



THÜRINGER

FDM-TAGE

2021

Datendokumentation

```
<title>
```

```
  A love note to the future!
```

```
</title>
```

```
<meta name="description" content="event">
```

```
<meta name="dates" content="21.06.21-25.06.21">
```

```
<meta name="author" content="Thueringer  
Kompetenznetzwerk Forschungsdatenmanagement">
```

Bauhaus-Universität
Weimar





How to deal with Research Data – Recommendations for the Data Life Cycle!

Moderator: Kevin Lang
Friday, 25.06.2021

Agenda

- 1) Recommendations for Handling Research Data
- 2) Projects, Developments and Events
- 3) Guest Lecture by Daniela Gawehns
- 4) Open Discussion

Introduction

About myself...

Kevin Lang, Master of Science

- 2011 – 2016: Bachelor Degree, Medieninformatik
- 2016 – 2018: Master Degree, Computer Science and Media
- Student Assistant activities:
 - Natural Language Processing, Machine Learning, Artificial Intelligence und Big Data
- since 2018:
 - Contact person for Research Data Management at Bauhaus-Universität Weimar
 - Member of the Thuringian Competence Network for Research Data Management (TKFDM)

Bauhaus-Universität Weimar

Stabsstelle Forschungsdatenmanagement

Recommendations for Handling Research Data


Why a Recommendation?

- Extension to the RD-Policy
- Explanations, Models and Principles
- Phases of a Research Project/
Data Life Cycle:
 - Planning, Processing, Publishing
and Archiving
- Support and Services



Support

Research Data Management



Research Data Management refers to all actions that are necessary to make digital research data permanently usable. The aim is to carefully organize and maintain the data generated in the research process and to make the research data freely accessible to all scientists in order to promote the public distribution of knowledge. This also supports also associated topics such as Open Data and Open Access.

The Alliance of the German Science Organization with its Principles and the German Research Foundation (DFG) with their guidelines have created the first fundamentals for handling research data. They include without limitation following points:

- Planning for research data management must already be taken into account in the project planning and application phase.
- The protection and free accessibility of the research data shall be granted provided that no scientific or legal interests are violated, for example in the form of data protection guidelines or obligations towards third parties
- The use of relevant standards and procedures should be followed in order to improve the quality of the data and to ensure its reusability

These guidelines are intended to help increase the recognition of researchers and the reputation of scientists. Many external funding organizations already demand to stick to the guidelines but also promote them in their projects.

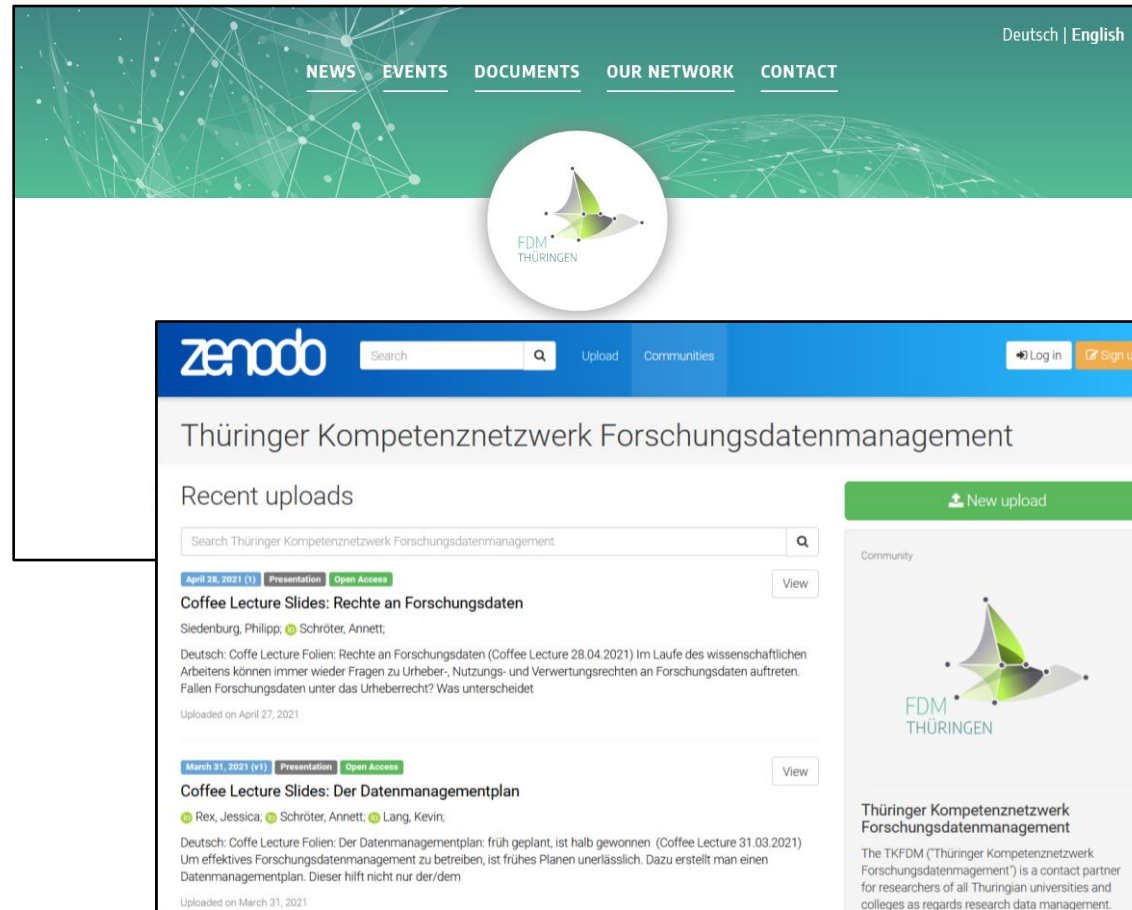
Research data generally refers to all data that can be generated, collected, processed or analyzed during the research process. This includes research results and any necessary tools or procedures that have arisen in the course of research work and to solve research tasks.

Research data management describes the entire planning, recording, processing, documentation, archiving and publication of the research data.

Kevin Lang
Bauhausstraße 11, Room 131
99423 Weimar
Phone: +49(0)36 43 158 37 (3)
Mail: kevin.lang@uni-weimar.de

- > The policy for handling research data at the Bauhaus-Universität Weimar
- > Organized Workshops
- > The Thuringian Competence Network for Research Data Management
- > The information website forschungsdaten.info
- > The new DFG-Code 2019 and Research Data Management
- > The National Research Data Infrastructure (Nationale Forschungsdateninfrastruktur, NFDI)
- > Free Online Courses

www.uni-weimar.de/rdm



The screenshot shows the Zenodo profile for the Thuringian Competence Network for Research Data Management. The header includes navigation links for NEWS, EVENTS, DOCUMENTS, OUR NETWORK, and CONTACT, along with language options (Deutsch | English). The profile features a search bar, a 'New upload' button, and a list of recent uploads. Two uploads are visible: 'Coffee Lecture Slides: Rechte an Forschungsdaten' (uploaded April 28, 2021) and 'Coffee Lecture Slides: Der Datenmanagementplan' (uploaded March 31, 2021). A community logo for FDM THÜRINGEN is also present.

www.forschungsdaten-thueringen.de

Local Services

- University Library
- Service Center for Computer Systems and Communication
- Research Department
- Bauhaus Research School
- Legal Office



 **BRS Onlineworkshop "Research Data Management: Organisation, Documentation, and Publishing"** ⓘ
wissenschaftliche Praxis
Termin: Donnerstag, 16. September 2021 10:00 - 12:00
Status: Nicht angemeldet · Anmeldefrist: Mittwoch, 15. September 2021 10:00

 **BRS Workshop | DISPUTATIONSTRAINING** ⓘ 📅 🗓️
Promotion
Termin: Dienstag, 12. Oktober 2021 09:00 - 16:00, +1 📅
Status: Nicht angemeldet · Anmeldefrist: Dienstag, 28. September 2021 23:59

 **BRS WORKSHOP | ONLINE | Writing an Exposé** ⓘ 📅 🗓️
wissenschaftliche Schreiben und Publizieren
Termin: Montag, 25. Oktober 2021 14:00 - 17:00, +1 📅
Status: Nicht angemeldet · Anmeldefrist: Montag, 18. Oktober 2021 23:59

Organizations



Rat für
Informations
Infrastrukturen



Deutsche Initiative für
Netzwerkinformation e. V.



The Data Management Plan

A DMP assists on how to plan and to document data within a project.

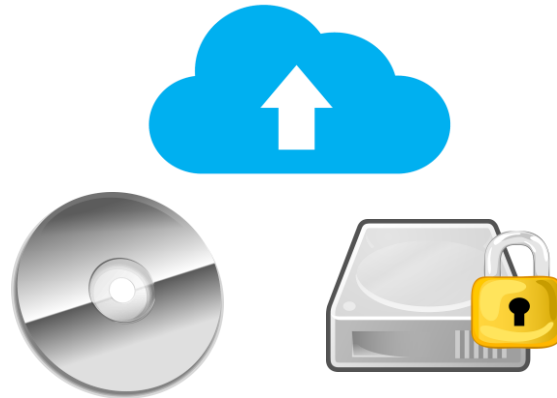
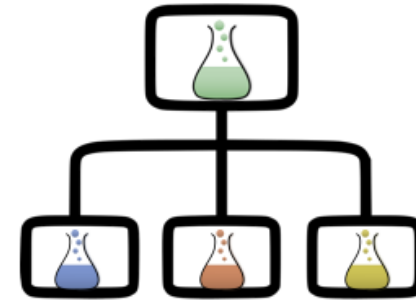
Reasons

- Coordination between project partners
- Understanding and reusing data
- Early identification of problems
- Basis for third party funding



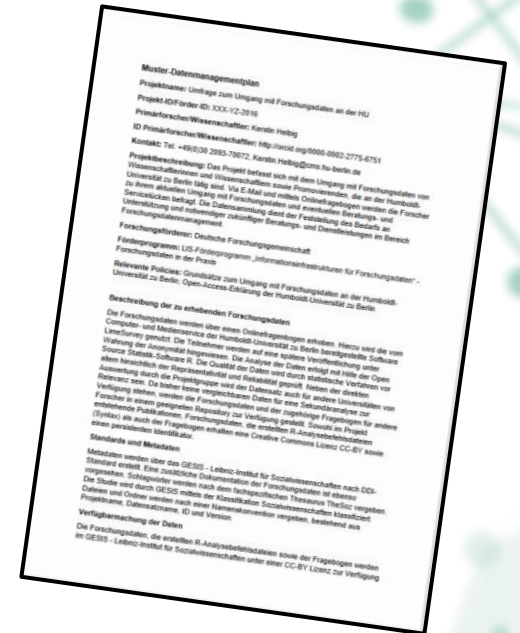
DMP: Contents

- Overview
- Data Stock
- Workflow
- Transfer
- Distribution
- Obligations
- Resources

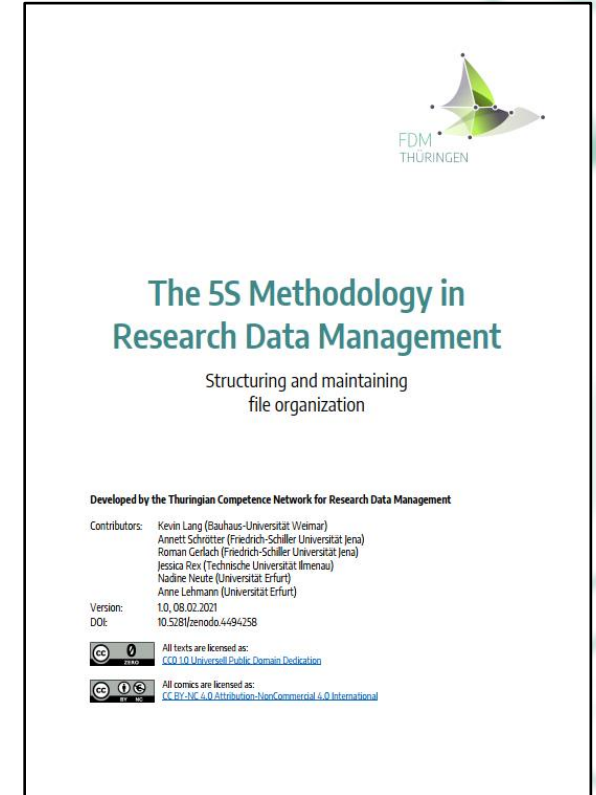
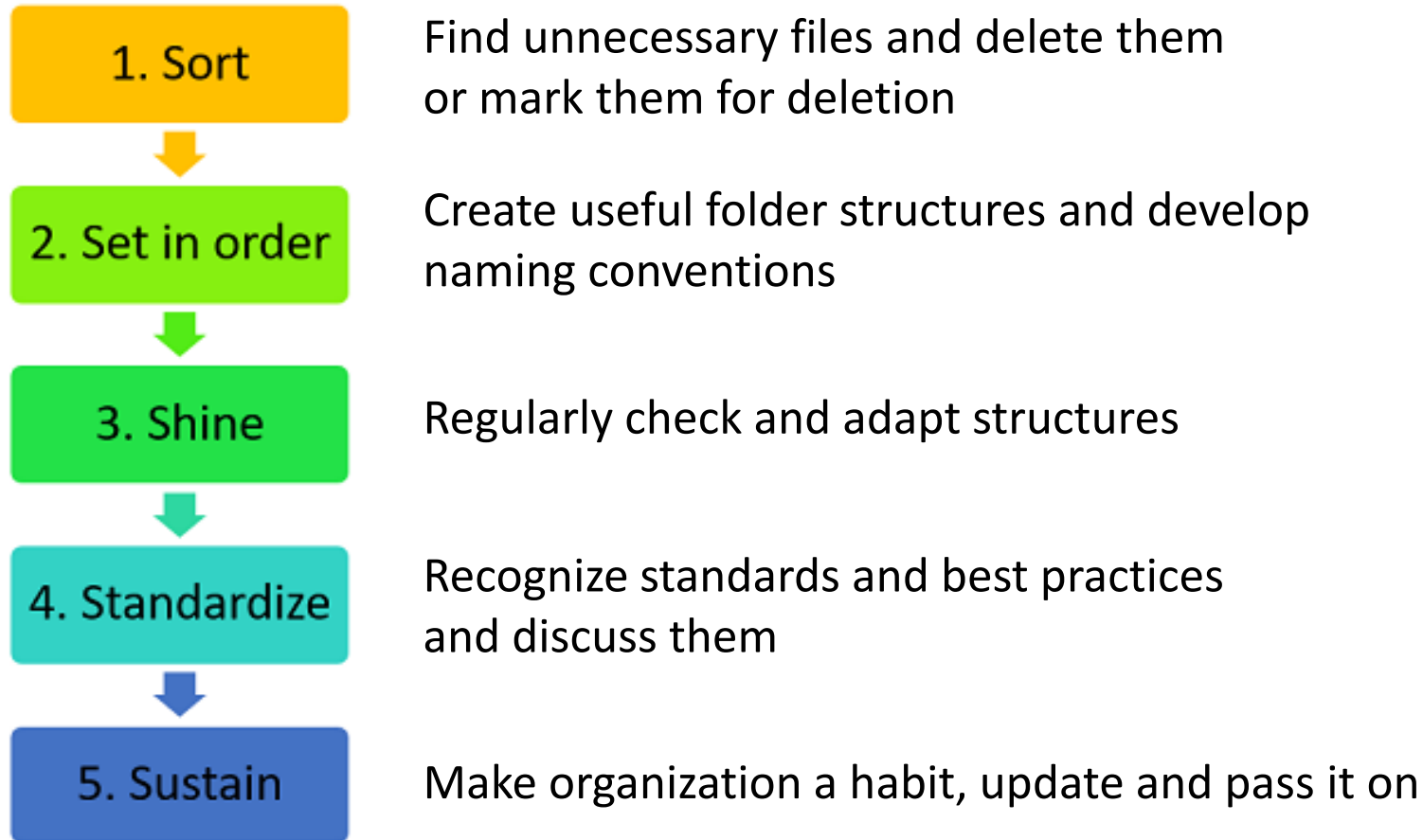


DMP: Template and Tools

- [Templates](#)
- [Research Data Management Organiser](#)
 - forschungsdaten.info
- [DMPonline](#)
- [Data Stewardship Wizard](#)
- [ARGOS](#)



Organizing Research Data (5S Data Model)



doi.org/10.5281/zenodo.4494257

Rights and Obligations

- Any restrictions by law or contracts
- Data protection of personal data
 - Datenschutzgrundverordnung (DSGVO),
Bundesdatenschutzgesetz (BDSG),
Thüringer Datenschutzgesetz (ThürDSG)
- Assessment by an ethics committee
- Rights of use of data (e.g. licences, consent forms, ...)
- Ownership rights



Documentation

- Mostly meta data about the origin and context of research data
- Can be described by the classical 5W1H questions:
 - Who? What? When? Where? Why? How?
- Should be machine readable by standardized terminologies:
 - e.g., [Schema.org](https://schema.org), [Dublin Core](https://www.dublincore.org/), [MARC](https://www.loc.gov/marc/) or [MODS](https://www.modsvocab.com/)
- Collaboratively managed wiki, readme files or simple documents (e.g., PDF format)



Versioning and Backup

- Research data must be protected against manipulation e.g. by backup and versioning systems
- Goals derived from information security: confidentiality, integrity and availability
- Automatic or manual versioning of data
- 3-2-1-0 Backup Rules:
 - 3 copies of a file
 - 2 different storage devices
 - 1 other location
 - 0 problems in recovery



Virtual Research Environments

- Dedicated work environments or portals to establish a uniform workflow, comply standards and exchange files
- Criteria: local/external hosting, possible organizational structures, storage capacity, user settings, access, licences, period of validity, subject-specific features, data protection, maintenance, costs...

- If possible open-source, e.g.:

- Nextcloud (+OnlyOffice)
- GitLab
- eLabFTW
- ...



GitLab



eLabFTW

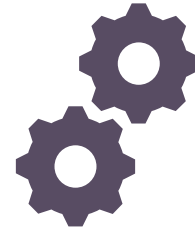


Publication by the FAIR principles



Findable

Data and metadata should be easy to locate, both by humans and by machines.



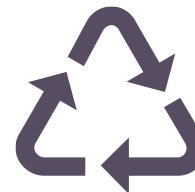
Interoperable

Data needs to be interoperable with applications or workflows for analysis, storage, and processing.



Accessible

Users need to know how they can access the data, possibly including authentication and authorisation.



Reusable

Data should be released with a usage licence. Origin and processing methods should be well-documented.

Subject-specific and general Repositories

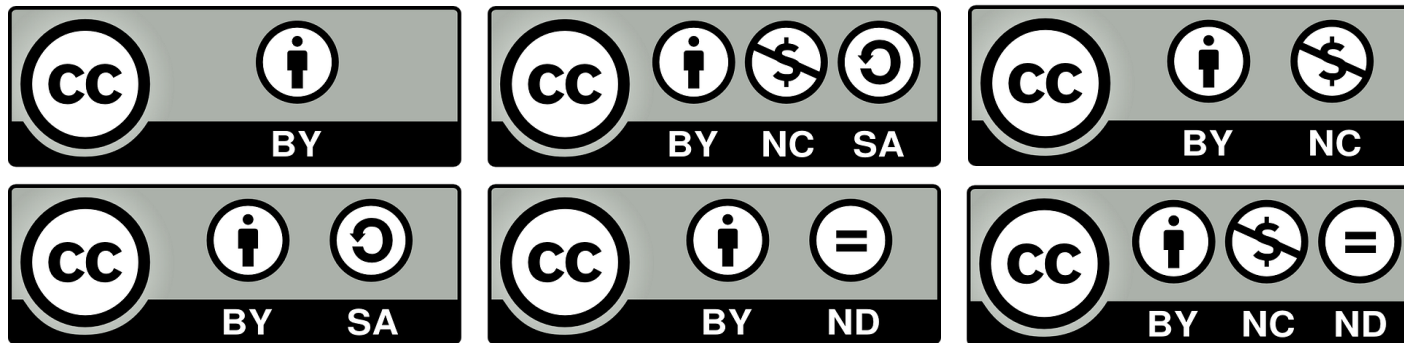


- Persistent identifiers
- Metadata
- Download
- Access
- Licences
- Versioning
- ...



Licensing

- Data usage agreement
- Choice based on the medium:
 - Creative Commons (CC) for creative data
 - Open Data Commons (ODC) for data collections
 - GNU General Public License (GPL) for software projects



Long Term Preservation

- Data should be preserved if it:
 - Provides evidence of research results
 - Is expensive to collect or not reproducible
 - Represents critical points in research
- At least 10 years in the institution or in a repository
- If possible open formats and fulfilling standards
- Maybe regulations for deletion after a period of time



Projects, Developments and Events

Strategie zur Digitalisierung (12/2017)

Topics about Research Data Management:

- Competence Network
- RD-Policy
- Open Data (Open Science)
- Digital Library Thuringia
- Infrastructure
- National Research Data Infrastructure



(aus [Thüringer Strategie zur Digitalisierung im Hochschulbereich](#), 2017, zuletzt aufgerufen am 17.06.2019)

Strategie zur Digitalisierung 2021-2025

New Topics about Research Data Management :

- Competence Network (TKFDM) and Thuringian Center for Learning Systems and Robotics (TZLR)
- Fields of action: Networking, Training, Open Science, Long-term Archiving, High-Performance Computing, Legal Support and Integration in Curricula
- Founding of a User Advisory Board



(aus [Thüringer Strategie zur Digitalisierung im Hochschulbereich](#), 2021, zuletzt aufgerufen am 17.05.2021)

TKFDM: Services

- **Consulting**
 - Data Management Plans, Data Protection, Applications, Web Services, Formats, ...
- **Training**
 - Thuringia-wide Information Events, Workshops, Train-the-Trainers or Coffee Lectures
 - Also possible on request
- **Networking**
 - Between Universities, Data Center, NFDI-Consortia and other Facilities



Materials

- Various information flyers
- Fact Sheets & Best-Practices
 - RDM Funding and Requirements, Research Data Repositories, Virtual Research Environments, Open-Source, Data Protection, ...
 - Quality Control Methods, eLabFTW, GitLab, LaTeX, BEXIS, RDM in courses, ...
- 23 Things about RDM
- ScaryTales
 - Based on the "Black Stories" card game
 - 50 stories about bad data management



FAIRest Dataset Award

- Award presentation on 21 June 2021 in Erfurt
- Which dataset fulfils the FAIR principles best?
- 2000€ price money and extras
- Assessment Tools:
 - [ARDC FAIR Assessment Tool](#)
 - [F-UJI Automated FAIR Data Assessment Tool](#)
- Winner:
 - ["Monitoring forest health using hyperspectral imagery: Does feature selection improve the performance of machine-learning techniques?"](#) submitted by Patrick Schratz (2019, FSU Jena)



January 29, 2020 Released Open Access

Monitoring forest health using hyperspectral imagery: Does feature selection improve the performance of machine-learning techniques?

294 views 123 downloads [See more details...](#)

Indexed in **OpenAIRE**

Publication date: January 29, 2020
DOI: [10.5281/zenodo.3630302](https://doi.org/10.5281/zenodo.3630302)

Keywords(s): [Machine Learning](#) [Ecological Modelling](#) [Remote Sensing](#) [Forest Health](#) [Hyperspectral](#) [Data](#)

License (for files): [Creative Commons Attribution 4.0 International](#)

Code, figures, appendices and the manuscript can be found in the corresponding [GitHub repository](#).
This RC is a static snapshot at the time of submission. The GitHub repository holds the latest version and may see changes after the publication was accepted.

Data sources and description

- [aoi.gpkg](#): Area of interest for downloading Sentinel-2 images. *Not used in the publication.* Source: Custom.
- [forest_mask.gpkg](#): A forest/non-forest mask of the Basque Country. *Not used in the publication.* Source: Custom.
- [hyperspectral.zip](#): Hyperspectral remote sensing data used to extract reflectance values on the tree level. Source: Custom.
- [plot-locations.gpkg](#): Spatial location of the plots used in the study. Source: Custom.
- [tree-in-situ-data-corrected.zip](#): Connected in-situ data containing defoliation information on the tree level. A correction of the spatial location was applied by the creators of the data. Source: Custom.
- [tree-in-situ-data.zip](#): First version of in-situ data containing defoliation information on the tree level. *Not used in the publication.* Source: Custom.

Future Events

- **TKFDM**
 - 30.06. Coffee Lecture "[Publikation von Forschungsdaten: Ein Gewinn für alle!](#)"
 - 07.07. Coffee Lecture "[Versionierung mit Git – Die Einführung](#)"
 - 21.07. Workshop "[Git und GitLab für Anfänger*innen](#)"
- **Research Facility/Bauhaus Research School**
 - 08.07. Workshop "[Forschungsdatenmanagement: Von der Planung und Organisation bis hin zur Veröffentlichung](#)"
 - 16.09. Workshop "[Research Data Management: Organisation, Documentation, and Publishing](#)"



Guest Lecture by Daniela Gawehns

Transparent Workflows with R - From Raw Data to Results!

Open Discussion

Thank you for your attention!

Sources

- Thuringian Competence Network for Research Data Management:
 - [Portal von TKFDM](#)
 - [TKFDM Community auf Zenodo](#)
 - [Research Data Scarytales](#)
- Material from other Websites:
 - [Open Data/EU Horizon 2020 on Labfolders](#)
 - [Organizing data folders with #5SDATA method \(RDA\)](#)
- Stockimages with CC0 Licence:
 - [pixabay.com](#)
 - [unsplash.com](#)