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)
Recommendations for Handling Research Data
The Research Data Policy

- **A Guideline for handling research data**
  - Released on February 12th, 2020: [Link](#)

- **Intention**
  - Create awareness about research data management (top down)
  - Support early planning
  - Make data findable, understandable and reusable
  - Work with founders (DFG, BMBF, Horizon 2020 Program, ...)

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

www.uni-weimar.de/rdm

www.forschungsdaten-thueringen.de
Local Services

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

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

1. Sort
   - Find unnecessary files and delete them or mark them for deletion

2. Set in order
   - Create useful folder structures and develop naming conventions

3. Shine
   - Regularly check and adapt structures

4. Standardize
   - Recognize standards and best practices and discuss them

5. Sustain
   - Make organization a habit, update and pass it on

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:
- Should be machine readable by standardized terminologies:
  - e.g., Schema.org, Dublin Core, MARC or MODS
- 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
  - ...
Publication by the FAIR principles

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

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

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

**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
- ...

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