HeFDI Data Talk

Date	Topic	Presenter
02 June 2023	HeFDI-Repositories – a service offer	Andreas Geißner (TU Darmstadt) Lydia Riedl (Philipps-Universität Marburg)

Abstract:

A research data repository is a platform that allows researchers to deposit, store, and share their data with others. We will discuss the importance of sharing research data openly and the benefits of using a research data repository. We will introduce the HeFDI repository landscape with a special focus on the institutional repository of the University of Marburg (data_UMR).

About the HeFDI Data Talks:

The HeFDI Data Talks are a bi-weekly open information and discussion event focused on data management in the context of science, in which relevant NFDI consortia as well as research data management services present themselves. The series discusses current topics and presents numerous – including local and regional – tools and services. The HeFDI Data Talks are an offer of the HeFDI Initiative (Landesinitiative HeFDI), which is funded by Hesse's Ministry for Science and Arts (HMWK).
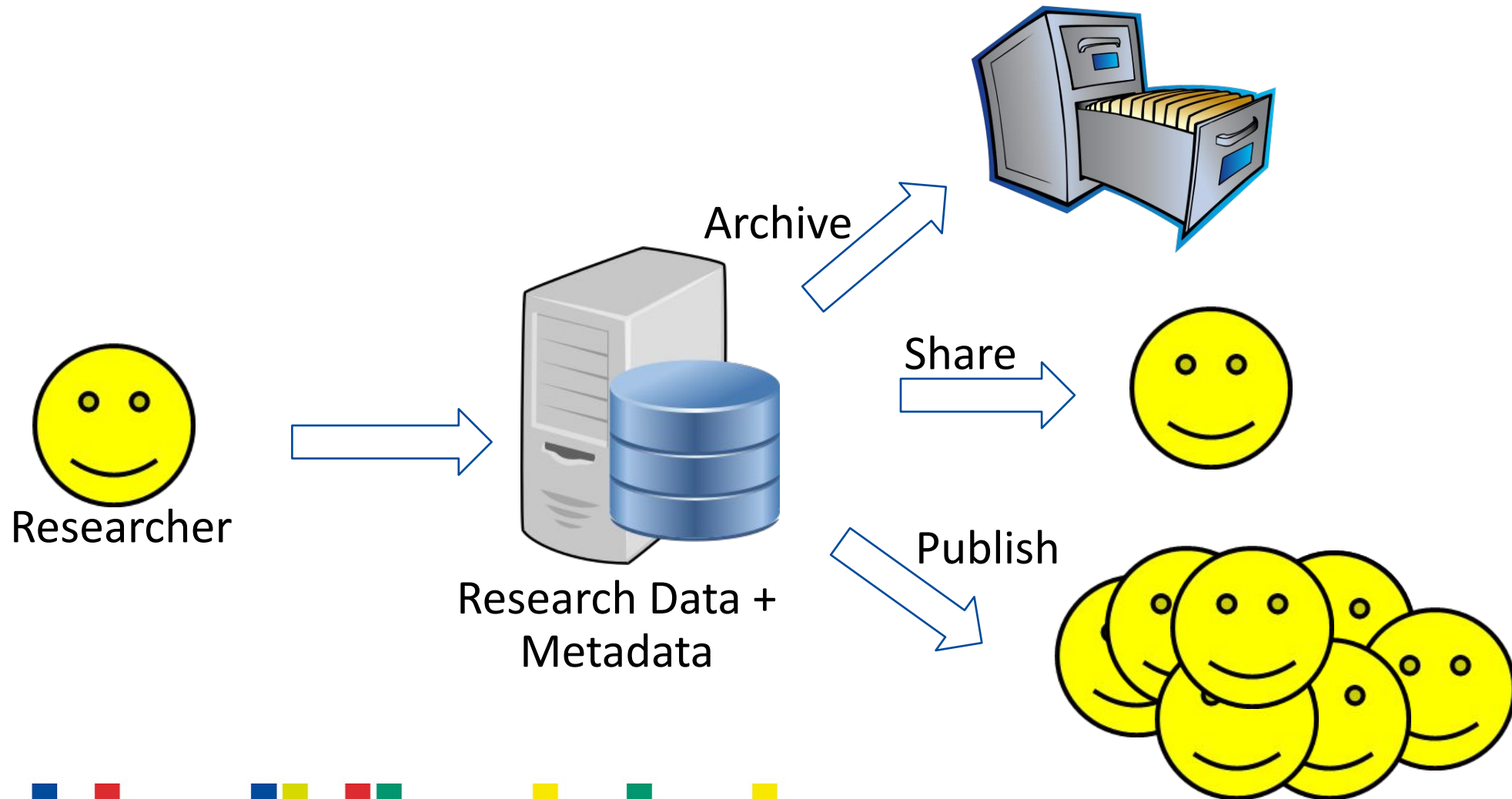
HeFDI-Repositories - a service offer

Andreas Geißner (TU Darmstadt)

Lydia Riedl (Philipps-Universität Marburg)

What is a research data repository?

Archive

Share

Publish

Researcher

Research Data +
Metadata

# Why: Information and knowledge transfer
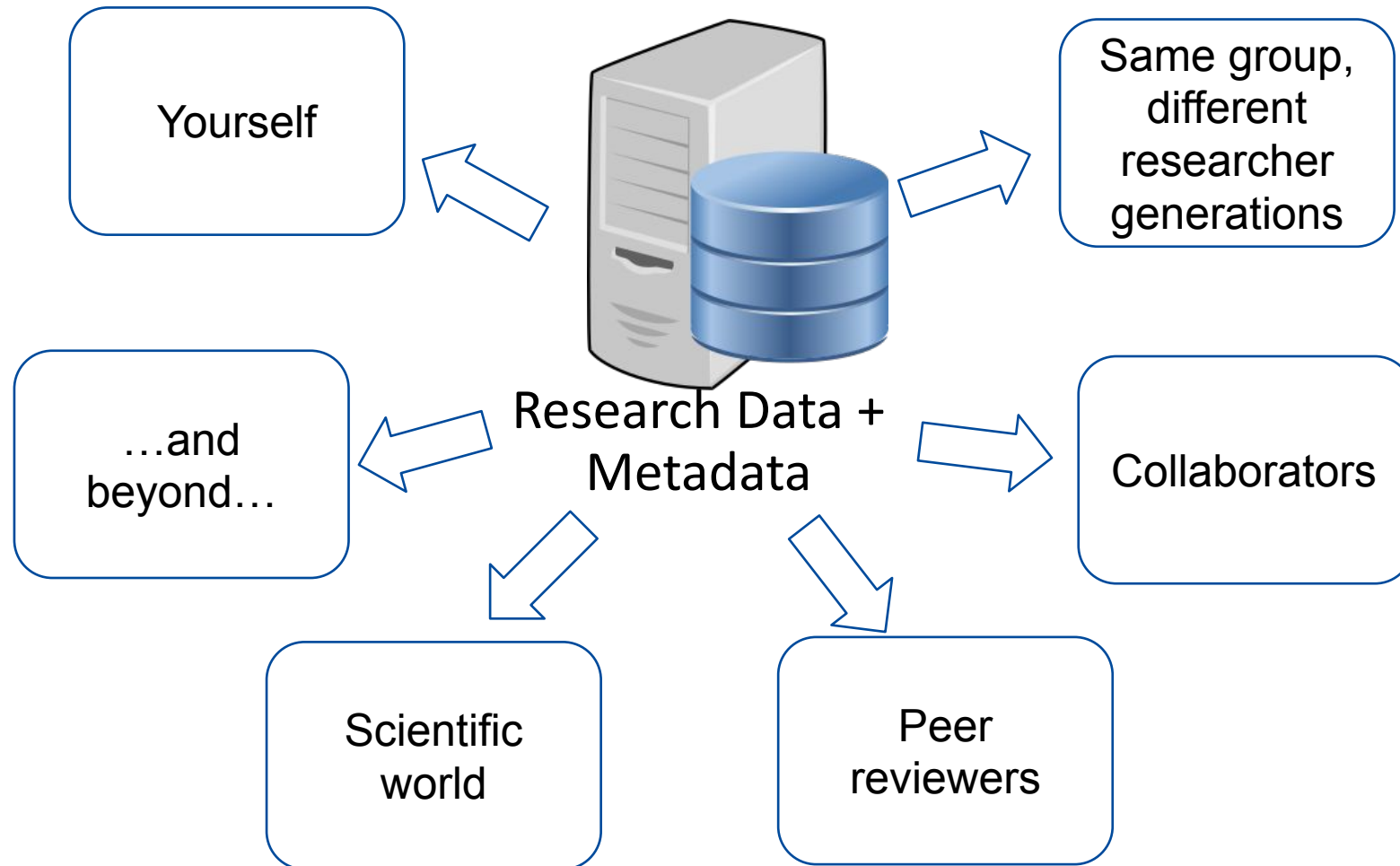
Why: Good scientific practice

Guideline 7: Cross-phase quality assurance	Guideline 13: Providing public access to research results	Guideline 17: Archiving
That includes "compliance with subject-specific standards and established methods […], the collection, processing and analysis of research data, the selection and use of research software […] and the keeping of lab notebooks." "The origin of the data […] and software used in the research process is disclosed and the reuse of data is clearly indicated."	„Where possible and reasonable, this includes **making the research data, materials and information** on which the results are based, as well as the methods and software used, **available** and fully explaining the work processes. **Software** programmed by researchers themselves is made **publicly available along with the source code.**"	„When scientific and academic findings are made publicly available, the **research data** (generally raw data) on which they are based **are generally archived in an accessible and identifiable manner for a period of ten years at the institution where the data were produced or in cross-location repositories**."

Life Cycle
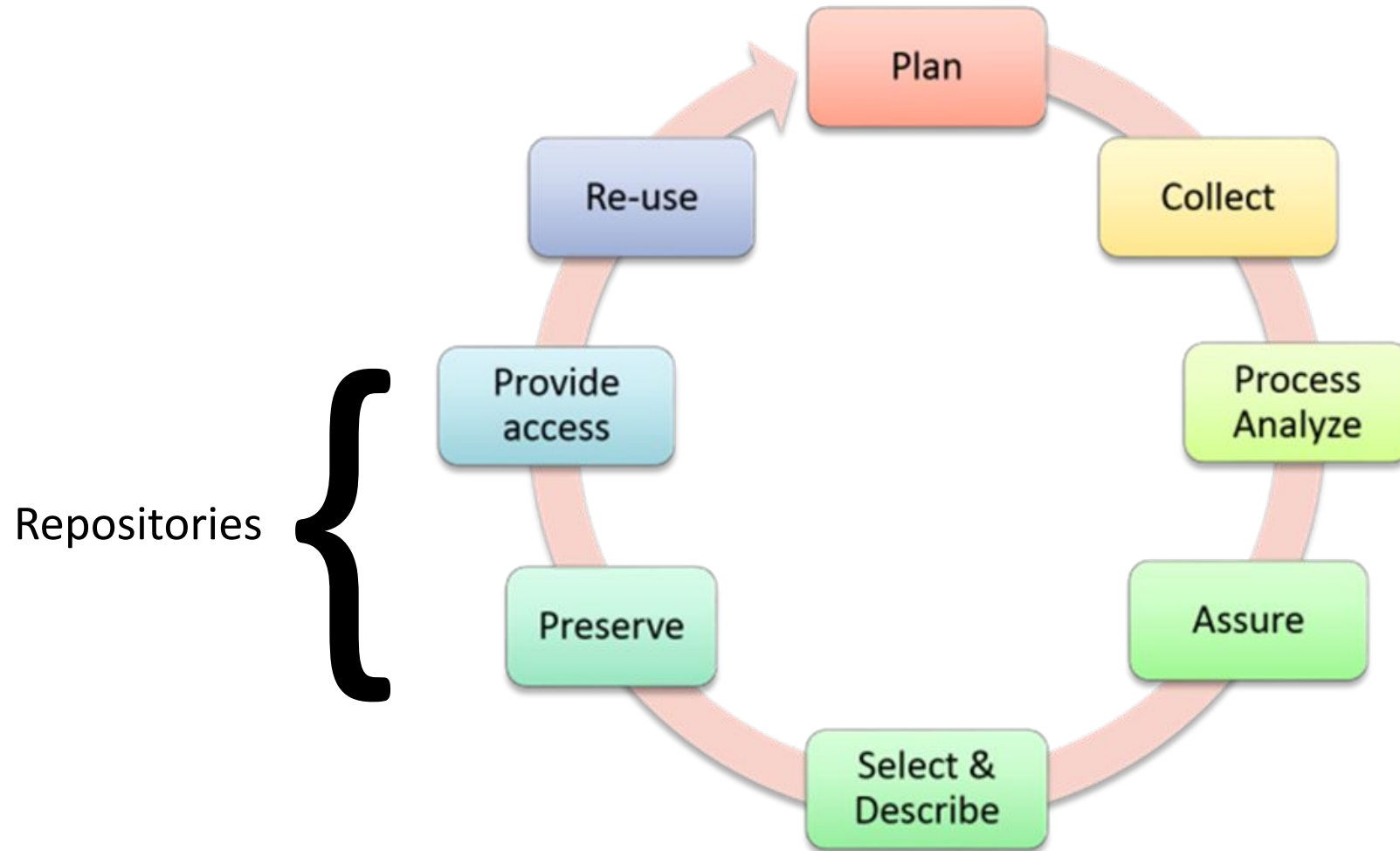
Typical features of data repositories

- Upload data and input metadata

- Assign persistent identifiers for data (e.g. DOI)

- Link persistent identifiers for author (e.g. ORCID)

- Data sharing in teams or upon request

- Assignment of licenses

- Download of public datasets

- Dataset versioning

- Search functions

- Indexed in data discovery systems

Types of research data repositories

- Usually differentiated by target audience

  - Who can submit datasets?

  - Researchers of which fields might be interested in the contents?

- Discipline-specific repositories

- Generic repositories

- Institutional repositories

Discipline-specific repositories

- Available for anyone in the field without regard of data origin

- High visibility in the respective field

- Fitting domain-specific metadata schemes

- Run by discipline experts

- Repository-specific extent of service (e.g. only publishing)

- Should be located at permanent institution

- Available for your field?

re3data.org

- Registry of Research Data Repositories
- >3000 repositories listed

- Identify discipline-specific repositories
- Other filters
  - Data types
  - Metadata standards
  - Possible licenses
  - Access restrictions
  - …

Generic repositories

- "Catch-all" repositories

- Not tailored to a specific discipline

- Not restricted to a specific data origin

- Generic metadata schemas

- Some large ones by public institutions

- Example: Zenodo

  - Hosted at CERN

  - Not for profit

Institutional repositories

- Subset of subject-agnostic repositories

- Run by research institutions

- Scope limited

  - Archive and publish data from that institution
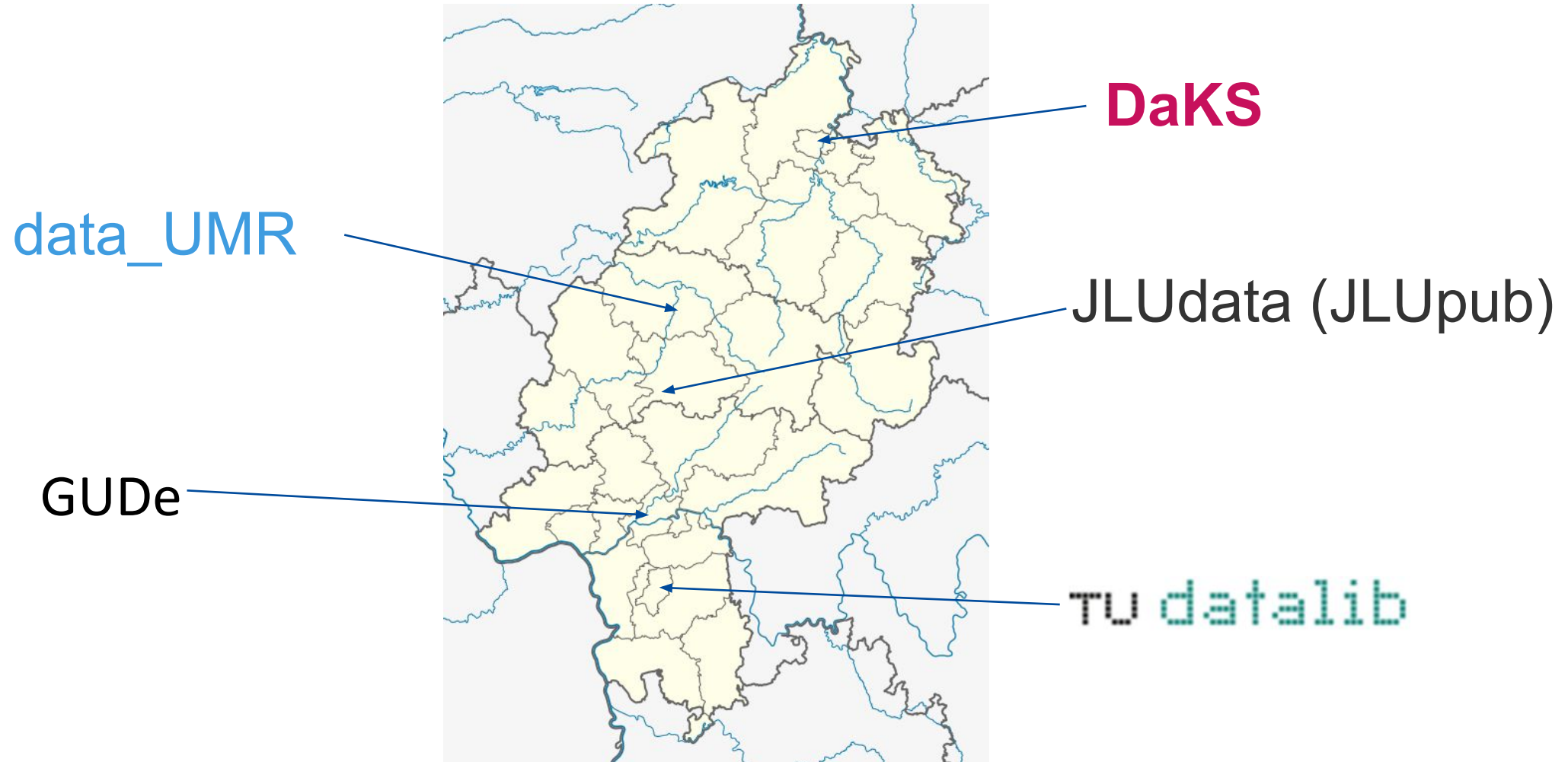
- Several institutional repositories in HeFDI context

An institutional repository is suitable for these purposes:

- Data for which no discipline-specific repository exists yet

- Scientists who prefer to store their data "at home"

- Differentiated assignment of rights

- Long-term storage without publication

HeFDI institutional repositories

DaKS

data_UMR

JLUdata (JLUpub)

GUDe

TU datalib

HeFDI institutional repositories

DaKS

data_UMR

JLUdata (JLUpub)

GUDe

TU datalib

Example data_UMR

Technical Basis of data_UMR

- Currently: DSpace 6
- Hosting by Philipps-University Marburg (Library + University Computer Center)
- Adaptations of the software and further developments are jointly discussed, processed and reused in HeFDI
- Metadata based on recommendations from DataCite (Institute for DOI registration) and DINI (German Initiative for Network Information e.V.)

Example data_UMR

Special service: This repository is curated.

FAIR principles
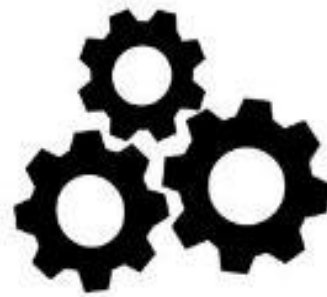
Wilkinson et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3, https://doi.org/10.1038/sdata.2016.18

Findability in data_UMR

Publication of research data:

- Visibility through indexing in relevant search engines
- Citability through DOI registration
- Rich metadata
- Human and machine readable metadata

Findability in data_UMR

Integration of optic flow into the sky compass network in the brain of the desert locust

Abstract

This repository accompanies the 2023 Frontiers manuscript 'Integration of optic flow into the sky compass network in the brain of the desert locust'. It contains physiological and meta data, code for their analysis, and code underlying the computational model and simulation.

Metadata

⬇ To the files

Date	2021-11-17
Authors	Zittrell, Frederick 🔵 Pabst, Kathrin 🔵
Contributors	Carlomagno, Elena Rosner, Ronny 🔵 Pegel, Uta Endres, Dominik M. 🔵 Homberg, Uwe 🔵
DOI	http://dx.doi.org/10.17192/fdr/76
License	3-Clause BSD License (NewBSD)

show more

Publication of research data:

- Visibility through indexing in relevant search engines
- Citability through DOI registration
- Rich metadata
- Human and machine readable metadata

Accessibility in data_UMR

Sharing of research data:

- Open data repository
- Differentiated rights assignment on file level
- Embargo possible
- Consulting and support

Preservation of research data:

- All file types: file formats suitable for archiving
- Preservation 10 years minimum
- Professional backup at the HRZ (University Computer Center)

Accessibility in data_UMR

Files

Name	Format	Size	Checksum (MD5)
README.txt	.txt	4.119Kb	c5724c15e5cfc79b69960e282ef9ca77
complete_dataset.bit	.bit	2 Bytes	9336ebf25087d91c818ee6e9ec29f8c1
available_recording_dates.csv	.csv	479.9Mb	50c431b12c3aec1fcbee09647298b21e
license_CC-BY-NC-4.0.txt	.txt	18.88Kb	d882379f6314cc023ed84088401bbde8

Sharing of research data:

- Open data repository
- Differentiated rights assignment on file level
- Embargo possible
- Consulting and support

Preservation of research data:

- All file types: file formats suitable for archiving
- Preservation 10 years minimum
- Professional backup at the HRZ (University Computer Center)

Accessibility in data_UMR

Create new policy for ITEM 2eb13f3b-f728-47c0-8853-272c95bc9c55

Name:

Description:

Select the action:

○ READ

○ WRITE

○ ADD

○ REMOVE

○ ADMIN

○ WITHDRAWN_READ

Select a group:

Administrator

Start Date:

Accepted format: yyyy, yyyy-mm, yyyy-mm-dd

End Date:

Accepted format: yyyy, yyyy-mm, yyyy-mm-dd

Sharing of research data:

- Open data repository
- Differentiated rights assignment on file level
- Embargo possible
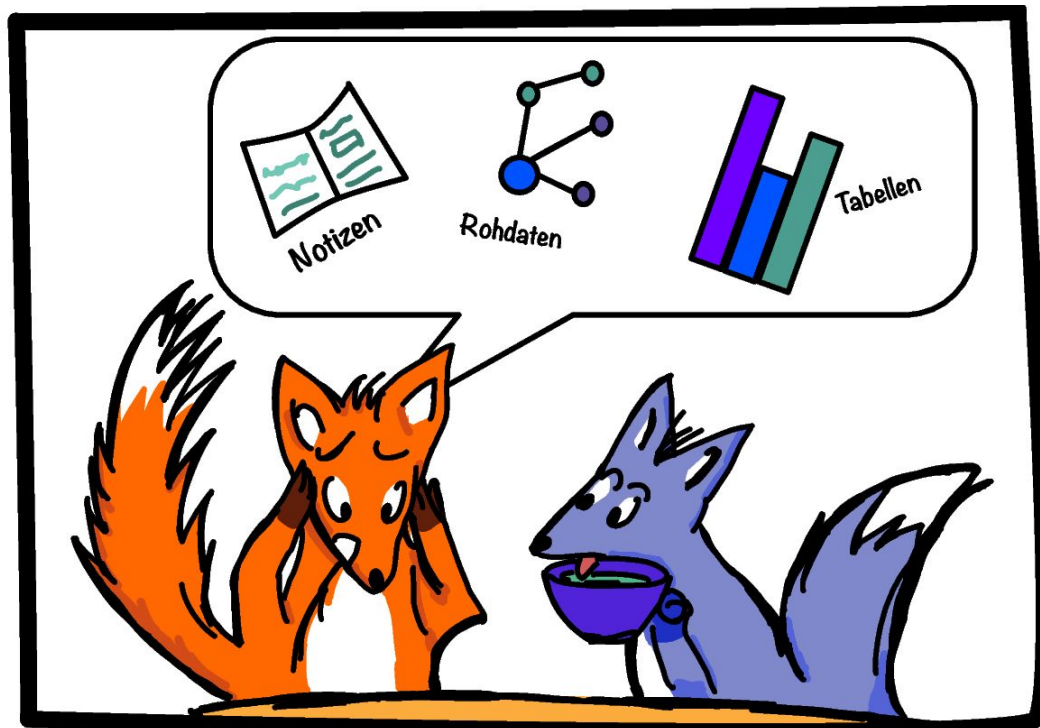- Consulting and support

Preservation of research data:

- All file types: file formats suitable for archiving
- Preservation 10 years minimum
- Professional backup at the HRZ (University Computer Center)

Interoperability in data_UMR

Publication of research data:

- Reference to associated publication possible

Interoperability in data_UMR

Publication of research data:

- Reference to associated publication possible

Relationship   Is Supplement To: (DOI) 10.1371/journal.pcbi.1008179
                Is Supplemented By: (URL) https://github.com/umr-ds/feature_pyramid_fusion

Reusability in data_UMR

Publication of research data:

- Clear regulation of use through licensing
- Documentation as detailed as possible: support through consulting

Reusability in data_UMR

Files

Name	Format	Size	Checksum (MD5)
README.txt	.txt	3.508Kb	e50dee947c42eadb04a4d26c1419d3c1
license_BSD-3-Clause.txt	.txt	1.728Kb	481b25e8d86a8b5fcfde6c38ff077960
Supplementary_Material.pdf	.pdf	469.4Kb	21b61a97965e7ec93821e908397a3b7c
model_code.tar	.tar	60Kb	73363f22ac24415fa60d074bd659f782
analysis_code.tar	.tar	140Kb	ca1d9459850d6ae2cdcd6a986643d336
data.tar	.tar	330Kb	35f958c9339ecc841129bf17f8f6df95
Manuscript.pdf	.pdf	12.49Mb	ee0b3a7a3d5db95c0b0a647fbd040d75
license_BSD-3-Clause.txt	.txt	1.612Kb	d7b203ca8a572d6115e3843841560476

Except where otherwise noted, this item's license is described as 3-Clause BSD License (NewBSD)

Version History

Version	Item	Editor	Date	Summary
12	dataumr/140.12*	Jonas Tschammer	2023-04-04T12:29:23Z	Updating Manuscript.pdf

Publication of research data:

- Clear regulation of use through licensing
- Documentation as detailed as possible: support through consulting

Reusability in data_UMR

In this work, a novel approach to measuring amenable deaths is introduced. The lowest age-specific mortality rates in the USA have been isolated to create normative life tables. The concept of normative life tables was first described in the context of the Global Burden of Disease Study at the University of Washington in Seattle, for measuring the general burden of disease in specific populations. Normative life tables provide an ideal life table for the USA, and shed light on shortcomings in states with comparatively high mortality rates. The normative life table approach is applied for a chronic and frequent health condition in the USA, namely COPD (Chronic Obstructive Lung Disease). The lowest COPD mortality rates in the USA for 2016 have been isolated to create normative COPD life tables. These normative life tables show the best practice for COPD in the USA. Excess deaths in COPD across the states are regarded as amenable deaths, i.e., deaths that with timely and effective medical interventions and public health efforts could have been prevented. California has the lowest proportion of amenable deaths due to COPD. Texas has moderate mortality rates for COPD, while Kentucky has the highest COPD mortalities in the USA, and therefore the highest proportion of amenable deaths in COPD. These changes are also reflected in the life expectancy of individuals with COPD. California has the highest life expectancy for individuals with COPD. In 2016, 50-year-olds with COPD in California were expected to live for an additional 21.01 years, while in Texas they had an additional 17.82 years to live and in Kentucky, only 11.94 years. The normative life table approach adds to current efforts by providing a fair way of measuring health care performance. It acts as an indicator of health care quality by measuring the share of amenable deaths that, with timely and effective medical interventions and public health efforts, could have been avoided.

Metadata

Date	2023-05-17	
Authors	Falk, Yesser	
Contributors	Supervisor: Mueller, Prof. Dr. Dr. Ulrich	
License	Copyright	
	show more	

Files

Name	Format	Size	Checksum (MD5)
ReadME.txt	.txt	3.463Kb	c576947bdfd641962e729b3be409b39c

Publication of research data:

- Clear regulation of use through licensing
- Documentation as detailed as possible: support through consulting

Reusability in data_UMR
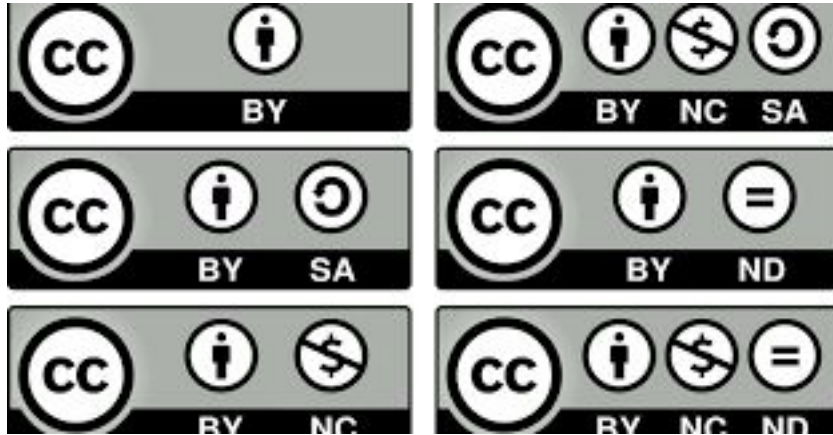
Publication of research data:

- Clear regulation of use through licensing
- Documentation as detailed as possible: support through consulting

Example data_UMR