officially issued on 28th August 2014, revised version in 2024
- One of the first German universities with such a policy



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Göttingen eResearch Alliance

FOCUS AREAS

RESEARCH DATA POLICY OF THE GEORG-AUGUST UNIVERSITY GOETTINGEN (INCL. UMG)

SEARCH DEUTSCH

Research data policy of the Georg-August University Goettingen (incl. UMG)

Excerpt from the Amtlichen Mitteilungen I der Georg-August-Universität Göttingen of 15.01.2024/Nr. 1, p. 1ff.

Preamble:

The Georg-August-University Goettingen is committed to diligently preserve results of scholarship, to produce novel results through research, and to make results accessible and reusable for academia and the wider society, now and for future generations. The management, protection, preservation and sustainable provision of research data must therefore be carried out in accordance with recognized standards, meet the FAIR data principles and fulfill legal and ethical obligations. The University acknowledges that the implementation of this guideline will depend on the settings and requirements of each subject area.

Guideline:

1. The University commits itself to the FAIR data principles and promotes and supports open access to research data.
2. Research data are analogue and digital objects which are collected, observed, simulated, derived, generated or analysed for the purpose of information retrieval in the course of research.
3. Management of research data includes their planning, collection, processing, documentation and preservation. It ensures the access to, and the reuse, reproducibility, and quality assurance of all research data underpinning research results.
4. Research data management is generally the responsibility of the person leading a project and the researcher who is acting in an individual capacity. A particular responsibility is the adherence to the DFG code of conduct "Guidelines for Safeguarding Good Research Practice" as well as standards in their subject area.
5. Research projects with research data require a data management plan that includes but is not restricted to the topics of access rights to research data and necessary precautions for handling them.
6. The University provides support and advice for research data management in the preparatory stages of research projects, during their conduct and after their completion, and provides appropriate training.
7. The University implements and maintains essential services for research data infrastructure that ensures adequate storage, technical availability and a citable publication of digital research data. Specific requirements have to be aligned among all stakeholders and may involve additional funding.
8. Storage and archiving of digital research data is carried out within the technological and informational infrastructure of the University or in acknowledged external or internal subject repositories, with specific consideration of the services of the National Research Data Infrastructure, NFDI.
9. The University and its researchers adhere in their research data management to given conditions of ethics, data protection, intellectual property, privacy and disclosure. This leaves regulations untouched that relate to an assessment of research data according to the German employee invention act and specific contractual agreements.
10. If exploitation or publication rights of data were transferred to third parties, it should be a precondition that research data remain openly and freely available for research purposes.

Amtliche Mitteilungen der Georg-August-Universität Göttingen vom 15.01.2024/Nr.1

Outdated version (2014): Amtliche Mitteilungen der Georg-August-Universität Göttingen vom 06.12.2016/Nr. 65

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

Research Data Policy of the Georg-August Universität Göttingen

- Officially issued on 28th August 2014, revised version in 2024
- One of the first German universities with such a policy
- Topics addressed:
  - Research Data, Research Data Management and its purposes
  - Data Management Plans
  - Support, training and provision of services
  - Storage location
  - Ethical and legal standards
  - Open Access
- eResearch Alliance: support and advice on the implementation of the RDP for the Göttingen Campus


GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN
Göttingen eResearch Alliance
FOCUS AREAS
Research data policy of the Georg-August University Goettingen (incl. UMG)
Preamble:
Guideline:
Outdated version (2014): Amtliche Mitteilungen der Georg-August-Universität Göttingen vom 06.12.2016/Nr. 65

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

Why Research Data Management?

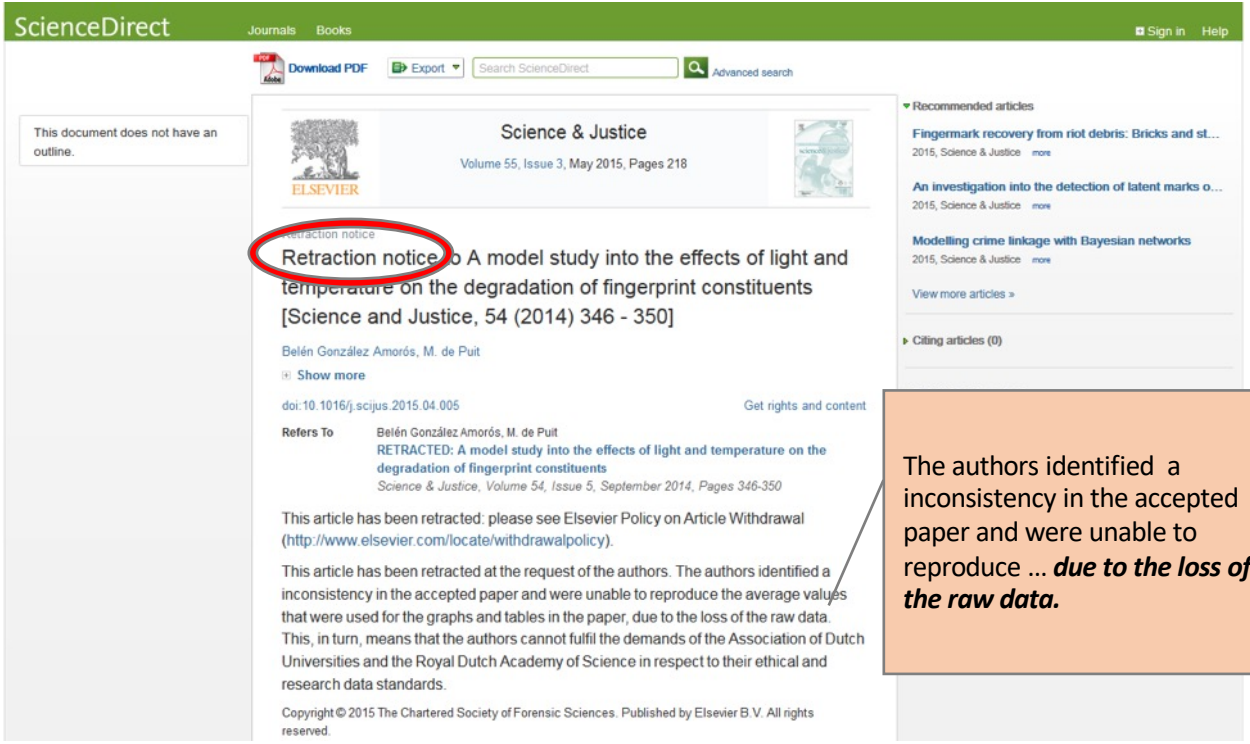
Why Research Data Management?

- **Improve your research**
  - Prevent data loss
  - Prevent unnecessary work
  - Better data quality
- **Good scientific practice**
  - Reproducibility, accountability and compliance
  - "Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin." (DFG, Proposals for safeguarding good scientific practice, 1998)
  - Requirement from funding agencies (e.g. EC Horizon Europe, DFG)
- **Data sharing with colleagues**
  - Research can be *very* expensive and often the only result may be data.
  - Data management costs are small in comparison to data creation costs.
  - Productive data sharing is simply a matter of efficiency.

DFG FOR 5522

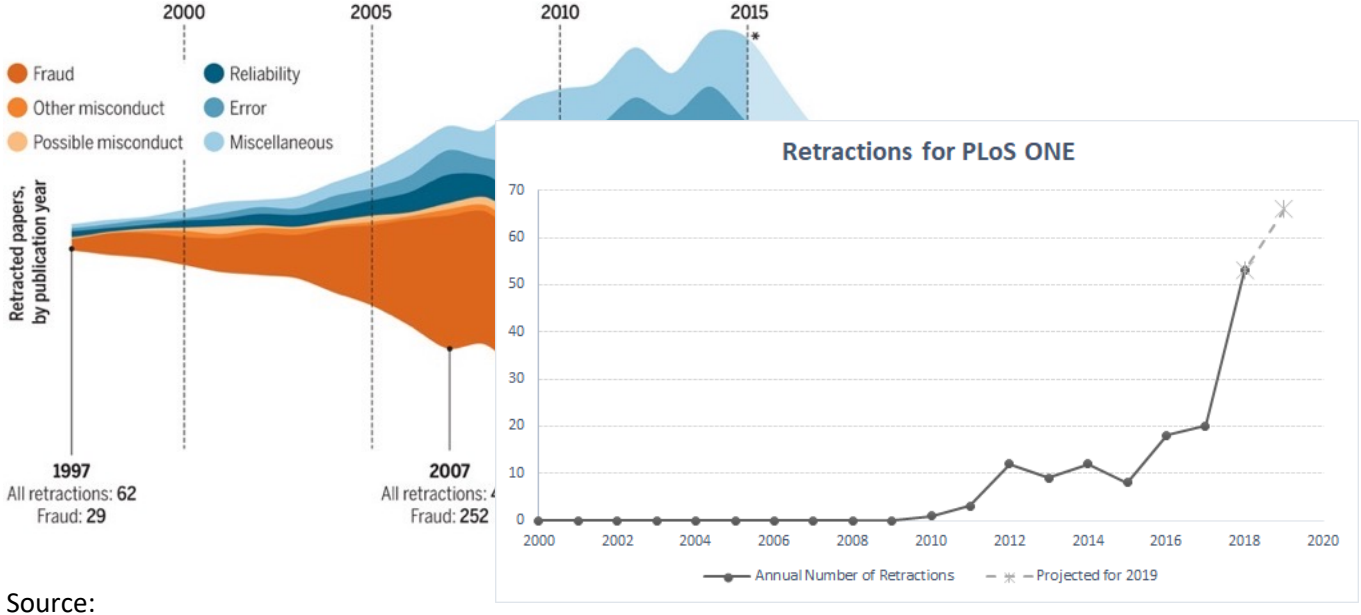
Why Research Data Management?

Why Research Data Management?



The authors identified a inconsistency in the accepted paper and were unable to reproduce … ***due to the loss of the raw data.***

Why Research Data Management?



Source:
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

Source:
Alison Abritis/RetractionDatabase.org

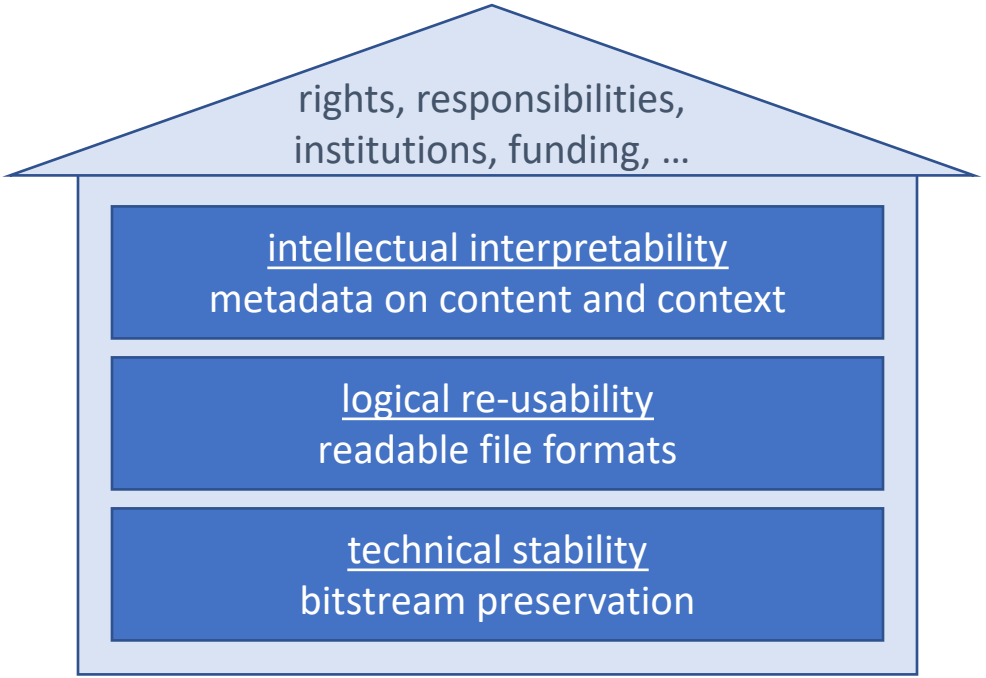
Why Research Data Management?

- **Improve your research**
- **Good scientific practice**
- **Data sharing with colleagues**
- **Data Publication**
  - Required by increasing number of journals
  - Obviate retraction
  - Get credit for your data!
- **Enable new kinds of research**
  - Feedback loops between empirical and modeling approaches
  - Initiating research questions in completely different fields

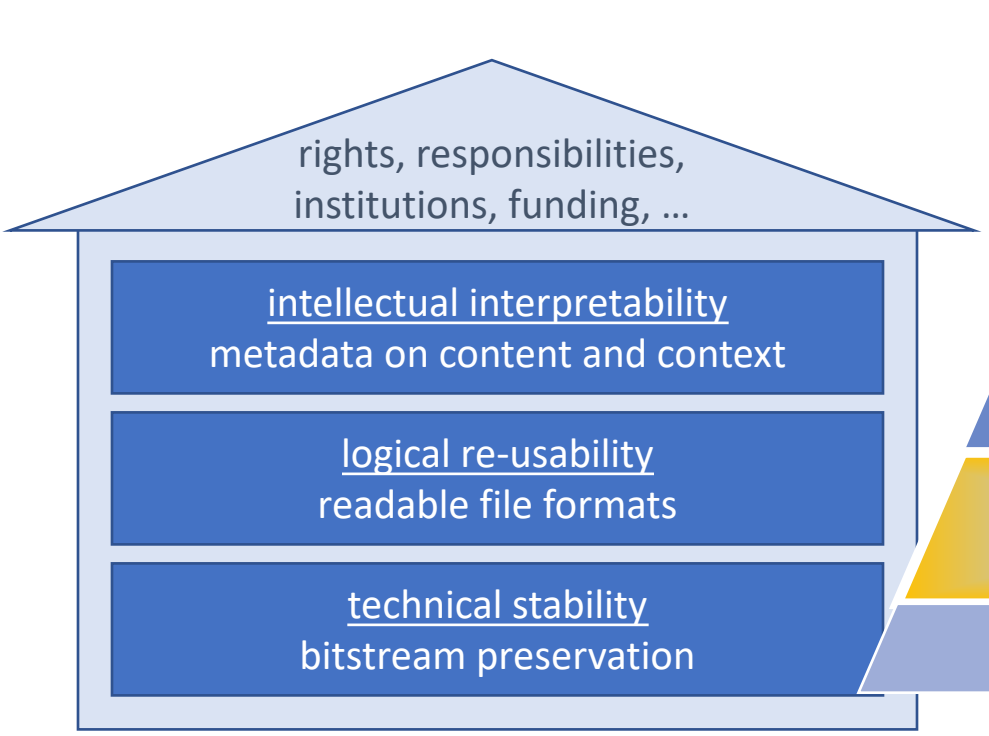
Publications are arguments
made by authors,
and ***data are the evidence***
used to support the arguments.

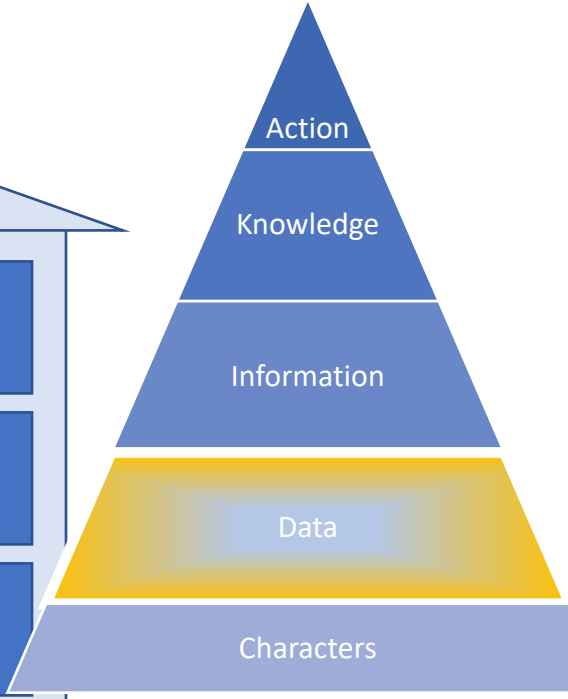
(*Christine Borgmann, 2014*)

Levels of Data Preservation

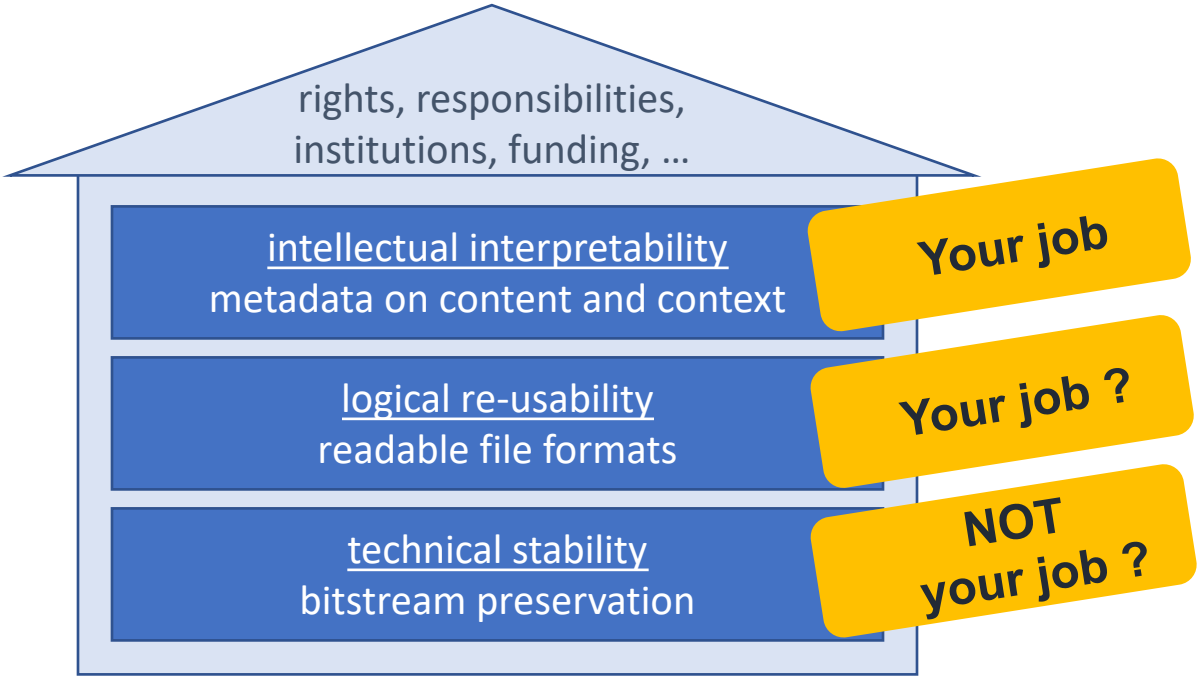


Levels of Data Preservation

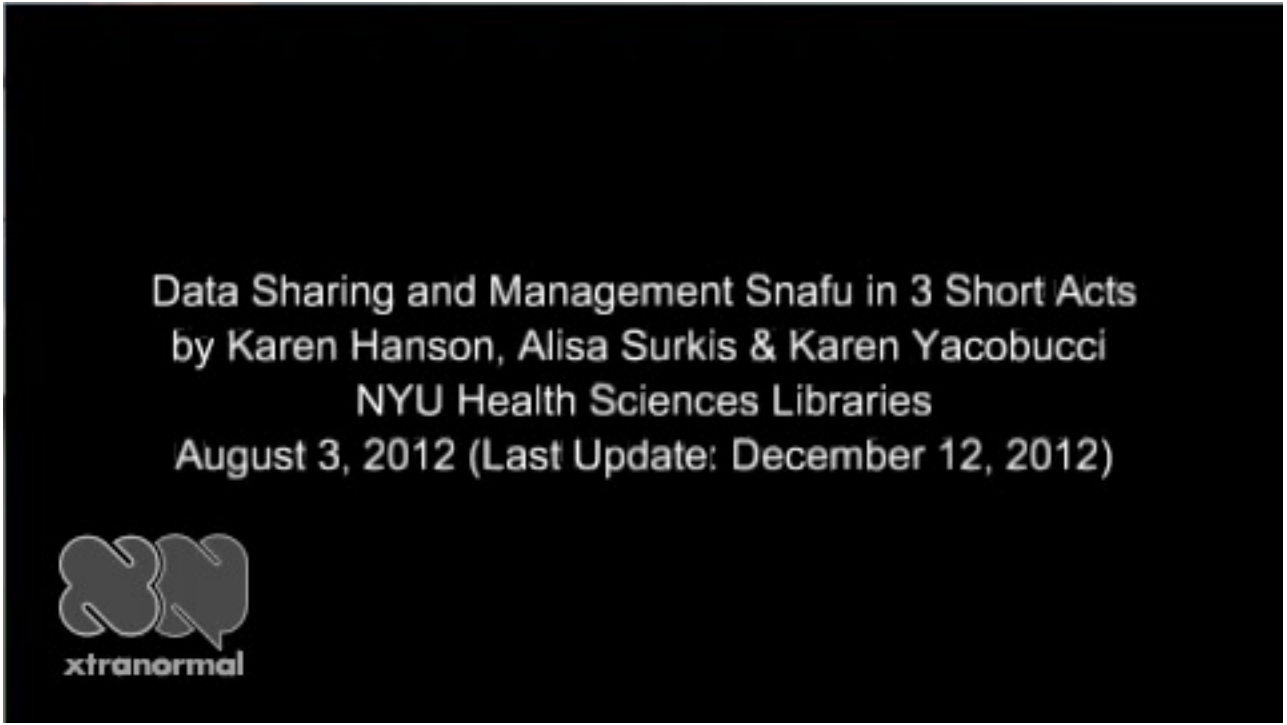




Levels of Data Preservation

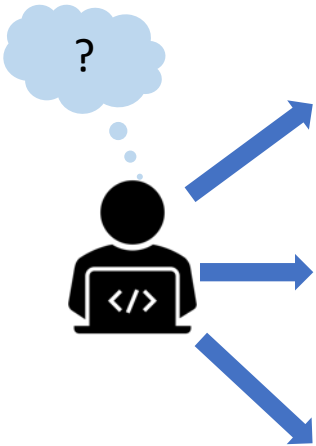


Data preservation motivation



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	Archival	Depositing
Storage Purpose	**Ability to restore data** in case of data loss or error propagation	Enable validation by peers through **persistent storage** of data used for research results / publication	Enable verification, citation & reuse of datasets (**data sharing**)
Data Characteristics	Duplication of **current work data** & intermediate work results	Archive format (e.g. zip) containing **all related & relevant data** / files (ideally incl. metadata)	Format specified by repository; **discipline-specific metadata** standards
Process Regularity	Regularly **during work phase** or project runtime	Once for each relevant dataset, usually **at the end of or after work phase**	Once for each selected dataset, **either during or after work phase**
Effort	Depends – e.g.: set up once, verify regularly	Establish predefined procedure with data archive (e.g. data center)	Process documented, sometimes guided by repository

Self-Assessment: Backup

Check for yourself:

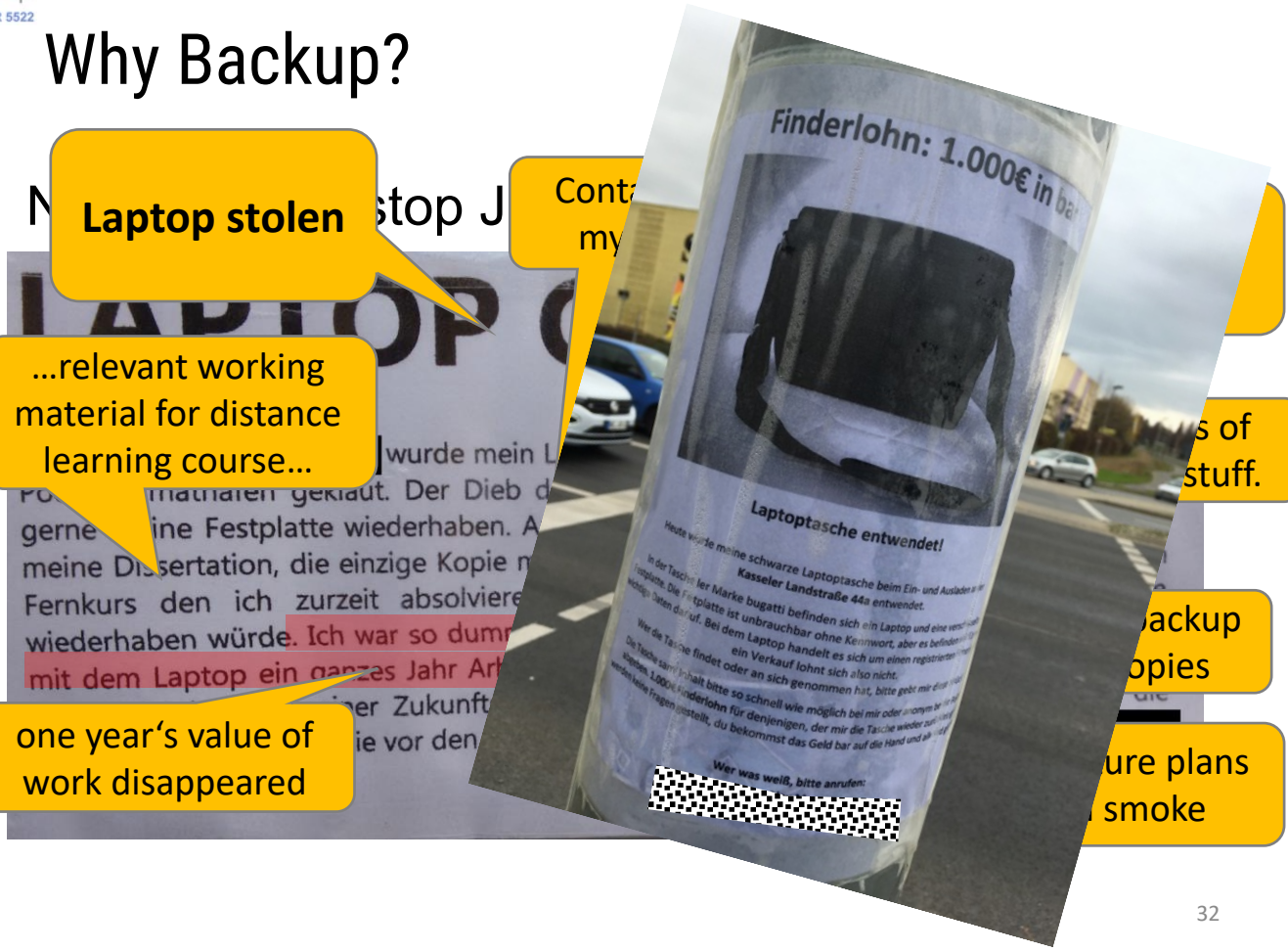
- Do you backup your research data? How?
- How often do you do it?
- Have you ever tried to recover a deleted file?
- Can you return to a previous version of a file?
- Who is responsible for Backup and Storage services at your institute?

Why Backup?

Laptop stolen

…relevant working material for distance learning course…

one year‘s value of work disappeared



Why backup?

Source: Gino on flickr

Source: Kjell Eson on flickr

Source: steviep187 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

Source: reuters.com, image by Sapeurs-Pompiers du Base-Rhin, 2021:

OVHCloud data center in Strasbourg, France

Why backup?

As ***you*** are responsible for your research methods and results, ***you*** are also responsible for your research data.

If you rely on others to store your data, make sure you know:

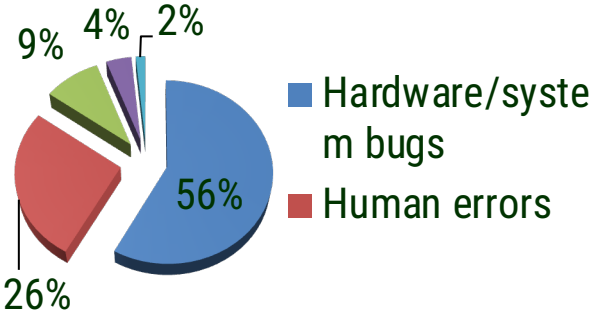
- where the data are stored
- what measures are in place for data safety *and* data security
- how you or someone else can access and retrieve backups
- whom to contact in case of emergency or for support

Sources of Data Loss

Sources of Data Loss

- Malware / Theft / Destruction
- Software failures
  - Program errors / bugs / software updates
  - Features
    (e.g.: Dropbox overwriting on synchronization)
- Hardware failures
  - Bad design / cheap parts / defects
  - Age
  - Dropped laptops / HDDs
  - Liquids (water, coffee, coke)
  - Lightning strikes / electric pulses
- Human errors
  - Accidental deletion
  - Missing knowledge

Source: a man working at home while eating breakfast by Socialeurope via flickr: https://www.flickr.com/photos/socialeurope/4303391587



Source: Kroll Ontrack, 2007, Robin Harris, http://www.zdnet.com/blog/storage/how-data-gets-lost/167

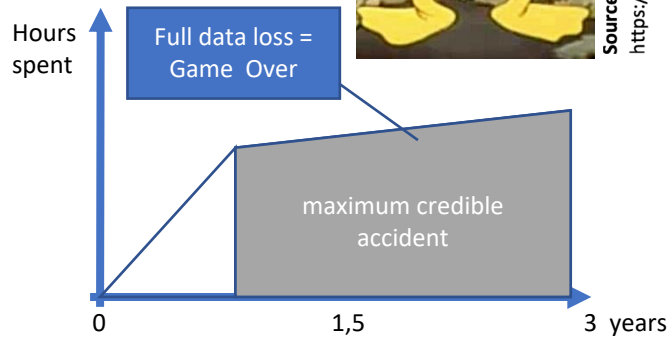
Further reading: disasters and tales of data loss, statistics on how data gets lost

Costs of data loss

Is backing up really worth the effort?

- PhD or postdoc salary costs for employer:
  € 79.800 – 86.100 / year *
- Estimated costs for losing data of one year's work:
  usually even higher

Source: https://de.wikipedia.org/wiki/Dagobert_Duck



➢ **Besides, you can lose a lot of time … and possibly your nerves**

Required investments:

- External hard drives start at € 50,-
- Backup Software is included in most modern operating systems

➢ **When will you start? When will you be required to?**

* DFG staff appropriation rates for 2024: https://www.dfg.de/de/formulare-60-12-246894

Backup Principles

- Create multiple backups
- Expect human errors (keep older versions)
- Do not use backup drives for sharing files
- Store backups physically separate from your PC / laptop
- Check your backups regularly
- Practice the worst case and make a full recovery dry-run
- Discuss the topic with friends to learn their best-practices
- Include required software, scripts, documentation in your planning

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

Backup Strategy Recommendations

1. Use an **institutional storage solution** for backing up your work and data on a daily basis (e.g. network drive, ownCloud)

2. Use your OS built-in backup software or an external software to **regularly back up your data to external media**
   - Predefine files/folders that are regularly backed up
   - Automate the process, eventually make a calendar entry

3. Keep one backup copy at a **remote place**

4. **Check your backups regularly** (e.g. once/month) on integrity by performing by random checks

5. Practice doing a **full recovery dry-run** at least once
   - Possibly have a workstation set up at work for practicing

Backup-Software

Operating system	Integrated Backup SW	Comment
Windows 7	File Recovery	• Requires configuration, to not only copy local libraries • Can create bootable image
Windows 8/10	File History („Dateiversions-verlauf“)	• Saves only local libraries • Can be configured by individual libraries and excluding folders • Cannot create bootable image
Mac OS	Time Machine	• Saves everything except excluded folders • Can use encryption • Can be used to recover a non-booting Mac
Ubuntu	Déjà Dup	• 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, Timeshift

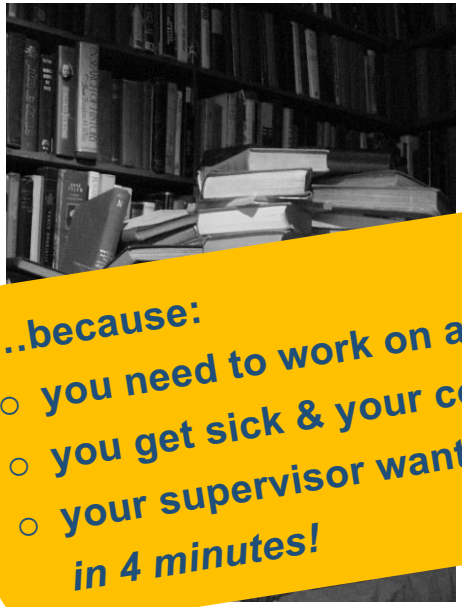
GWDG Storage Services

Name	Backup	Archival	Sharing	Comment
Fileservice / Active Directory	Yes	No	Possibly	Network drives, e.g. P: , possibly more Are backed up automatically
IBM Tivoli Storage Manager (TSM)	Yes	Partially	No	Offer for institutes to centrally back up servers or local workstations
CrashPlanProE	Yes	No	No	Individual Backup-Solution GWDG License: €29,- per year
ownCloud	Partially	No	Yes	Free storage: 50 GB, extendable
GRO.data	Partially	Partially	Yes	Primarily for data exchange and data storage for subsequent publication
HSM	No	Yes	No	For archiving data from finished projects
GitLab	No	Partially	Yes	Versioning; not for large data amounts

More info: https://info.gwdg.de/dokuwiki/doku.php?id=en:services:storage_services:start

Organisation of Data

Why Organize?



austinevan/1225274637/

chy/6829994084

…because:

- **you need to work on a different task for 2 weeks**
- **you get sick & your colleagues need to finish your joint publication**
- **your supervisor wants your results from 4 months ago –**
  in 4 minutes!

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

File Naming Conventions

To stay organized, you should define:

- A self-describing folder structure or tagging scheme
- What information should be in filenames
- How filenames should be structured
- How to group and label related files
- How to refer to and exchange files

Use what works for YOU

AND STICK TO IT !

… especially when working in a team!

Self-speaking file name:

```
20240613_RDM_CRC1073_v42.pptx
```

vs. short file name:

~~`CRC1073_final.pptx`~~

Avoid special characters

~~`„ “, ‘ ´ ` {} < > : ; / \ ? ! $ & ~ *`~~

File Naming Policy Example

- All data files must be stored using this naming scheme:

```
[experiment_datetime]_[experiment_name]_[experiment_part]_[data_type]_[execution_no].[file_extension]
```

- All script files must be stored using this naming scheme:

```
[experiment_datetime]_[experiment_name]_[experiment_part]_[script_version_no].[file_extension]
```

- All analyses files must be stored using this naming scheme:

```
[experiment_datetime]_[experiment_name]_[experiment_part]_[analysis_datetime]_[analysis_type].[file_extension]
```

- All dates in filenames must be in ISO 8601 format: YYYY-MM-DD
- All data files, script files, analyses files and image files are stored in separate folders: `data`, `scripts`, `analyses`, `images`
- For publishing datasets, zip archives are created containing the relevant files and according to the following naming scheme:

```
[experiment_datetime]_[experiment_name]_[experiment_part]_[archiving_date].zip
```

Versioning

```
20240613_RDM_CRC1073_v13.pptx
20240613_RDM_CRC1073_v13final.pptx
20240613_RDM_CRC1073_v13new-final.pptx
20240613_RDM_CRC1073_v13final-finalv1.pptx
20240613_RDM_CRC1073_revised_v01a.pptx
```

Best practice:

- Before editing a file, save it under a new name as a new version
- Use consecutive **version numbers** and eventually **author initials**
  - No „final“ or other unreliable descriptors in filenames
  - Rather **use folders** to mark/sort different purposes and avoid confusion
- If you collaborate, agree on a common naming system for versioning
- ***OR:*** Use a versioning system
  - e.g. gitlab, ownCloud or GRO.data

Version control with git

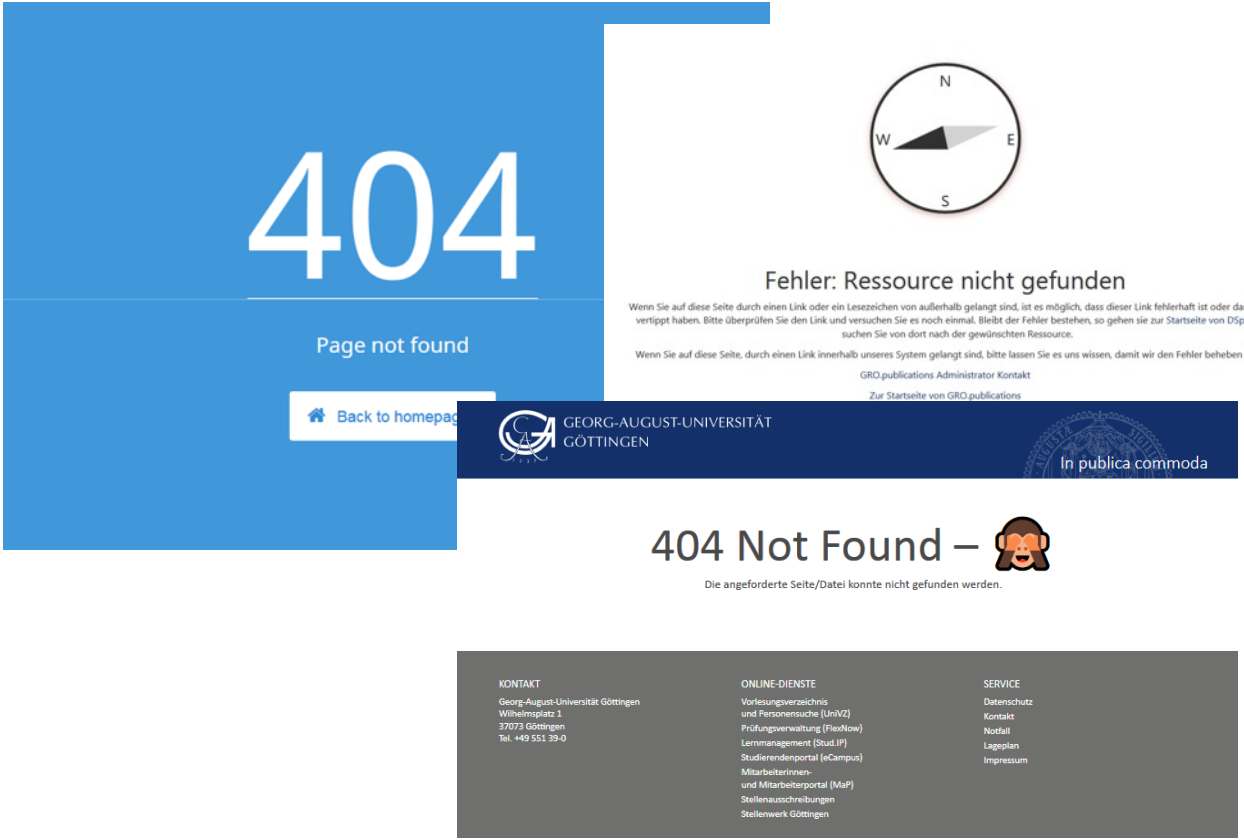
- Mainly used in software development
  - many functionalities for the support of development processes
- Also usable for versioning of documents

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

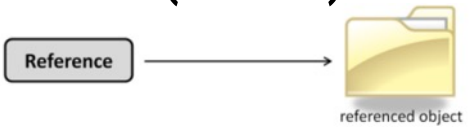
More information: https://info.gwdg.de/docs/doku.php?id=en:services:email_collaboration:gitlab:start

Persistent Identifiers (PIDs)



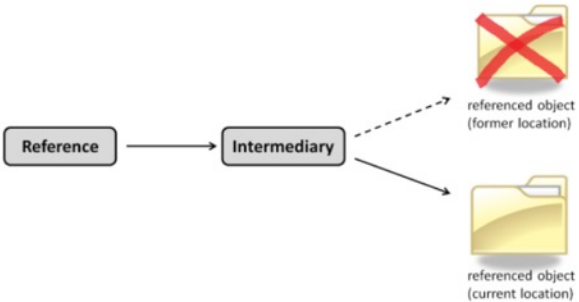
What are persistent Identifiers (PIDs) ?

Common referrers or links like URLs point directly to the location of an object:



When the storage 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 link and the referenced object:



This intermediary monitors any movement or changes performed on the object and always directs requests to the current storage location.

Adapted from: Kálmán, Tibor: Nachhaltige Referenzierung von digitalen Objekten mit Hilfe von persistenten Identifikatoren (Persistent Identifiers), nestor/DigCurV School 2012, 22-24 October 2012, http://hdl.handle.net/11858/00-ZZZZ-0000-0007-5E80-3

What are persistent Identifiers (PIDs) ?

- Prevention of dead links
- Unique naming (referencing) of a digital ressource (e.g. journal articles or research data)
- Assignment of a stable and uniquely referencable 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 https://orcid.org/0000-0001-2345-6789

How do I get a PID?

- Publications: From the publishing journal or repository
- Data: Public depositing in a repository
- At the Göttingen Campus:
  - GRO.data (Campus-Repositorium), GRO.publications
  - SUB Göttingen: GOEDOC, University press
  - GWDG: ePIC PID-Service

More information:
https://www.eresearch.uni-goettingen.de/knowledge-base/howto/getting-an-identifier/

Metadata

Explain these data

AOC	Caesalpinia sappan	2	Dyed silk A	8.0		3	0.12	2.1		0.61		3.8
		2	Dyed wool A	9.8		3	0.16	2.0		0.46		4.9
		2	Pigment A	19.6	0.8	2.5	0.14	5.4	0.1	0.46		3.6
	Carthamus tinctorius	2	Dyed silk A	19.3		2.5	0.06	2.9		0.44		6.6
		2	Dyed wool A	12.2		2.5	0.11	2.8		0.73		4.3
		2	Prepigment on paper	19.7	0.4	2.5	0.05	3.2	0.1	0.40		6.2
	Curcuma longa	2	Dyed silk A	8.7		3	0.12	6.6		0.18		1.3
		2	Dyed silk N	8.5		3	0.18	3.8		0.61		2.2
		2	Dyed wool A	12.3		2.5	0.09	7.7		0.09		1.6
		2	Pigment A	8.6	0.4	3	0.20	4.2	0.9	0.37		2.1
	Gardenia augusta	2	Dyed silk A	6.4		3	0.18	1.1		0.80		5.6
		2	Dyed silk N	9.2		3	0.30	2.8		0.42		3.3
		2	Dyed wool A	7.2		3	0.26	3.1		0.43		2.3
		2	Pigment A	16.3	0.4	2.5	0.10	11.4	0.3	0.16		1.4
	Laccifer lacca	2	Dyed wool A	5.1		3	0.29	0.9		1.24	1.1	4.5*
		2	Pigment A	6.1	0.3	3	0.32	0.8	0.6	9.99	3.0	2.0*
	Lithospermum erythrorhyzon	2	Dyed silk A	6.5		3	0.29	5.4		0.40		1.2
		2	Dyed silk N	8.6		3	0.18	4.0		0.40		2.2
		2	Pigment A	16.9	1.6	2.5	0.17	4.3	0.3	0.49		4.0
	Philodendron amurens	2	Dyed silk N	4.7		3.5	0.34	3.7		0.39		1.3
		2	Pigment A	7.2	0.1	3	0.18	0.7	0.4	1.72	1.0	7.0*
	Rhamnus catharticus, immature	2	Dyed silk A	5.2		3	0.31	0.7		1.83	1.2	4.4*
		2	Pigment A	11.2	0.1	2.5	0.15	1.7	0.4	1.06	1.5	6.5
	R. catharticus, ripe	2	Dyed silk A	10.7		3	0.09	4.2		0.60		2.5
		2	Dyed wool A	7.2		3	0.18	2.9		0.57		2.5
	Rubia tinctorum	2	Dyed silk A	3.9		3.5	0.37	2.7		0.72		1.5
		2	Dyed wool A	5.1		3	0.19	2.7		0.45		1.9
		2	Pigment A	6.8	0.8	3	0.31	1.3	1.2	5.56	2.3	3.0*
	Sophora japonica	2	Dyed silkA	3.4		3.5	0.54	2.3		0.52		1.5
		2	Dyed wool A	5.0		3.5	0.32	2.2		0.51		2.2
		2	Pigment A	9.6	1.6	3	0.25	1.2	0.1	1.52	2.0	4.9*
EU-Artech	Reseda luteola	2	Dyed wool A	2.0		4	0.51	2.1		0.55		0.9
		2	Pigment AC	7.8	0.6	3	0.15	0.9	0.1	1.08	1.0	8.3*
		2	Pigment AL	7.5	0.7	3	0.08	0.8	0.1	1.37	1.0	7.4*
		2	Pigment AP	5.5	0.5	3	0.28	1.2	1.1	5.67	2.6	2.2*
Schweppe	Rocella tinctoria	2	Dyed wool A	12.1		2.5	0.28	3.0		0.72		4.1
ISOBW	ISOBW1	2	Dyed wool	30.5		—	0.02	7.0		0.11		4.3
	ISOBW2	2	Dyed wool	24.9		—	0.03	5.9		0.17		4.2
	ISOBW3	2	Dyed wool	6.6		—	0.19	0.6		1.45	0.8	8.2*
CMN	Acer	1	Leaf	16.4		2.5	0.10	2.3		1.03	2.4	6.9*
	Aster	1	Leaf	23.3		2	0.04	1.6		0.76		14.2
	Dryopteris	1	Leaf	26.4		1.5	0.05	1.3		0.98		20.8
	Scirpus	1	Grass	9.2		3	0.27	0.4		2.48	1.0	9.6*
	Trifolium	1	Leaf	17.1		2.5	0.09	0.9		1.46	1.3	13.2*

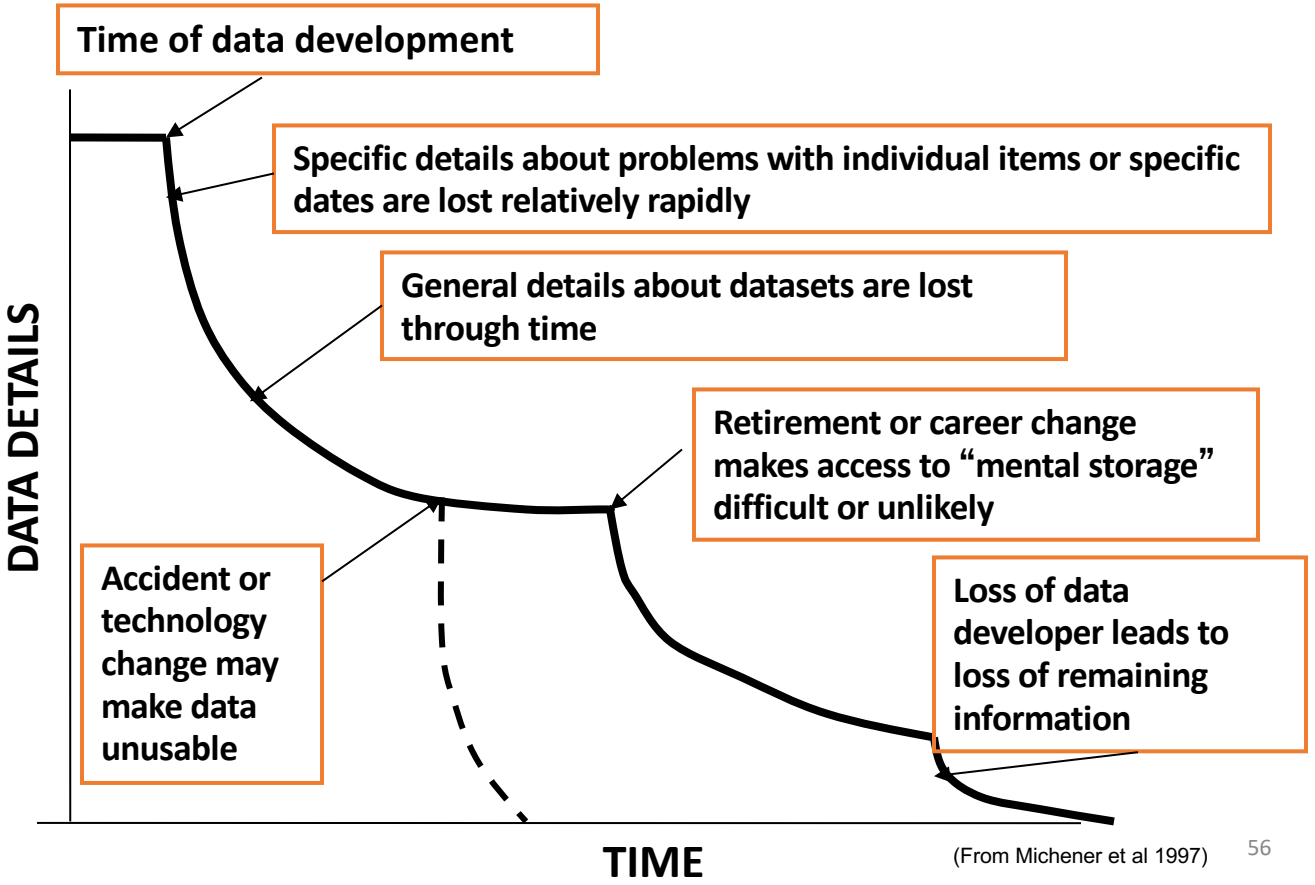
54 Continued

Explain these data

Source	Sample	Trial	Type	ΔE_{Air} Mean	SD	BW	$MCDM_{Air}$:ΔE_{Air} Max	$MCDM_{Air}$ Max	ΔE_{Anoxia} Mean	SD	$MCDM_{Anox}$:ΔE_{Anox} Max	$MCDM_{Anox}$ Max	ΔE_{R}
AOC	Caesalpinia sappan	2	Dyed silk A	8.0		3	0.12		2.1		0.61		3.8
		2	Dyed wool A	9.8		3	0.16		2.0		0.46		4.9
		2	Pigment A	19.6	0.8	2.5	0.14		5.4	0.1	0.46		3.6
	Carthamus tinctorius	2	Dyed silk A	19.3		2.5	0.06		2.9		0.44		6.6
		2	Dyed wool A	12.2		2.5	0.11		2.8		0.73		4.3
		2	Prepigment on paper	19.7	0.4	2.5	0.05		3.2	0.1	0.40		6.2
	Curcuma longa	2	Dyed silk A	8.7		3	0.12		6.6		0.18		1.3
		2	Dyed silk N	8.5		3	0.18		3.8		0.61		2.2
		2	Dyed wool A	12.3		2.5	0.09		7.7		0.09		1.6
		2	Pigment A	8.6	0.4	3	0.20		4.2	0.9	0.37		2.1
	Gardenia augusta	2	Dyed silk A	6.4		3	0.18		1.1		0.80		5.6
		2	Dyed silk N	9.2		3	0.30		2.8		0.42		3.3
		2	Dyed wool A	7.2		3	0.26		3.1		0.43		2.3
		2	Pigment A	16.3	0.4	2.5	0.10		11.4	0.3	0.16		1.4
	Laccifer lacca	2	Dyed wool A	5.1		3	0.29		0.9		1.24	1.1	4.5*
		2	Pigment A	6.1	0.3	3	0.32		0.8	0.6	9.99	3.0	2.0*
	Lithospermum erythrorhyzon	2	Dyed silk A	6.5		3	0.29		5.4		0.40		1.2
		2	Dyed silk N	8.6		3	0.18		4.0		0.40		2.2
		2	Pigment A	16.9	1.6	2.5	0.17		4.3	0.3	0.49		4.0
	Philodendron amurens	2	Dyed silk N	4.7		3.5	0.34		3.7		0.39		1.3
		2	Pigment A	7.2	0.1	3	0.18		0.7	0.4	1.72	1.0	7.0*
	Rhamnus catharticus, immature	2	Dyed silk A	5.2		3	0.31		0.7		1.83	1.2	4.4*
		2	Pigment A	11.2	0.1	2.5	0.15		1.7	0.4	1.06	1.5	6.5
	R. catharticus, ripe	2	Dyed silk A	10.7		3	0.09		4.2		0.60		2.5
		2	Dyed wool A	7.2		3	0.18		2.9		0.57		2.5
	Rubia tinctorum	2	Dyed silk A	3.9		3.5	0.37		2.7		0.72		1.5
		2	Dyed wool A	5.1		3	0.19		2.7		0.45		1.9
		2	Pigment A	6.8	0.8	3	0.31		1.3	1.2	5.56	2.3	3.0*
	Sophora japonica	2	Dyed silkA	3.4		3.5	0.54		2.3		0.52		1.5
		2	Dyed wool A	5.0		3.5	0.32		2.2		0.51		2.2
		2	Pigment A	9.6	1.6	3	0.25		1.2	0.1	1.52	2.0	4.9*
EU-Artech	Reseda luteola	2	Dyed wool A	2.0		4	0.51		2.1		0.55		0.9
		2	Pigment AC	7.8	0.6	3	0.15		0.9	0.1	1.08	1.0	8.3*
		2	Pigment AL	7.5	0.7	3	0.08		0.8	0.1	1.37	1.0	7.4*
		2	Pigment AP	5.5	0.5	3	0.28		1.2	1.1	5.67	2.6	2.2*
Schweppe	Rocella tinctoria	2	Dyed wool A	12.1		2.5	0.28		3.0		0.72		4.1
ISOBW	ISOBW1	2	Dyed wool	30.5		—	0.02		7.0		0.11		4.3
	ISOBW2	2	Dyed wool	24.9		—	0.03		5.9		0.17		4.2
	ISOBW3	2	Dyed wool	6.6		—	0.19		0.6		1.45	0.8	8.2*
CMN	Acer	1	Leaf	16.4		2.5	0.10		2.3		1.03	2.4	6.9*
	Aster	1	Leaf	23.3		2	0.04		1.6		0.76		14.2
	Dryopteris	1	Leaf	26.4		1.5	0.05		1.3		0.98		20.8
	Scirpus	1	Grass	9.2		3	0.27		0.4		2.48	1.0	9.6*
	Trifolium	1	Leaf	17.1		2.5	0.09		0.9		1.46	1.3	13.2*

Continued

Information Entropy



(From Michener et al 1997)

Explain Your Data

- Why?
  - Make data ***FAIR***: Findable, Accessible, Interoperable, Reusable!
  - Not only for others, but also and mainly **for yourself**!
- How?
  - Directly write down which **methods/materials** you used. Write down what fails and what was successfully analysed.
  - Write down **time, place, persons involved** in creation of data.
  - Include title, name of **primary and processed data**.
  - **Add a text file** with this information to each data file/folder ***or:*** keep an **overview spreadsheet**
  - **Do not change/erase your original notes** but add more infos chronologically (with date of insertion).

Excursion:
FAIR data principles

FAIR data Principles

Set of guiding principles for research data

- Goal: make data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable
- Address data producers and data publishers to promote maximum use of research data
- FAIR refers to both humans and machines
- 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

More information: *https://www.force11.org/group/fairgroup/fairprinciples*

FAIR data Principles

Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

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

F4. metadata specify the data identifier.

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.

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.

Re-usable:

R1. meta(data) have 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 their provenance.

R1.3. (meta)data meet domain-relevant community standards.

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

FAIR Data ≠ Open Data !

- FAIR Data aims at maximum reusability of data, this includes:
  - specific licensing of data to specify the usage conditions
  - possibility of authentication and authorization
  - metadata and identifiers for findability and reusability

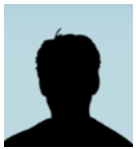
- Open Data (just) means that the data are freely accessible and available, but usually makes no claim regarding
  - findability
  - retrievability
  - interoperability
  - reusability
  - metadata
  - licensing and citation requirements

What are 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

What are metadata?

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



Timo Henne
henne@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 spreadsheet with test data for training purposes

Used random number generator to modify original field data

July 26 2016

At my office Windows PC

To be used in training workshop

- Include:
  - **Description** of the item
  - **Methodology** and **instrumentation**
  - **Units** of measurement
  - **References** to related data
  - **Definitions of** jargons, acronyms, code
  - **Technical information** about the file

"Metadata describes objects in a structured and standardised way…"

Many existing metadata standards, e.g.:

Dublin Core Metadata Element Set (15 optional elements)

ID:	identifier
Technical Data:	format, type, language
Content:	title, subject, coverage, description
Persons & Permissions:	creator, publisher, contributor, rights
Provenance:	source, relation
Life cycle:	date

Can be extended to 55 elements (DCMI Metadata Terms):

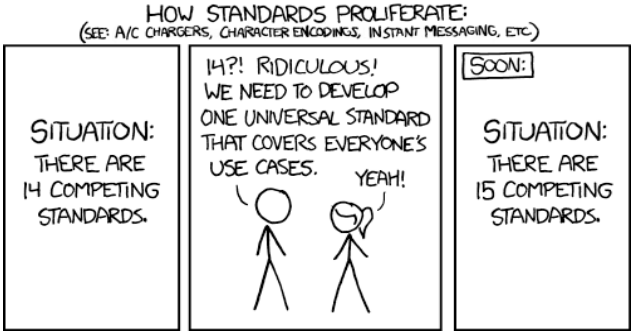
abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy, alternative, audience, available, bibliographicCitation, conformsTo, created, dateAccepted, dateCopyrighted, dateSubmitted, educationLevel, extent, hasFormat, hasPart, hasVersion, instructionalMethod, isFormatOf, isPartOf, isReferencedBy, isReplacedBy, isRequiredBy, issued, isVersionOf, license, mediator, medium, modified, provenance, references, replaces, requires, rightsHolder, spatial, tableOfContents, temporal, valid

- <oai_dc:dc>
  - <dc:title>
    Sociology of Religion: Exercises Using General Social Surveys, 2000-2002 [Instructional Materials]
  </dc:title>
  <dc:creator>Nelson, Edward E.</dc:creator>
  <dc:subject>Bible</dc:subject>
  <dc:subject>Christianity</dc:subject>
  <dc:subject>church attendance</dc:subject>
  <dc:subject>instructional materials</dc:subject>
  <dc:subject>instructional modules</dc:subject>
  <dc:subject>pornography</dc:subject>
  <dc:subject>prayer</dc:subject>
  <dc:subject>religion</dc:subject>
  <dc:subject>religious attitudes</dc:subject>
  <dc:subject>religious behavior</dc:subject>
  <dc:subject>religious beliefs</dc:subject>
  <dc:subject>religious fundamentalism</dc:subject>
  <dc:subject>social issues</dc:subject>
  <dc:subject>sociology</dc:subject>
  <dc:subject>ICPSR.X.A.3</dc:subject>
  <dc:subject>ICPSR.XVI.A</dc:subject>
  - <dc:description>
    These instructional materials were developed from GENERAL SOCIAL SURVEYS, 1972-2002: [CUMULATIVE FILE], compiled by James A. Davis, Tom W. Smith, and Peter V. Marsden. The data file (an SPSS portable file) and accompanying documentation are provided to assist educators in instructing students about religion and social issues in the United States in the late 20th and early 21st centuries. An instructor's handout has also been included. This handout contains the following sections, among others: (1) an exercise using General Social Surveys data to create and validate a measure of religiosity, and then to relate the measure to other social variables, (2) an exercise using General Social Surveys data to explore the relationship between religiosity and other social variables using crosstabulation (focusing on two- and three-variable relationships) and to explore the concepts of explanation, spuriousness, and replication, and (3) an exercise using General Social Surveys data to create a measure of religious fundamentalism and to explore the relationship between this measure and various forms of religious behavior and opinions on social issues. The data contain information on the attitudes of a national probability sample of adults 18 years of age and older on a range of social and political issues. For this instructional subset, some variables were recoded and some new variables were created to facilitate analysis. Variables in the dataset include responses to questions on family and gender roles, abortion, sex and sexual materials, personal morals and social mores, social control, general political attitudes, and socioeconomic status.
  </dc:description>
  <dc:date>2005-01-07</dc:date>
  <dc:type>survey data</dc:type>
  <dc:identifier>3719</dc:identifier>
  <dc:identifier>10.3886/ICPSR03719.v2</dc:identifier>
  <dc:source>personal interviews</dc:source>
  <dc:coverage>United States</dc:coverage>
  <dc:coverage>2000--2002</dc:coverage>
  - <dc:rights>
    ICPSR metadata records are licensed under a Creative Commons Attribution-Noncommercial 3.0 United States License (http://creativecommons.org/licenses/by-nc/3.0/us/).
  </dc:rights>
</oai_dc:dc>

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*

Example Metadata standards

- HDF5

HDF5 is a data model, library, and file format for storing and managing data. It supports an unlimited variety of datatypes, and is designed for flexible and efficient I/O and for high volume and complex data. HDF5 is portable and is extensible, allowing applications to evolve in their use of HDF5.

- NeXus

NeXus is an international standard for the storage and exchange of neutron, x-ray, and muon experiment data. The structure of NeXus files is extremely flexible, allowing the storage of both simple data sets, such as a single data array and its axes, and highly complex data and their associated metadata, such as measurements on a multi-component instrument or numerical simulations.

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

Example Metadata standards

- CHEMINF

The Chemical Information Ontology (CHEMINF) aims to establish a standard in representing chemical information. In particular, it aims to produce an ontology to represent chemical structure and to richly describe chemical properties, whether intrinsic or computed.

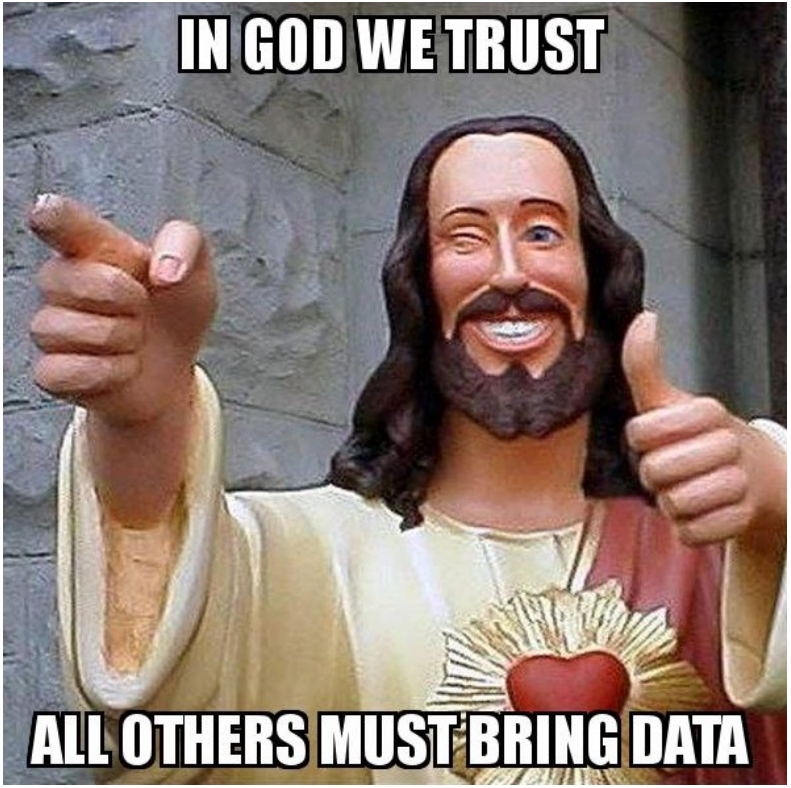
- BioAssay Ontology

The BioAssay Ontology (BAO) describes chemical biology screening assays and their results including high-throughput screening (HTS) data for the purpose of categorizing assays and data analysis.

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

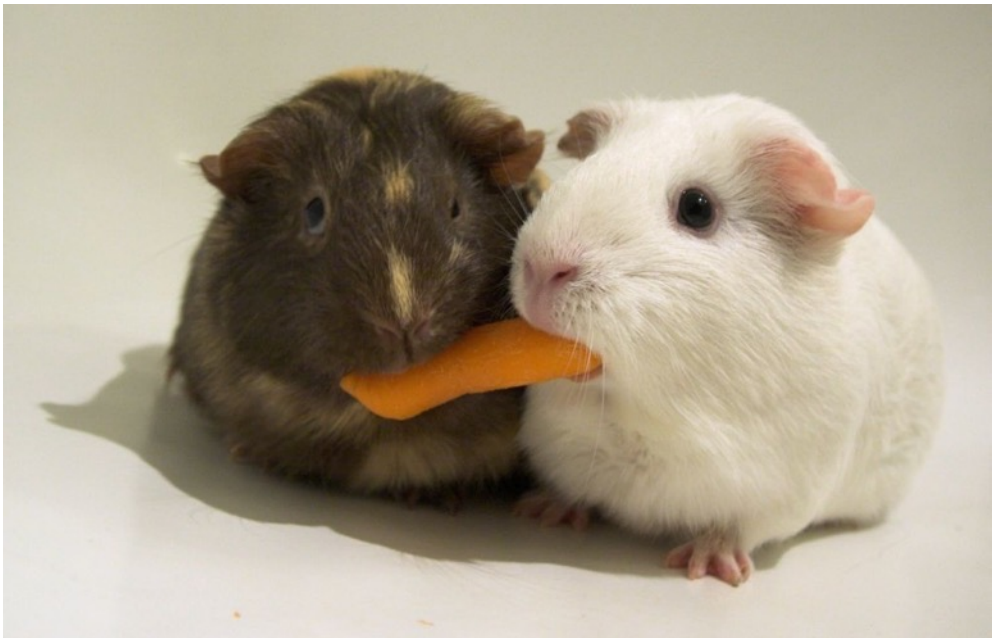
Data sharing

Data sharing - motivation



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

… but active, open sharing? for free??



Source: Sharing by ryancr via flickr

Why share your data?

Reputation

- Get credit for high quality research
- Increased understanding of your methods
- Allows work to be verified by others
- Recognition for contribution to research community
- Extend research beyond your discipline

Funding

- Making data and/or publications available may be a requirement of your funding body
- It may make your funding proposal more attractive when sharing data is not essential

Why share your data?

Impact

- Sharing makes your data:
  - Easier to find
  - Easier to access
- Open data/publications leads to increased citations

Source: Richard Matthews, flickr: dart (2011) online at: https://commons.wikimedia.org/wiki/File:Darts_in_the_middle_of_a_dartboard.jpg?uselang=de

Reuse

- Starting point for a complementary study
- Test data for new software and algorithms
- Teaching purposes
- Contexts not currently envisioned

Data Sharing – Concerns

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

Self-use

No documentation

Work in progress

Theft and misuse

Un-importance

Value over time?

Embargo!

Do it for yourself!

„Working data set“

Trust law & science

Future is unpredictable

Data Sharing - Credits?

- Well documented research data

  helps your own (future) research
- Shared data may serve as

  facilitator for cooperation
- Increased accessibility and usability will

  enable reuse and citations
- Public data and open access will

  extend the range of your data and research

Data Sharing – Real Barriers

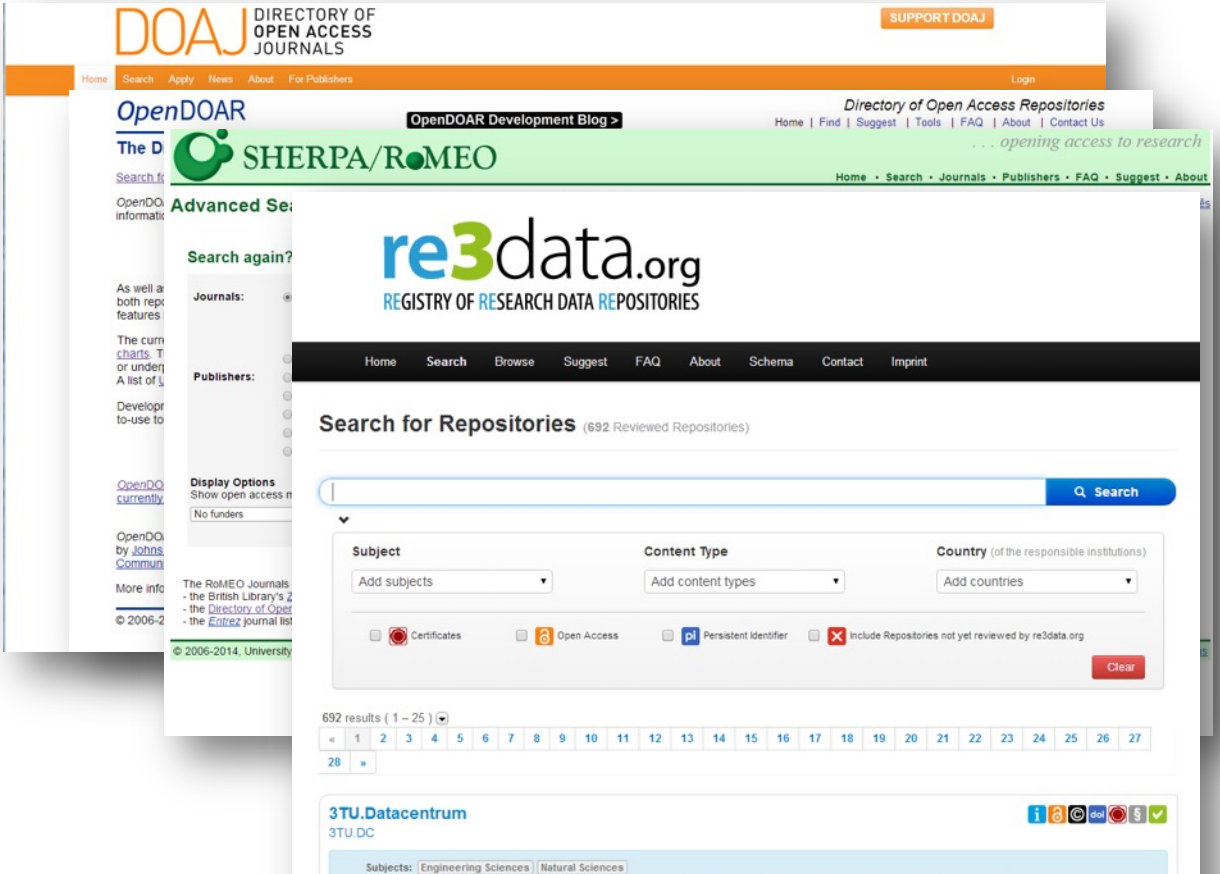
- Place
  - No sharing tradition
  - No repository
  - No expertise
- Funds
  - No money
- Rights
  - No carte blanche

Source: Simatai_Great_Wall by Arian Zwegers on Wikimedia Commons

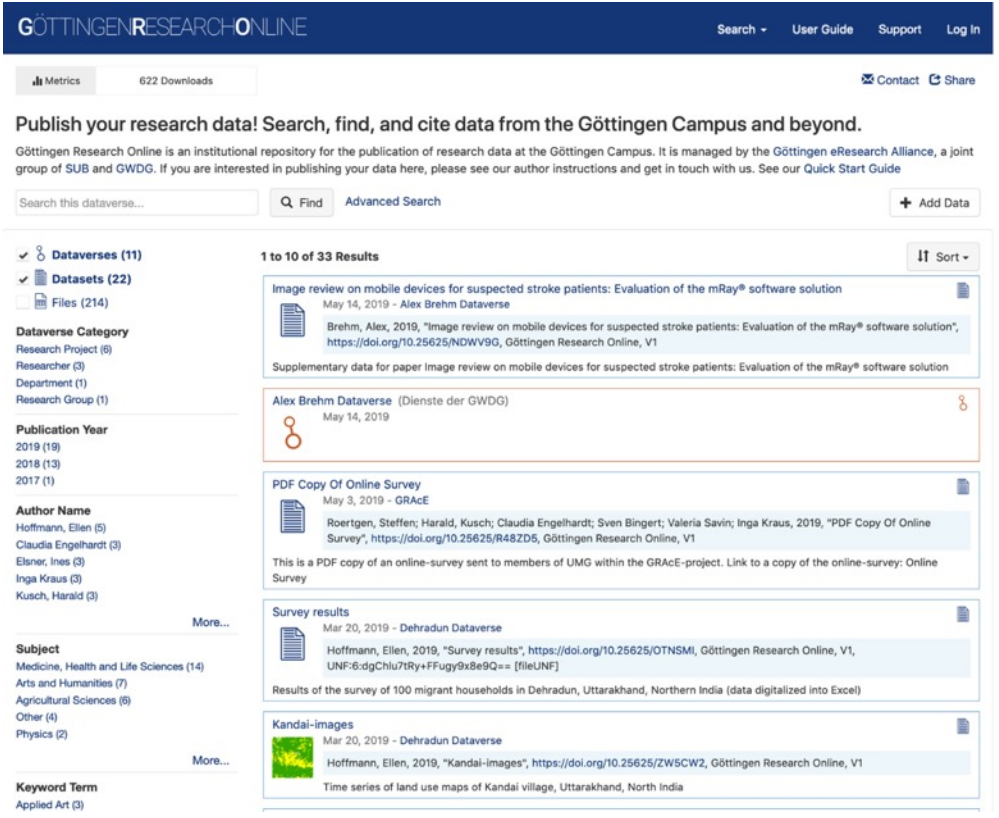
Modes of Sharing

Transfer Way	Access Mode	Use Condition
peer-to-peer	restricted	none
webspace	on demand	agreement
repository	embargo	license
	open	

Finding OA journals and repositories



GRO.data: Research Data Repository



https://data.goettingen-research-online.de/

zenodo

Search

Upload Communities

Log in Sign up

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)

Zenodo now supports DOI versioning!

Read more about it, in our newest blog post.

Using GitHub?

Just Log in with your GitHub account and click here to start preserving your repositories.

Zenodo in a nutshell

- **Research. Shared.** — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- **Citeable. Discoverable.** — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- **Communities** — create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
- **Funding** — identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- **Flexible licensing** — because not everything is under Creative Commons.
- **Safe** — your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.

Read more about Zenodo and its features.

Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv

Login

Search... All fields Search

Help | Advanced Search

arXiv is a free distribution service and an open-access archive for 2,085,125 scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics. Materials on this site are not peer-reviewed by arXiv.

Subject search and browse:

Physics Search Form Interface Catchup

News

Read about recent news and updates on arXiv's blog. (View the former "what's new" pages here). Read robots beware before attempting any automated download.

COVID-19 Quick Links

See COVID-19 S

- arXiv
- medRxiv a

Important: e-pr they should not practice or healt media as establi in the field.

Physics

- Astrophysics (**astro-ph** new, recent, search)
  includes: Astrophysics of Galaxies; Cosmology and Nongalactic Astrophysics; Earth and Planetary As Instrumentation and Methods for Astrophysics; Solar and Stellar Astrophysics
- Condensed Matter (**cond-mat** new, recent, search)
  includes: Disordered Systems and Neural Networks; Materials Science; Mesoscale and Nanoscale Phy Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity
- General Relativity and Quantum Cosmology (**gr-qc** new, recent, search)

Cornell University

arXiv.org blog

news from arXiv.org

Home > about arXiv > New arXiv articles are now automatically assigned DOIs

New arXiv articles are now automatically assigned DOIs

Update: As of Feb 2022, all arXiv articles now have DOIs.

New articles submitted to arXiv are now automatically assigned DOIs that align with their arXiv ID. This makes research articles more discoverable across search engines because associated metadata is made available to the community in a reusable format.

DOIs (digital object identifiers) are unique and unchanging, just like the original arXiv ID number already assigned to every arXiv article. However, because DOIs are used across many different platforms, they enable greater interoperability with other services.

"Sure, the arXiv papers have had a persistent identifier for years, the arXiv identifier, and that's a very good thing not only for citability. But with the assignment of DOIs, arXiv now becomes even more visible as an important element of the publication ecosystem," said Dr. Irina Sens, Deputy Director and Head of Library Operations at the Technische Informationsbibliothek in Germany, which organizes arXiv's largest member consortium. "The DataCite metadata schema enables connections to other persistent identifiers, for example researchers' ORCIDs and institutional RORs, and other content like associated research data – and offers easy tracking of research outputs through simple user interfaces. That's a huge step forward."

arXiv's DOI prefix with DataCite is 10.48550

DOIs align with arXiv IDs

Publishing / sharing research software

- Any scripts, methods implementations or programs that you use or develop in your research can be important to understand, verify, reproduce or reuse your results
- If possible, you should also share or publish such software in a similar way as your research data (including metadata and documentation), in order to:
  - ensure your software is citable, preserved, and accessible to support scientific reproducibility, replication, and transparency.
  - gain appropriate credit for your work
  - help your research community by enabling reuse
  - fulfil increasing requirements by funders and journals
- Where and how do you do that?

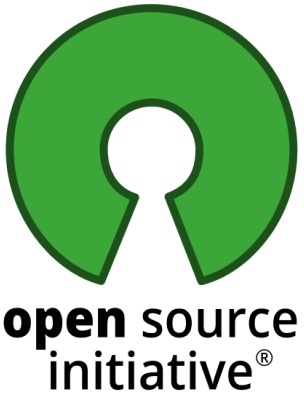
Publishing software: where

	Main purposes	Examples	Pros
Source code repository	Accessibility, cooperation, development	Github, Codeplex, BitBucket, Sourceforge	ongoing development, own organization & structure, workflow integration
Software journal	Scholarly attribution, credit, citation	JOSS (OA), JORS (SSI), RescienceC (OA), PLOS One, Computer Physics Communication, SoftwareX (Elsevier)	Visibility, DOI assignment, review process
General application archive	Transparency, replicability, reuse	Zenodo, Figshare	Versioning, DOI assignment
Disciplinary archive	Reuse	CRAN, NanoHub	Visibility, possibly across communities
Institutional archive	Preservation	GRO.data	Easy access, institution‘s commitment

Publishing research software: how

OSI's Open Source Definition

- free redistribution
- source code availability
- derivatives allowed
- no limitations of who may use it or for what
- no additional license in place
- license must not depend on distribution format, technology, presence of other works



Source: https://github.com/readme/guides/open-source-licensing

Publishing research software: how

1. Create a software citation file
   - can be human readable CITATION files or machine readable
   - should include:
     - author(s)
     - title of the software package or code
     - link to the code location
     - DOI or other unique identifier
     - version number
     - release date
2. Choose a license

Research software licenses

- Why do I need a license?
  - As creator, you are the sole copyright holder of your work
  - Others are legally not allowed to reuse your software, even if it is freely available
  - A license allows you to define *additional rights and obligations* regarding your work that go beyond the copyright you own, and it can protect yourself
- What options are there in licenses?
  - adhere to similar terms: Copyleft (GPL, AGPL, LGPL), CC-BY-SA
  - attribution & protection: MIT, BSD, Apache, CC-BY
  - no commercial use: CC-BY-NC
  - no claims: Public Domain, CC0, Unlicense

Source: https://github.com/readme/guides/open-source-licensing

Publishing research software: how

1. Create a software citation file
2. Choose a license
3. Include a license in your software
   a) can be included in citation file
   b) or create a separate LICENCE.txt
4. Choose your channel(s) or place(s) for publication
   a) Code repository
   b) Software journal
   c) Software archive

Publishing software: Further reading

- https://libguides.mit.edu/software
- https://www.software.ac.uk/top-tip/which-journals-should-i-publish-my-software
- https://www.tudelft.nl/en/library/research-data-management/r/publish/publish-research-software
- https://blog.tib.eu/2024/02/29/software-journals-der-einfluss-von-software-auf-die-forschung/
- https://github.com/readme/guides/open-source-licensing
- https://choosealicense.com/
- https://www.tldrlegal.com/

Other services on campus

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
Overleaf (ShareLaTeX)	GWDG	Collaborative LaTeX editing in the browser

GWDG services

SERVICES

Storage Services
- Data Archiving
- Backup
- File Service
- GWDG ownCloud
- Cryptshare

E-Mail Collaboration
- E-Mail-Service (MS Exchange 2016)
- Spam and Virus Filtering
- Mailing Lists
- MS Sharepoint
- Managed Services
- Project Management Service
- GWDG Pad
- ShareLaTeX
- Rocket.Chat
- GWDG Web-Office
- GitLab

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

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

Application Services
- Persistent Identifier (PID)
- Library Service Aleph
- Library Service Koha
- Databases
- Application/Event Management
- Plagiarism Prevention
- Online Surveys
- Bioinformatics Programs
- Statistics Programs
- Jupyter
- Pseudonymisation and Data Trusteeship

General Services
- Videoconferencing
- Identity and Access Management [MPG]
- Identity and Access Management [Uni]
- Single Sign-on (SSO) /Authentication and Authorization Infrastructure (AAI)
- URL Shortener
- Software and Licence Management
- Computer Lending Pool
- General Services - 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

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

https://www.gwdg.de/services

Further information

https://reproducible-science-curriculum.github.io/rr-jupyter-workshop/

https://www.forschungsdaten.info/

https://www.dfg.de/foerderung/info_wissenschaft/2022/info_wissenschaft_22_25/index.html

https://www.eresearch.uni-goettingen.de/consulting-and-training/training-material/

Göttingen eResearch Alliance (eRA)

Team

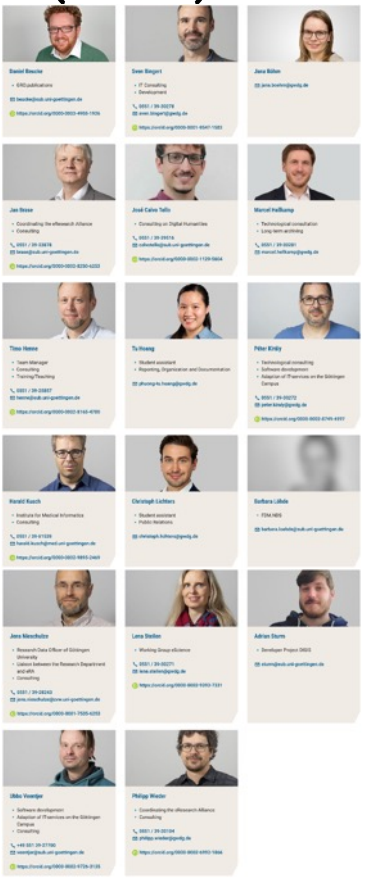
- Diverse scientific backgrounds
  - Mainly in natural sciences and humanities
- Jointly run by

GWDG Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

SUB NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

- Partner: University Research & Transfer Department, Medical Informatics
- High expertise on *e*Research topics

→ *„We are no experts in your discipline, but we can relate to your data management requirements.“*



What eRA can do for you

- Consultations / Support
  - Research Data Management
  - Publication strategies
  - Digital methods, software and technologies to enhance a research project
  - Information hub for experts & expertise on the whole campus
- Training
  - Discipline- or project-specific or general
  - Information material / knowledge base
- Collaboration
  - Liaising project partnership
  - Project as a service

www.eresearch.uni-goettingen.de

The deeper meaning of Research Data Management

Source: cmhughes on pgfplots

CONTACT:

info@eresearch.uni-goettingen.de

www.eresearch.uni-goettingen.de