Chair of Architectural Informatics
TUM School of Engineering and Design
Technical University of Munich

Technical University of Munich

How to teach good research data management to next generation researchers?

Syed Ashfaq Hussain Shah
Prof. Dr. Ing. Frank Petzold (AI, TUM)
NFDI4Ing Conference 2023
September 27, 2023

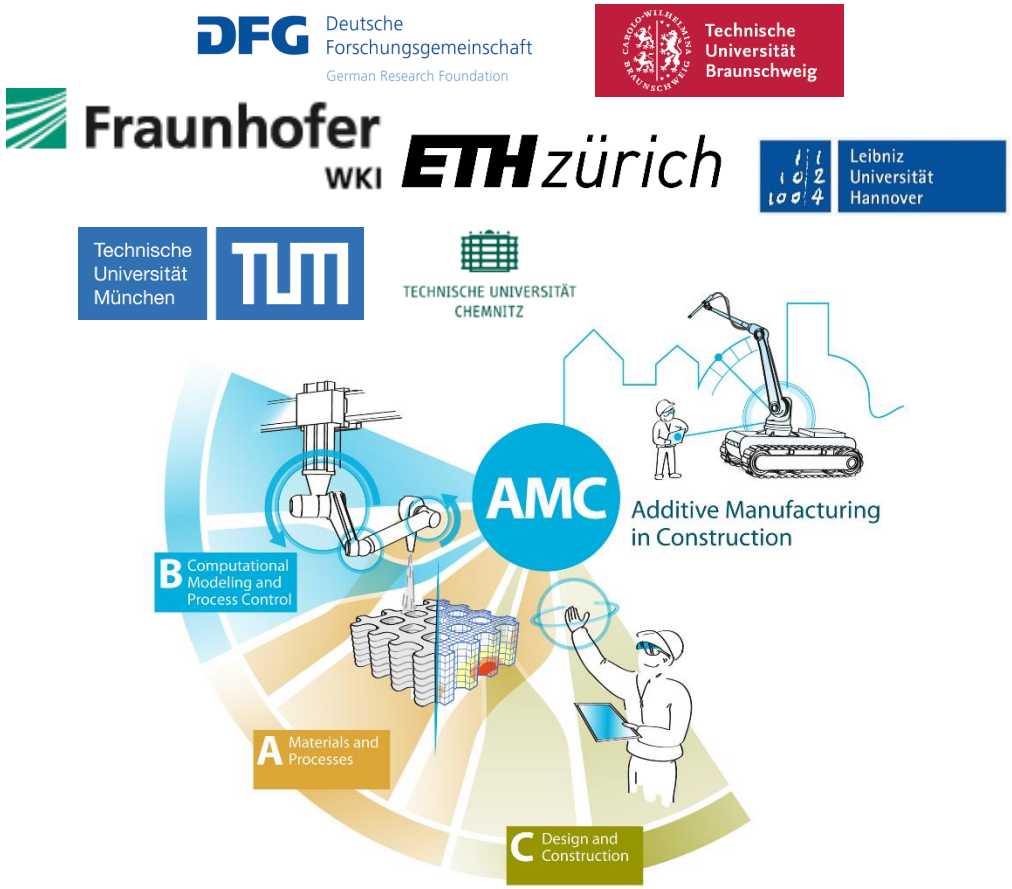
Agenda

- Introduction
- Agenda of Data Management
- Teaching and guiding methodologies
- Guiding materials
- Means of distributions
- Interactive sessions
- On boarding and migration processes
- Processes to create guiding contents
- Events to improve guidance
- Actions to improve RDM practices
- Tactics to enforce compliance for RDM
- Conclusion

Image source: https://unsplash.com/

Introduction: TRR277 AMC (Collaborative Research Center)

- Information infrastructure project
- Hosted by 2 universities at 2 distinct locations, multiple participating institutes
- 3 focus areas 22 interdisciplinary research groups, over 120 members
  - Materials and processes
  - Computational modelling and control
  - Design and construction
- Collaborated with Leibniz-Rechenzentrum (LRZ), Gausß-IT center, libraries of the both universities and ProLehre at TU Munich
- Approaches developed for in person, digital medium as well as for hybrid situations e.g. in Lockdown time.



Credit: https://amc-trr277.de/

Introduction: Good Research Data Management (RDM)

- Compliant with: -
  - FAIR principles
  - Open data/ Open Science practices
  - Reproducible results (where apply)

- Resilient and mitigate changing compliance
  - (Inter) National funding body/ research council requirements
    - DFG Good Research Practice

  - Other stake holders specific requirements
    - Research organising body
    - Publishers, (Scientific) Community

DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Adopted by the Senate of the DFG at September 30, 2015

Deutsche Forschungsgemeinschaft

DFG Guidelines on the Handling of Research Data

Release date: 21th December 2021

Handling of research data

Checklist for planning and description of handling of research data in research projects

Introduction: Samples of questions

- What is research data management (RDM) and why one should care?
- What are these terms vocabulary, metadata, ontology, DMP?
- What makes research data for project?
- What information relating to practices is part of research work?

- Are there ways to speed up the upload?
- Is the cable network better than wireless network?
- How to install required software?

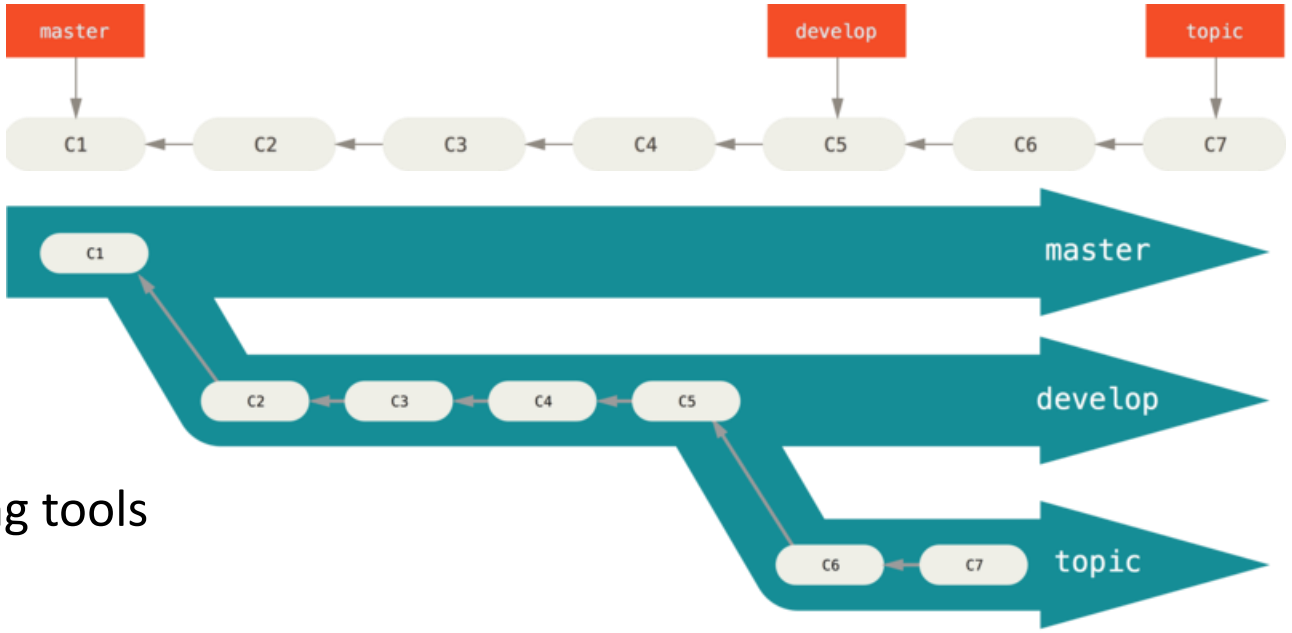
Introduction: Samples of questions contd…

- What to do when data is on a terminal which cannot be connected with network/ internet?
- I do not have storage device with sufficient capacity to port data on an internet connected terminal?
- How to organise data in collaborative research?
- I have data generated over time. I can not find where it should be uploaded?
- How the provided system works e.g. to record the practice, upload the data.
- I could not understand what you just demonstrated/ explained.

Agenda of data management

- Fundamentals of data management
  - Naming convention
  - Versioning
  - Identification of independent component and organisation
  - Documentation
    - Logging
    - README/ Docs/ Comments
  - Packaging/ Bundle
  - Unique ID

  - Digital means
    - Purpose/ use based data processing tools

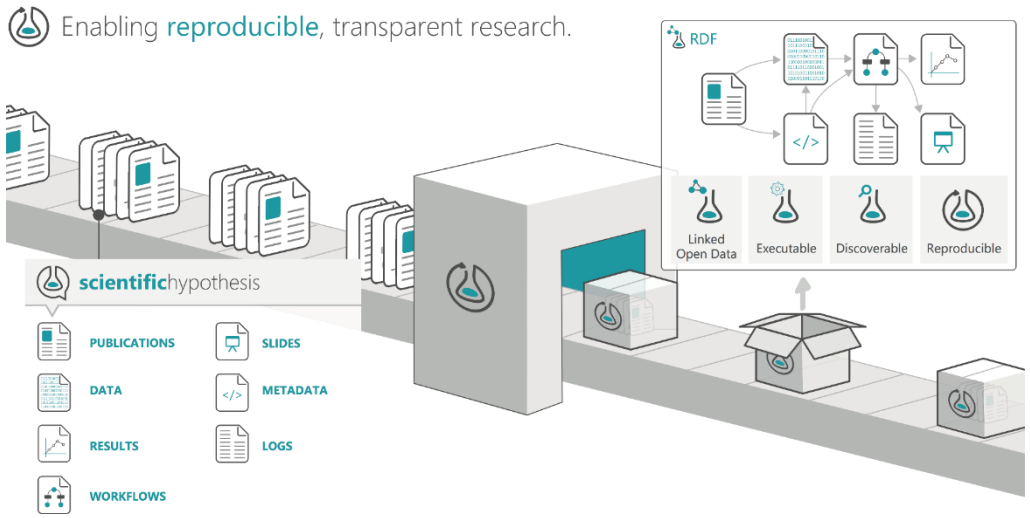


https://git-scm.com/book/en/v2/Git-Branching-Branching-Workflows

Agenda of data management contd…

- Research data management + Fundamentals of data management
  - Documentation of practices
    - Metadata
    - Data management plan (DMP)
    - User/ Technical guide
  - Maintenance of provenance and provisioning
  - Compliance with long term archive compliant format
  - Data anonymisation and pseudonymisation
  - Universally unique IDs.
    - ORCID …
    - DOI …
  - Attribution e.g. Licencing
  - Compliant system

  - A well thought change control agenda

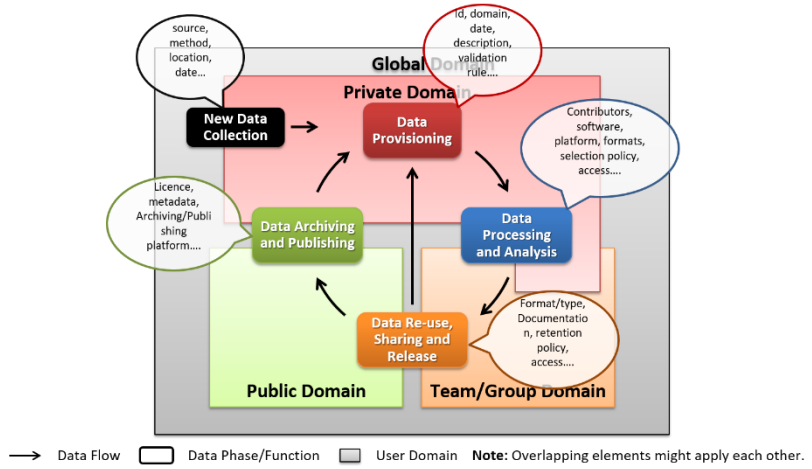


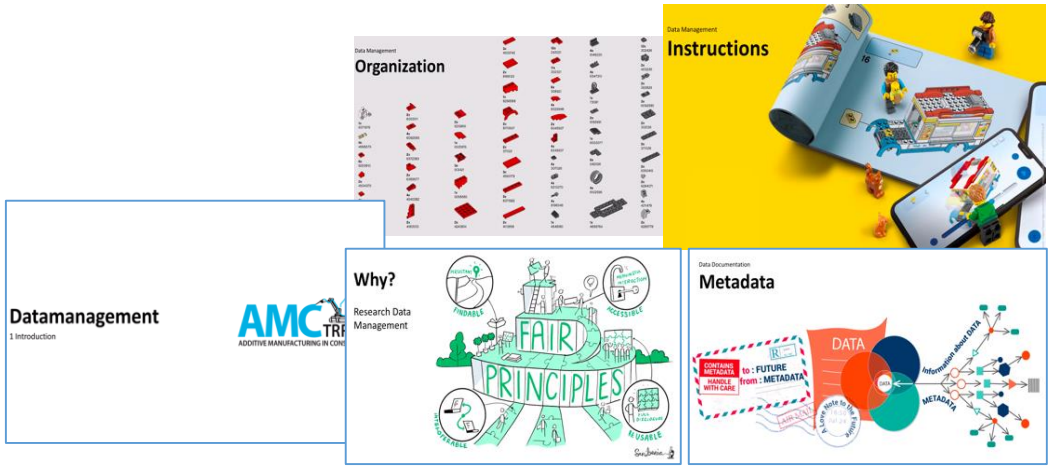
https://www.researchobject.org/

Teaching and guiding methodologies

- Policy and definition of general rules
- Custom Data Management Plan (DMP)
- Guiding materials
  - Printable materials
  - Multimedia contents
- (Simplified) data models, templates
- Conducted one to one/ group (in person/ online) sessions
- Dedicated weekly hours for support
- Evaluation, feedback and follow ups

- Complemented with: -
  - Screenshots, screencasts
  - diagrams, illustrations and animations
  - Glossaries and definitions

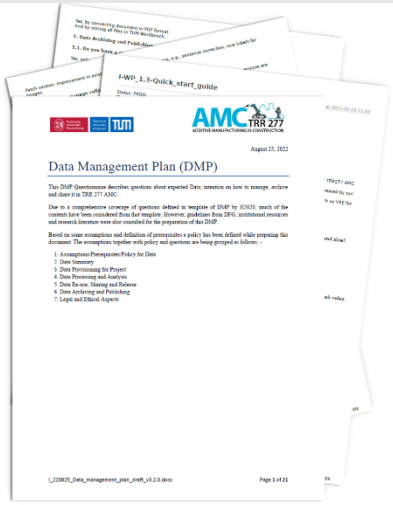


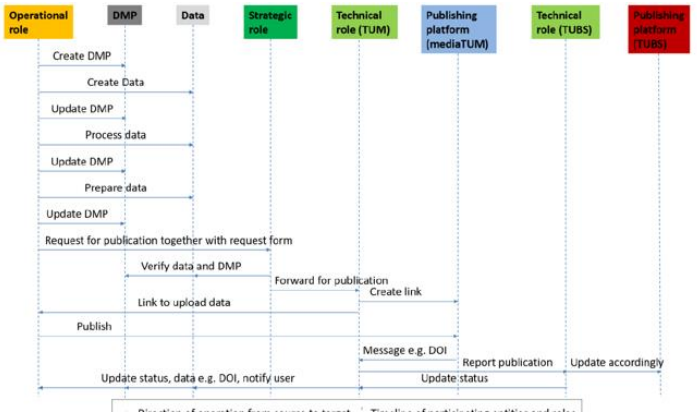


Policy and definition of general rules

- Research relevant data and outcomes
- Official RDM platform and infrastructure
- Official Data Management Plan (DMP)
- Roles and responsibilities of the participants
- Types of research outcomes/ data
- Data file types/ formats to communicate research outcome and long term archiving
- Common/ mandatory metadata standard
- Naming conventions including versioning

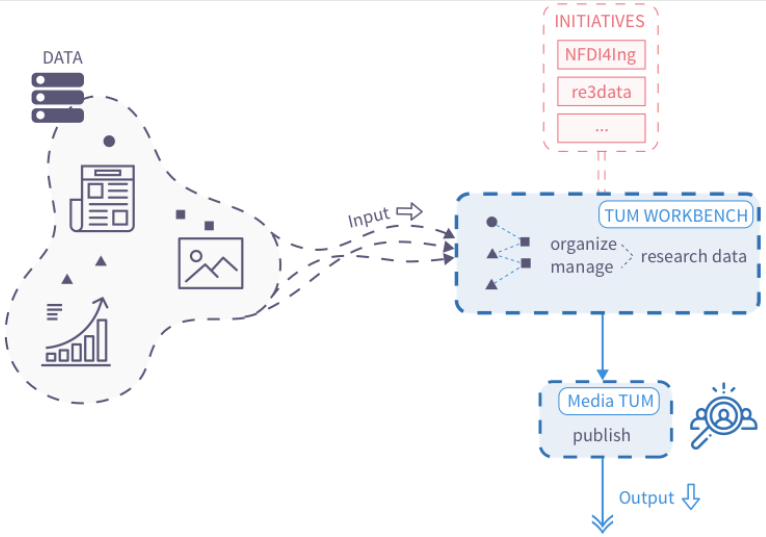
- Use cases of research outcomes and DMPs
- Workflows for publishing

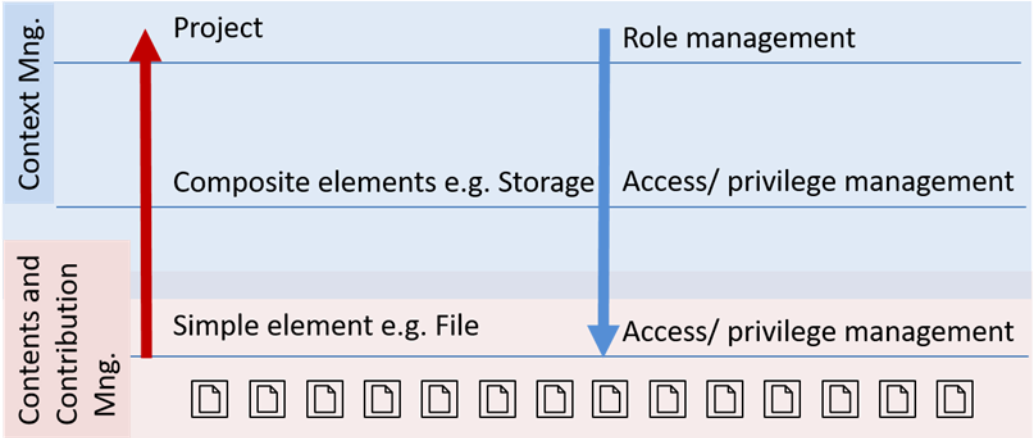




Guiding materials: Official RDM platform and infrastructure

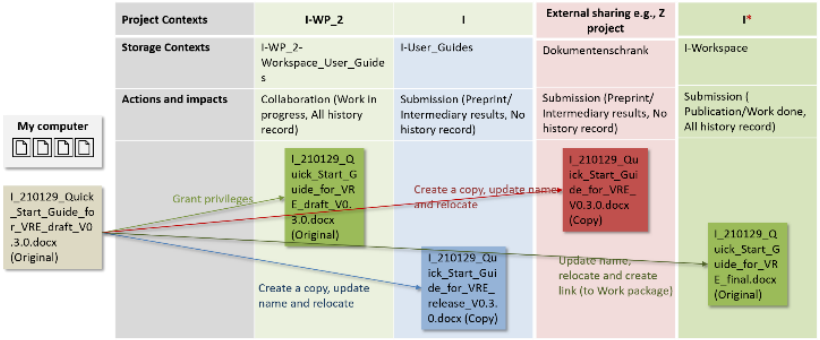
- User competency requirements
- Functional requirements
- Hardware/ Software requirements
- Conceptual/ theoretical foundations
  - Common entities and functions
  - User interface design/ webpages
  - Navigation and exploration workflows
  - Definition of tools and controls
  - Access rights and privileges
  - Data and context management
  - Data input and persistence procedures
  - Built-in RDM features e.g. UIDs, backups, communication and messages, metadata





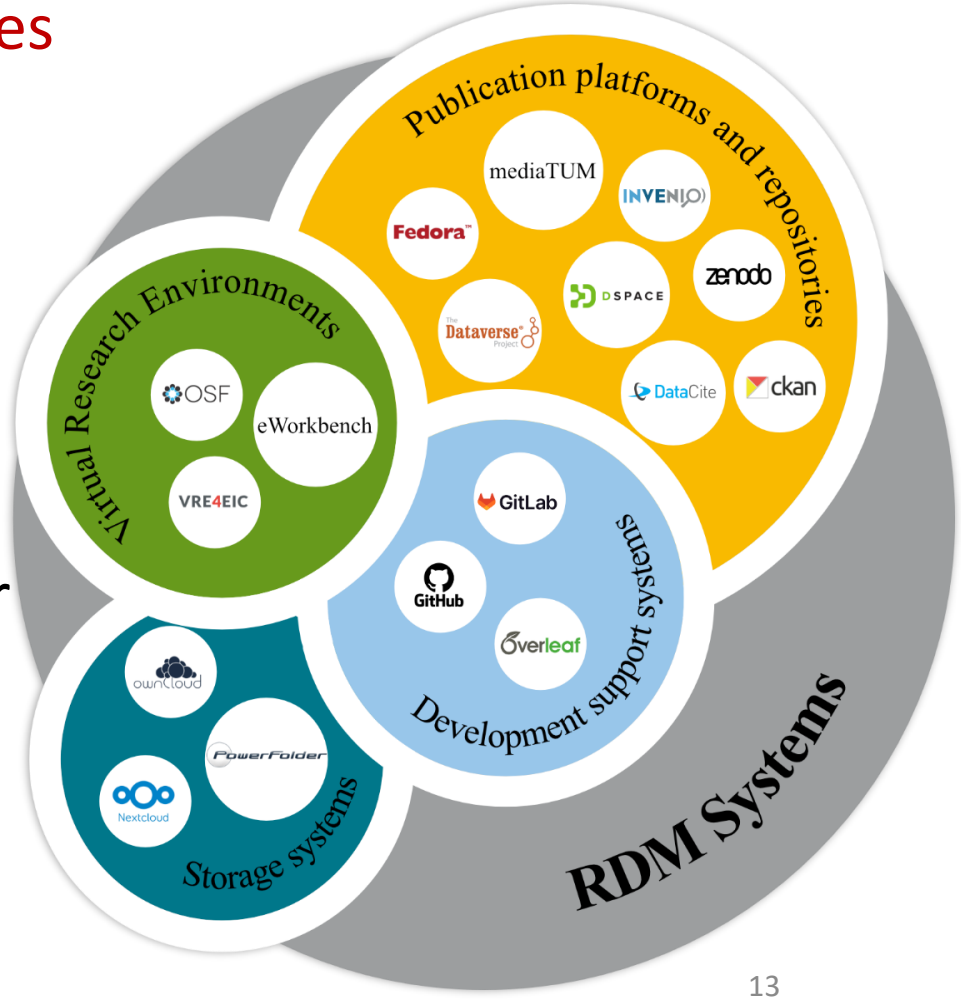
Guiding materials: Official RDM platform and infrastructure contd…

- User scenario
- Practical part
  - User interaction schemes and outcomes
  - Metadata
  - Data management plan (DMP)
  - Network and communication protocols
  - Integration and interaction with external systems
- Conventions and best practices
- Documentation possibilities
- Sharing and release scenarios
- Configuring functions and features
- Frequently asked questions (FAQ)
- Access to software, support and services



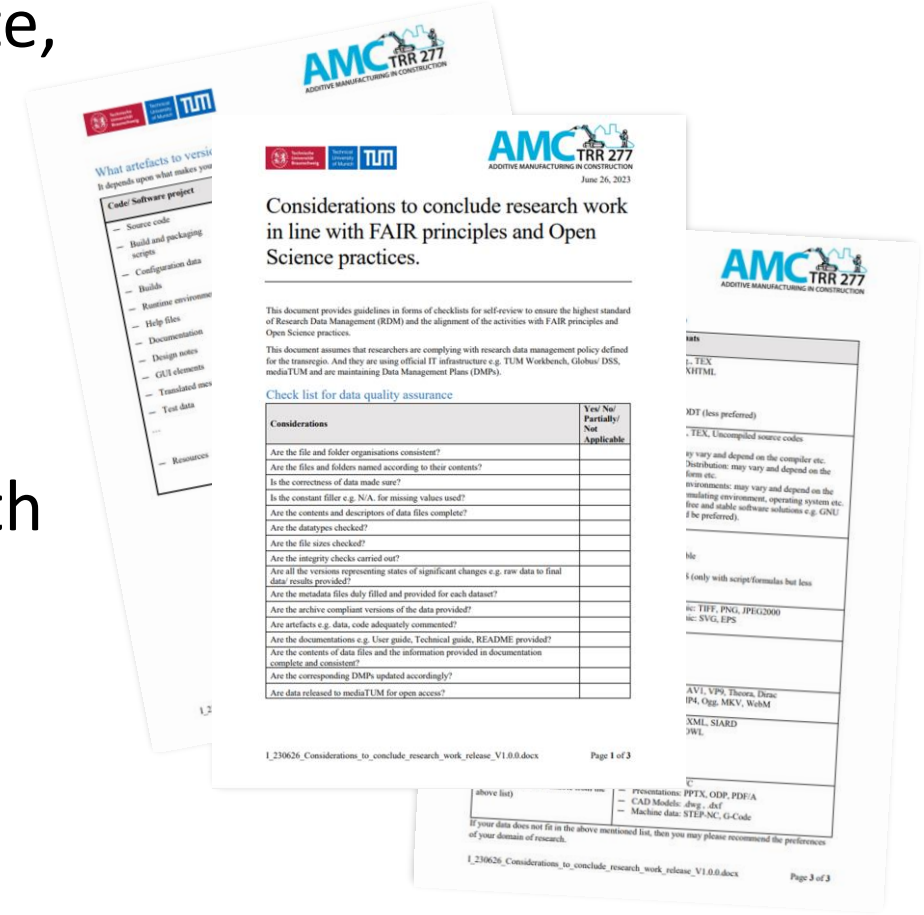
Guiding materials: Research practices

- Data identification based of its roles, states, types
- Folder/ file organisations and structures
- Identification of atomic/ independent workload
- Collaborative work e.g. task based
- Distribution of larger/ collaborative work packages
- Specific and general rules for decision: -
  - e.g. research data, information necessary for reproducibility, to register data in RDM system, significant change in data to record
- Research and data support systems and their roles



Guiding materials: Research practices contd…

- Adoption of common standards e.g. DataCite, DOI, ORCID…
- Filling DMPs
- Applications of DMP for atomic workload to the whole CRC
- Concluding research task/ project in line with FAIR principles and Open Science practices.



AMC TRR 277
ADDITIVE MANUFACTURING IN CONSTRUCTION

June 26, 2023

Considerations to conclude research work in line with FAIR principles and Open Science practices.

This document provides guidelines in forms of checklists for self-review to ensure the highest standard of Research Data Management (RDM) and the alignment of the activities with FAIR principles and Open Science practices.

This document assumes that researchers are complying with research data management policy defined for the transregio. And they are using official IT infrastructure e.g. TUM Workbench, Globus/ DSS, mediaTUM and are maintaining Data Management Plans (DMPs).

Check list for data quality assurance

Considerations	Yes/ No/ Partially/ Not Applicable
Are the file and folder organisations consistent?	
Are the files and folders named according to their contents?	
Is the correctness of data made sure?	
Is the constant filler e.g. N/A. for missing values used?	
Are the contents and descriptors of data files complete?	
Are the datatypes checked?	
Are the file sizes checked?	
Are the integrity checks carried out?	
Are all the versions representing states of significant changes e.g. raw data to final data/ results provided?	
Are the metadata files duly filled and provided for each dataset?	
Are the archive compliant versions of the data provided?	
Are artefacts e.g. data, code adequately commented?	
Are the documentations e.g. User guide, Technical guide, README provided?	
Are the contents of data files and the information provided in documentation complete and consistent?	
Are the corresponding DMPs updated accordingly?	
Are data released to mediaTUM for open access?	

I_230626_Considerations_to_conclude_research_work_release_V1.0.0.docx Page 1 of 3

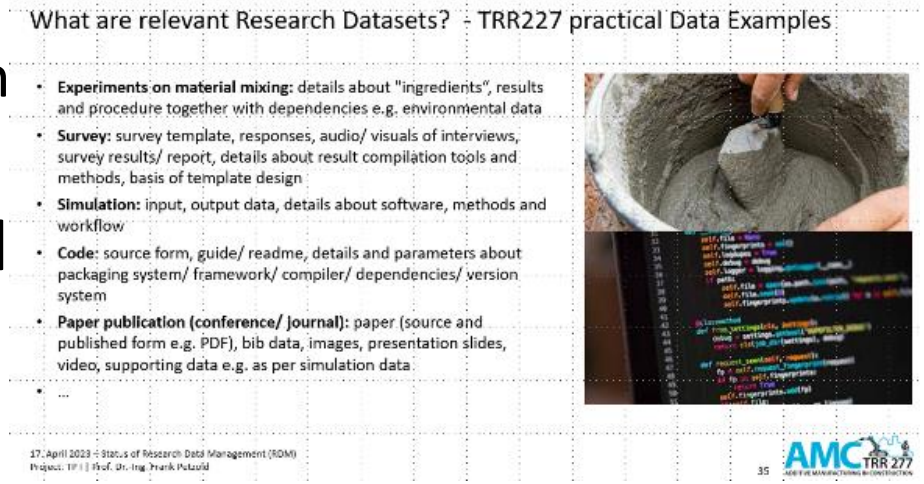
Guiding materials: Categories

- Introductory
- User role specific
- Task/ activity specific
- Topic/ Tool specific
- Use case specific
- Learning phase/ pace specific

Guiding materials: Supplementary materials

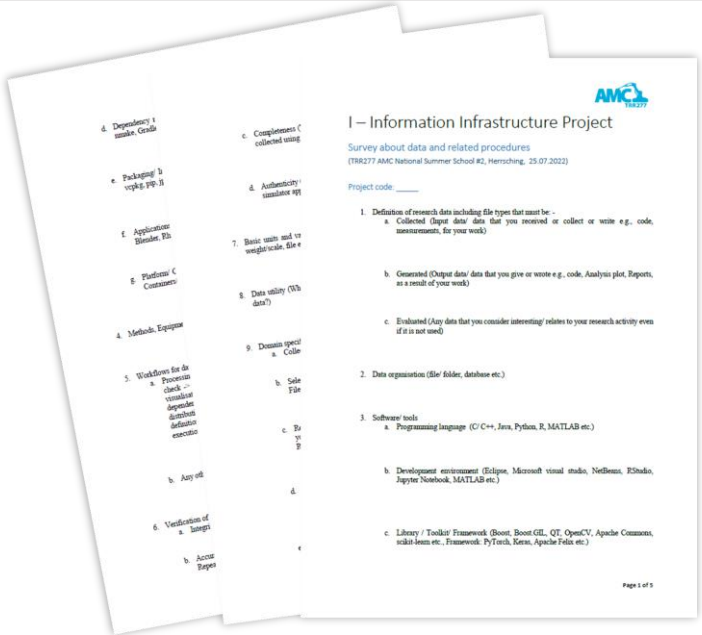
- Standard reusable contents
  - e.g. metadata templates, folder/ file structures, publication templates, labels to improve naming convention, DMP
- Practical relevant research data examples
  - e.g. lab experiment, material mixing, survey, simulation, code, journal/ conference publication data management plan
- Collections of standard external contents and resources
  - e.g. List of licences, tools, metadata standards

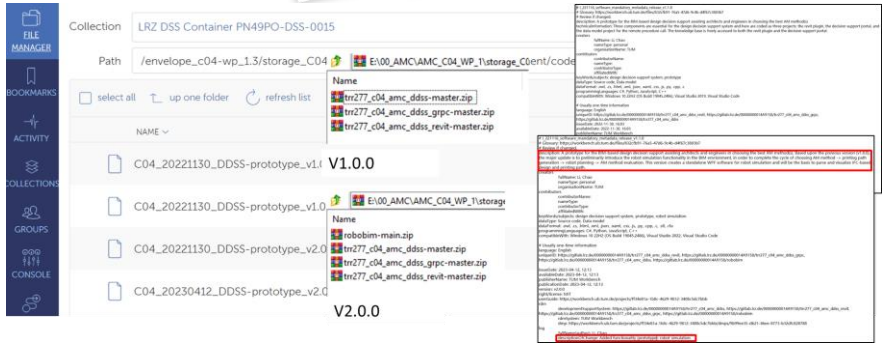
draft	raw	dev
final	process/ process_name	master
preprint	release/ publish	build
release	release_candidate	release
Publish	stable_release	release_candidate
		stable_release



Guiding materials: Supplementary materials contd …

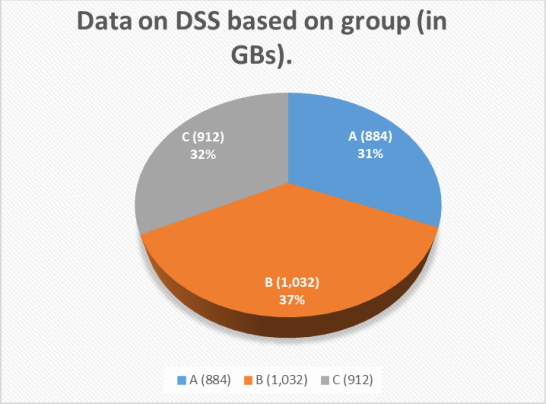
- Comprehensive survey
- Recommendation and hints about tools, best practices and workflows
  - Coding e.g. build and packaging, dependency management, compiler, execution environments
  - Data/ metadata creation e.g. collection, selection, retention, transformation, basic units and value ranges
  - Data verification/ quality assurance measures and strategies e.g. integrity, accuracy, completeness, authenticity checks
  - Data analysis, experiments, digital representation
  - Data acquisition, integration, anonymization, pseudonymization, release, archiving



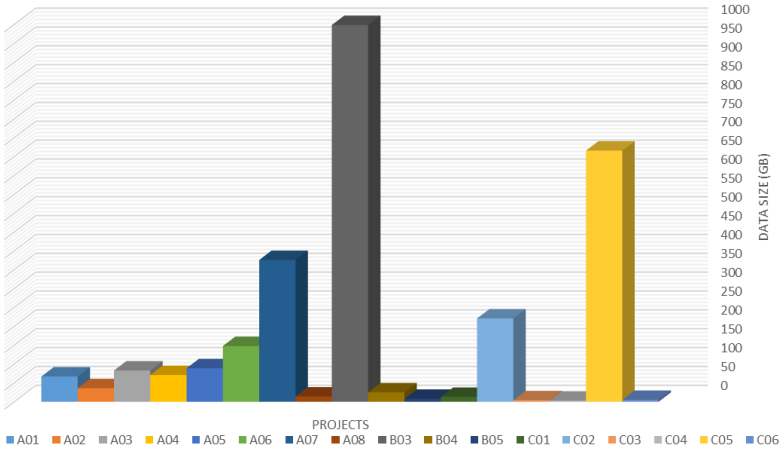


Guiding materials: Supplementary materials contd …

- Use cases of collaboration and cooperation
- Systems and tools for automation
  - Guides for additional tools and systems
- Detailed review, evaluation and feedback reports and follow ups
- Considerations/ Checklist to conclude research work

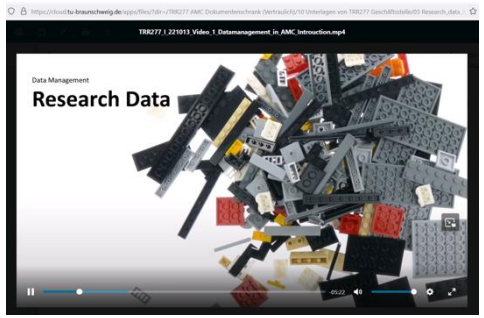


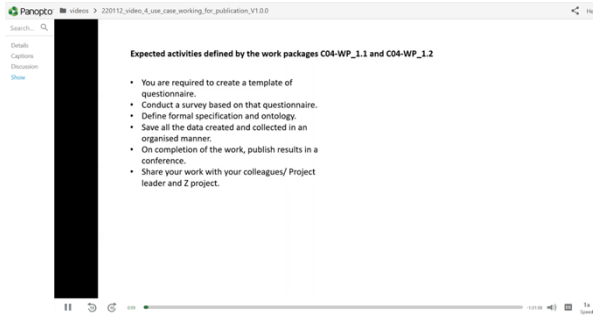
Data on DSS (total: 2,828 GBs)

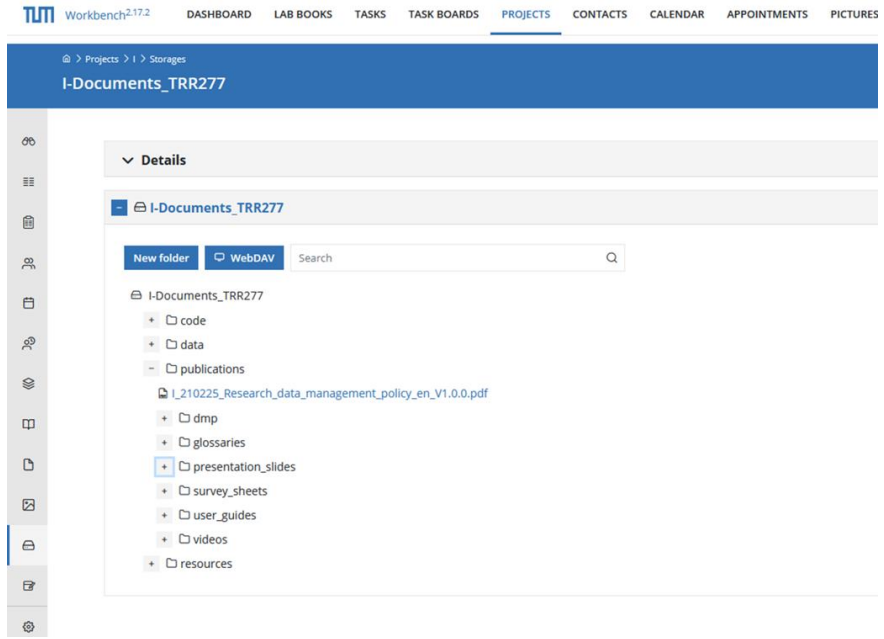


Means of distribution, communication and imparting trainings

- Online streaming
  - Panopto
  - TUBS Cloud/ NextCloud
- File hosting (downloadable)
  - Official platform (TUM Workbench)
  - TUBS Cloud/ NextCloud
- Live sessions (One to one/ Group/ Collective)
  - Online e.g. via Zoom
  - In person
- Official platform based communication tools
- Institutional email and communication services





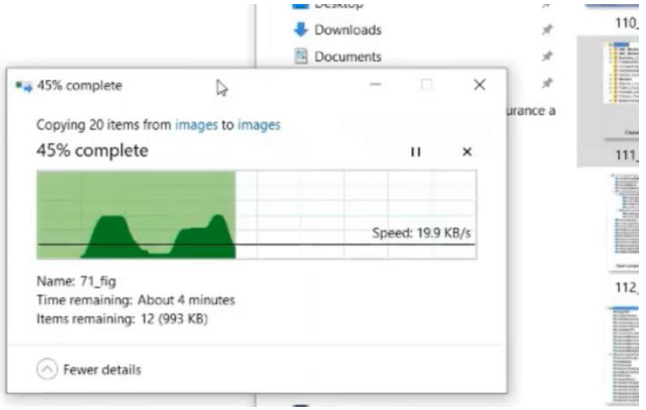


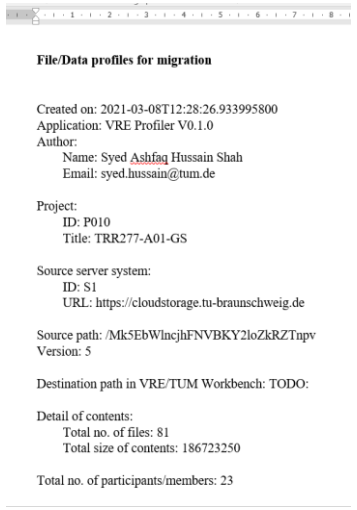
Interactive sessions and workshops' themes

- Introduction to the VRE system and concepts
- User and data migration
- VRE for data managers/ RDM
- RDM and its basics
- RDM tools and systems
- How to use (tools specific sessions)
- Software/ tools test sessions
- Workflows e.g. for publication, experiments, collaboration
- Practical research data examples
- Do it together session
- Weekly consultation session
- Issue/ Topic/ Task specific consultation sessions
- Q & A sessions
- Reviews and feedbacks sessions
- Consultations based on the reviews and feedbacks
- Quarterly progress meeting
- Summer school

On boarding and migration processes

- Profiling of existing data
  - Human readable format (.docx/HTML)
  - Machine readable format (JSON)
- Verification of data profiles
- Test and allocation of resources
  - Project structures
  - Storage structures
  - Access rights
  - Metadata templates
- Interactive do it together migration sessions





File/Data profiles for migration

Created on: 2021-03-08T12:28:26.933995800
Application: VRE Profiler V0.1.0
Author:
Name: Syed Ashfaq Hussain Shah
Email: syed.hussain@tum.de

Project:
ID: P010
Title: TRR277-A01-GS

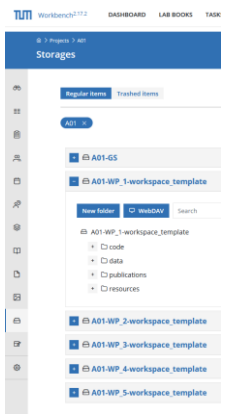
Source server system:
ID: S1
URL: https://cloudstorage.tu-braunschweig.de

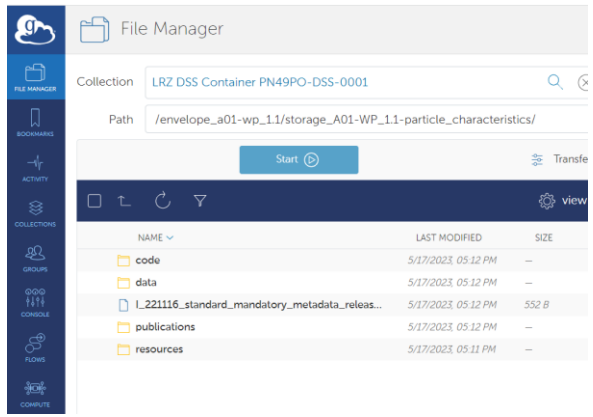
Source path: /Mk5EbWlncjhFNVBKY2loZkRZTnpv
Version: 5

Destination path in VRE/TUM Workbench: TODO:

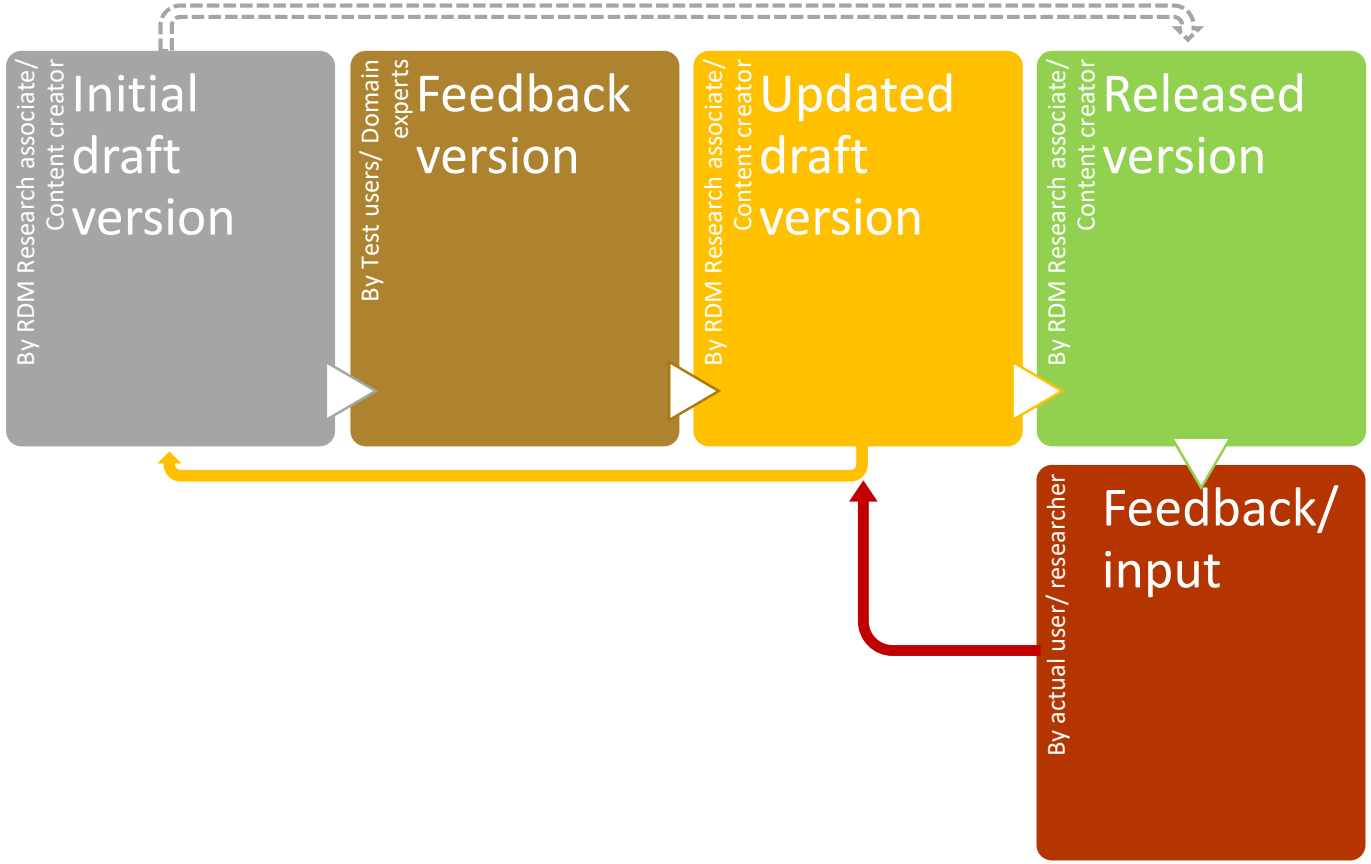
Detail of contents:
Total no. of files: 81
Total size of contents: 186723250

Total no. of participants/members: 23

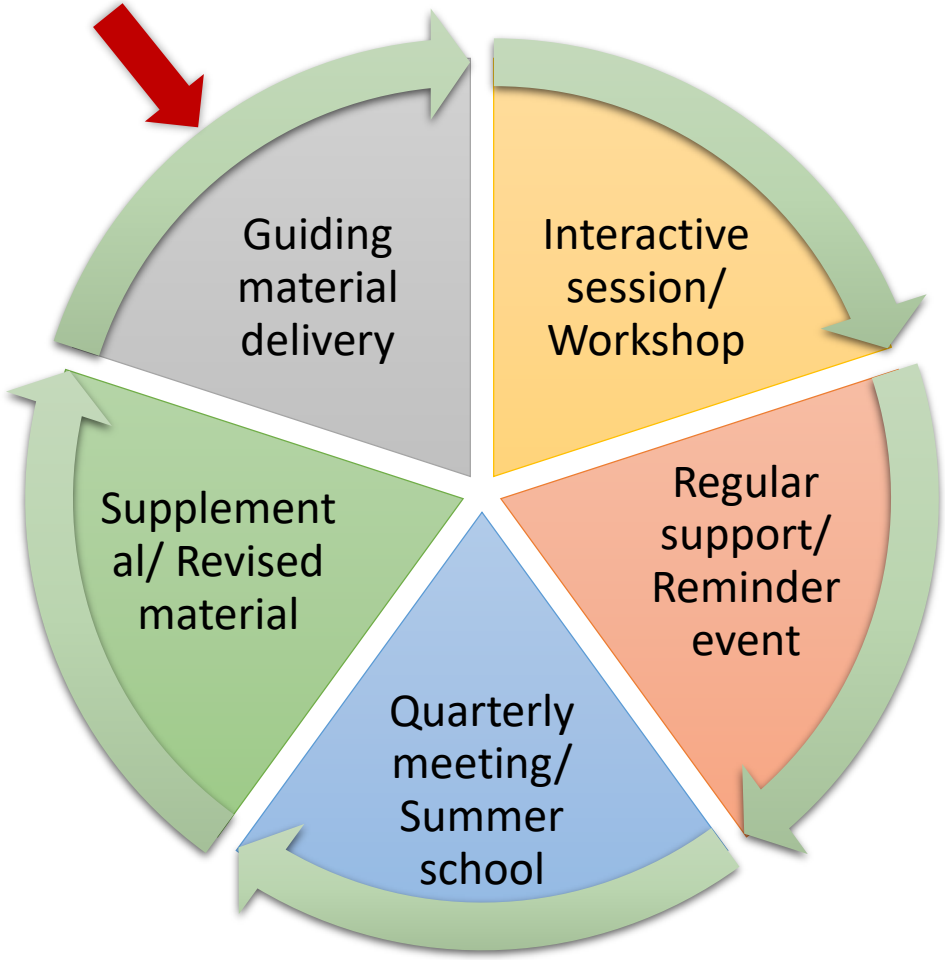




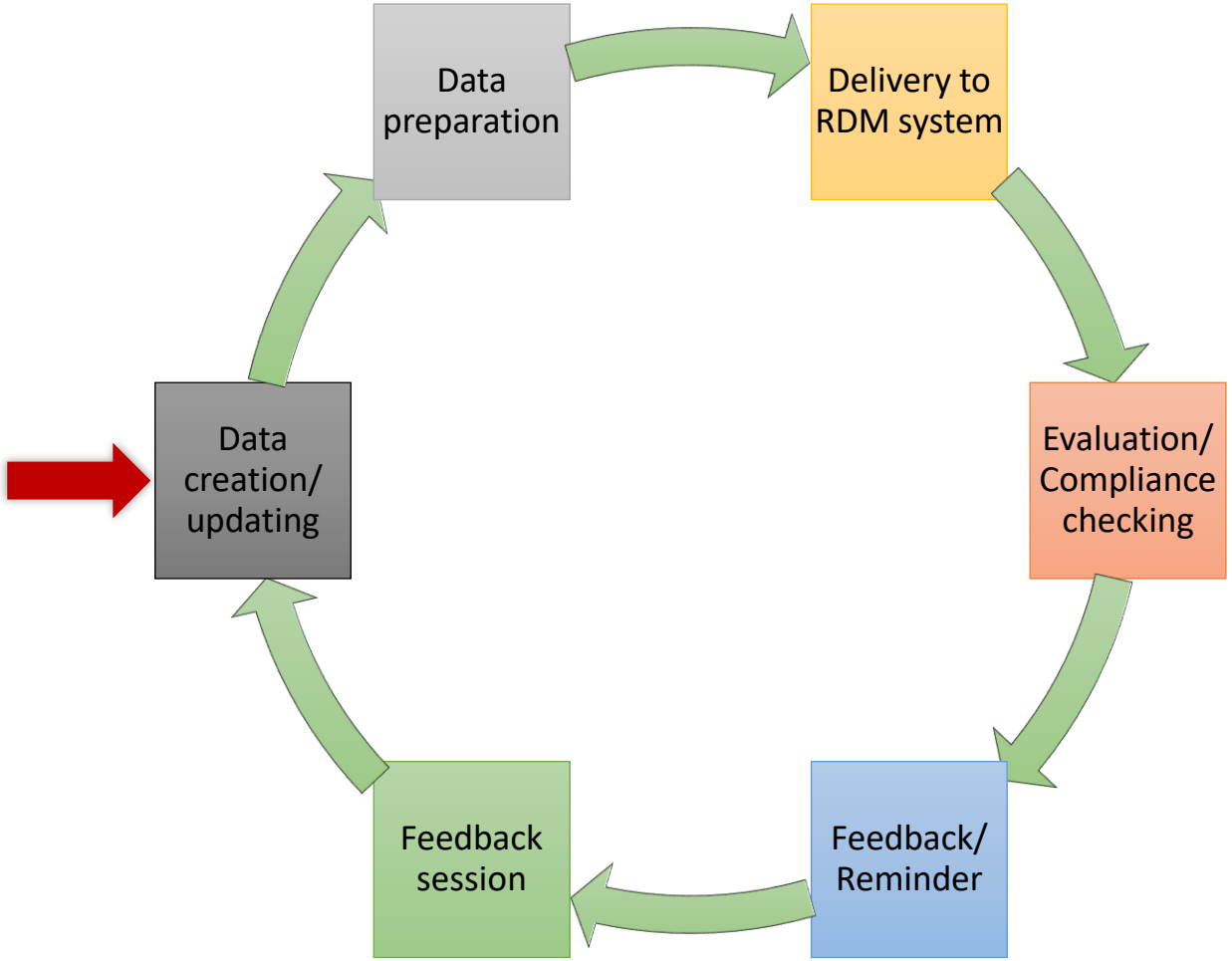
Processes and phases to create contents



Events to improve guidance and understanding

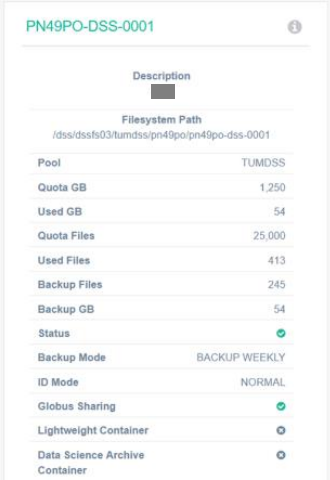


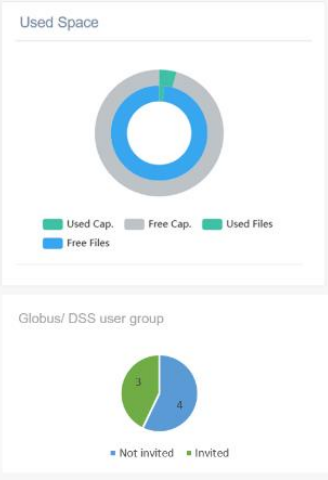
Processes and actions to improve RDM practices

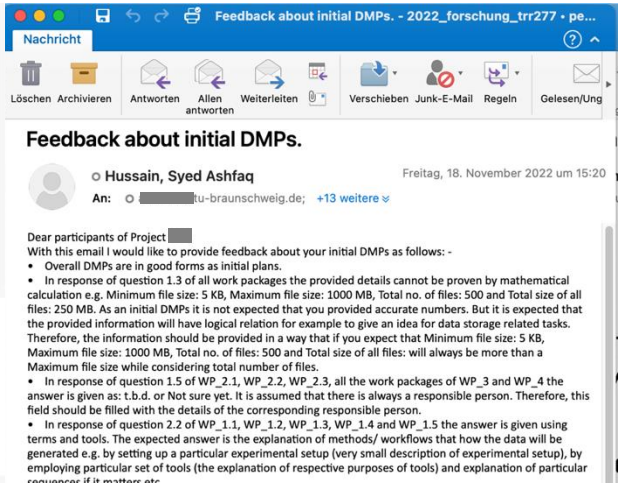


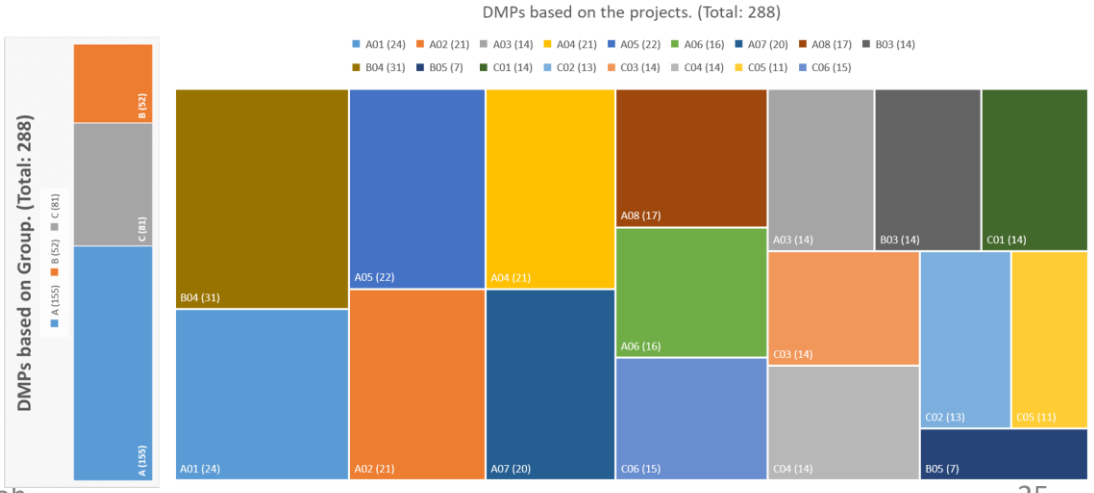
Key compliance strategies

- Provision of initial structures and templates for a smooth and easy kick start
- Examples of best practices
- Detailed interactive sessions
- Interactive do it together sessions
- Q & A rounds
- Detailed reviews and feedbacks
- Interactive review and feedback sessions
- Supplementary materials
- Comprehensive survey about RDM practices
- Collaborative work/ assistance in case of difficulties
- Continuously improving guiding materials and IT solutions.









Tactics to enforce compliance for RDM

- Obligation through Research data policy
- Advantages and confidence in official platforms
- Influence of research organising board/ committee
- Influence through research supervisor/ supervisory board
- Reminders during collective meeting events, through communication channels before and after the schedule is missed.
- Presentation of results, summaries and updates during meetings/ collective events
- Presentation of feedback and evaluation reports
- Narrating adverse impacts and implications of non compliance
  - e.g. data loss, consequences of further approvals

Future work

- Marginalising and estimating the weight of impact.
- Further features' definition and development for automation.
- Evaluation of different approaches.

- It is a process of continuous improvement with retrospectives.

Thank you for your attention

Questions?

Acknowledgements

The work is being conducted as part of the collaborative research centre ‘Additive Manufacturing in Construction - The Challenge of Large Scale. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project Number 414265976 - TRR 277".