Research Data Management

Basics and Services

Holger Gauza, ZDV University of Tübingen

Tübingen Research Data Management Days
July 14th & 15th 2022

Overview

- What are research data?
- What is research data management?
- Importance of research data management
- Central aspects
- Service landscape

What are research data?

Research data are data produced during scientific work.

1. Primary data: Directly collected form the data source or an object / sample itself
    - Measurements
    - A/V Material
    - Born digital or digitalised text

2. Secondary data: Data derived from primary data
    - Aggregated data
    - Transcription
    - Translation

What are research data?

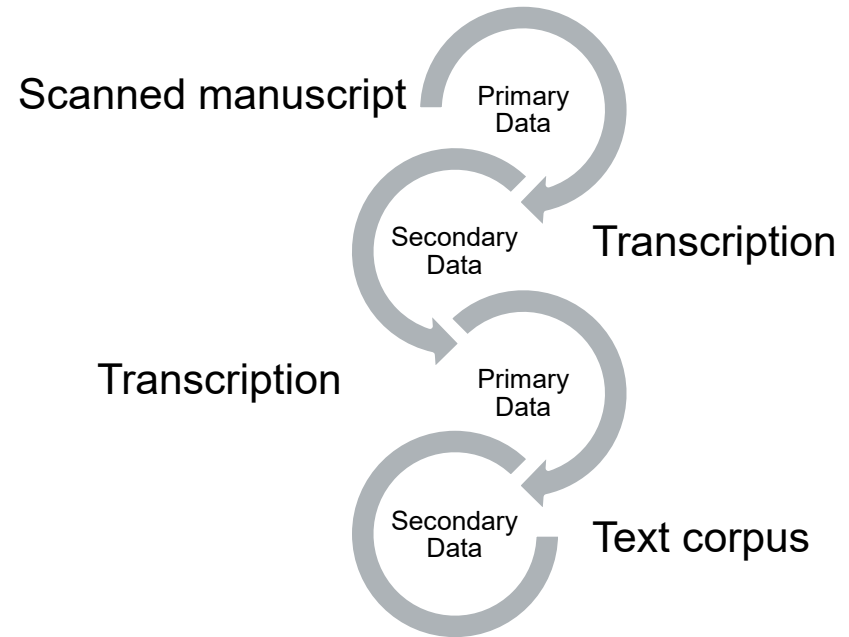Research data are data produced during scientific work.

3. Tool Data: Tools used to analyse data
    - Algorithms
    - Scripts
    - Software

4. Metadata: Information about data
    - Information about data analysis
    - Information about data context
    - Information about granted rights
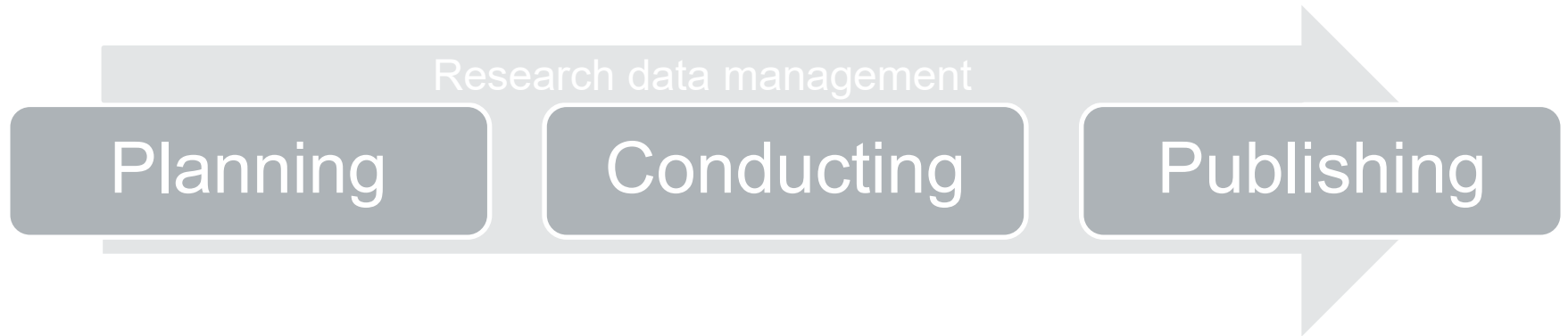
What are research data?

- Heterogeneous and domain specific group
    - Objects
    - Statistics
    - …
- Input to further research reduces costs
- Data must be accessible and reusable

Scanned manuscript
Primary Data
Transcription
Secondary Data
Transcription
Primary Data
Secondary Data
Text corpus

What is research data management?

Research data management is a process that aims at making research data long term available, reusable and reproduceable.

Research data management

| Planning | Conducting | Publishing |

What is research data management?

Research data management is a process that aims at making research data long term available, reusable and reproduceable.
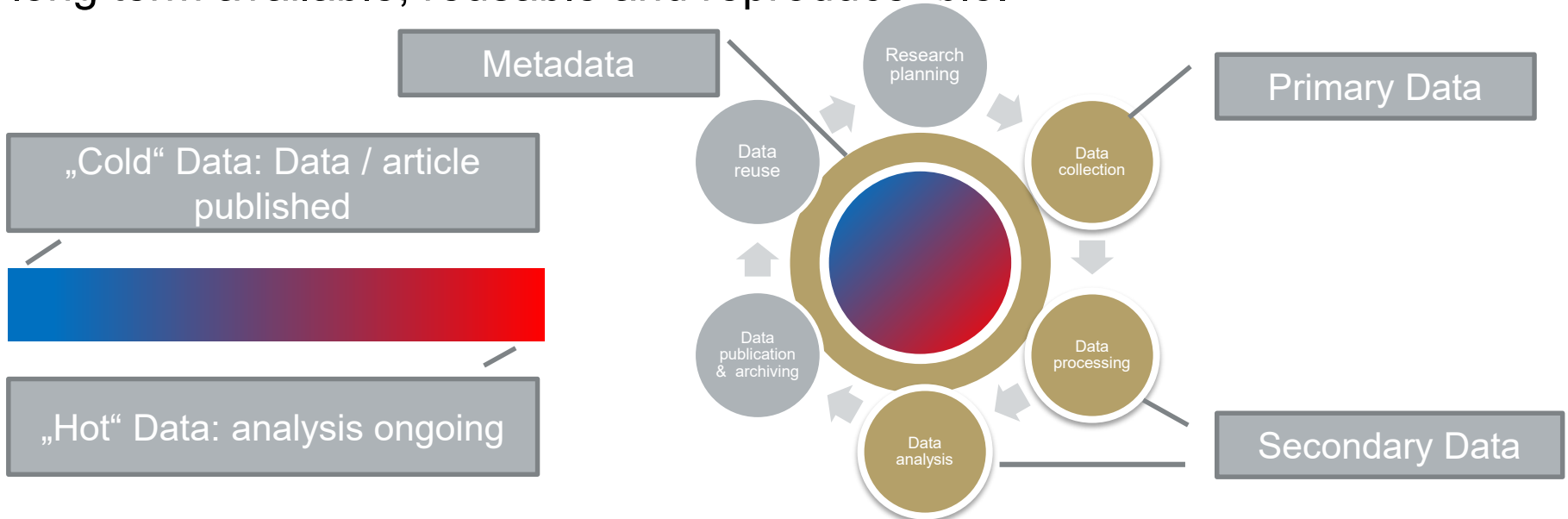
Why is research data management important?

Individual reasons for research data management:

1. Research data management saves time

   • No reverse engineering of your own results

   • Keep track of your data

2. Increases transparency

   • Origin and data provenience

   • Workflows and data processing

3. Reduces risk of data loss

   • Institutional storage instead of thumb drives

...DEIN Backup?

Wo sind deine Forschungsdaten in 10-Jahren?! CC BY 4.0

Why is research data management important?

Individual reasons for research data management:
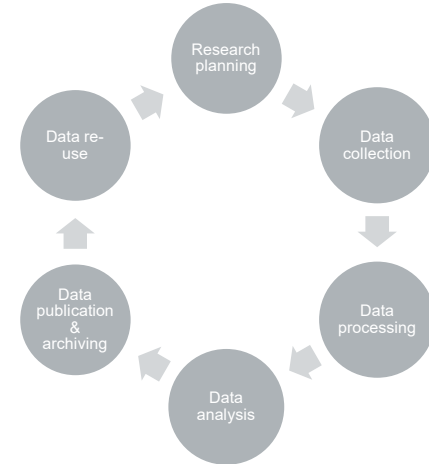
4.   Improves citations

 •   Data as citable as a (journal) publication

 •   Increases visibility of your scientific work

5.   Makes data available for the community

 •   Reuse of existing data

 •   Facilitates new insights into old data

 •   Easy collaboration

Why is research data management important?

Research data management is also expected:

1. Funding agencies

   - Horizon Europe: "No opting out of RDM. Projects generating research data MUST manage their data responsibly and in line with FAIR principles."

   - DFG: "In the future, the handling of research data will be given more attention than before in the review and evaluation process."
     Additionally, "[…] the DFG encourages applicants to list corresponding contributions in their curriculum vitae."

2. Institution policies

   - Tübingen University guidelines: "Responsibility for research data and compliance with discipline-specific standards lies with the scientists as producers."

Central aspects

FAIR Data Principles serve as guideline throughout the life cycle.

Findable
1. Unique identifier for (meta)data
2. Rich metadata description
3. Metadata include unique identifiers for data
4. (Meta)data are indexed in searchable resource

FAIR Data Principles
🔍 Findable ⚙ Interoperable
⊕ Accessible🔄 Reusable

For a complete description and definition see FAIR Principles - GO FAIR (go-fair.org)

Central aspects

FAIR Data Principles serve as guideline throughout the life cycle.

Accessible
1. (Meta)data are easily retrievable via unique identifiers and interfaces
2. Metadata are accessible, even when the data are deleted

FAIR Data Principles

🔍 Findable   ⚙ Interoperable

⊕ Accessible 🔄 Reusable

For a complete description and definition see FAIR Principles - GO FAIR (go-fair.org)

Central aspects

FAIR Data Principles serve as guideline throughout the life cycle.

Interoperable

1. (Meta)data use a broadly applicable knowledge representation (e.g. file format)
2. (Meta)data use vocabularies
3. (Meta)data include references to other (meta)data

FAIR Data Principles
🔍 Findable   ⚙ Interoperable
⊕ Accessible 🔄 Reusable

For a complete description and definition see FAIR Principles - GO FAIR (go-fair.org)

Central aspects

FAIR Data Principles serve as guideline throughout the life cycle.

Reusable
1. (Meta)data are richly described
    with accurate and relevant
    attributes
    - Detailed provenance
    - Meet domain-relevant
      community standards
    - Data usage licences

FAIR Data Principles
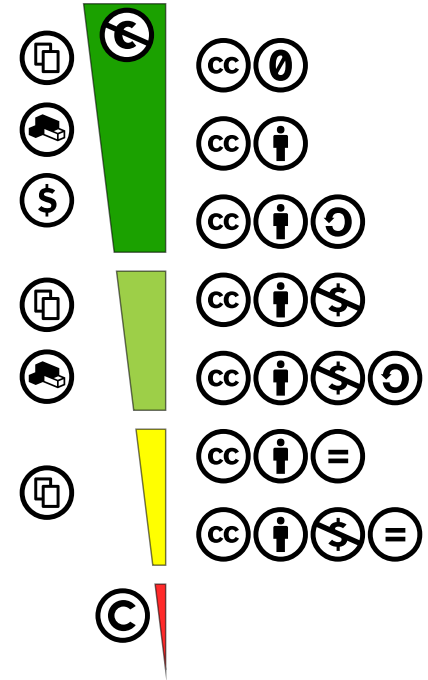🔍 Findable   ⚙ Interoperable
⊕ Accessible 🔄 Reusable

For a complete description and definition
see FAIR Principles - GO FAIR (go-fair.org)

Central aspects

Data usage licences:

- Standardized licences
- Creative Commons Licenses
  - CC 0: No restrictions
  - CC BY: Give credit
  - CC BY ND: Give credit, don't modify
- Peace of mind about data usage
- More information here

Shaddim; original CC license symbols by Creative Commons, Creative commons license spectrum, CC BY 4.0

Central aspects

Persistent unique identifiers:

- ORCID: Open Researcher and Contributor ID
- 0000-0003-0191-3680
- Unique identifier for people
- CV, publications, funding, …

- DOI: Digital Object Identifier
- https://doi.org/10.17192/bfdm.2021.3.8348
- Unique identifier for publications
- More stable than URLs

Service landscape

Research data management is not only required but also supported by various institutions:

- National network
- Trainings, standards, services
- Up to 30 consortia for different research areas
  - FAIRmat (physics), MaRDI (mathematics)
  - Text+, NFDI4Culture

Service landscape

Research data management is not only required but also supported by various institutions:

- 4 Science Data Centers sponsored by Baden-Württemberg
- Trainings, standards, services
- Economics, bioinformatics, material science, literature research

Service landscape

Research data management is not only required but also supported by various institutions:

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

- Digital Humanities Center
  - Consulting & service
  - FDAT: Research data repository
- ZDV for storage solutions

Examples

Bob who?

A paper from 2015 describes interesting data

- Contact info of the corresponding author: <u>bsmith@university.edu</u>
- Mailer daemon as replay

- Contact info of the corresponding author: iD 1234-4321-5678-8765
- New affiliation & contact info

👍

Examples

Data where?

Bob is happy to help but …

- The data were stored on the institution`s server
- As Bob left the university, the data were not moved to the new servers

- Bob uploaded the data on a public repository
- A DOI leads to the data

FAIR Data Principles
🔍 Findable ⚙ Interoperable
⬇ Accessible 🔄 Reusable

Examples

Data what?

The data are there but …

- The were stored in a proprietary format that requires an expensive software to read

- Bob converted the data into an open format

FAIR Data Principles
🔍 Findable   ⚙️ Interoperable
⬇ Accessible 🔄 Reusable 👍

Thanks for your attention!