HeFDI Data Talk

| Date | Торіс | Presenter | L 51 . |
|-----------------|--|--------------------------------------|--------|
| 26 January 2024 | DFG requirements on research data management | Dr. Ortrun Brand, HeFDI Coordination | HeFD |

Abstract:

When you write a DFG proposal today, you have to comment on research data management. The DFG now expects you to describe and plan how you will handle research data - tailored to your specific project. The management of data can also be in focus, especially during the review and inspection process.

But what needs to be taken into account? What support is available? What questions need to be answered regarding data management - and who can I ask for support?

In this talk, Dr. Ortrun Brand, coordinator of the state initiative HeFDI - Hessian Research Data Infrastructures, presents the key points of the DFG specifications, necessary considerations and practice examples as well as further support opportunities.

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).

DOI-Link: <u>https://doi.org/10.5281/zenodo.10572083</u>; License information: Creative Commons Attribution 4.0 International (<u>CC BY 4.0</u>)





DFG requirements on research data management HeFDI Data Talk 26th January 2024

Dr. Ortrun Brand, HeFDI Coordination

Based on a workshop from Dr. Nina Dworschak, Goethe-Universität Frankfurt/HeFDI

Similar Talk in German from December 2022: https://doi.org/10.5281/zenodo.6567185

All published HeFDI Data Talk materials in our HeFDI-Community on Zenodo: <u>https://zenodo.org/communities/hefdi</u>



Agenda

- 1. Preliminary remarks
- 2. Your framework: DFG guidelines & checklist
- 3. Your DFG proposal for RDM what do you need to consider?
 - 1. Data description
 - 2. Documentation & data quality
 - 3. Storage and technical backup
 - 4. Legal aspects
 - 5. Data exchange/accessibility
 - 6. Responsibility and resources
- 4. Bad & good practice
- 5. Special case of Collaborative Research Centers INF projects
- 6. Wrap-Up: Assistance and support services



1. Preliminary remarks

- DFG only! Not all research funding proposals/third-party funding formats
- However, the DFG sets the pace and many other funding programs are oriented to it, e.g. LOEWE, BMBF
- Subject-specific considerable differences as to what the status quo is in each case, what is considered "professionally appropriate" -> specialist societies, review boards, NFDI consortia (see nfdi.de)



- New DFG Code of Conduct from July 2019
- Legally binding implementation at all universities was necessary to maintain eligibility for funding, by mid-2023
- Includes new guidelines for dealing with research
- 19 guidelines, eleven of which relate to the research process
- Research data management relevant in eight of eleven guidelines on the research process



Guidelines for Safeguarding Good Research Practice

Code of Conduct





Guidelines with RDM relevance

Guideline 7: Quality across all phases

Guideline 10: Legal and ethical framework conditions, utilization

Guideline 11: Methods and status

Guideline 12: Documentation

Guideline 13: Providing public access to research results

Guideline 14: Authorship

Guideline 15: Publication medium

Guideline 17: Archiving

Guideline 12:

 Generally document properly, including research data and software

Guideline 13:

- Publish the research data on which the publications are based (whenever possible)
- Compliance with the FAIR principles
- Accessibility via recognized archives and repositories
- Software publication

Guideline 17:

- Those research data on which published results are based must be stored in an accessible and traceable manner for a period of at least 10 years
- At the institution where they were generated or in accredited repositories



Data reuse

How do I find a suitable repository for my data?

- DFG recommends to contact a suitable research data repository as early as possible, already during the planning phase / when writing your proposal
 - Metadata and costs in your proposal can only be explained on this basis!
- Have a look at 'your' nfdi-consortium
- or contact your local rdm service center
- or see re3data.org





What is a repository?

 A repository is a storage location for digital objects that makes them available to a public or limited circle of users

Repositories can be **distinguished** according to

- nature of stored digital objects (publications or research data)
- discipline of data (institutional, subjectspecific, generic)
- storage period of data (e.g. 10 years according to good scientific practice or long term >10 yrs

DEUTSCH LOGIN Service Center E-Research data_UMR | Research data repository Image: Submit Discover Browse About

data_UMR

Submit

If a collection fitting in with your submission's research area exists already, or if you would like make a submission into the Generic Collection ("Allgemeine Sammlung"), then you can <u>start your submission</u> immediately. If there exists no suitable collection yet, please proceed by <u>creating a new collection 才</u>). (fyou would like to learn more about the submission procedure, please refer to the user manual (PDF, German).

 \rightarrow

Search

, Enter a search term

Browse

 $\begin{array}{l} \text{Submit Date} \rightarrow \\ \hline \\ \text{Subject} \rightarrow \\ \hline \\ \text{Faculties} \rightarrow \\ \hline \\ \text{DFG-Classification} \rightarrow \\ \hline \\ \hline \\ \text{Collections} \rightarrow \end{array}$

Recently Added

| Identif | ring and Counting Avian Blood Cells in Whole Slide Images via Deep Learning $ ightarrow$ $_{\rm rr,Markus}$ |
|------------|---|
| A virtu | l library of small molecules mimicking dipeptides $ ightarrow$ |
| Kolb, Pete | |
| Multisp | ectral analysis-ready satellite data for three East African mountain ecosystems $ ightarrow$ |
| | |

Three types of repositories:

- Generic
 - See e.g. https://zenodo.org/
- Subject-specific
 - See https://www.re3data.org/
 - See NFDI consortia
- Institutional
 - See e.g. the HeFDI repositories based on DSpace <u>https://t1p.de/hefdi-repos</u>

| HeFDI - Hessian R | esearch Data Infrastructures |
|--|---|
| 6 HeFDI Federal State | Initiative • HeFDI Data Events HeFDI Information Materials & Data FAQ • HeFDI Data Services • |
| ome > HeFDI > HeFDI Data Serv | ices > HeFDI Data Repositories |
| HeFDI Data Services | Repositories |
| HeFDI Data Repositories | Since 2019, institutional research data repositories have been offered at three of the participating uni |
| HeFDI Data Management Plans: RDMO | operating the repositories, the HeFDI partners rely on a division of labor: Philipps-Universität Marbur and Technische Universität Darmstadt (TUDa) host repositories for other Hessian universities and en technical operation and 3rd-level support. Justus Liebig University is also establishing an institution |
| Sync & Share | repository solution. |
| HeFDI Data Versioning - GitLab Services | Through cooperation in setting up and operating the repository, maximum interoperability is aimed thus as isolated solutions within Hesse are avoided. The metadata used in the repository is based on devolved and used in the CDU within the set of each set of the CDU and the CDU and the Set ODU. |
| ELN Finder | developed and used in HeFDI, which is based on recommendations from DataCite (Institute for DOI R and DINI (German Initiative for Network Information e.V.). Adaptations and further developments of t repositories are jointly discussed, processed and reused in HeFDI. This creates considerable synergie |
| | savings. The technical basis is the open source repository software DSpace. It is one of the most widely used i |
| | software worldwide and is maintained and further developed by a large, international community. |
| | Through this common repository solution, all HeFDI universities have the option to promptly impler |
| | solution for the storage and, if necessary, publication of research data at their own institution. They easily comply with the DFG's new guidelines (PDF). |
| | easily comply with the <u>brossnew Budelines</u> (rbr). |
| | Your local repository |
| | Repository of Frankfurt University of Applied Sciences |
| | Repository of Goethe University Frankfurt |
| | Repository of Hochschule Darmstadt |
| | Repository of Hochschule Fulda |
| | Repository of <u>Hochschule Geisenheim</u> Repository of Hochschule Rhein-Main |
| | Repository of Justus-Liebig-Universität Gießen: JLUdata |
| | Repository of Philipps-Universität Marburg: data_UMR |
| | Repository of Technische Universität Darmstadt: TUdatalib |
| | Repository of Universität Kassel: DaKS |

The repositories of the Goethe University Frankfurt and the Technische Hochschule Mittelhessen are still being established and will be listed here as soon as they are available.





mandatory in all proposals! Contact your local RDM service center!

Checkliste zum Umgang mit Forschungsdaten

1. Datenbeschreibung

Auf welche Weise entstehen in Ihrem Projekt neue Daten? Werden existierende Daten wiederverwendet? Welche Datentypen, im Sinne von Datenformaten (z. B. Bilddaten, Textdaten oder Messdaten) entstehen in Ihrem Projekt und auf welche Weise werden sie weiterverarbeitet? In welchem Umfang fallen diese an bzw. welches Datenvolumen ist zu erwarten?

2. Dokumentation und Datenqualität

Welche Ansätze werden verfolgt, um die Daten nachvollziehbar zu beschreiben (z. B. Nutzung vorhandener Metadaten- bzw. Dokumentationsstandards oder Ontologien)? Welche Maßnahmen werden getroffen, um eine hohe Qualität der Daten zu gewährleisten? Sind Qualitätskontrollen vorgesehen und wenn ja, auf welche Weise? Welche digitalen Methoden und Werkzeuge (z. B. Software) sind zur Nutzung der Daten erforderlich?

3. Speicherung und technische Sicherung während des Projektverlaufs Auf welche Weise werden die Daten während der Projektlaufzeit gespeichert und gesichert? Wie wird die Sicherheit sensibler Daten während der Projektlaufzeit gewährleistet (Zugriffs- und Nutzungsverwaltung)?

4. Rechtliche Verpflichtungen und Rahmenbedingungen

Welche rechtlichen Besonderheiten bestehen im Zusammenhang mit dem Umgang mit Forschungsdaten in Ihrem Projekt? Sind Auswirkungen oder Einschränkungen in Bezug auf die spätere Veröffentlichung bzw. Zugänglichkeit zu erwarten? Auf welche Weise werden nutzungs- und urheberrechtliche Aspekte sowie Eigentumsfragen berücksichtigt? Existieren wichtige wissenschaftliche Kodizes bzw. fachliche Normen, die Berücksichtigung finden sollten?

5. Datenaustausch und dauerhafte Zugänglichkeit der Daten

Welche Daten bieten sich für die Nachnutzung in anderen Kontexten besonders an? Nach welchen Kriterien werden Forschungsdaten ausgewählt, um diese für die Nachnutzung durch andere zur Verfügung zu stellen? Planen Sie die Archivierung Ihrer Daten in einer geeigneten Infrastruktur? Falls ja, wie und wo? Gibt es Sperrfristen? Wann sind die Forschungsdaten für Dritte nutzbar?

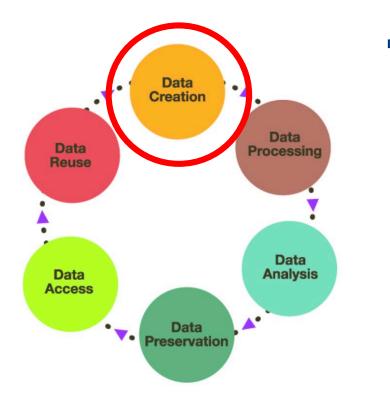
Verantwortlichkeiten und Ressourcen

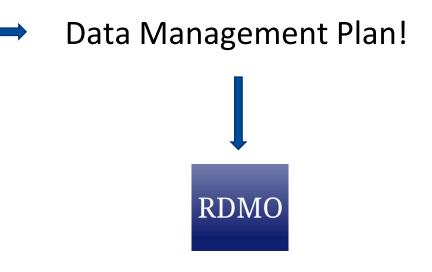
Wer ist verantwortlich für den adäquaten Umgang mit den Forschungsdaten (Beschreibung der Rollen und Verantwortlichkeiten innerhalb des Projekts)? Welche Ressourcen (Kosten; Zeit oder anderes) sind erforderlich, um einen adäquaten Umgang mit Forschungsdaten im Projekt umzusetzen? Wer ist nach Ende der Laufzeit des Projekts für das Kuratieren der Daten verantwortlich?

Quelle: DFG, <u>https://t1p.de/50h4</u>, Stand 19.05.2022







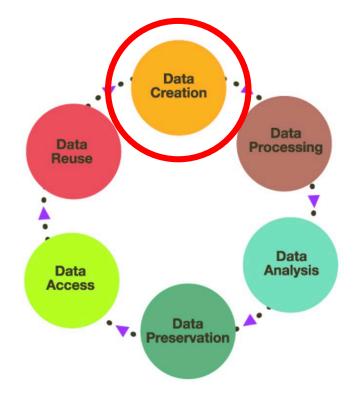


Research Data Management Organiser <u>https://t1p.de/hefdi-rdmo</u>



Data Description

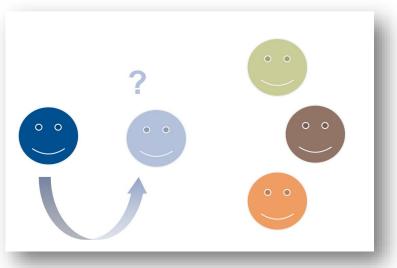
- How does your project generate new data?
- Is existing data reused?
- Which data types (here: FORMATS), e.g. image data, text data, measurement data, are created in your project and how are they processed?
- To what extent is this data generated and what volume is to be expected?





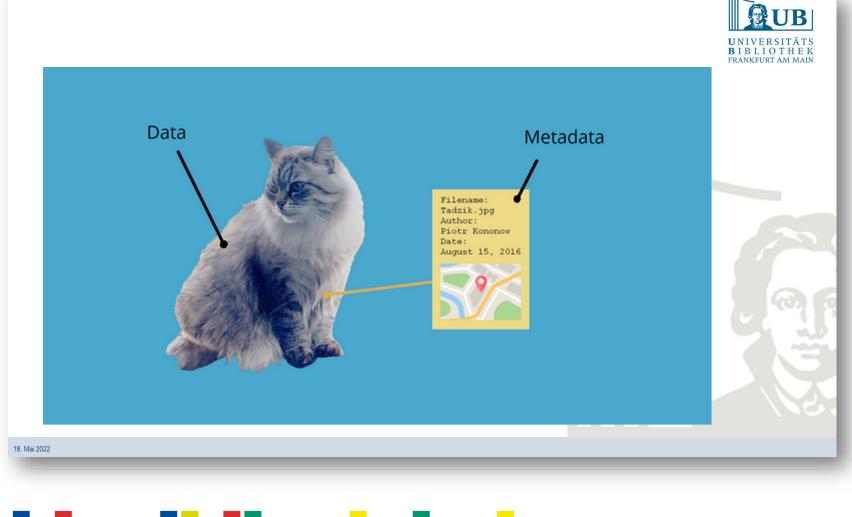
Documentation and data description

- What approaches are taken to describe the data in a comprehensible manner (e.g. use of existing metadata and documentation standards)?
- What measures are taken to increase data quality?
- Are quality controls planned, and if so, which ones?
- What digital methods and tools are required to use the data?



Data documentation:

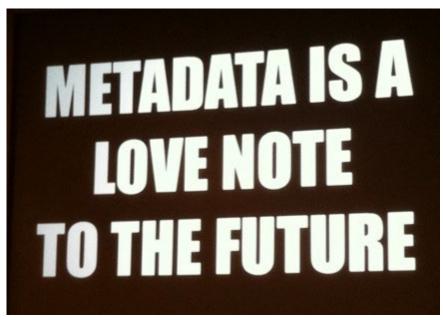
- Who can understand my data?
- Who can open and re-use my data?
 - today?
 - in 6 months?
 - in 10 yrs?





Data description – metadata

- Metadata help to understand research data
- Structured description of research data
- machine-readible —> research data become Findable in data bases
- Without metadata, research data won't be understandable
- Rich and correct metadata are a strong provision to good scientific practice

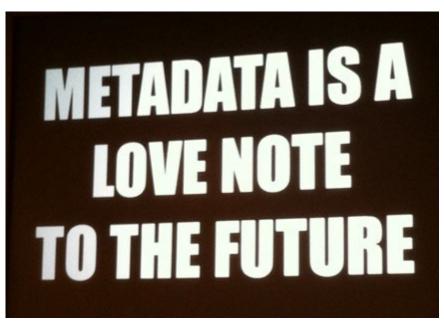


Source: cea + from The Netherlands, CC BY 2.0, Wikimedia Commons, https://upload.wikimedia.org/wikipedia/commons/d/df/Metadata_is_a_ love_note_to_the_future_%288071729256%29_%28cropped%29.jpg



How do I describe my research data properly with metadata?

- Who (Co-Authors)?
- Where?
- When?
- Why?
- With which tools?
- Which parameters?
- Usage options? (Licenses, DOIs, ...)
- What? (scientific object)



Source: cea + from The Netherlands, CC BY 2.0, Wikimedia Commons, https://upload.wikimedia.org/wikipedia/commons/d/df/Metadata_is_a_ love_note_to_the_future_%288071729256%29_%28cropped%29.jpg



What metadata standard should I use?

| Discipline | Standard | |
|---|---|--|
| Cross-disciplinary | Dublin Core, MARC21 | |
| Natural sciences | ICAT Schema | |
| Geosciences/Geo Data | ISO 19115, Darwin Core, FGDC | |
| Arts, Cultural Sciences | CDWA, LIDO, CIDOC-Framework | |
| Climate, wheather data | Conventions for Climate and Forecast Metadata | |
| Social Sciences, Humanities | Data Documentation Initiative (DDI) | |
| Text data (commenting, annotation) | PS Guideline TEI | |
| Visualization, Exchange of data on neutrons, xrays, myons | NeXus | |

- See https://forschungsdaten.info/themen/beschreiben-und-dokumentieren/metadaten-und-metadatenstandards/
- contact your NFDI Consortium
- look at https://fairsharing.org/search?fairsharingRegistry=Standard

Metadata Annotation Tools?

FAIR Data Assessment Tools

See https://www.dcc.ac.uk/guidance/standards/metadata/tools See <u>https://forschungsdaten-thueringen.de/entry/fair-assessment-tools-en.html</u>

| DCC ecause good rese | arch ne | North Spord data | | | |
|---------------------------|---------|--|--|--|--|
| tuod | ~ | Hore - Guidesce - Metadata - Dacobrasy Metadata | | | |
| eva | ÷ | List of Metadata Tools | | | |
| rents | ÷ | | | | |
| vices | ÷ | AgriMetamaker | | | |
| dance | ^ | A service to facilitate the publication of metabolis in the XRIBS Reporting: It continues to the the XRIBS Application Photies, which draws from the Dublin Core and ApMES standards. | | | |
| fing Papers | | ANZ-MEST - Metadata Entry and Search Tool A cestive who exploted to traditional management on largeing with priviles available for two selections or Marine Community Profile. | | | |
| to Guides | | AVM Adobe Metadata Panels | | | |
| Studies | | A set of metidade a Patients A set of metidate bank text as be added to Added Creative Suite 4 applications to allow WAR-compliant metidate to be entered directly into impose. | | | |
| y Analysis | ~ | Second and the | | | |
| data Jplinary adata | î | AVM Web Tool A web-based tool for missentialing an RM4 computer XMP projects for insertion into an image file. | | | |
| dion Lifecycle el | | Bio-Formats Bio-Formats rests providery microscopy image data and metadata, and converts them to CME/TPF, a combination of TIVF and CME/XML. | | | |
| Management 6 | | CF Compliance Checker | | | |
| arch | | A stilly that checks netCDP files for CP-compliance. | | | |
| arch | Ŭ. | CIE2Cell | | | |
| cations | ~ | A tool to generate the geometrical assay for various electronic structure codes from a CPF Bis. | | | |
| | ~ | CIM Tools This for where an utilitating CM metacate encoded in XM, format, | | | |
| metion for | × | CKAN Tor which utilizes the CCXF standards, CXXH is a gowerful data management system that makes data accessible – by providing toxis to streamine guidebing, sharing, finding and using data. | | | |
| | | to the spectra of u.v. reasons U.V.A. a powerz learning power to management power to management power on a spectral of provide power power, some, nong and u.V.A. a powerz learning power to spectra of the politike power of the politike power and the power of the politike power and the power of the politike power and the politike politike power and the politike | | | |
| | | CMOR - Climate Model Output Rewriter A set of Clease Inclines, with bindings is both types and FORTIMI K0, that can be used to produce Of-compliant netCDF files. | | | |
| e sake estings | | 1 2 3 4 5 6 7 Ngch | | | |



AN OVERVIEW OF FAIR ASSESSMENT TOOLS!

This overview presents a selection of FAR assessment tools. The categories also use the arrowyn FAR but stand for: Fully Configurable Tools, Automatic Tools, Improved Survey Tools and Regular List Tools. The classification of tools into these categories is not always 100% clear, especially when certain properties overlap. In ergeparation for the poster, an evaluation of all known FAR Resessment tools was carried out between july and August 2022. Each tool was evaluated by three independent people. The categories used included the time taken to carry out the FAR assessment, the presence of feedback and the required prior knowledge. With regard to the technical specifics, the research revealed that only generic tools are currently on offer.

The poster was created for the "3. Sächsische FDM-Tagung".

Following tools weren't evaluated:

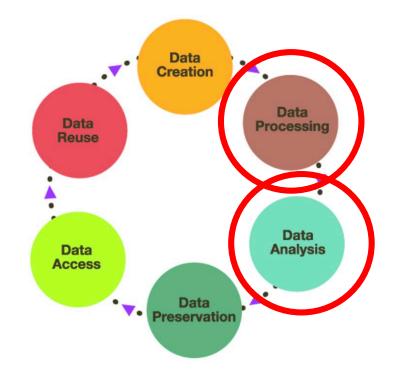
» FAIRdat (unfinished, beta version of SATISFYD?)
 » FAIR-Aware (awarenness, not evaluation questions for dataset)
 » 5-Star Data Rating Tool (inaccessible)

On Zenodo lies the poster and the data table in different formats: https://doi.org/10.5281/zenodo.7022037

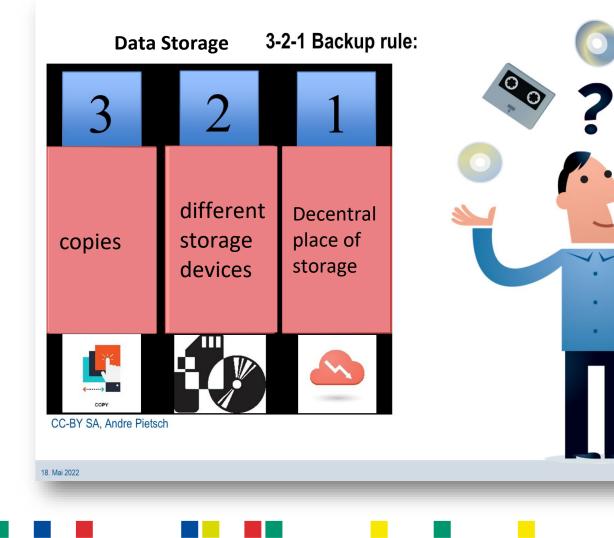


Storage during the project phase

- How is your data stored, and where?
 - use your public infrastructure!
- infomation security / data protection for sensitive data (rights of access)



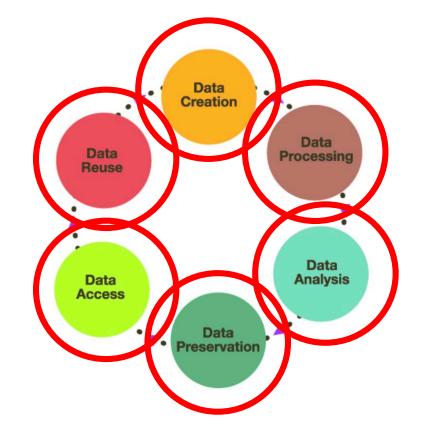
www.digitalbevaring.dk



Legal aspects

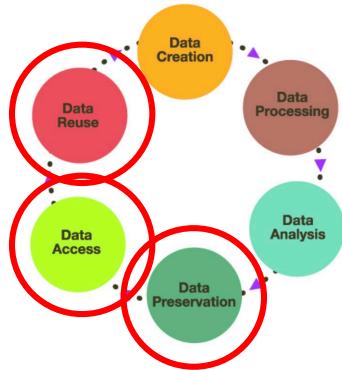
- What are the legal conditions with regard to your research data?
- Are there any restrictions concerning re-use and accessibility?
- data & software licensing / copyright, rights of use

see HeFDI legal aspects on research data https://doi.org/10.5281/zenodo.4625417



Data exchange and permanent accessibility

- Which data is **particularly suitable for reuse** in other contexts/by third parties?
- What **criteria** are used to **select** research data in order to make it available for re-use by others?
- Do you plan to archive your data in a suitable infrastructure? If so, how and where? Are embargo periods in place?





Data formats suitable for long term archiving



| Data type | recommended formats | less or non suitable formats |
|-----------------------------|---|---------------------------------------|
| Audio | . <u>wav</u> / . <u>flac</u> | .mp3 |
| Computer-aided Design (CAD) | . <u>dwg</u> / . <u>dxf</u> / . <u>x3d</u> / . <u>x3db</u> / . <u>x</u> | <u>3dv</u> - |
| Datenbanken | . <u>sql</u> / . <u>xml</u> | .accdb / .mdb |
| Rastergrafiken & Bilder | . <u>tif</u> (unkomprimiert) / .jp2 / .j | <u>og2</u> .gif / .jpeg / .jpg / .psd |
| | / . <u>png</u> | |
| Statistische Daten | . <u>por</u> / <u>.csv</u> | .sav (SPSS) |
| Tabellen | . <u>csv</u> / . <u>tsv</u> / . <u>tab</u> | .xls / .xlsx / .xlx |
| Texte | . <u>odf</u> / . <u>rtf</u> / . <u>txt</u> / PDF/A | .docx / .doc / PDF |
| Vektorgrafiken | . <u>svg</u> / . <u>svgz</u> | .cdr |
| Video | . <u>mp4</u> / . <u>mkv</u> / . <u>mj2</u> / | .mov / .wmv |
| | . <u>avi</u> (unkomprimiert) | |

18. Mai 2022

See HeFDI-recommendations on data formats in <u>https://gitlab.ulb.tu-darmstadt.de/hefdi/hefdi-datenformate</u>, based on <u>https://documentation.library.ethz.ch/display/DD/Archivtaugliche+Dateiformate</u>

NFDI – National Research Data Infrastructure

- The NFDI systematically develops structures for sustainably secure and make accessible research data
- driven by the scientific community, in strong cooperation with infrastructures
- 26 consortia, disciplin-specific / data specific
- Previous isolated solutions and parallel developments are to become integrated – Research Data Commons...
- Systematically connect with European Open Science Cloud (EOSC)



DFG currently formulates more clearly than ever before that NO singular development is desired, but INTEGRATION in / connection to existing infrastructures

- turn to local service points, local/regional infrastructure, NFDI consortium; recognized data centres/repositories, international specialist/offers
- use existing databases, repositories, tools that are permanently available

Information infrastructures for Open Science Horizontal and vertical integration

Ensure horizontal and vertical integration

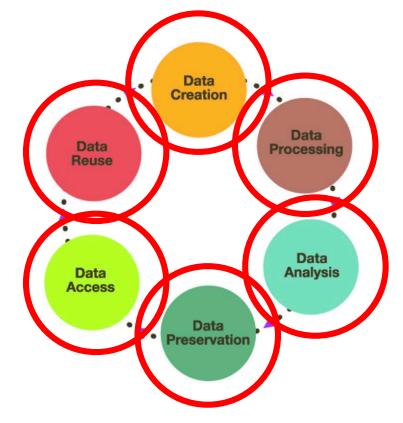
- Cross-connections between subject-specific data repositories, but also between repositories for data, publications, software or CRIS systems – horizontal integration
- As mirrored at DFG funding portfolio



Quelle: Screenshot from Winkler, Kathrin (DFG) (2022): "Funding Strategy in Germany: The Significance of Horizontal and Vertical Integration (Impulse lecture)" EOSC Tripartite Event Germany "NFDI & EOSC – connecting the communities", 24.11.2022, https://cloud.nfdi.de/f/431deba6771b4f6e8b93/

Responsibility and resources

- Who is responsible for the adequate handling of research data (description of roles and responsibilities within the project)?
- What resources (costs, time or other) are required to implement adequate handling of the research data in the project?
- Who is responsible for curating the data after the end of the project?



What costs can researchers **NOT** claim from the DFG for data management?

Specifically, there is an expectation that local data backup and archiving of published results will be guaranteed for the purpose of recall in cases of misconduct.

This includes a modern IT structure that makes it possible to process even extensive data sets, as well as basic services for the local storage and documentation of research data. **Costs that serve these purposes are part of the basic equipment and cannot be approved**



See also https://t1p.de/dfg-rdm-funding



What costs can researchers claim from the DFG for data management?

-> gain access to research data

—> to process and prepare the research data generated in the project so that it can be used by others

--> costs to transfer the data to a public repository.

The following can be applied for:

- personnel costs for processing the data,
- user fees, membership fees
- or costs incurred when **using established infrastructures** (other costs).



Source: kschneider2991, CCO, via Wikimedia Commons, https://upload.wikimedia.org/wikipedia/commons/3/37/Money-2180330_1920.jpg

Important!

- The prerequisite for the financing of user fees is the existence of <u>a publicly accessible, transparent cost-performance catalog</u> of the repository.
- The purpose of use and the distinction from basic costs must be clear from the information in the application...

Bad practice

- Data management = methods
- Backup to local hard disks
- Sharing = passing on via Dropbox or similar.
- No backup in the infrastructure of the university / institution; no explanation of what this is based on (existing offers / service points / contact)
- No reference to NFDI, federal state initiatives, existing RDM-networks...
- If reference to personal data -> no mentioning of data protection
- Equate open access publication with and data management

... good practice

- Describe data, name metadata/schemas, ontologies, controlled vocabularie
- Head for a Data management plan
- Awareness that RDM is about transparency, replicability and reuse of data collected with public funds (FAIR)
- Aim for data publication, name repo contact in advance!
- Data literacy / training

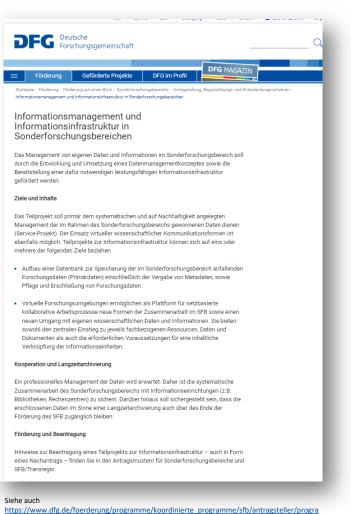
 \checkmark

...

Reference to NFDI, to state initiatives

4. Collaborative Research Center – INF-Project

- Own sub-project,
- equipment often 1 data manager/steward
- Works for other projects often Zproject
- Data management in the CRC is to prepare data for subsequent use, internal interoperability
- Project-specific
- Close cooperation with local infrastructure required
- No hardware this is basic equipment
- No IT specialist for the SFB



mmelement inf/index.html

5. Wrap Up and Support

| What | URL | |
|--|---|--|
| DFG-Website on RDM | https://t1p.de/dfg-rdm | |
| Guidelines of professional associations | https://t1p.de/dfg-rdm-specific | |
| HeFDI | <u>www.hefdi.de</u> | |
| local Service Centers on RDM | https://t1p.de/hefdi-local | |
| HeFDI Data Services (Repositories, RDMO, Git,) | https://t1p.de/hefdi-data | |
| HeFDI FAQ | https://t1p.de/hefdi-faq | |
| Disciplin-specific information on forschungsdaten.info | https://www.forschungsdaten.info/ wissenschaftsbereiche/ | |





www.hefdi.de hefdi@uni-marburg.de

