Data Management Guidelines:
How To Handle Data

January 10, 2023

Norbert Riefler

Preface

‘On July 20, 1969, Neil Armstrong climbed out of his spacecraft and placed his feet on the moon. The landing was broadcast live all over the world and was a significant event in both scientific and human History. Today, we can still watch the grainy video of the moon landing but what we cannot do is watch the original, higher quality footage or examine some of the data from this mission. This is because much of the data from early space exploration is lost forever. (...) The tapes were likely wiped and reused for data storage sometime in the 1970s.’

- Kristin Briney[3]

This reader holds for all employees of the Leibniz-Institute for Materials Engineering (IWT) and serves as a reference for how to save data. It delivers reasons for the question of why one should think about data management and gives instructions and examples. Data management in the information age is an absolutely required skill and a base to cope with the wealth of data to avoid an information overload. It helps organizing everyone's own data, but it is also mandatory for data sharing so that others can understand the structure and are able to interpret the information given by your data. In this way, your data meet the FAIR standard (Findable, Accessible, Interoperable and Reusable), basis for the looming new discipline of data science.

There are many topics in data management like data governance and security or storage management, but we are treating here only a condensed extract, tailored to data handling in our department. Therefore, this reader might help you to clarify what must be done to be a data steward.

Contents

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

1   Why Data Management

Save time
Planning your data management needs will save you time and resources, e.g., by faster locating data. And well managed data requires less preparation for sharing or writing a paper.

Preserve your data
Depositing your data in a file server (repository) safeguards your investment of time and resources while preserving your research contribution for you and others to use.

Increase your research impact
Making your data available to other researchers can impact discovery and relevance of your research.

Maintain data integrity
Managing and documenting your data throughout its life cycle will allow you and others to understand and use your data in the future.

Meet grant requirements
Many funding bodies now require that researchers deposit data collected as part of a research project.

Promote new discoveries
Sharing your data with other researchers can lead to new and unanticipated discoveries and provide research material for those with little or no funding.

Support open access
Be a catalyst for research and discovery. Show your support for open access by sharing your data.

2   Data Management Plans

The handling of the data, created during a project, contains several issues which are described in a Data Management Plan (DMP). A DMP contains structured information about the research process of the corresponding project. But these information has to be given long before a project is started: Applicants have to create a DMP in every proposal, required by all important research funding organizations (DFG, BMBF, etc.). So in the process of writing, applicants start to think about how they will treat data. The following topics belongs to typical DMPs and might be of interest for you; more details can be found in [1, 4]:

1. Administrative information: project name, kind of funding, time period.

2. Methods of data generation: simulations, experiments (devices), data size, kind of data documentation.

3. Data safety: location, interval and capacity of storage device, who gets access.

4. Archiving: which data are archived where with what kinds of metadata.

5. Data sharing: which repository, license condition, required metadata.

6. Resources and responsibility: who is responsible for processes, IT, defining defaults and formats, monitoring; required personal resources; costs.

Please see the Appendix for more details; some exemplary DMPs can be found here: `https://www.cms.hu-berlin.de/de/dl/dataman/arbeiten/dmp_erstellen`

3 File Organization

The following explains what and how to save your files on our servers:

- Access to the VT-Server is established with your Web browser by typing `http://134.102.40.154` or `http://vt-nas.mvt-bremen.de` in the address field, via mapping the VT-Server to the Windows Explorer as an own drive, or by using the Synology Drive Client (see [5]) with some more and very helpful options like, e.g., document versioning. The **maximum size** of uploaded data should not exceed **100Gbyte per file**. Access privilege is administrated by Stefan Endres (s.endres@iwt.uni-bremen.de), Arvind Chouhan (a.chouhan@iwt.uni-bremen.de) and Nils Ellendt (ellendt@iwt.uni-bremen.de).

- The server of the manufacturing technology (FT) division is accessed via **\\fts-ags.iwt.uni-bremen.de** or **\\fts-ags** or **\\134.102.244.1** in the windows explorer. Access privilege is administrated by the IWT IT division.

3.1 Different Data Types

Depending on the origin of the data, we distinguish between two data types:

- Primary data (sometimes called raw data) is generated by measurement devices, sensors/cameras, simulation programs, observations etc. and are untreated; it is collected for the first time by the researcher.

- Secondary (or analyzed) data are processed primary data. (In social science, secondary data are already collected or produced data by others).

Very large primary data cannot be saved on our Electronic Lab Notebook (ELN)[1], but the origin of that data are discussed there together with information about the storage location (e.g. somewhere on a file server). The decision about the limit, which data is considered as large, is:

- data $<$ 100 MByte is stored and documented on the ELN

---

[1]The use of the ELN is currently only mandatory for the VT.

- 100 MByte $<$ data $<$ 1TByte is stored on the VT-Server and documented on the ELN

- data $>$ 1 TByte is stored on USB-Disks and documented on the ELN

It is absolutely required that **all data from external sources** (e.g. by an USB stick) have to be checked for viruses manually, if the auto scan doesn't come up, before they are used/stored on hard disk.

3.2   File and Folder Naming

Spend time planning out file naming conventions in the beginning of a project. How do you or others will look for and access files at a later date? Do you think about them by type, location, study or something else?

Naming conventions:

- File names are self-explanatory - they tell what data is in!

- Do not use spaces or special characters - use CamelCase and hyphen '-' or underscore '_' as separator:
  NO   'name date v1.txt'
  YES 'name_date_v01.txt'
  YES 'nameDateV01.txt'
  YES 'simpleFoam_u0.2m-s_etha2mPas_v01'

- Develop a file naming scheme that includes information about the data. Example: [Date]_[Run]_[SampleType]

- Consider sorting when deciding what element of the file name will go first:
  Use 'YYYY-MM-DD' to save dates.
  Use leading zeros:
  NO   'ProjID_v1.csv' ... 'ProjID_v11.csv'
  YES 'ProjID_v01.csv' ... 'ProjID_v11.csv'   (sequence up to 99)
  YES 'ProjID_v001.csv' ... 'ProjID_v111.csv' (sequence up to 999)

3.3   File Formats

As technologies changes, researchers should plan for both hardware and software obsolescence and consider the longevity of their file format choices to ensure long term readability and access.
File formats more likely to be accessible in the future have the following characteristics:

- Non-proprietary

- Open, documented standard

- Common usage by research community

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

- Standard representation (ASCII, Unicode)

- Unencrypted

- Uncompressed (if not too large...)

Examples of **preferred** file format choices include:

- Open Document Files: .odt (Text), .ods (Spreadsheet), .odp (Presentation); primary format of OpenOffice/Libre Office, standardized in ISO 26300

- Office Open XML: .docx (Text), .xlsx (Spreadsheet), .pptx (Presentation); primary format of Microsoft Office, standardized in ISO 29500

- Ritch Text Format: .rtf; proprietary, but well documented format

- Plain Text or Markup Language: .txt, .md (Markdown), .html (Hypertext), .tex (LaTeX), .csv (comma separated values)

- Portable Document Format: .pdf; standardized in ISO 32000

- Bibliographic Data: .enl (EndNote; this is the standard software for literature management at the IWT); otherwise, please use human readable formats such as .bib (BibTex), .ris (Research Information System Format)

- Video files in data formats .mp4 and .mpg generated by codecs H.264 and H.265; standardized format (ISO 14496-10) for video compression; **DO NOT USE proprietary codecs such as Quicktime (.mov, .qt)**

- Images: .jpg is the standard file format of most cameras, it is readable by almost any image-related tools, but is not ideal in all cases (e.g. line art). It offers no transparency. Moreover, it is a lossy format which can lead to a permanent degradation of the image, according to the chosen compression. Thus it should be avoided if this is not tolerable. The succession format JPEG2000 .jp2 allows lossless compression and thus may be used for image archiving. Due to licensing issues it is not widely used. The .tiff (Tagged Image File Format) is an important format for file exchange with publishers. It allows lossless compression and high depth (32 bit) as well as CMYK color. It may be used for image archiving, however, the file size remains comparably large. The 'Portable Network Graphic' format .png is a raster image format with good lossless compression. It yields larger file sizes than .jpg for photos but is very well suited for line art.

- Drawings: It is preferable to use vector based file formats for sketches and drawings, as these allow for an unlimited scaling of the image without affecting the quality. For large printouts, vector graphics are significantly smaller than raster graphics. However, vector-based images are not handled well by all word processors and even if they are usable, may yield varying results on different computers. It is recommended to prepare a vector

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

graphic in an open format (.svg, scalable vector graphics) to ensure long-term readability and it export it to the file format best suited for your word processing software (e.g. .emf for Word, .eps or .pdf for LaTeX).

- Binary data: sometimes, it is most efficient to save data in a binary format, as text-based representations would dramatically increase the file size. In this case, it is vital to document how the data is stored and preferably give an example of how it can be read again.

If you consider exporting your (measurement) data into a format with the above characteristics, keep a copy in the original software format unless you can assure that all data and metadata is correctly converted. If you deposit your data in a repository, your files may be migrated to newer formats, so that they're usable to future researchers.

3.4 Folder Hierarchy

The specific value of data has a wide variance. Many measurements like those from preliminary experiments were very important to design your experiments, but they are more or less worthless in respect of publications. In contrast, some of your data are highly valuable and find their way into a publication. Due to these differences, the location where you save your data has different valence, according to these three data kinds:

I. Data with a low valence, see: 3.4.1 Daily Data

II. Data which are going to be published, see: 3.4.2 Data for Publication

III. Data of accepted publications: These data are stored on a special area within your file-server where you can only write your data one time – you cannot change or delete anything there. It is like a mail box: Once you have thrown in your tax declaration, it is over. This data area is for accepted publications only where really everything is all right!

3.4.1 Daily Data

Every day you generate data, for instance by measuring, simulations, preparing a paper etc. The VT-Server serves as a file storage for your data, and all these daily data are saved under `Mitarbeiter` and your `<family name>`:
`/VT-Server/<department>/Mitarbeiter/<family name>/`
The names enclosed by `<department>` have to be chosen from our departments:

- MPS (MehrPhasenStrömung)

- RST (Reaktive Sprüh-Technik)

- SPK (SPrühKompaktieren)

according to your affiliation. An example directory is:
`/VT-Server/MPS/Mitarbeiter/riefler/`

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

Here, you save the following data:

→ **For Experimental Data:**

  – All experiments of the project

  – Relation between these experiments (which parameters are varied) in a JSON based text file, see subsection 4.2

  – intermediate results

  – List of all used devices with adjustment parameters

→ **Experimental Setups:**

  – Drawings of all parts

  – A list of all required parts and devices

→ **For Simulation Data:**

  – The required initial values and mesh information to recalculate all simulations of the project

  – Relation between these simulations (which parameters are varied) in a JSON based text file, see subsection 4.2

  – Intermediate results

  – List of the used software together with used source code

  – All required input files to reproduce simulations

There is no prespecified directory structure for your daily data, so that's up to you. However, these data under your name are distinct from project data like, e.g., important experimental findings or simulations which are published, which are described in the following.

3.4.2 Data for Publication

Here, you save all data relevant for every publication within this project. Depending on the size of the data, and the decision within your department (see above), primary data files larger than, say, 10 GByte must not be saved within the directory specified here. Instead, a reference has to be given within the documentation in the experiment description on the ELN about the corresponding origin (e.g. high-speed camera measurement or CFD simulation) and where exactly the primary data are stored. The decisive **basic principle** for all your raw and primary data is: **All data must be understandably documented in its origin!**

When you prepare a document, say a presentation or a project report, the data are saved (according to a strict scheme given later) under this directory:

```
/VT-Server/<department>/Projekte/<funding>/<project name>
```

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

$\rightarrow$ **Resulting Documents:**

- Documents from your scientific work: Papers, Dissertations, Presentations, Reports, Proposals, etc.

- All results (figures, tables, movies, animations, ...) and the underlying measured or simulated data together with evaluation programs

Funding sources (`<funding>`) are for instance: DFG, AiF, BMBF, Industrie, ERC, etc. Last, the directory name for published works in a project, the `<project name>`, is compound on:
`<year of project start>_<department>_<project handle>_<family name>` See subsection A.1 for an example, and the two examples given here:

- `/VT-Server/MPS/Projekte/DFG/2016-MPS-Tropfengenerator_Riefler/...`

- `/VT-Server/RST/Projekte/DFG/2018_RST_Flatbandpotential_Naatz/...`

The doubling of 'MPS' and 'RST' comes due to an anticipated common file server which hosts data of all IWT departments.

Papers and Dissertations

In agreement with all IWT departments, the directory naming scheme for published data of the File-Server is based on departments, the funding source and then the corresponding project name (see above for `<project name>` definition):
Text and data of a paper are saved, e.g., in that directory:
`/VT-Server/<department>/Projekte/<funding>/<project name>/Papers/`
with a directory name like that:
`<year>_<department>_<paper-name>_<first author>`
Text and data of a dissertation are saved, e.g., in that directory:
`/VT-Server/<department>/Projekte/<funding>/<project name>/Dissertation/`
with a directory name like that:
`<year>_<department>_<dissertation_title>_<author>`
In these directories, you save all the data of your **paper** in subdirectories, whose names are given in the left row of table 1, with the corresponding files explained on the right.

In case of a **dissertation**, you save your presentation of the doctoral examination together with the manuscript under the directory `'01_Manuscript+Presentation'`. If you do not have supplementary information for your paper/dissertation, or you did not write specific computer programs for your research, the `'05_ProgramSources'`, `'07_Supplement'` and `'08_Misc'` directories must still be there but with no entries.

An important directory is the `'02_Figures'`. There, you save every figure in the manuscript in good quality, and you also save a *.csv file with the data in the figure, or you save a file which generates the corresponding figure – for instance a Matlab file – from the primary data in the `'04_RawData'` directory. This data directory includes the measured or simulated data in *.txt

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

Table 1: The data of every paper has to be saved in eight subdirectories; further remarks:

*When the paper is ready for press the corresponding author loads the last and revised version (i.e. only the real used but absolutely complete data) into the protected directory

**If the data reproduced in a figure/table is from distributed primary data directories, it is sufficient to save here only the generating program

`00_FinalPublication`*	• The very last version of your paper as a pdf with volume, year and page numbers on the VT-Server into `'VT-Publikationen'` • EndNote entry for IWT-WiKo including DOI number, and an End-Note *.ris ASCII export fulfilling the FAIR principles
`01_Manuscript`	• the *.pdf and *.docx or *.tex
`02_Figures`**	• one sub-directory for every figure (`figure_01`, ...) with data points as *.csv or *.txt, or the generating file/program for the figure points (Matlab, Python, Origin,...); the results are from data stored in or extracted from the `04_PrimaryData` directory
`03_Tables`**	• one sub-directory for every table (`table_01`, ...) with data stored in as a csv-file
`04_PrimaryData`	• measured or simulated data, each sub-directory with a 'readme.json' file describing the data • ELN pdf pages with QR-code (see below)
`05_ProgramSources`	• e.g. OpenFOAM-Solver, Python/Matlab code, etc. • case files (mesh) for OpenFOAM, Fluent, ABAQUS etc.
`06_References`	• the references *.bib or *.enl together with *.ris • Cited papers/books (if publication rights are permitted)
`07_Supplement`	• Supplementary files
`08_Misc`	• Reviewer comments / rebuttals

or *.csv format from the measuring device or the simulation software, and a 'readme.json' file generated by the Readme-File-Creator described below with information and metadata about your measured / simulated data. The relation of each data point in a figure to the primary data must be given either implicit, for instance in the Matlab file, or better explicit in the 'readme.json' file.

To complete your publication, every cited paper should be added to the directory `'06_References'` for internal use. All too often the ideas of a paper cannot be tracked and reproduced because there are references which you can't get because your institution do not offer access to that specific journal, or the reference are from a master thesis which is not available. Therefore, saving all references with suitable file names for a direct relation – together with all the previously mentioned data and information – presents your paper in its entirety. Above

that, EndNote allows a comfortable direct linking to the cited publications, i.e. you should add these links in your '*.enl' file. The same is possible within BibTeX. However, if your paper will be transferred on a public repository, only free publications can be uploaded because many publishing companies, like Springer and Elsevier, do not allow to provide papers from their journal or books.

Writing a paper is a multistep process. With the decision of the appropriate journal, the process comes to a crucial point where all the data have to be saved in the structure given above and in the following screenshot, and the review process starts. If the paper is accepted, all changes are integrated and the final pdf with volume, year and page numbers is released, then the last actions are:

- Check if every ELN-file which describes the experiment/simulation is saved.

- The very last step: Make a copy of the complete paper data into the directory `'Published'` on the file server. It's a special directory: You can copy something there, but afterwards you cannot change anything. It's like a post box: Once you have thrown something in, it is gone!

Posters and Presentations

Posters or presentations are saved under:
`/VT-Server/<department>/Projekte/<funding>/<project name>/Posters/`
`/VT-Server/<department>/Projekte/<funding>/<project name>/Presentations/`
The directory structure is similar to that of papers, which means that you have to save the used measured or simulated data as well. Some directories have of course differing names and different content like, e.g. `'01_Poster+Abstract'` which includes poster and abstract for the conference. In case you have held an oral presentation, the first two directories are named `'00_FinalPresentation'` and `'01_Presentation+Abstract'`.

Table 2: The data of every poster has to be saved in seven subdirectories; further remarks:

*When the poster was presented, the corresponding presenter loads the last and revised version (i.e. only the real used but absolutely complete data) into the protected directory

**If the data reproduced in a figure/table is from distributed raw data directories, it is sufficient to save here only the generating program

`00_FinalPoster`*	• The very last version of your poster as a pdf with volume, year and page numbers on the VT-Server into `'VT-Publikationen'` • EndNote entry for IWT-WiKo including DOI number, and an End-Note *.ris ASCII export fulfilling the FAIR principles
`01_Poster+Abstract`	• the *.pdf and *.pptx or *.tex
`02_Figures`**	• one sub-directory for every figure (`figure_01`, ...) with data points as *.csv or *.txt, or the generating file/program for the figure points (Matlab, Python, Origin,...); the results are from data stored in or extracted from the `04_PrimaryData` directory
`03_Tables`**	• one sub-directory for every table (`table_01`, ...) with data stored in as a csv-file
`04_PrimaryData`	• measured or simulated data, each sub-directory with a 'readme.json' file describing the data • ELN pdf pages with QR-code (see below)
`05_ProgramSources`	• e.g. OpenFOAM-Solver, Python/Matlab code, etc. • case files (mesh) for OpenFOAM, Fluent, ABAQUS etc. • "readme.json" file with metadata
`06_References`	• the references *.bib or *.enl together with *.ris • Cited papers/books (if publication rights are permitted)
`07_Supplement`	• Supplementary files
`08_Misc`	• Conference program

4 Documentation and Metadata

Data documentation and the generation of metadata is essential to understand your data in detail, and also help other researchers find, use and properly cite your data.

Various and particular metadata standards are available for each discipline, and you have to find them by yourself. Further general guidelines are provided below.

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

4.1 Important things to do while you collect or create your data

- Make a note of all file names and formats associated with the project, how the data is organized, how the data was generated (including any equipment or software used), and information about how the data has been altered or processed.

- Include an explanation of codes, abbreviations, or variables used in the data or in the file naming structure.

- Keep notes about where you got the data so that you and others can find it.

4.2 Things to document about your data

The used metadata scheme follows the recommendation of the world wide DataCite consortium (TIB Hannover, Britisch Library, Denmark Technical Information Center, ETH Zurich, CalTech, Australian National Data Service, ...). These data are stored in a 'readme.json' file in **EVERY DIRECTORY** of your data! Please, use our 'Readme File Creator' described below to generate this file. Explanation of some extracted properties are given in the following; please refer to Table 3 in [2] for more information:

ID 1: Identifier

Number used to identify the data. Can be a DOI (Digital Object Identifier) in case of a paper, or an ID from the ELN (eLabFTW generates a long number for every experiment, see below)

ID 2: Creator

Name of the person responsible for the described data (ORCID).

ID 3: Title
May be the title of a dataset, or the name of a piece of software, or the title of a paper.

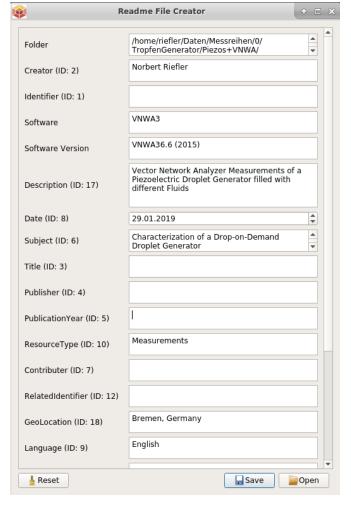### ID 4: Publisher

The name of the entity that produces, holds, archives, publishes prints, distributes, releases or issues the resource (e.g. 'Leibniz Institute for Materials Engineering IWT').

Figure 1: Data Input Tool: Readme-File-Creator

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

ID 5: Publication Year

The year when the data was or will be made publicly available.

ID 8: Date

Key date associated with the data, e.g. project start, end date, data modification, data release date, or time period covered by the data

ID 9: Language

Language(s) of the intellectual content of the resource, when applicable

ID 10: Resource Type

A description of the resource; mostly 'DataSet', but may be 'Audiovisual', 'High-Speed Images', 'DataPaper', etc.

ID 19: Funding Reference

Organizations or agencies who funded the research

ID 16: Rights

Any known intellectual property rights held for the data. The best choice is either the Creative Commons license 'CC-BY' (BY attribution) or 'CC-0' for data and 'CC Version 4.0' for written documents.

ID 17: Description

How the data was generated, including equipment or software used, experimental protocol, other things you might include in a lab notebook

ID 18: Geolocation

When the data relates to a physical location, record information about its spatial coverage

This tool displayed in figure 1 can be downloaded from the VT-Server (/VT-Allgemein/Software/ReadmeFile-Creator) for every operating system to simplify metadata entry (figure on the right). You can create automatically a file called 'readme.json' with your metadata, embedded in a JSON (Java Scipt Object Notation) format, which is used as documentation as well as metadata for Data Science Methods. This file is stored in every directory which contains data or program code.

The same DataCite metadata structure can be included into the ELN (see 5).

5   Electronic Lab Notebooks (ELNs)

Electronic Lab Notebooks (ELNs) enable researchers to organize and store experimental procedures, protocols, notes and data using their computer or mobile device. ELNs can offer several advantages over the traditional paper notebook in documenting research during the active phase of a project, including searchability within and across notebooks, secure storage with multiple redundancies, remote access to notebooks, and the ability to easily share notebooks among team members and collaborators.

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

5.1 eLabFTW

This is an OpenSource, generic, browser based ELN, developed/initialized by Nicolas Carpi (Institut Pasteuer, Paris) 2012. Data are stored in MySQL/MariaDB. It is maintained by many developers and is used worldwide (Berkley, Indian Institute of Technology, KIT, ...).

- Access: `https://eln.mvt-bremen.de/` or `https://134.102.38.139`

- Free definable status of the experiments ('finished', 'running', ...)

- Definition of templates for a step-by-step procedure description of measurements

- Every experiment gets a unique ID and can be summarized easily as a pdf

- Free definable categories/tags

- Graphical text editor to describe your experiments and simulations

- File attachments of data with preview of common formats (pdf, tiff, png, …)

- linking of experiments

- Time stamp service (RFC 3161, e.g. DFN)

- Data import/export (csv, zip, json, …)

- access using, e.g., Python via an Application Programming Interface (API)

You have to create a pdf for every experiment simply by a click ('Make a pdf') in the edit mode (see '`04_PrimaryData`' in Table 1. This kind of documentation includes a unique QR-code and is required for data which is published! The QR-code enables a direct link to the data together with unique ID.

The **maximum size** of uploaded data should not exceed **100Mbyte per file**. Further hints and tricks may be found in the seafile document (please add comments there if you explore new ways of usage):

`https://seafile.zfn.uni-bremen.de/f/fe685882eab14a2e853c/`

6 Git

Git is a web-based tool that provides a repository for the version control. It is used to manage, plan, create, verify and monitor your software developments. Furthermore, it can be used as a document versioning system. For example, during the creation of a larger piece of writing (a dissertation, papers, proposals, ...), you can enter the actual state of the document, and all the previous versions are stored as well. With that, every changed sentence can be tracked and reconstructed.

We have actually two Git instances on IWT:

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

- GitLab as part of the VT-Server for non-public documents

    $\rightarrow$ registration by the admins mentioned above

- Public Git: `https://github.com/Leibniz-IWT/`

    $\rightarrow$ registration: email 'IWT Organization access request' to: github@iwt.uni-bremen.de

7   EndNote

We have EndNote as literature management software for everyone (single user system). End-Note can be integrated into Word to manage cited references in publications. It also serves as a literature database for papers, textbooks, presentations etc. And it is used as the literature database for all publications from the IWT.

8   Further Tools for Data Management

- Search for Data Repositories: `https://www.re3data.org/`

- `https://zenodo.org/` $\rightarrow$ scientific data, publications, reports, presentations, videos, etc.

- `https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats`

    $\rightarrow$ list of recommended data formats

- `http://rd-alliance.github.io/metadata-directory/standards/` $\rightarrow$ Metadata Discipline Standard Formats

- Renaming Tools: PSRenamer, ExifToolGUI

References

[1] Deutsche Forschungsgemeinschaft (DFG). *Umgang mit Forschungsdaten*. 2021. URL: `https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/index.html`.

[2] DataCite Metadata Working Group. *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data*. Version 4.3. 2019. URL: `https://doi.org/10.14454/7xq3-zf69`.

[3] Kristin Briney. *Data Management for Researchers*. Pelagic Publishing, 2015. ISBN: 9781784270117.

[4] Leibniz Universität Hannover. *Leitfaden zur Erstellung eines Datenmanagementplans*. Version 2.3. Aug. 18, 2020. URL: `https://www.fdm.uni-hannover.de/fileadmin/fdm/Dokumente/Leitfaden_DMP_LUH_v2.3.pdf`.
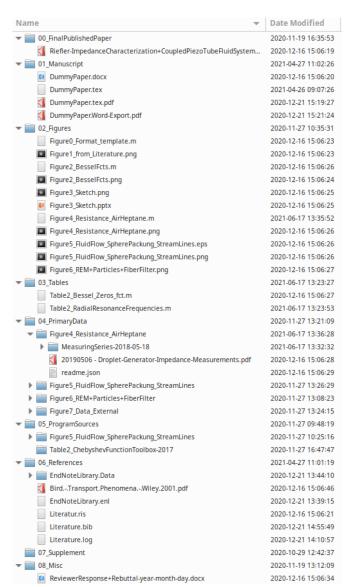
Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

[5]   Synology®. *Download-Zentrum*. URL: `https://www.synology.com/de-de/support/download/RS3618xs?version=7.1#utilities`.

A   Appendix

A.1   How to Store Papers – 'Dummy Paper'

The following screenshot shows a complete data structure of a dummy paper. You may find this dummy paper on our VT-Server here (`http://gofile.me/6C5nG/0uMzVrlwb`), together with these guidelines. Basically, the idea is that every data in the paper, whether it is a figure or a table, can be related to the original source of the measurement or the simulation. This is realized in the Dummy Paper by Matlab scripts, but can be done using other means as well (Python, Excel). The point is: No matter how, but you have to give a relation between the data saved in the `04_PrimaryData` directory and their representation in the publication. All figures and their generating scripts – here Matlab files – are given together with the measurement data so that running a Matlab (Octave) script within that directory generates the according figure. In the end, someone else can reproduce your results given in your publication. In case of other data like SEM images for instance, you can refer to them by a PDF generated by the ELN including a link to the experiment stored there.

Name	Date Modified
▼ 📁 00_FinalPublishedPaper	2020-11-19 16:35:53
Riefler-ImpedanceCharacterization+CoupledPiezoTubeFluidSystem...	2020-12-16 15:06:19
▼ 📁 01_Manuscript	2021-04-27 11:02:26
DummyPaper.docx	2020-12-16 15:06:20
DummyPaper.tex	2021-04-26 09:07:26
DummyPaper.tex.pdf	2020-12-21 15:19:27
DummyPaper.Word-Export.pdf	2020-12-21 15:21:24
▼ 📁 02_Figures	2020-11-27 10:35:31
Figure0_Format template.m	2020-12-16 15:06:23
Figure1_from_Literature.png	2020-12-16 15:06:23
Figure2_BesselFcts.m	2020-12-16 15:06:26
Figure2_BesselFcts.png	2020-12-16 15:06:24
Figure3_Sketch.png	2020-12-16 15:06:25
Figure3_Sketch.pptx	2020-12-16 15:06:25
Figure4_Resistance_AirHeptane.m	2021-06-17 13:35:52
Figure4_Resistance_AirHeptane.png	2020-12-16 15:06:26
Figure5_FluidFlow_SpherePackung_StreamLines.eps	2020-12-16 15:06:26
Figure5_FluidFlow_SpherePackung_StreamLines.png	2020-12-16 15:06:26
Figure6_REM+Particles+FiberFilter.png	2020-12-16 15:06:27
▼ 📁 03_Tables	2021-06-17 13:23:27
Table2_Bessel_Zeros_fct.m	2020-12-16 15:06:27
Table2_RadialResonanceFrequencies.m	2021-06-17 13:23:53
▼ 📁 04_PrimaryData	2020-11-27 13:21:09
▼ 📁 Figure4_Resistance_AirHeptane	2021-06-17 13:36:28
▶ 📁 MeasuringSeries-2018-05-18	2021-06-17 13:32:32
20190506 - Droplet-Generator-Impedance-Measurements.pdf	2020-12-16 15:06:28
readme.json	2020-12-16 15:06:29
▶ 📁 Figure5_FluidFlow_SpherePackung_StreamLines	2020-11-27 13:26:29
▶ 📁 Figure6_REM+Particles+FiberFilter	2020-11-27 13:08:23
▶ 📁 Figure7_Data_External	2020-11-27 13:24:15
▼ 📁 05_ProgramSources	2020-11-27 09:48:19
▶ 📁 Figure5_FluidFlow_SpherePackung_StreamLines	2020-11-27 10:25:16
Table2_ChebyshevFunctionToolbox-2017	2020-11-27 16:47:47
▼ 📁 06_References	2021-04-27 11:01:19
▶ 📁 EndNoteLibrary.Data	2020-12-21 13:44:10
Bird.-.Transport.Phenomena.-.Wiley.2001.pdf	2020-12-16 15:06:46
EndNoteLibrary.enl	2020-12-21 13:39:15
Literatur.ris	2020-12-16 15:06:21
Literature.bib	2020-12-21 14:55:49
Literature.log	2020-12-21 14:10:57
📁 07_Supplement	2020-10-29 12:42:37
▼ 📁 08_Misc	2020-11-19 13:12:09
ReviewerResponse+Rebuttal-year-month-day.docx	2020-12-16 15:06:34

Figure 2: Complete data structure of a dummy paper.

The 'Dummy Paper' contains a Word and a LaTeX template in '`01_Manuscript`'. The EndNote or BibTeX literature references are in '`06_References`'. Both, the EndNote file '*.enl' or the BibTeX file '*.bib', should include links between the references and the freely accessible PDFs of the papers and books also saved in '`06_References`'.

17

Data Management Guidelines:
How To Handle Data

IWT
Leibniz-Institut für
Werkstofforientierte
Technologien

A.2 How to Write a DMP

A.2.1 Motivation

A Data Management Plan (DMP) documents the thoughts of an applicant about the data which will be generated during the applied research project. Research funders (DFG etc.) expects that publicly promoted projects makes their data accessible to the public, where a simple publication was sufficient in the past.

There is no unified structure of DMPs due to the vast number of different research tasks. In the following, the main elements of a DMP from the TIB Hannover are listed, which can be found also in the recommendation of the DFG:

(`https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf`)

A.2.2 Elements of a DMP

0) **Administrative information**

- Project name
- Project participants
- Project description
- Project ground (doctorate, third party funding, etc.)
- Project duration
- Version of this DMP

1) **Methods and kinds of data gathering**

- Are there primary data generated, or will be secondary data used?
- Which data types / formats are generated and processed?
- How large are the data?
- Which equipment (instruments, hardware, software, etc.) will be used?
- How are the data organized? (Directory of file oriented structure? Version control?)
- How is the research process and the data been documented?
- Which (technical) standards will be used for description/documentation (metadata, classification)?
- How are the metadata been generated (e.g. automatically, manually, after a guideline, self-defined)?

2) **Backup and data safety**

- Where are the data stored?

- Which capacity is required?

- Interval of data safety?

- Are there any protective measures required for sensitive data?

- Are there any third party of project partners (e.g. in joint projects) who need access to the data?

3) Archiving

- Which data is archived?

- On which data carrier?

- Are there any requirements for the operators of the infrastructure? E.g. data curation?

- Which metadata have to be provided to find the archived data?

- Which information is additionally required to understand the context of the data?

- How long are the data archived?

- Are there any legal technicalities for the data archiving?

- What are the costs for which service?

4) Data sharing and publication

- Are there data which must be shared with others?

- What are the possible systems for data sharing?

- Which metadata and documentation are additionally required for third parties to use the data?

- Where (data repository, data journal) will the data be published, and how (e.g. open access, with embargo time, restricted access)?

- What are the license condition for the published data? (E.g. 'CC BY 4.0')

5) Resources and responsibilities

- How is the distribution of responsibility governed in this project?

- How is responsible for the data management (processes, IT, guidelines, formats, monitoring)?

- Required personal resources for a successful implementation/realization?

- What are the costs within the project phase and possibly thereafter?

- Which infrastructure resources are required, with additionally costs?