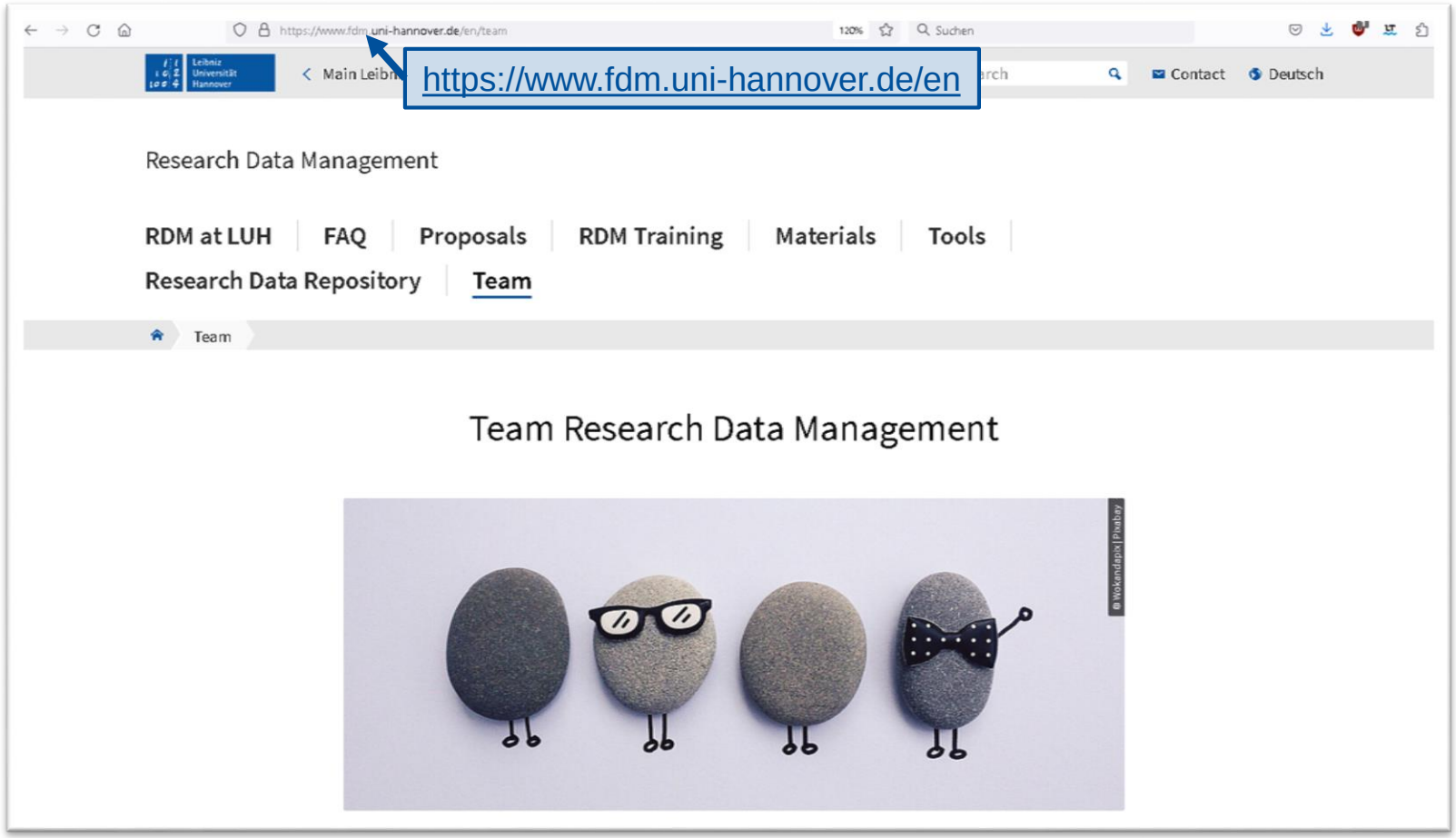
Managing digital research data

Basics, tips & tricks

Team Research Data Management

last updated: August 2024

Welcome!

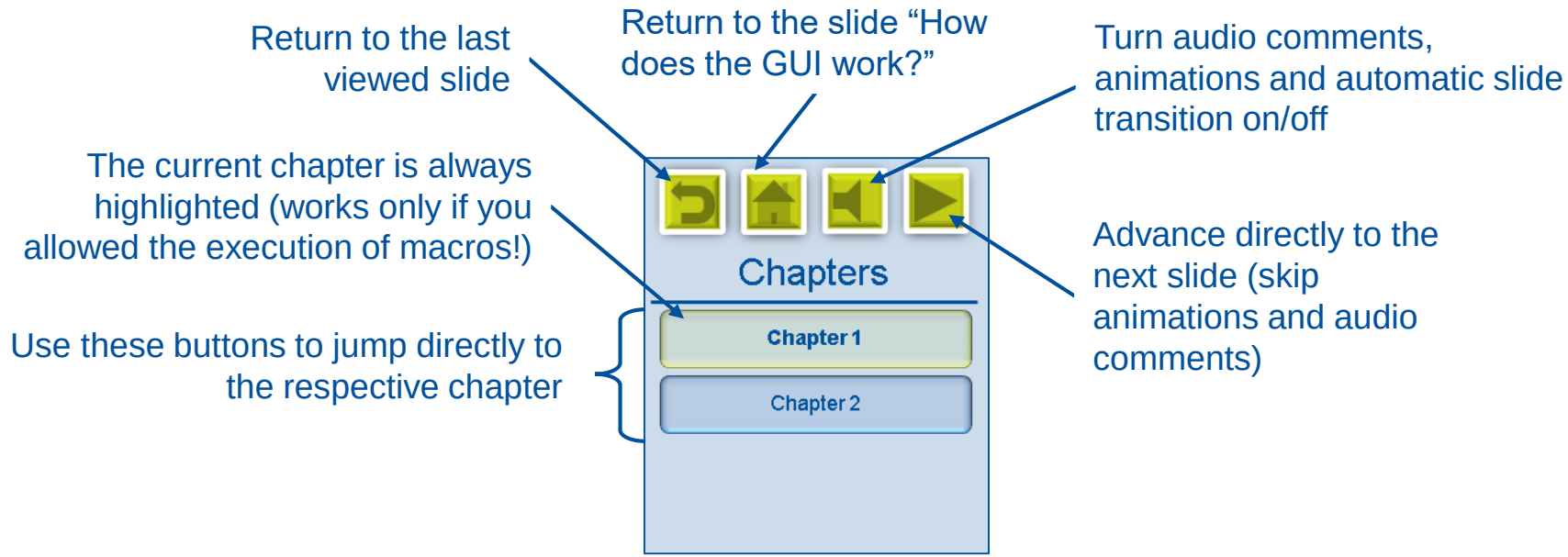


https://www.fdm.uni-hannover.de/en

visit the website of the Research Data Support Team

Chapters

How does the GUI work?



On many slides you will find clickable buttons, which will lead to additional information on external websites. These may contain video tutorials, legal texts, checklists and many more.

go to website

Contents and aspiration of this course

Aspiration of this course: Giving an introducing overview over the most important aspects of the handling of digital research data. Addresses persons with little prior knowledge.

go to the advanced courses

Image by Menno de Jong: https://pixabay.com/de/photos/lemur-neugierig-halfaap-guck-guck-3295891

Introduction

What are research data?

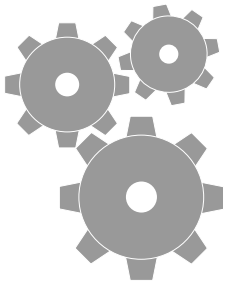
What is research data management (RDM)?

Why is data management important?

What are research data?

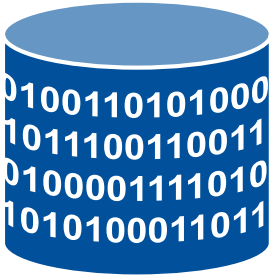
documentation

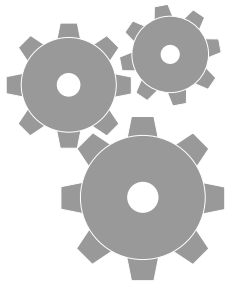
raw data



processing

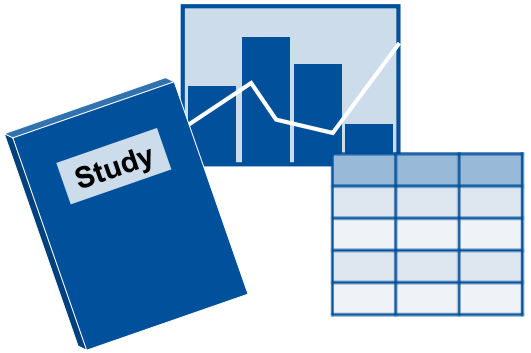
processed data





analysing

results



What is research data management (RDM)?

planning

processing

administrating

structuring

documenting

selecting

archiving

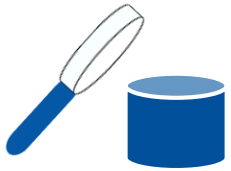
publishing

Image by Tima Miroshnich: https://www.pexels.com/photo/focused-professional-man-using-laptop-7567529/

i

RDM: Conscious and systematic handling of data from the planning stage until the end of a project and perhaps even beyond.

Why is data management important?

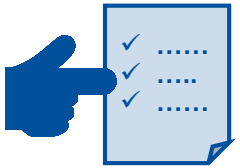


✓ you keep an overview

✓ team work is easier

✓ you can safeguard high quality standards in research

✓ you save time and avoid stress



✓ you comply with official requirements

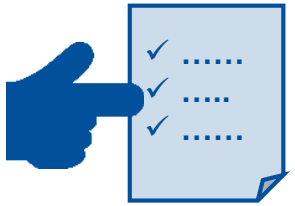
further reading

The Thuringian Competence Network for Research Data Management compiled some “Research Data Scarytales”. These are true stories about data management failures and their consequences.

go to the „Scarytales“

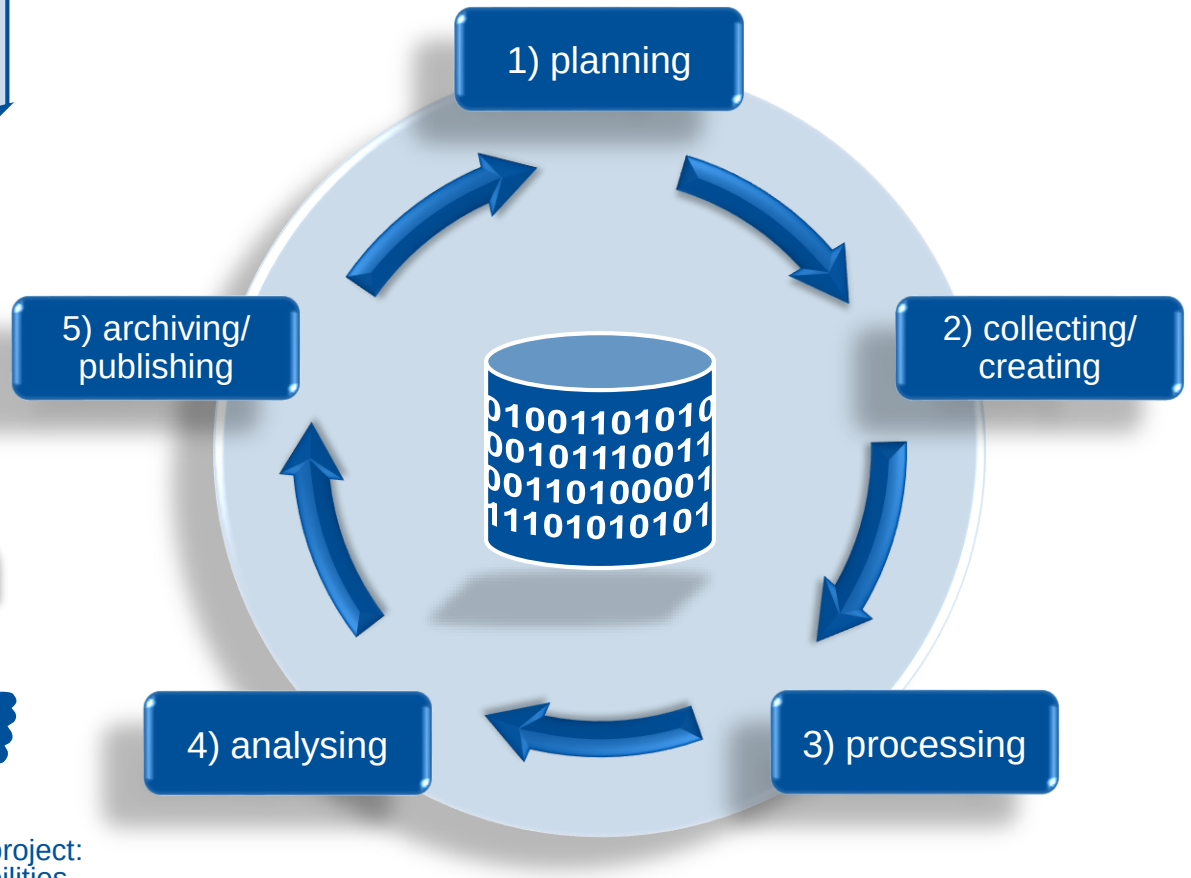
Planning your data management

What should I consider?



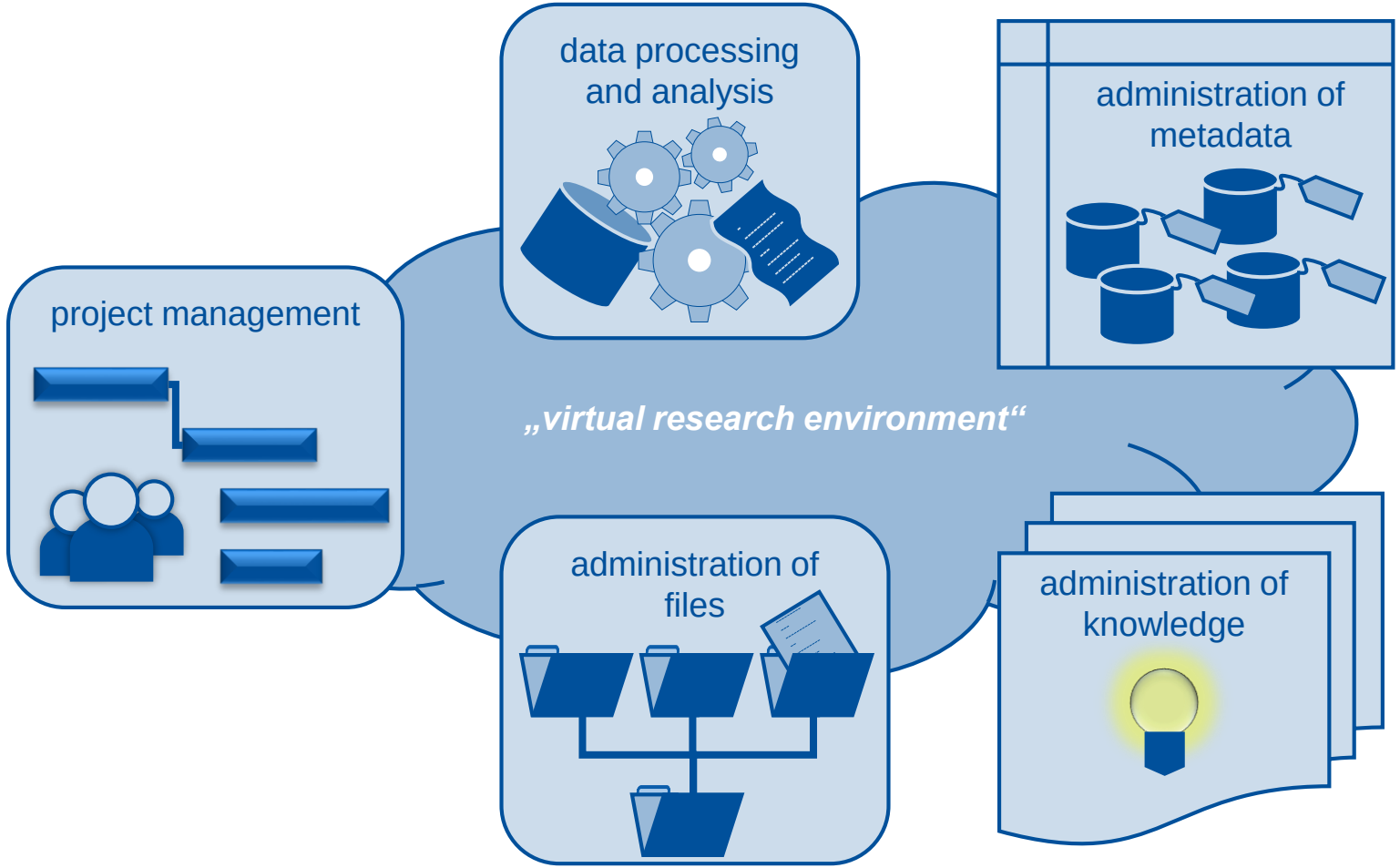
specifications, guidelines, laws

people involved in the project: workflows & responsibilities



- documentation
- structure & overview
- data storage & transmission
- protection against loss and misuse
- selection

Adequate infrastructure



People and departments involved

within the project


PIs


PhD students & postdocs


data manager


students


external partners

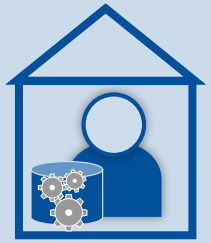
Who has which abilities?

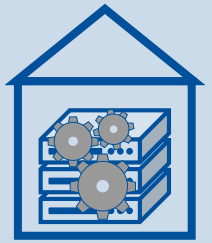
Who should be involved when and how?

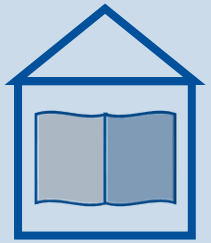
Who fulfils which tasks within the project?

How do you organise team work and interchange?

at the university


your institute´s IT admin


central IT department


library

research service

Developing a data management plan (DMP)

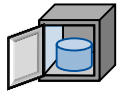
DMP: A structured document containing detailed information on data handling.
→ May and should be constantly updated and complemented during the project

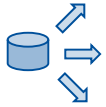
typical structure of a DMP:


administrative information

data collection and basic methodology


storage, backup and security


archiving


sharing and publishing data

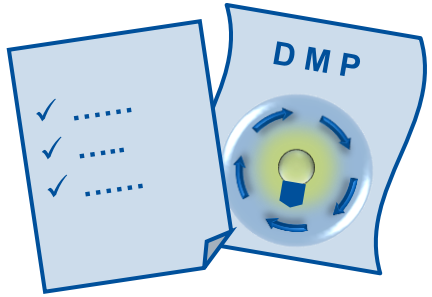

resources und responsibilities

further reading

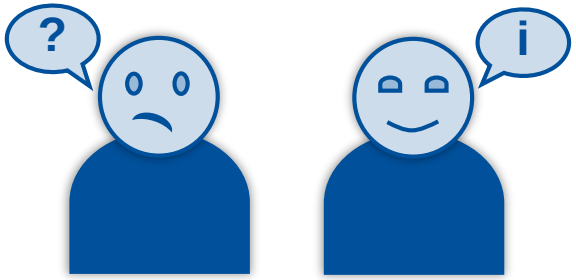
The Digital Curation Centre´s website on data management plans provides tools, checklists and guidance.

go to the DCC website

Tools for developing a data management plan



templates, checklists, examples (can be found in disciplinary web portals and on the websites of some funding bodies)



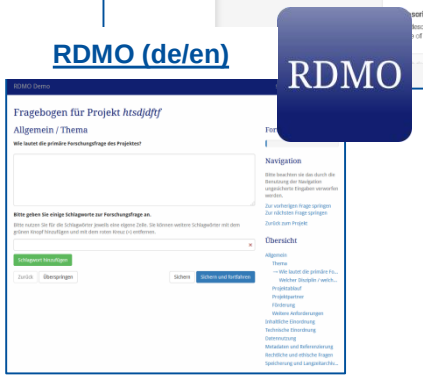
individual counselling by RDM support staff

go to the website of the Research Data Support Team at LUH

argos (en)

DMPonline (en)

RDMO (de/en)



commented online editors

Specifications, guidelines and laws

Laws frequently affecting research practice

RDM is good scientific practice!

Guidelines for handling research data at LUH

Funder requirements

Why funders require RDM

Recommendations for RDM statements in funding proposals

Laws frequently affecting research practice

Image by succo: https://pixabay.com/de/photos/hammer-waage-gericht-justiz-recht-802301

i

We have compiled the most important information on legal issues on our website on the subpage "Legal & ethical aspects in handling research data".

to subpage „Legal & ethics“

§

Privacy and data protection laws

to the official version of the GDPR (Official Journal of the EU)

to NDSG (niedersächsisches Vorschrifteninformationssystem, in German)

for detailed information (in German) see:

Advanced course “Umgang mit personenbezogenen Forschungsdaten. Rechtliche Grundlagen, Methoden und Hilfsmittel.”

go to the advanced course

§

Intellectual property and patent laws

to the German copyright law (Bundesamt für Justiz)

to the German patent law (Bundesamt für Justiz)

further reading (in German):

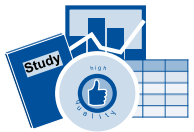
website of the Federal Ministry of Education and Research (BMBF): Urheberrecht in der Wissenschaft: Was Forschende und Lehrende wissen sollten. At LUH, Dezernat 4 offers counsel on patent applications.

to the BMBF website

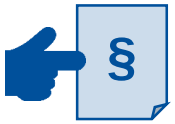
to the Dez.4 website “Patents and Startups“

RDM is good scientific practice!

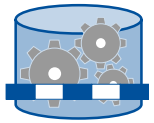
RDM is especially relevant for these 6 out of 19 guidelines:



Guideline 7: Cross-phase quality assurance



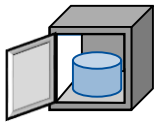
Guideline 10: Legal and ethical frameworks, usage rights



Guideline 11: Methods and standards

Guideline 12: Documentation

Guideline 13: Providing public access to research results



Guideline 17: Archiving

i

You can read the guidelines themselves and accompanying disciplinary comments in the DFG portal “Research Integrity” (English version not finished, yet).

to the DFG portal

!

Please note as well:

- DFG guidelines on the handling of research data
- disciplinary guidelines, requirements, policy papers etc.

You can find both on this website:

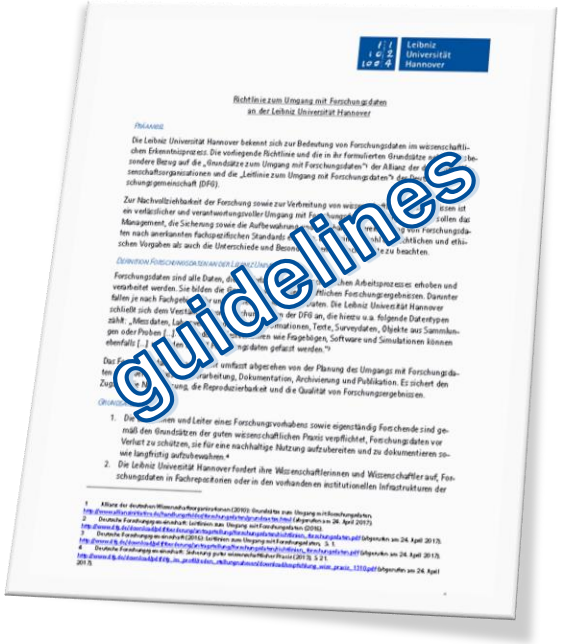
to the DFG website „Handling of Research Data“

Guidelines for handling research data at LUH

principles:

- how to handle research data:
  - protect against losses
  - process for sustainable (re-)use
  - document
  - archive (long-term)

  → head of project is responsible!



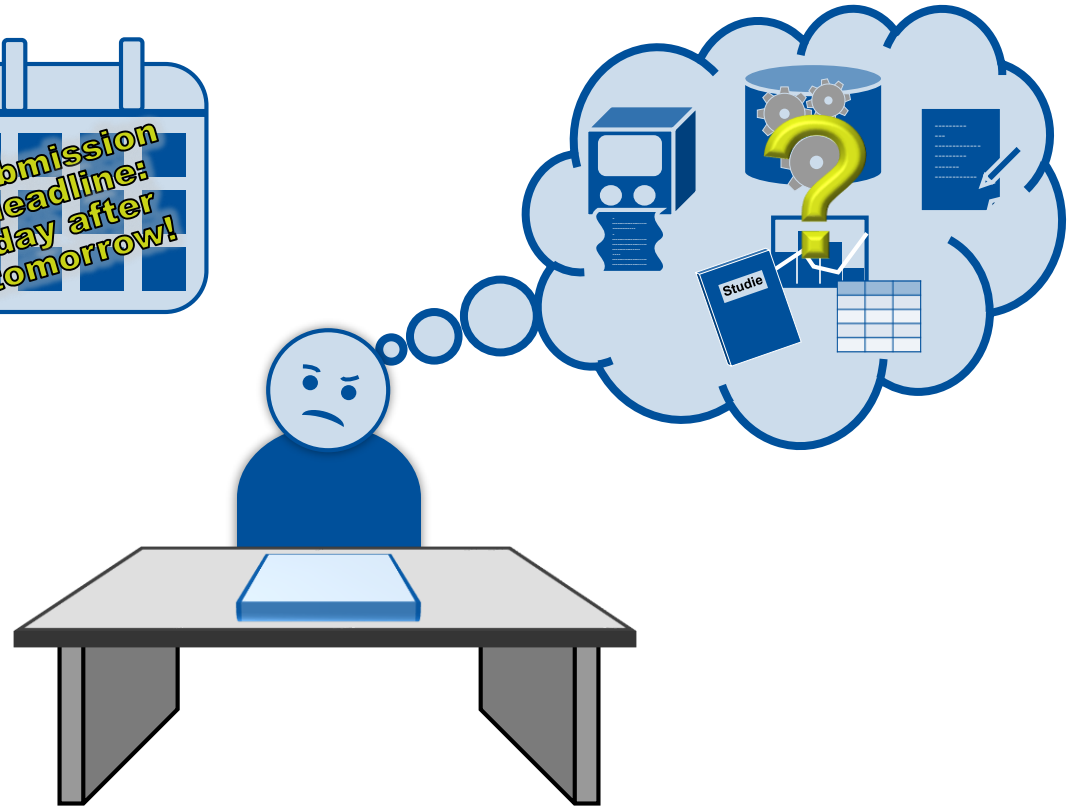
- publish research data according to the FAIR principles in (disciplinary) repositories (request)
- develop project-internal RDM policies and data management plans (recommendation)
- integrate RDM into teaching → faculties (recommendation)

§ You can find the official LUH website with the guidelines here:

got to guidelines

Funder requirements

the classic: „Oh oh, we still have to write something about research data...“



Application guide

------ --- ------ ----- - --
------ -- --- -- ------ -- --------

Data handling

------ --- ------ ---------- -- ---
-- ------ -- -------- ----------

i

Dezernat 4 is happy to support when it comes to drafting RDM-related proposal chapters. The sooner you come up to us, the more efficient our counselling can be. Please do not expect text blocks for copy & paste!

to the Dezernat 4 website (competences overview)

Why funders require RDM

funders' key concerns:
- make sure that results can be validated
- avoid multiple funding

important prerequisites:
- data are accessible long-term
- data are readable and comprehensible long-term

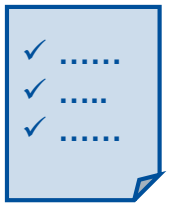
video tip

This video perfectly illustrates a "worst practice scenario"…

Karen Hanson, Alisa Surkis, Karen Yacobucci (2012): Data Sharing and Management Snafu in 3 Short Acts. CC-BY 3.0 unported, doi: 10.5446/31036

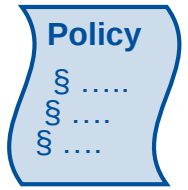
to the video in the AV-Portal of TIB

Recommendations for RDM statements in funding proposals



commitment to relevant policies

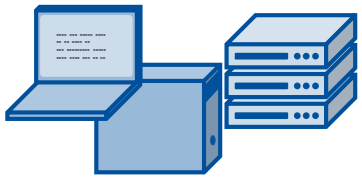
- of the funder (e.g. good scientific practice (DFG), H2020 Open Data Pilot)
- the own university (for German universities, see list on forschungsdaten.org)
- the own discipline (see overview on the respective DFG website, at the bottom)



in case of joint projects: intention to write a project-internal RDM policy

intention to develop a data management plan

intention to elaborate data according to the FAIR principles and to make accessible for re-use open access



information on used (IT) infrastructure: what is available at your own university and beyond?

further reading

We compiled further tips and infos on our website. We also provide a manual for drafting internal RDM guidelines (in German).

go to the proposal infos | see the manual

Science Europe issued an extensive RDM guide. It also contains DMP tips for both, applicants and reviewers .

go to the Science Europe guide

Collecting, storing and processing data

Re-using existing data

Data storage and transmission

LUIS services for data storage and backup

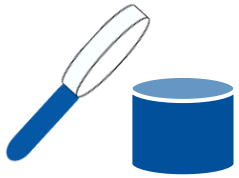
Structured deposition and file naming

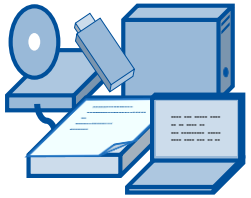
Handy tools to make live easier…

Special services und programmes

Sharing data (also with external partners)

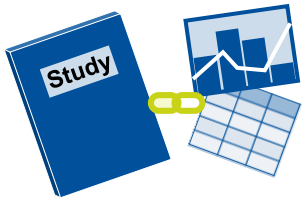
Re-using existing data





your own data from previous works

your institute´s data storage



links to supplement data in journal articles

search engines of repositories

!

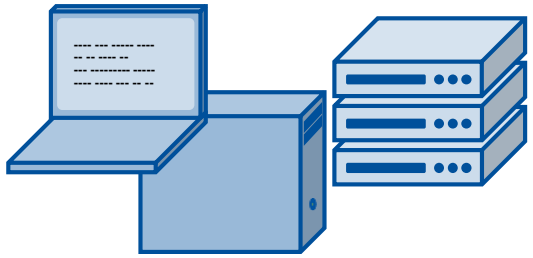
Check before generating new data:

- Are there any existing data you could include in your own analysis (e.g. to broaden your database or for reference and comparison purposes)?
- If the answer is yes, do you have access to these data and all necessary permissions to process them?
- If the answer is yes, can you read and understand the data?

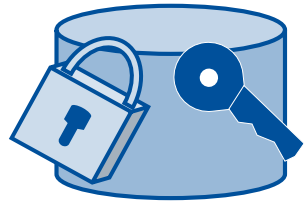
→ At this point you already notice what is important when making your own data available for re-use later on…

Data storage and transmission

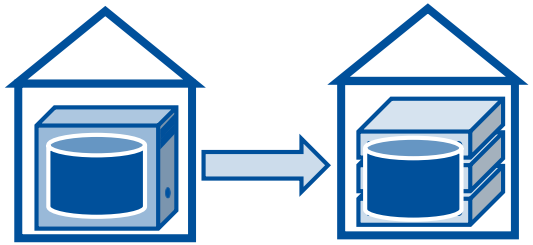
! Make sure you dispose of adequate infrastructure and sufficient capacities before generating any data!



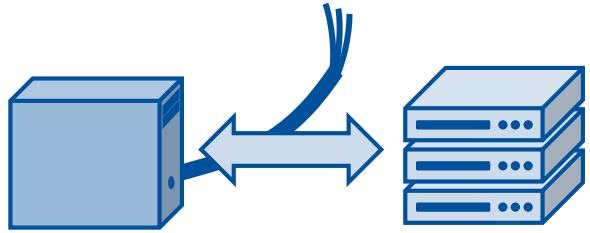
sufficient storage capacity?



adequate security levels of storage systems?

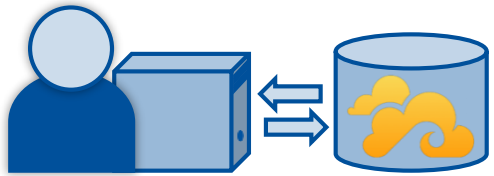


is an automated daily backup guaranteed?
(two copies in different locations)



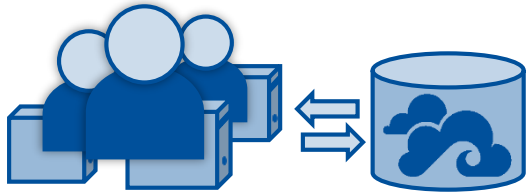
sufficient bandwidth?

LUIS services for data storage and backup



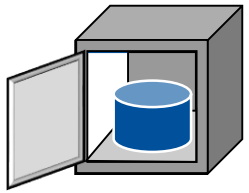
personal “Cloud-Seafile”
(storage + data synchronisation)

go to service description (German)



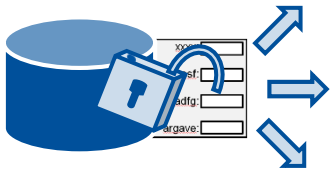
“Projekt-Seafile”
(storage + data synchronisation)

go to service description (German)



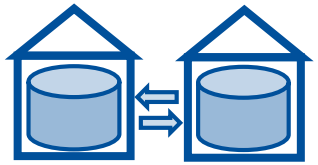
LUH data archive

go to service description



LUH data repository

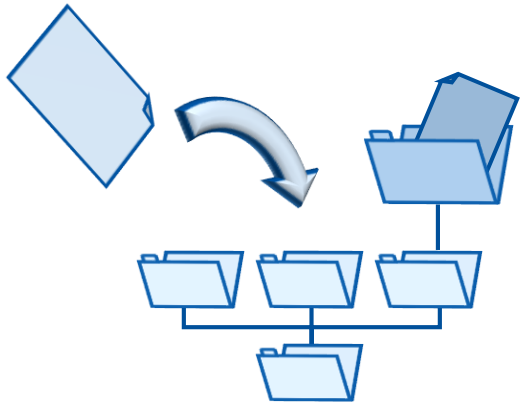
go to service description



Backup & Restore
(for servers of institutes)

go to service description (German)

Structured deposition and file naming



date YYMMDD | **…** | **editor** | **version**

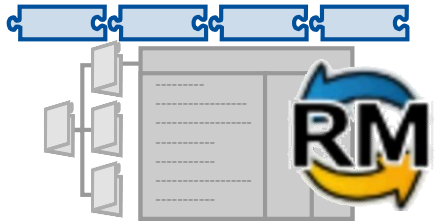
Example: 210519_WorkshopRDM_slides_draft_VS_v03.pptx

i

recommendations:

- consistently name files and folders according to a defined scheme
- use name components that provide meaningful information about the file´s content
- start with the component most often used for sorting
- avoid spaces and special characters

Handy tools to make live easier…

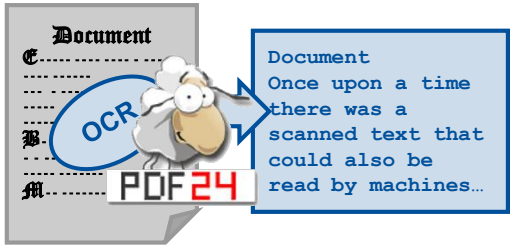


systematically name lots of files at once using RenameMaster

go to download page

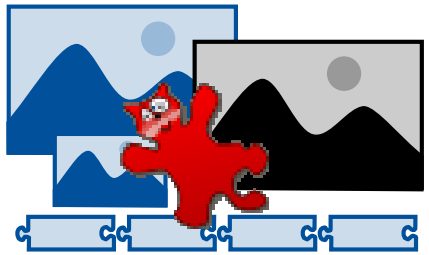
clean up tables using OpenRefine

go to download page



use PDF24´s OCR to make scanned texts searchable

go to download page



edit many images at once following a defined pattern (e.g. changing names, colours or size) using IrfanView

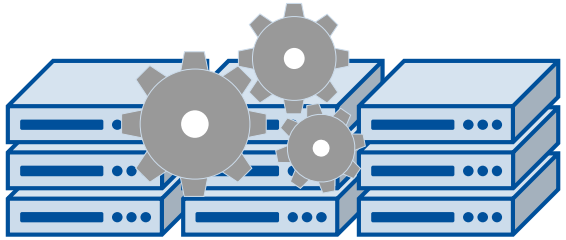
go to download page

i

Find more tools in the “tool box“ on our website (currently in German only).

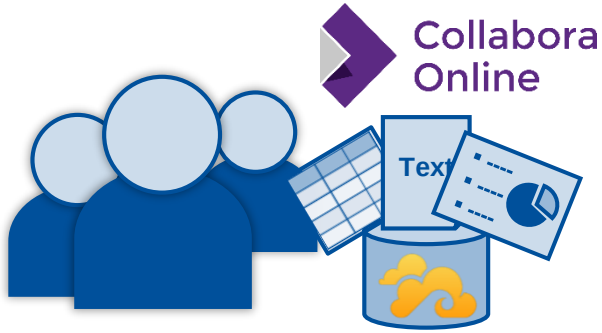
go to tool box

Special services und programmes



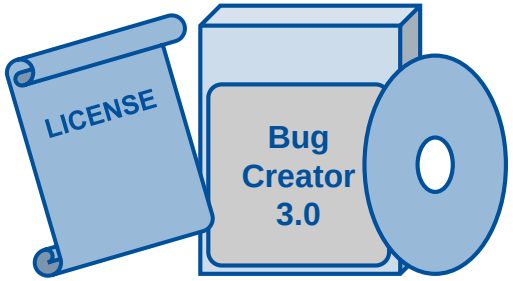
Scientific Computing
(using the LUIS computing cluster)

go to service description



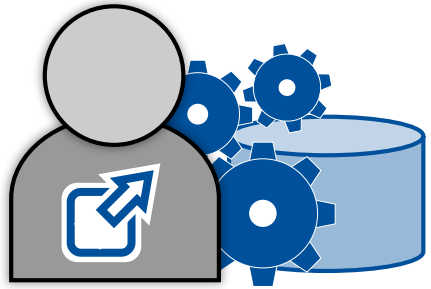
multi-user online editing of documents (also simultaneously) using Collabora

go to service description (German)



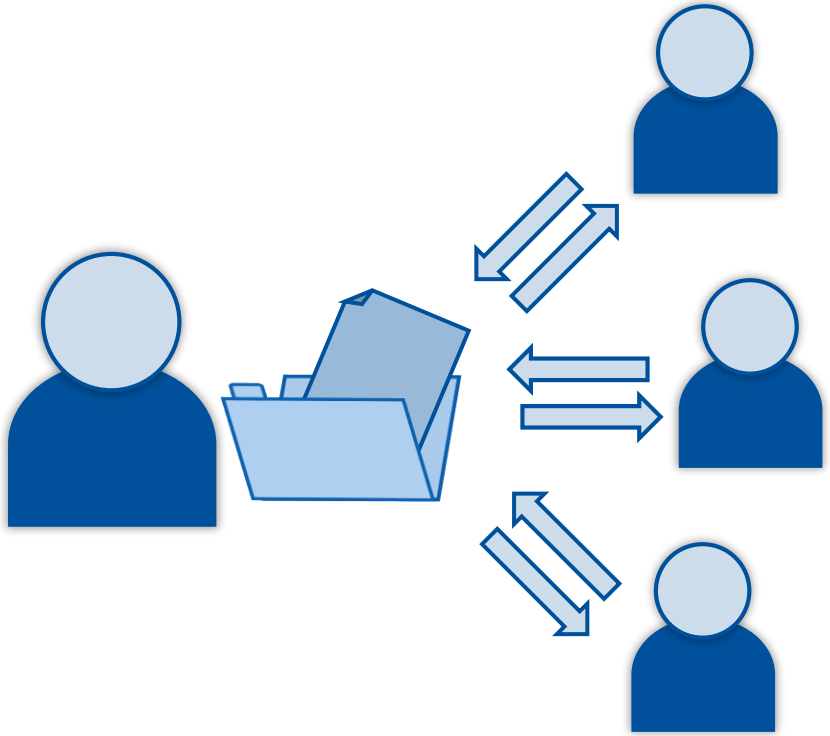
LUIS central software and license acquisition

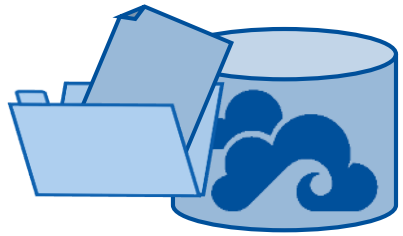
go to software catalogue (German)



data generation or processing by external service providers (e.g. transcription service, call centre, aerial photography etc.)

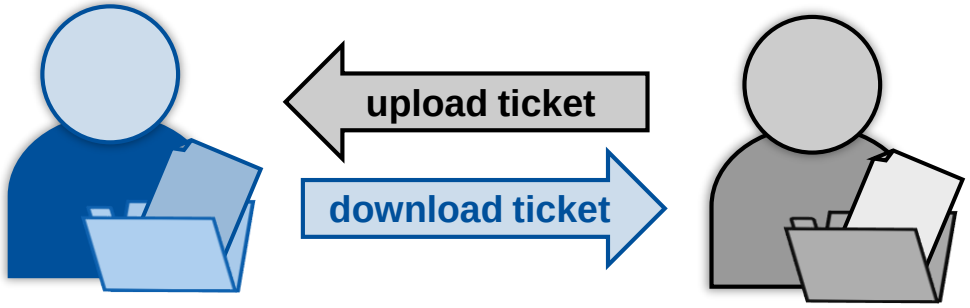
Sharing data (also with external partners)





creating share links to files and folders
in the “Projektablage” or “Cloud-Seafile”

go to Seafile user manual



LUIS Download Ticket Service
(as an alternative to big e-mail attachments)

go to service description (German)

Documenting your data processing

Why documentation is so important

The significance of metadata

General and specific context information

How metadata is generated and which forms they can take

Documentation – possible instruments

Why documentation is so important

!

Without good documentation you are in danger of …

- …failing to find specific data
- …being unable to retrace data creation and processing steps
- … being unable to interpret data due to lacking context information
- …mixing up files (outdated or contending versions)
- …not being able to interchange data with others or to combine them with data from other sources

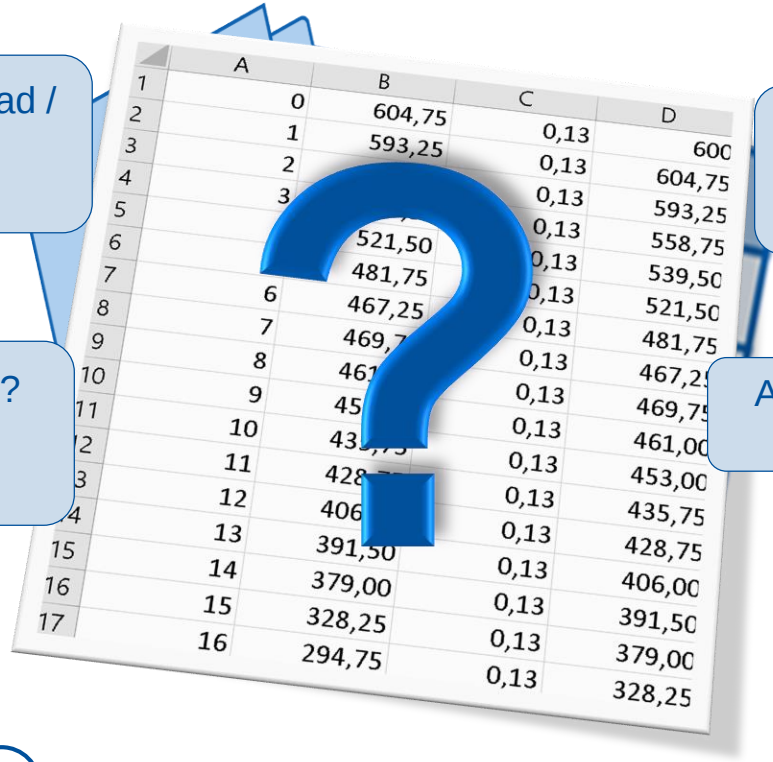
An adequate documentation is good research practice!

The significance of metadata

What is this?
(context / description)

Who produced this?
(authors)

Am I able to open / read / edit it?
(file format)



Where does it originally come from?
(source / storage location / identifier)

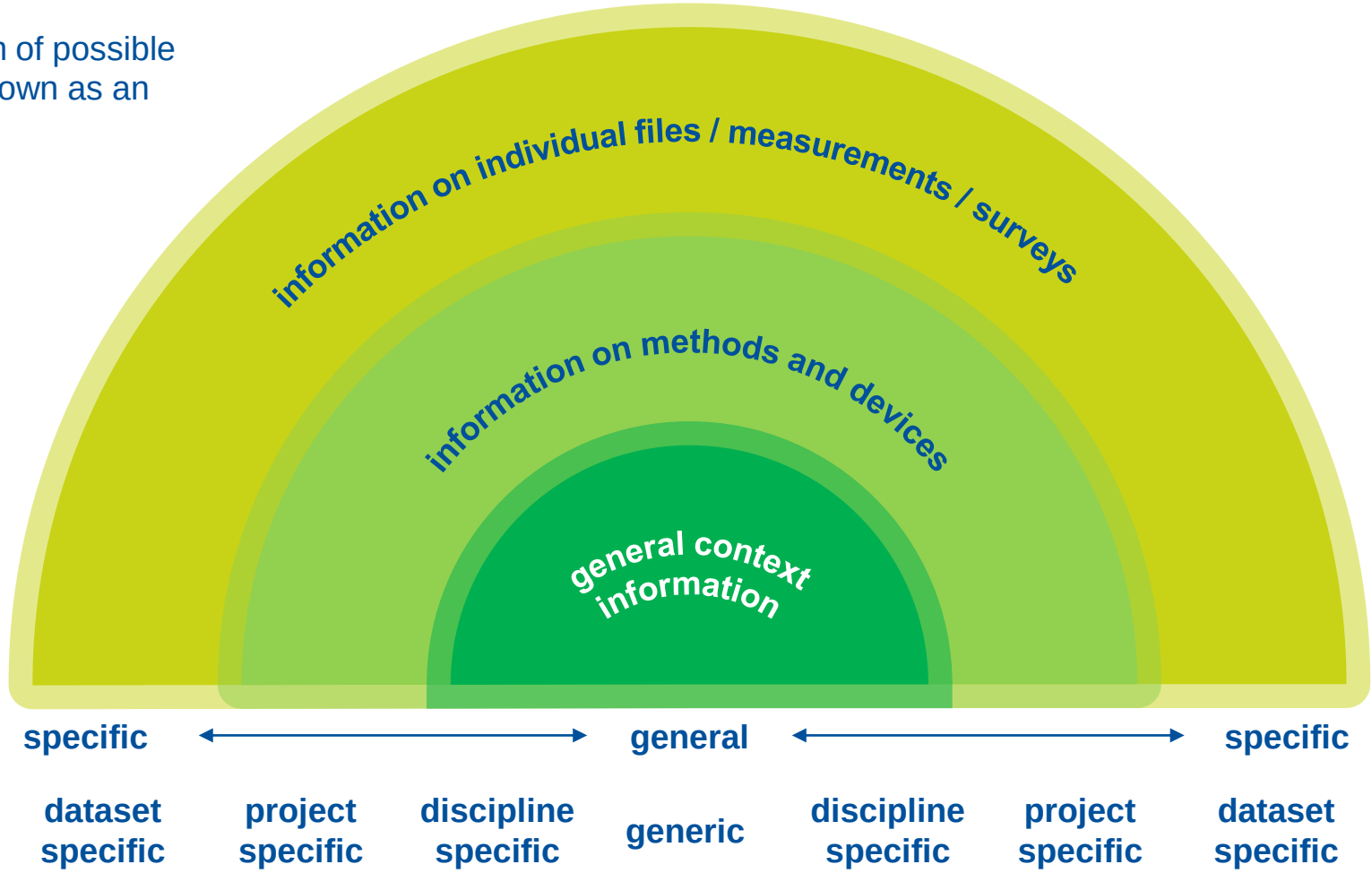
When was it produced?
(date of generation / publication)

Am I allowed to (re-)use it?
(license)

i Metadata are structured information necessary to understand and (re-)use the data themselves.

General and specific context information

the spectrum of possible metadata shown as an onion model

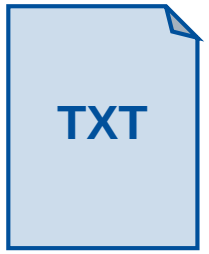


How metadata is generated and which forms they can take

How metadata is generated

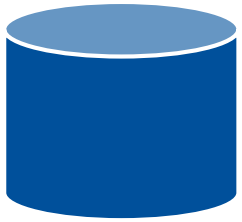
manual input in a structured form

automatic generation by devices



text files (e.g. „readme“ files)

spreadsheets

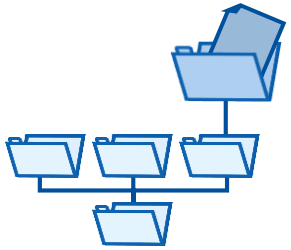


databases

embedded into the data files (e.g. file header, properties of image files)

structured xml files

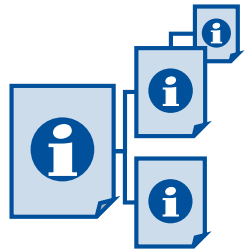
Documentation – possible instruments



consistent folder system

metadata database

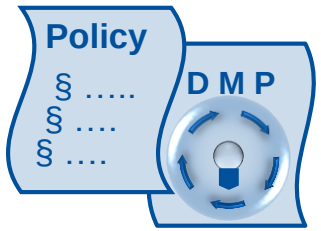
example: ResourceSpace (LUIS Projektablage)



wiki

example: DokuWiki (LUIS Projektablage)

lab notebook (on paper or electronic)



internal RDM policy & data management plans (DMP)

further reading

ZB MED (ed.) 2021: Electronic laboratory notebooks in the context of research data management and good research practice. A guide for the life sciences. DOI: 10.4126/FRL01-006425772.

go to guide

Harvard Medical School: Website about Electronic Lab Notebook including a comparison grid.

go to website

Protecting your data

Handling passwords: NOT like this, please!

Generating secure passwords

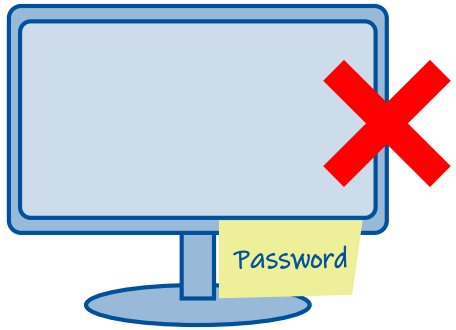
Rights management

Encrypting files and folders

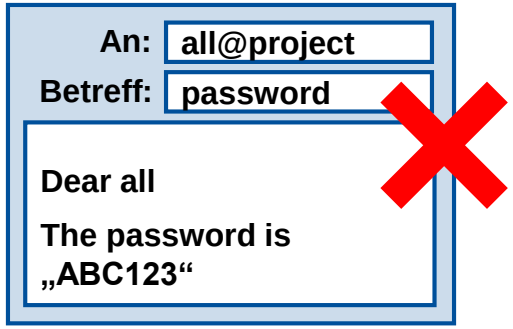
Encryption using VeraCrypt

Physical access restrictions

Handling passwords: NOT like this, please!



Do not stick your password on your screen!



Never send passwords as plain text via e-mail!

Do not use the same password for different accounts!

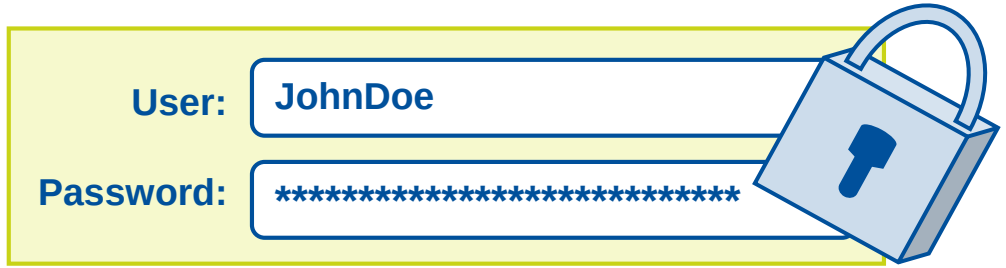
tool tip

You can manage your passwords in a reasonably secure way using the password manger KeePassXC (not recommended for highly sensitive passwords). In this case you only have to memorise one master password

Tip: Save your password database to your personal LUH cloud storage, which can be accessed from different devices.

go to download page

Generating secure passwords



further reading

ConnectSafely (May 2020): Tips for Strong, Secure Passwords & Other Authentication Tools. Website

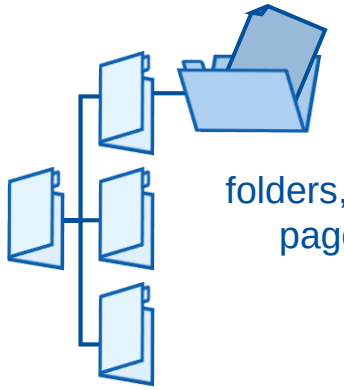
go to website

tips for secure passwords:

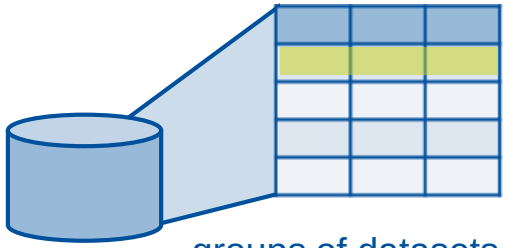
- at least 20 characters
- hard to guess (do not use birthdays etc.)
- phrase instead of a single word
- no obvious sense (no common proverbs etc.)
- altered spelling (e.g. „tr!ck“ instead of „trick“)
- combining several languages

Avoid language specific characters, such as ö,ä,ü,ß etc. in passwords. These characters may not be available on devices with the keyboard layout of another country. You may thus be unable to enter your password!

Rights management



folders, files, pages



groups of datasets in databases

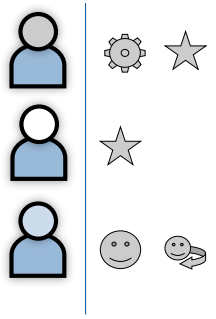
rights (examples)

- read (just look)
- write (edit)
- create (generate new item)
- move
- delete
- download

user groups (examples)

- IT admins
- heads of project
- staff members
- external partners

group assignment



! Reliably configuring and managing user rights requires IT expertise and possibly special software. Take this into account when planning your project and your resources!

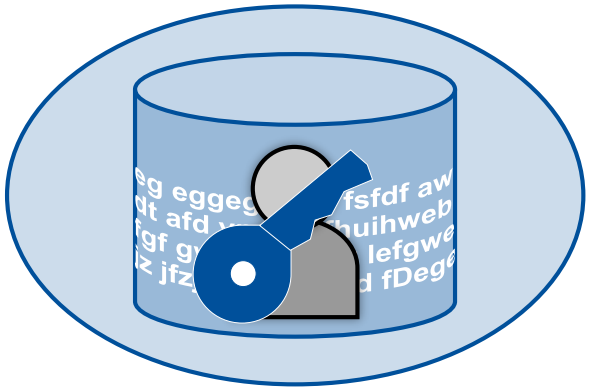
LOG

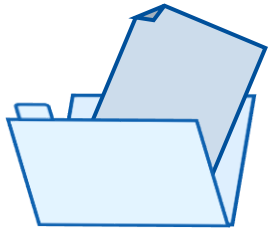
date	time	user	action
.....	....	.....	.....
.....	.....	....	.....
.....	....	.....	.....
.....	.....	....	.....

Encrypting files and folders

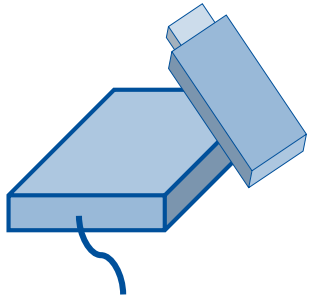
! The more sensitive the data, the more important is *continuous* encryption!

recording devices

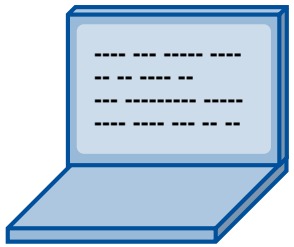




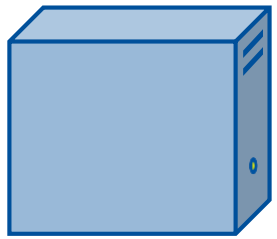
individual files and folders



mobile storage media



laptops



workspace PCs

server

further reading

Aseem Kishore (21 July 2019): How To Encrypt Zip Files. Online article at Help Desk Geek.

go to online article

WinRAR documentation.HELP!: Archive name and parameters dialog: general options. Set password.

go to documentation.HELP!

Dave McKay (23 October 2023): What Is Encryption, and How Does It Work? Online article at How-To Geek.

go to online article

Encryption using VeraCrypt

Generally speaking, encryption by software is more reliable than encryption by hardware!

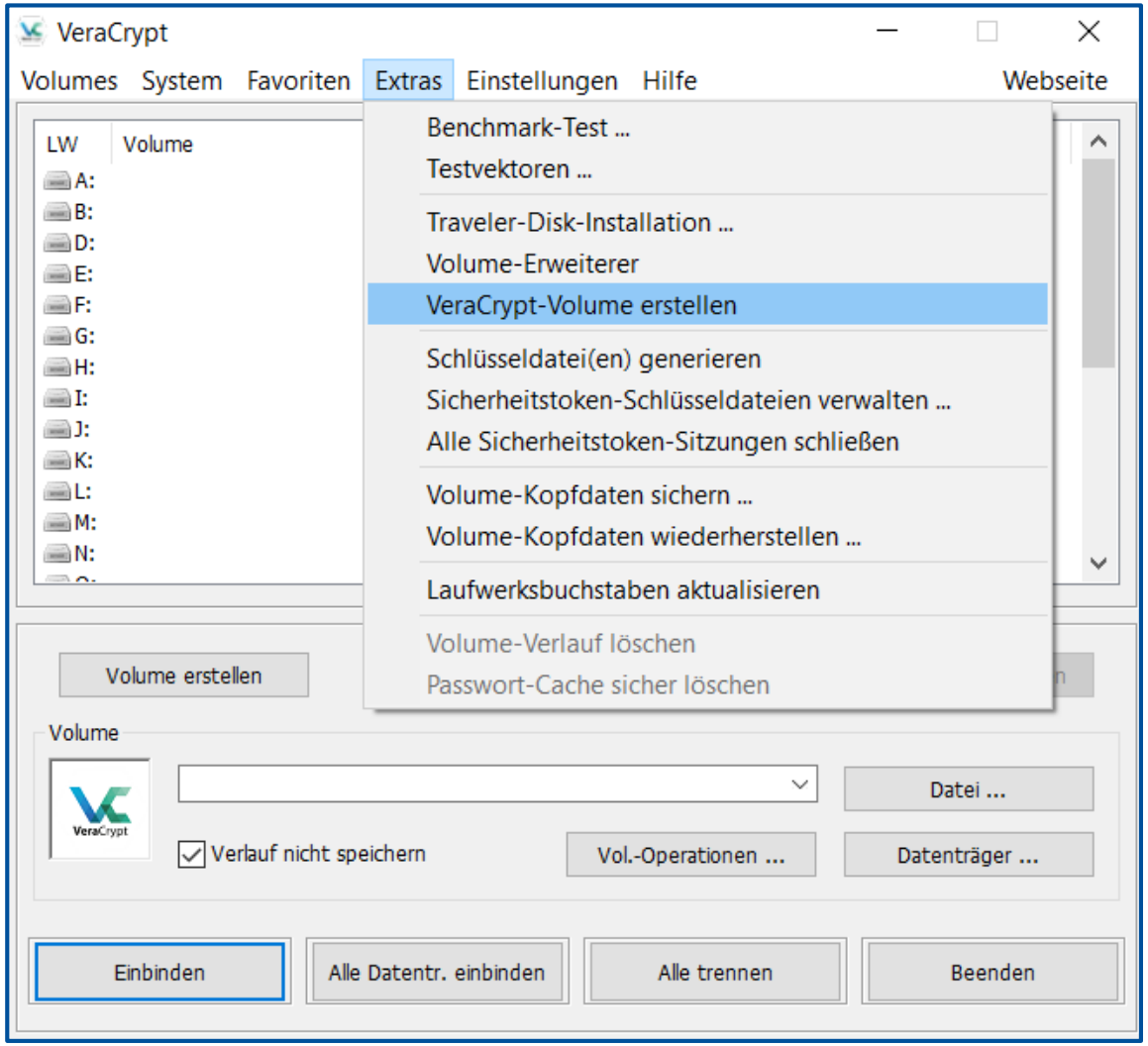
tool tip

VeraCrypt is a powerful, user-friendly, open source programme capable to securely encrypt individual folders, partitions and entire drives. It is available for all common operating systems.

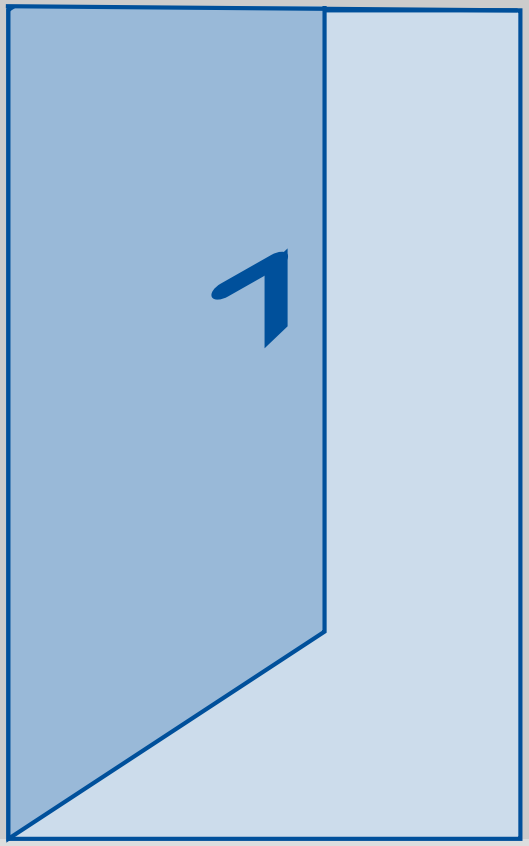
go to download page

Techlore`s YouTube video explains well the advantages, features and configuration choices of VeraCrypt.

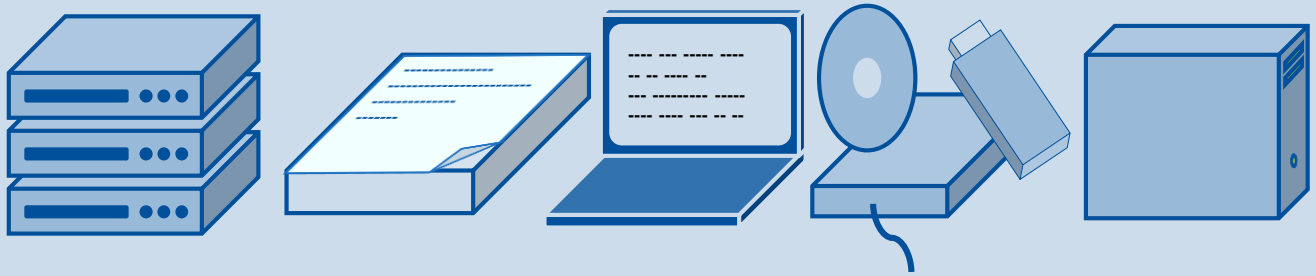
go to video



Physical access restrictions



- Which media contain the originals or copies of the data?
- Where are these media located?
- Who has legally access (including janitors, cleaning service etc.)?
- How easy would it be to gain access illegally (burglary)?



Choose storage locations whose security and access restrictions are adequate to the sensibility of your data!

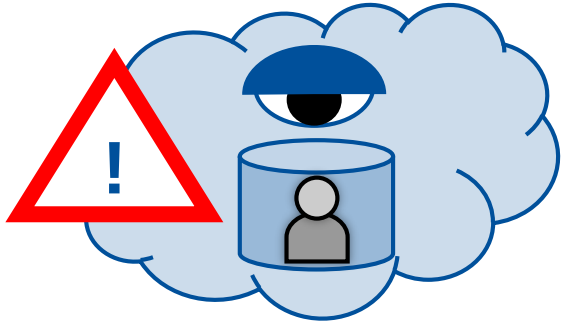
further reading

A recent example:
Maïa de La Baume (2020): Thieves use lockdown as cover for EU Parliament burglaries. (online article from 1 July 2020 at POLITICO.eu)

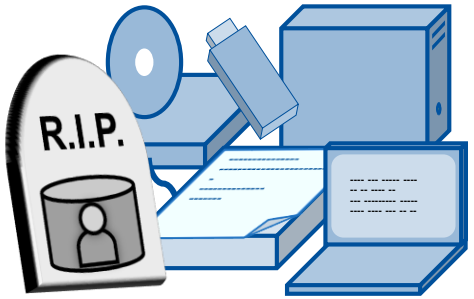
go to article

Additional organisational measures

Train and sensitise all persons involved in data processing.



Avoid “free” commercial cloud storages, file hoster etc.



See for a professional (!) physical destruction of defect media and media no longer needed.

If evitable, do not store data on mobile devices. If you must, remove the data from these devices as soon as possible.

Check regularly if everybody follows the rules. Define workflows and responsibilities!

i

You can hand over electronic media designated to be destroyed to Sachgebiet 12 (IuK) in the main building “Welfenschloss” (R318A).

further reading

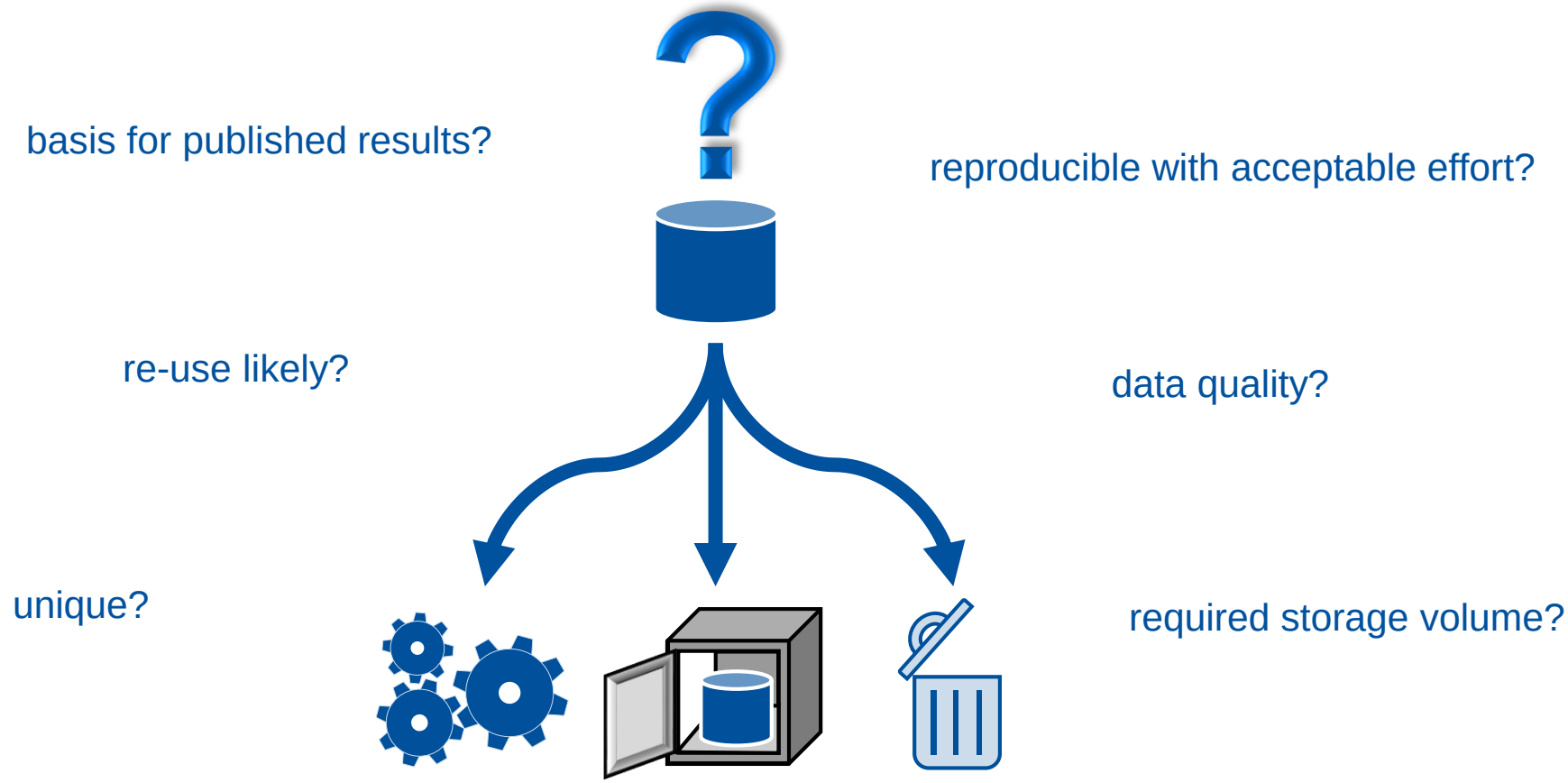
Surveillance Self-Defence (2019): Your Security Plan. Online article, 1 October 2019

go to article

Archiving and publishing your data

Which data to keep long-term?

! You do not have to keep all data. However, you should decide consciously and with good reason what you do or do not want to keep.



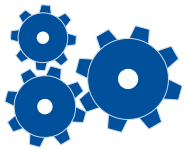
Refining data according to the FAIR principles

Findable (others can find my data)

- rich metadata in searchable public registries (e.g. the search databases of repositories)
- permanently retrievable via a unique identifier (e.g. a DOI)

Accessible (others have access to my data)

- preferably online access via standard protocols (e.g. http(s) / ftp)
- transparent access conditions



Interoperable (my data is compatible to other data and can be processed by machines)

- systematically elaborated and documented according to disciplinary standards
- machine-readable data and metadata in common and preferably open file formats
- references to other related data

Re-usable (other can utilise my data for their own purposes)

in addition to all points listed above:

- adequate documentation and precise attributes (data is comprehensible)
- distinct license (conditions for re-use are defined in a legally binding manner)

further reading

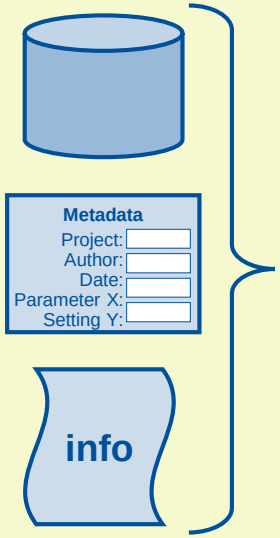
Angelina Kraft (2017): The FAIR Data Principles for Research Data. Article in the TIB blog, 12 September 2017.

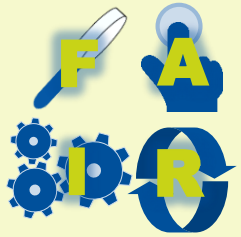
go to TIB-Blog

GoFAIR: FAIR Principles (Website).

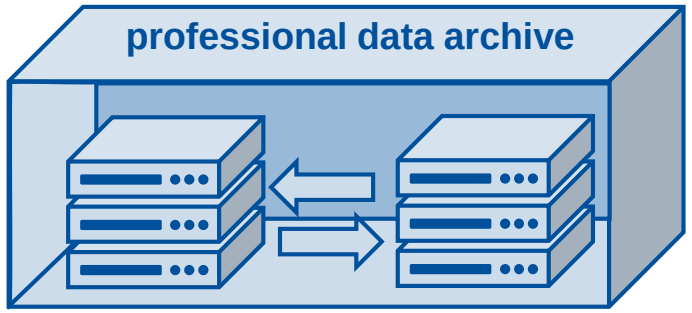
go to GoFAIR website

Correctly archiving data





- appropriate file formats
- structured according to disciplinary or data-specific standards
- machine-readable, if possible



go to LUH data archive

i

what a data archive should offer:

- servers located in appropriate rooms
- redundant data storage
- periodic replacement of media
- maintenance by professionals
- bit stream preservation guaranteed for at least 10 years
- access limited to data owners

i

You can find lists of file formats appropriate for long-term storage all over the internet, e.g. on this website of Duke University Libraries:

go to list

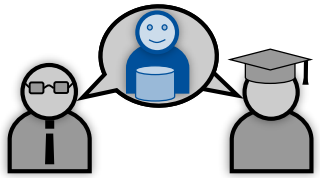
The Research Data Alliance (RDA) compiled a catalogue of disciplinary metadata standards.

go to catalogue

!

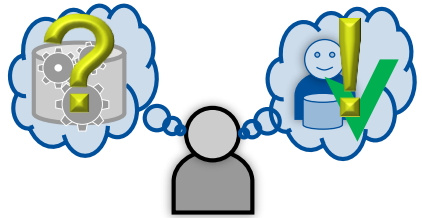
CDs, external USB drives and sticks etc. are NOT appropriate!

Publishing high-quality data opens up opportunities!



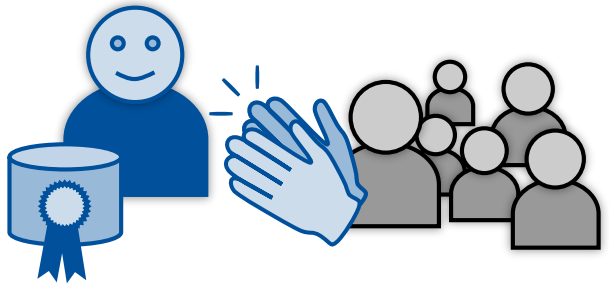
increased impact on the scientific discourse

special trust in the integrity of your research



others follow your example

additional citations when others re-use your data



additional impact and recognition (if data are relevant and of high quality)

gaining profile beyond one´s own disciplinary community

further reading

Max Planck PhDnet: Incentives to adopt open science practices in your daily research. Online article, 30 June 2020.

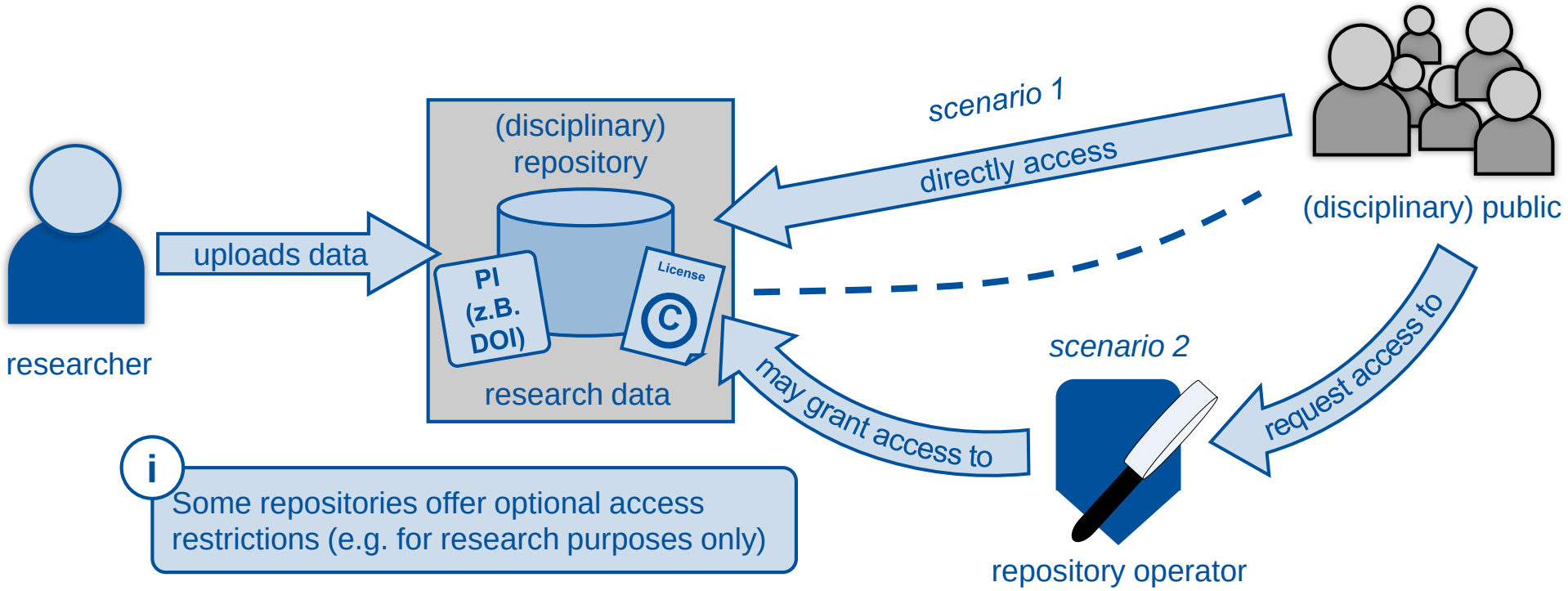
go to website

further reading

open-access.net: Open Access and Research Data. Undated online article.

go to website

Making data available via a repository



Choosing a suitable repository

go to re3data.org

- How is access regulated (open, restricted)?
- How is data re-use regulated (e.g. Creative Commons licenses)?
- Do the data get persistent identifiers (e. g. a DOI)?
- Is the repository certified (e. g. Data Seal of Approval)?
- Does the repository have its own policy?

!

Also pay attention to these points:

- Who is allowed to *upload* data?
- Which data and metadata formats are accepted?
- Which (perhaps optional) services do they offer?
- How much do they charge for hosting your data?

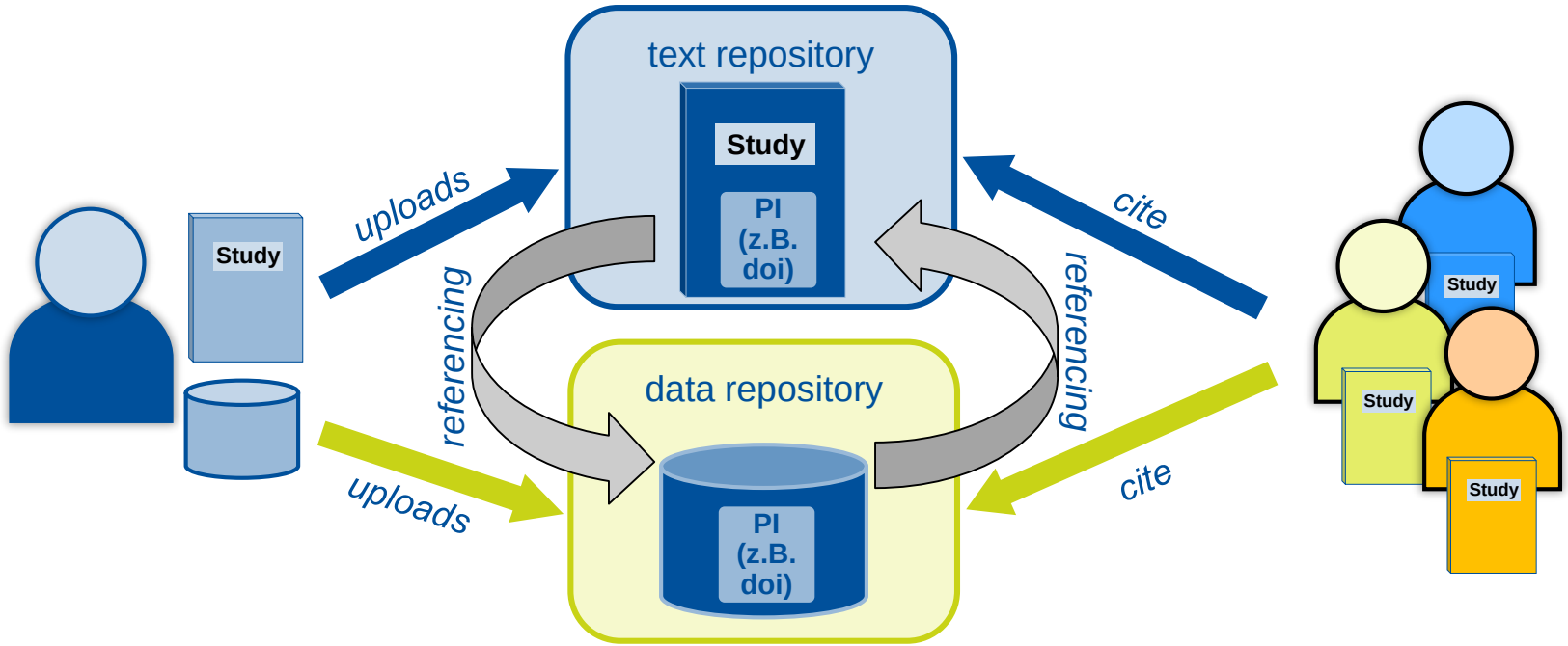
→ Preferably, clarify these points BEFORE the start of your project. You will then be able to allocate or apply for the necessary funds, and you can generate your data straight away in the required formats.

i

LUH provides its own institutional data repository

go to service description

Linking journal articles with related data



further reading

Diana Kwon (2019): The Push to Replace Journal Supplements with Repositories. Online article, 19 August 2019, on the website of „The Scientist“.

go to article

LUH also provides its own institutional text repository for open access publications.

go to text repository

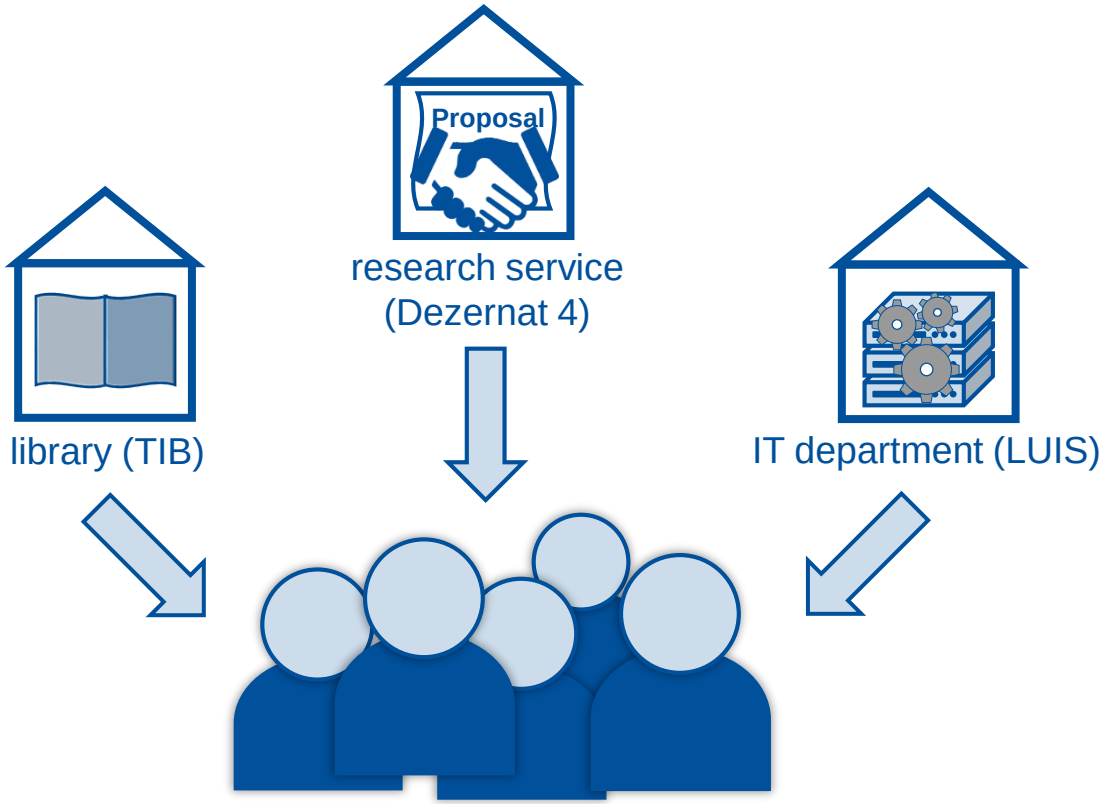
Supporting services and initiatives

The Research Data Support Team of LUH

External information and support

RDM working groups and initiatives - national and international

The Research Data Support Team of LUH



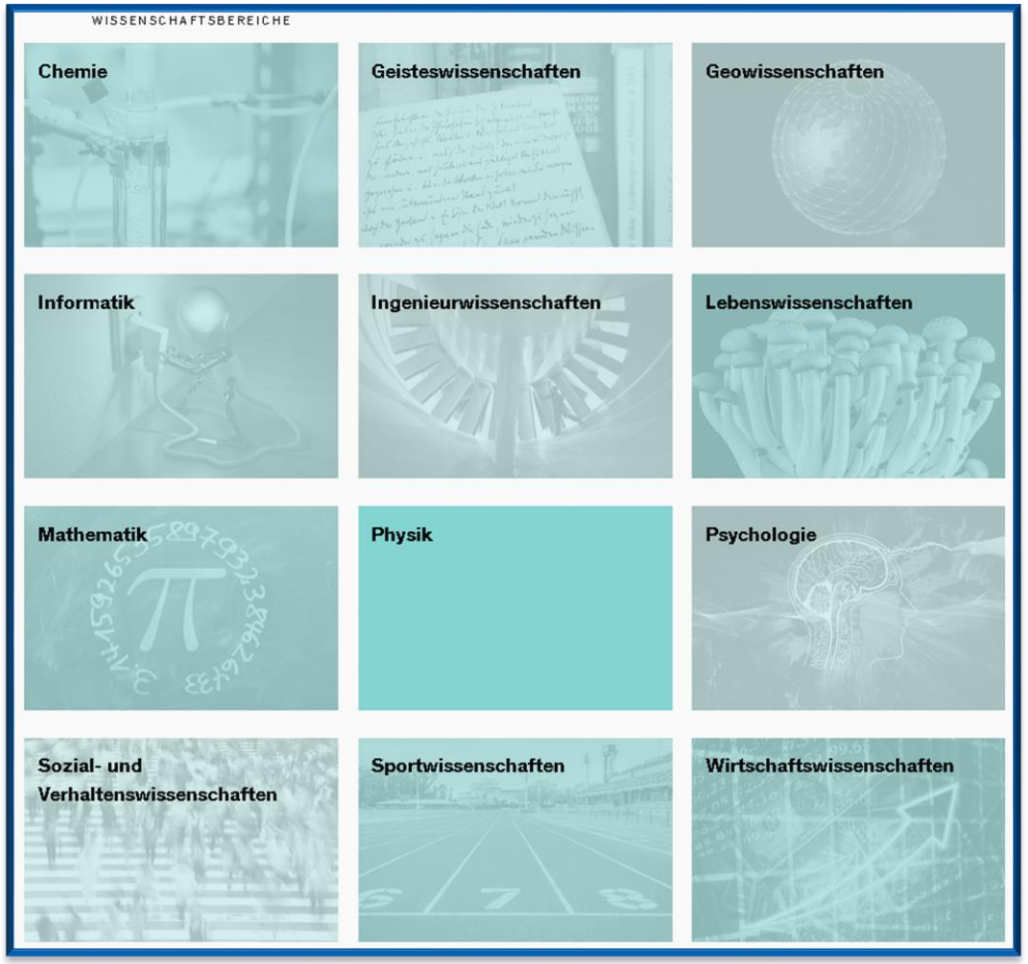
We train and counsel LUH members on the following topics, among others:

- documentation and publication of data
- DMP and RDM policies
- RDM statements in funding proposals
- LUH services and infrastructures for data management
- legal issues (in cooperation with the legal department and the data protection office)
- practical implementation of data management in research processes

Please note as well the additional information, guides and training courses on our website.

go to website of the Research Data Support Team

External information and support



go to overview at forschungsdaten.info (German)

Disciplinary services may provide the following:

- counselling of projects from a disciplinary perspective
- discipline-specific info materials
- discipline-specific services, tools and infrastructure (e.g. a DMP tool, data processing, quality checks, archiving and publishing data)
- networking in the disciplinary research community
- developing and defining disciplinary standards and policies
- representing the interests of your discipline towards policy makers and funding bodies

Get information on services for your discipline and current developments!

RDM working groups and initiatives - national and international

DINI/nestor-AG Forschungsdaten
(RDM support staff at research institutions in German-speaking countries)

go to the working group´s website (German)

Research Data Alliance (RDA)
(international network of RDM specialists with a special focus on technical aspects)

got to RDA website

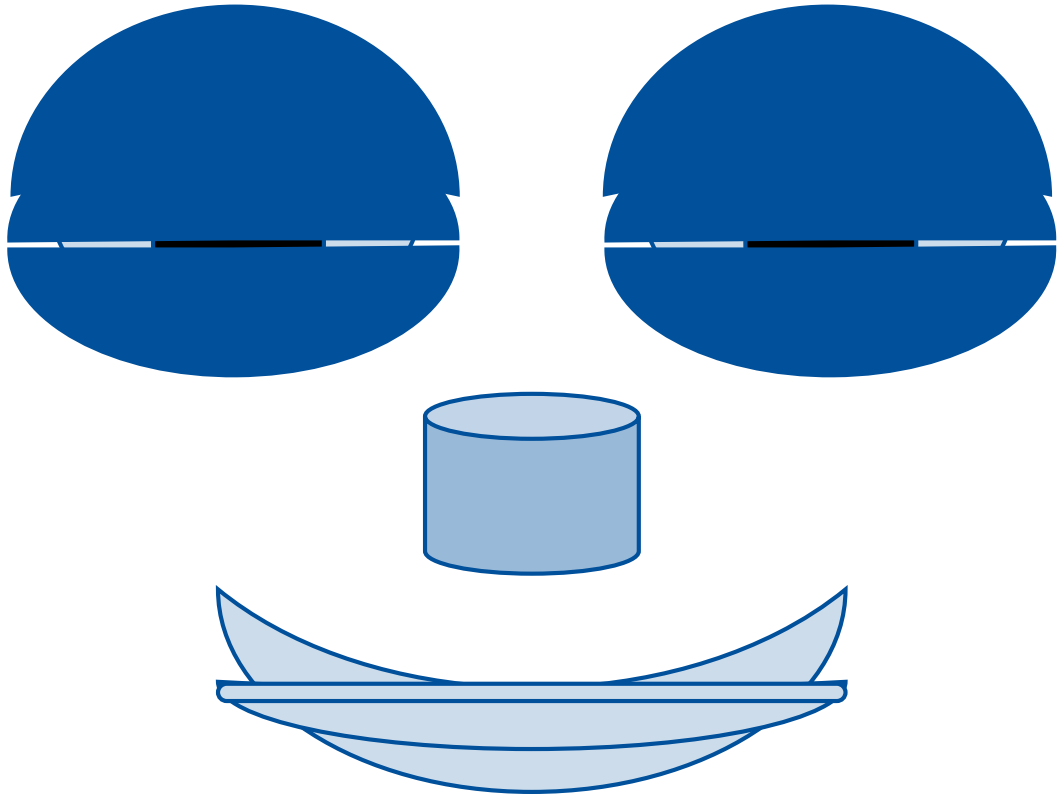
NFDI consortia

The federal government and the state governments jointly fund the development of a national research data infrastructure (NFDI). Its core are disciplinary consortia composed of researchers and infrastructure providers. These consortia are tasked with developing disciplinary standards, establishing services and organising the research community.

All researchers are invited to participate in the work and to join in the discussions!

LUH's NFDI website | official NFDI website | DFG`s NFDI website

Thanks for your attention and have a nice evening!



Zzzz

zzzzzz

zzzz