






IWT

Leibniz-Institut für 
Werkstofforientierte 
Technologien 

Leibniz
Leibniz
Gemeinschaft

Data Management Guidelines: Guter Umgang mit Daten

14. Januar 2025

Norbert Riefler

Vorwort

„Am 20. Juli 1969 kletterte Neil Armstrong aus seinem Raumschiff und setzte seine Füße auf den Mond. Die Landung wurde live in die ganze Welt übertragen und war ein bedeutendes Ereignis in der Geschichte der Wissenschaft und der Menschheit. Heute können wir uns immer noch das grobkörnige Video der Mondlandung ansehen, aber wir können uns nicht das Originalmaterial in höherer Qualität ansehen oder einige der Daten dieser Mission untersuchen. Das liegt daran, dass viele der Daten aus der frühen Weltraumforschung für immer verloren sind. (...) Die Bänder wurden wahrscheinlich irgendwann in den 1970er Jahren gelöscht und für die Datenspeicherung wiederverwendet“.

- Kristin Briney[5]

Dieser Leitfaden (<https://github.com/Leibniz-IWT/DataManagementGuidelines>) gilt für alle Mitarbeitenden des Leibniz-Instituts für Werkstofforientierte Technologien (L-IWT) und dient als Nachschlagehilfe für die Datensicherung. Es geht dabei um die Frage, warum wir über Datenmanagement nachdenken sollten, und enthält Anleitungen und Beispiele. Datenmanagement ist im Informationszeitalter eine unabdingbare Voraussetzung und Grundlage dafür, die Fülle an Daten zu bewältigen. Es hilft, die eigenen Daten zu organisieren, ist aber auch für die Weitergabe von Daten unerlässlich, damit andere die Struktur verstehen und die Informationen aus Ihren Daten korrekt deuten können. Auf diese Weise entsprechen Ihre Daten dem **FAIR**-Standard (Findable, Accessible, Interoperable and Reusable), der Grundlage für die neue Disziplin der **Datenwissenschaft**.

Es gibt viele Themen im Bereich des Datenmanagements, wie z.B. Data Governance und Sicherheit oder Speichermanagement, aber wir behandeln hier nur einen komprimierten Auszug, der auf den Umgang mit Daten in unserem Institut zugeschnitten ist. Dieser Reader soll Ihnen helfen zu klären, was derzeit unter dem Begriff **Data Steward** zu verstehen ist. Er lehnt sich an die von der Leibniz-Gemeinschaft heraus gegebenen Leitlinien an [3].

Inhaltsverzeichnis

1	Warum Datenmanagement?	3
2	Pläne zur Datenverwaltung	3
3	Dateiorganisation	4
3.1	Verschiedene Datentypen	4
3.2	Benennung von Dateien und Ordnern	5
3.3	Dateiformate	6
3.4	Ordnerhierarchie	7
3.4.1	Tägliche Daten	8
3.4.2	Daten zur Veröffentlichung	9
4	Dokumentation und Metadaten	14
4.1	Wichtige Information beim Sammeln oder Erstellen von Daten	14
4.2	Dokumentation allgemeiner Daten im DataCite-Schema	14
4.3	Erzeugen eines Metadaten-Schema	16
5	Electronic Lab Notebook	16
5.1	Allgemeine Informationen	16
5.2	IWT-Vorgaben zur Dokumentation und Struktur	17
6	Git Software Repositorium	18
7	FAIR Data und Open Science	18
8	EndNote	19
9	Weitere Tools für Datenmanagement	19
	Literatur	19
A	Anhang	21
A.1	Speicherung der Daten einer Veröffentlichung - „Dummy-Paper“	21
A.2	Schreiben eines Datenmanagementplans	22
A.2.1	Motivation	22
A.2.2	Elemente eines DMP	22
A.3	Glossar	24

1 Warum Datenmanagement?

Zeitersparnis

Wenn Sie Ihre Datenverwaltung planen, sparen Sie Zeit und Ressourcen, z.B. durch schnelleres Auffinden von Daten. Gut verwaltete Daten erfordern weniger Vorbereitungen für die Weitergabe oder das Verfassen eines Artikels.

Aufbewahrung Ihrer Daten

Wenn Sie Ihre Daten auf einem Dateiserver (Repository) ablegen, sichern Sie die von Ihnen geleistete Arbeit und bewahren Ihren Forschungsbeitrag für sich und andere auf.

Erhöhte Sichtbarkeit Ihrer Arbeit/Forschung

Daten anderen Forschenden zur Verfügung stellen erhöht die Sichtbarkeit und damit die Relevanz Ihrer Forschung.

Datenintegrität erhalten

Die Verwaltung und Dokumentation Ihrer Daten während deren Werdegangs ermöglicht es Ihnen und anderen, Ihre Daten in der Zukunft zu verstehen und weiter zu verwenden.

Einhaltung der Anforderungen von Mittelgebern

Viele Mittelgeber (z.B. DFG) verlangen, dass die im Rahmen eines Forschungsprojekts gesammelten Daten archiviert werden.

Förderung neuer Erkenntnisse

Die gemeinsame Nutzung Ihrer Daten mit anderen Forschenden kann durch Methoden des maschinellen Lernens zu neuen und unerwarteten Entdeckungen führen und Forschungsmaterial für diejenigen bereitstellen, die über wenig oder keine finanziellen Mittel verfügen.

Unterstützen Sie den öffentlichen Zugang zu Daten

Seien Sie Impulsgeber für die Forschung. Zeigen Sie Ihre Unterstützung für die offene Wissenschaft, indem Sie Ihre Daten zur Verfügung stellen.

2 Pläne zur Datenverwaltung

Der Umgang mit den Daten im Rahmen eines Projekts beinhaltet verschiedene Aspekte, die in einem Datenmanagementplan (DMP) beschrieben werden. Ein DMP enthält strukturierte Informationen über den Forschungsprozess des jeweiligen Projekts. Diese Informationen müssen jedoch vor Projektbeginn vorliegen: Antragsteller müssen in jedem Antrag einen DMP erstellen, der von allen wichtigen Forschungsförderorganisationen (DFG, BMBF, etc.) gefordert wird. Die Antragstellenden machen sich also schon beim Schreiben des Antrags Gedanken darüber, wie sie mit den Daten umgehen wollen. Die folgenden Themen gehören zu den typischen DMPs und könnten für Sie von Interesse sein:

1. Administrative Informationen: Projektname, Art der Finanzierung, Zeitraum.

2. Methoden der Datenerzeugung: Simulationen, Experimente (Geräte), Datenumfang, Art der Datendokumentation.
3. Datensicherheit: Speicherort, Speicherintervall und -kapazität, wer hat Zugriff.
4. Archivierung: welche Daten werden wo mit welchen Metadaten versehen.
5. Gemeinsame Nutzung von Daten: welches Repository, Lizenzbedingung, welche erforderlichen Metadaten
6. Ressourcen und Zuständigkeit: wer ist zuständig für Prozesse, IT, Festlegung von Vorgaben und Formaten, Überwachung; erforderliche personelle Ressourcen; Kosten.

Weitere Einzelheiten finden Sie in [1, 6] sowie im Anhang; einige beispielhafte DMP sind auch hier zu finden:

https://www.cms.hu-berlin.de/de/dl/dataman/arbeiten/dmp_erstellen

3 Dateiorganisation

Im Folgenden wird erläutert, welche Dateien Sie wie auf unseren Servern speichern:

- Der Zugang zum VT-Server erfolgt über Ihren Webbrowser durch Eingabe der Webadresse in das Adressfeld, über die Zuordnung des VT-Servers zum Windows-Explorer als eigenes Laufwerk, oder über den Synology Drive Client (siehe [8]) mit einigen weiteren und sehr hilfreichen Optionen wie z.B. der Versionierung von Dokumenten. Die **maximale Größe** der hochgeladenen Daten sollte **100Gbyte pro Datei** nicht überschreiten. Die Zugriffsberechtigung wird von Stefan Endres (s.endres@iwt.uni-bremen.de), Arvind Chouhan (a.chouhan@iwt.uni-bremen.de) und Nils Ellendt (ellendt@iwt.uni-bremen.de) verwaltet.
- Der Server der Abteilung Fertigungstechnik ist ebenfalls mittels Browser oder Windows-Explorer erreichbar. Die Zugriffsberechtigung wird von der IT-Abteilung des IWTs verwaltet.

3.1 Verschiedene Datentypen

Je nach Herkunft der Daten wird zwischen zwei Datentypen unterschieden [7]:

- Primärdaten (manchmal auch Rohdaten genannt) werden von Messgeräten, Sensoren/Kameras, Simulationsprogrammen, Beobachtungen usw. erzeugt und sind unbearbeitet; sie werden von Forschenden zum ersten Mal erzeugt.
- Sekundärdaten (oder analysierte Daten) sind Daten, die aus der Verarbeitung von Primärdaten hervorgehen. (In anderen Wissenschaftsbereichen sind Sekundärdaten bereits von anderen bearbeiteten oder erzeugten Daten).

Sehr große Primärdaten können nicht auf unserem Elektronischen Labor-Notizbuch (ELN)¹ gespeichert werden, aber die Herkunft dieser Daten wird dort zusammen mit Informationen über den Speicherort (z.B. wo genau sie auf einem Dateiserver liegen) erklärt. Dabei gelten folgende Regeln:

- Daten < 100 MByte werden im ELN gespeichert und dokumentiert
- 100 MByte < Daten < 1TByte werden auf dem VT-Server gespeichert und im ELN dokumentiert
- Daten > 1 TByte werden auf USB-Disks gespeichert und im ELN dokumentiert

Es ist unbedingt erforderlich, dass **alle Daten aus externen Quellen** (z.B. von einem USB-Stick) manuell auf Viren geprüft werden müssen, wenn der automatische Scan nicht funktioniert, bevor sie auf der Festplatte verwendet/gespeichert werden.

3.2 Benennung von Dateien und Ordern

Nehmen Sie sich zu Beginn eines Projekts Zeit für die Festlegung von Dateibenennungen. Wie werden Sie oder andere später nach Dateien suchen und auf sie zugreifen? Denken Sie dabei an den Typ, den Speicherort, die Studie oder ähnliches.

- Dateinamen sind selbsterklärend – sie beschreiben die Daten, um die es sich handelt!
- Verwenden Sie keine Leerzeichen oder Sonderzeichen, sondern nutzen Sie CamelCase, Bindestrich '-' oder Unterstrich '_' als Trennzeichen:

NO 'name date v1.txt'

YES 'name_date_v01.txt'

YES 'nameDateV01.txt'

YES 'simpleFoam_u0.2m-s_etha2mPas_v01'

- Entwickeln Sie ein Dateibenennungsschema, das Informationen über die Daten enthält.
Beispiel:

[Date]_[Run]_[SampleType]

- Berücksichtigen Sie die Sortierung, wenn Sie entscheiden, welches Element des Dateinamens an erster Stelle steht:

Verwenden Sie 'JJJJ-MM-TT', um Daten zu speichern.

Verwenden Sie führende Nullen:

NO 'ProjID_v1.csv' ... 'ProjID_v11.csv'

YES 'ProjID_v01.csv' ... 'ProjID_v11.csv' (Reihenfolge bis 99)

YES 'ProjID_v001.csv' ... 'ProjID_v111.csv' (Reihenfolge bis 999)

¹Die Nutzung des ELN ist derzeit nur für die VT verpflichtend.

3.3 Dateiformate

Aufgrund des technologischen Wandels veralten sowohl Hardware als auch Software. Deshalb müssen die von Ihnen gewählten Dateiformate eine langfristige Lesbarkeit bzw. Zugriff gewährleisten.

Dateiformate, die auch in Zukunft zugänglich sein werden, haben die folgenden Eigenschaften:

- Nicht-proprietär
- Offener, dokumentierter Standard
- Gemeinsame Nutzung durch die Forschungsgemeinschaft
- Standarddarstellung (ASCII, Unicode)
- Unverschlüsselt
- Unkomprimiert (wenn nicht zu groß...)

Beispiele für **bevorzugte** Dateiformate sind:

- Offene Dokumentdateien: .odt (Text), .ods (Tabellenkalkulation), .odp (Präsentation); primäres Format von OpenOffice/Libre Office, standardisiert in ISO 26300
- Office Open XML: .docx (Text), .xlsx (Tabellenkalkulation), .pptx (Präsentation); primäres Format von Microsoft Office, genormt in ISO 29500
- Rich Text Format: .rtf; proprietäres, aber gut dokumentiertes Format
- ASCII-Text oder Auszeichnungssprache: .txt, .md (Markdown), .html (Hypertext), .tex (LaTeX), .csv (Komma-getrennte Werte)
- Portable Document Format: .pdf; genormt in ISO 32000
- Bibliografische Daten: .enl (EndNote; Standardsoftware für die Literaturverwaltung am IWT); ansonsten verwenden Sie bitte lesbare Formate wie .bib (BibTeX) oder .ris (Research Information System Format)
- Videodateien in den Datenformaten .mp4 und .mpg, die mit den Codecs H.264 und H.265 erzeugt wurden; standardisiertes Format (ISO 14496-10) für die Videokomprimierung; verwenden Sie bitte **keine proprietären Codecs wie etwa Quicktime (.mov, .qt)**
- Bilder: .jpg ist das Standard-Dateiformat der meisten Kameras. Es kann von fast allen Bildbearbeitungsprogrammen gelesen werden, ist aber nicht in allen Fällen ideal (z.B. Strichzeichnungen), weil es ein verlustbehaftetes Format ist, das je nach gewählter Komprimierung zu einer dauerhaften Verschlechterung des Bildes führen kann. Das Nachfolgeformat .jp2 (JPEG2000) ermöglicht eine verlustfreie Komprimierung und kann daher für die Bildarchivierung verwendet werden, ist aber aufgrund Lizenzproblemen nicht weit

verbreitet. Das "Tagged Image File Format".tiff ist ein wichtiges Format für den Dateiaustausch mit Verlagen. Es ermöglicht eine verlustfreie Komprimierung und eine hohe Farbtiefe (32 Bit) sowie CMYK-Farben. Es kann für die Bildarchivierung verwendet werden, allerdings bleibt die Dateigröße vergleichsweise groß. Das "Portable Network Graphic"Format .png ist ein Rasterbildformat mit guter verlustfreier Komprimierung. Es ergibt größere Dateigrößen als .jpg für Fotos, ist aber sehr gut für Strichzeichnungen geeignet.

- Für Skizzen und Zeichnungen sind vektorbasierte Dateiformate vorzuziehen, da diese eine unbegrenzte Skalierung des Bildes ohne Qualitätseinbußen ermöglichen. Bei großen Ausdrucken sind Vektorgrafiken deutlich kleiner als Rastergrafiken. Allerdings werden vektorbasierte Bilder nicht von allen Textverarbeitungsprogrammen gut verarbeitet, und selbst wenn sie verwendbar sind, können sie auf verschiedenen Computern unterschiedliche Ergebnisse liefern. Es wird empfohlen, eine Vektorgrafik in einem offenen Format (.svg, scalable vector graphics) zu erstellen, um eine langfristige Lesbarkeit zu gewährleisten, und sie in das für Ihr Textverarbeitungsprogramm am besten geeignete Dateiformat zu exportieren (z. B. .emf für Word, .eps oder .pdf für LaTeX).
- Binäre Daten: Manchmal ist es am effizientesten, Daten in einem binären Format zu speichern, da textbasierte Darstellungen die Dateigröße drastisch erhöhen würden. In diesem Fall ist es wichtig, zu dokumentieren, wie die Daten gespeichert werden, und vorzugsweise ein Beispiel dafür zu geben, wie sie wieder gelesen werden können.

Wenn Sie in Erwägung ziehen, Ihre (Mess-)Daten in ein Format mit den oben genannten Eigenschaften zu exportieren, bewahren Sie eine Kopie im ursprünglichen Softwareformat auf, es sei denn, Sie können sicherstellen, dass alle Daten und Metadaten korrekt konvertiert wurden. Wenn Sie Ihre Daten in einem Repository hinterlegen, können Ihre Dateien in neuere Formate migriert werden, sodass sie auch für zukünftige Forschende nutzbar sind.

3.4 Ordnerhierarchie

Der spezifische Wert von Daten hat eine große Streuung. Viele Messungen, z.B. aus Vorversuchen, waren evtl. für die Planung Ihrer Experimente sehr wichtig, finden aber keine Verwendung in Veröffentlichungen. Im Gegensatz dazu sind einige Ihrer Daten sehr wertvoll und finden ihren Weg in eine Veröffentlichung. Aufgrund dieser Unterschiede hat der Ort, an dem Sie Ihre Daten speichern, je nach diesen drei Datenarten eine unterschiedliche Wertigkeit:

- I. Daten mit einer geringen Wertigkeit, siehe: 3.4.1 Tägliche Daten
- II. Daten, die wahrscheinlich für eine Veröffentlichung bestimmt sind, siehe: 3.4.2 Daten zur Veröffentlichung
- III. Daten von akzeptierten Publikationen: Diese Daten werden in einem speziellen Bereich des File-Servers gespeichert. in den Sie Ihre Daten nur einmal schreiben können – Sie

können dort nichts ändern oder löschen. Es ist wie ein Briefkasten: Sobald Sie Ihre Steuererklärung eingeworfen haben, ist es vorbei. Dieser Datenbereich ist nur für akzeptierte Veröffentlichungen gedacht, bei denen wirklich alles in Ordnung ist!

3.4.1 Tägliche Daten

Jeden Tag generieren Sie Daten z.B. durch Messungen, Simulationen, Erstellung eines Papers usw. Dafür sind die NAS-File-Server vorgesehen. All diese täglichen Daten werden unter Mitarbeiter und <Familiennamen> gespeichert, z.B.:

/VT-Server/<Abteilung>/Mitarbeiter/<Familiennamen>/

Die von <Abteilung> eingeschlossenen Namen müssen aus unseren Abteilungen gewählt werden, z.B.:

- LFM (Labor für Mikrozerspannung)
- MPS (MehrPhasenStrömung)
- RST (Reaktive Sprüh-Technik)
- SM (Strukturmechanik)
- SPK (SprühKompaktieren)

je nach Ihrer Zugehörigkeit. Ein Beispielverzeichnis ist:

/VT-Server/MPS/Mitarbeiter/riefler/

Hier speichern Sie die folgenden Daten:

→ **Experimentelle Daten:**

- Alle Experimente des Projekts
- Zusammenhang zwischen diesen Experimenten (welche Parameter wurden variiert) in einer JSON (Java Script Object Notation)-basierten Textdatei, siehe 4.2 Dokumentation allgemeiner Daten im DataCite-Schema weiter unten
- Zwischenergebnisse
- Liste aller verwendeten Geräte mit Einstellparametern

→ **Experimentelle Aufbauten:**

- Skizzen aller Teile
- Liste aller benötigten Teile und Geräte

→ **Für Simulationsdaten:**

- Die erforderlichen Anfangswerte und Netzinformationen, um alle Simulationen des Projekts neu zu berechnen

- Relation zwischen diesen Simulationen (welche Parameter variiert werden) in einer JSON-basierten Textdatei, siehe 4.2 Dokumentation allgemeiner Daten im DataCite-Schema weiter unten
- Zwischenergebnisse
- Liste der verwendeten Software zusammen mit dem verwendeten Quellcode
- Alle erforderlichen Eingabedateien zur Reproduktion der Simulationen

Es gibt keine vorgegebene Verzeichnisstruktur für Ihre täglichen Daten, das bleibt Ihnen überlassen. Diese Daten unter Ihrem Namen unterscheiden sich jedoch von Projektdaten wie z.B. wichtigen experimentellen Ergebnissen oder Simulationen, die veröffentlicht werden und die im Folgenden beschrieben werden.

3.4.2 Daten zur Veröffentlichung

Hier speichern Sie alle Daten, die für jede Veröffentlichung innerhalb dieses Projekts relevant sind. Abhängig von der Größe der Daten und der Entscheidung innerhalb Ihrer Abteilung (siehe oben) dürfen Primärdatendateien, die größer als z.B. 10 GByte sind, nicht in dem hier angegebenen Verzeichnis gespeichert werden. Stattdessen muss innerhalb der Dokumentation in der Experimentbeschreibung im ELN ein Hinweis auf die entsprechende Herkunft (z.B. Hochgeschwindigkeitskammermessung oder CFD-Simulation) gegeben werden und wo genau die Primärdaten gespeichert sind. Das **entscheidende Grundprinzip** für alle Ihre Roh- und Primärdaten ist: **Alle Daten müssen in ihrer Herkunft nachvollziehbar dokumentiert sein!**

Wenn Sie ein Dokument vorbereiten, z.B. eine Präsentation oder einen Projektbericht, werden die Daten (nach einem strengen Schema, siehe A.1 Speicherung der Daten einer Veröffentlichung - „Dummy-Paper“) als Beispiel in diesem Verzeichnis gespeichert:

/VT-Server/<Abteilung>/Projekte/<Finanzierung>/<Projektname>

→ **Ergebnisdokumentation:**

- Dokumente aus Ihrer wissenschaftlichen Arbeit: Papers, Dissertationen, Präsentationen, Berichte, Proposals, etc.
- Alle Ergebnisse (Abbildungen, Tabellen, Filme, Animationen, ...) und die zugrundeliegenden gemessenen oder simulierten Daten zusammen mit Auswertungsprogrammen

Finanzierungsquellen (<Finanzierung>) sind z.B.: DFG, AiF, BMBF, Industrie, ERC, etc. Zuletzt wird der Verzeichnisname für veröffentlichte Arbeiten in einem Projekt zusammengefügt aus diesen Komponenten:

<Jahr Projektstart>_<Abteilung>_<Projektbezeichnung>_<Familiennamen>. Siehe A.1 für ein Beispiel sowie die beiden hier angeführten Beispiele:

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

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

Die Dopplung von MPS und RST ist auf einen gemeinsamen Dateiserver zurückzuführen, auf dem in Zukunft die Daten aller IWT-Abteilungen gespeichert werden sollen.

Papers und Dissertationen

In Absprache mit allen IWT-Abteilungen basiert das Verzeichnisbenennungsschema für veröffentlichte Daten des VT-Servers auf den Abteilungen, der Finanzierungsquelle und dann dem entsprechenden Projektnamen (siehe oben für die Definition von <Projektname>):

Texte und Daten eines Papers werden z.B. in diesem Verzeichnis gespeichert:

Texte und Daten eines Papers werden z.B. in einem Verzeichnis gespeichert:

/VT-Server/<Abteilung>/Projekte/<Förderung>/<Projektname>/Paper/

mit diesem Verzeichnisnamen:

<Jahr>_<Abteilung>_<Paper-Name>_<Erstautor>

Texte und die Daten einer Dissertation werden dagegen so abgelegt:

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

Mit einem Verzeichnisnamen wie:

<Jahr>_<Abteilung>_<Dissertationstitel>_<Autor>

In diesen Verzeichnissen speichern Sie alle Daten Ihrer Arbeit in Unterverzeichnissen, deren Namen in der linken Zeile in Tabelle 1 angegeben sind. Die Namen der Unterverzeichnisse sind immer alle auf Englisch (wegen der Eindeutigkeit) und in der linken Zeile aufgeführt, die zugehörigen Dateien werden rechts erklärt:

Im Fall einer Dissertation speichern Sie Ihre Präsentation der Promotionsprüfung zusammen mit dem Manuskript unter dem Verzeichnis "01_Manuscript+Presentation". Wenn Sie keine ergänzenden Informationen zu Ihrer Publikation/Dissertation haben oder keine spezifischen Computerprogramme für Ihre Forschung geschrieben haben, müssen die Verzeichnisse "05_ProgramSources", "07_Supplement" und "08_Misc" noch vorhanden sein, aber ohne Einträge. Die Änderungen am Manuskript aufgrund des Revision-Prozesses können durch Anlegen von Unterverzeichnissen "01_Manuscript/01_Submission" sowie "01_Manuscript/02_Revision" samt Speichern der jeweiligen Manuskript- und Rebuttal-Datei nachvollziehbar gemacht werden.

Im Verzeichnis "02_Figures" speichern Sie jede Abbildung aus dem Manuskript in guter Qualität und zusätzlich eine *.csv-Datei mit den Daten der Abbildung. Oder Sie speichern eine Skript-Datei, die aus den Primärdaten im Verzeichnis "04_PrimaryData" die entsprechende Abbildung generiert - zum Beispiel eine Matlab-Datei. Dieses Datenverzeichnis enthält die gemessenen oder simulierten Daten im *.txt- oder *.csv-Format aus dem Messgerät oder der Simulationssoftware und eine weiter unten (4.2 Dokumentation allgemeiner Daten im DataCite-Schema) beschriebenen, z.B. vom Readme-File-Creator erzeugte "readme.jsonDatei mit Informationen und Metadaten zu Ihren gemessenen / simulierten Daten. Der Bezug jedes Datenpunktes in einer Abbildung zu den Primärdaten muss entweder implizit, z.B. in der Matlab-Datei, oder besser explizit in der "readme.jsonDatei angegeben werden.

00_FinalPublication*	<ul style="list-style-type: none"> Die allerletzte Version Ihrer Arbeit als pdf mit Band, Jahr und Seitenzahlen auf dem File-Server unter 'Publikationen' EndNote Eintrag für IWT-WiKo mit DOI-Nummer (Digital Object Identifier), sowie ein EndNote *.ris ASCII-Export, der den FAIR-Prinzipien entspricht
01_Manuscript	<ul style="list-style-type: none"> die *.pdf and *.docx or *.tex
02_Figures**	<ul style="list-style-type: none"> In Unterverzeichnis für jede Abbildung (z.B. "figure_01", ...) mit Datenpunkten als *.csv oder *.txt, oder die erzeugende Datei bzw. das erzeugende Programm für die Abbildungspunkte (Matlab, Python, Origin, ...); die Ergebnisse stammen aus Daten, die im Verzeichnis 04_PrimaryData gespeichert oder daraus extrahiert wurden
03_Tables**	<ul style="list-style-type: none"> Ein Unterverzeichnis für jede Tabelle ("Table_01", ..., usw.) in dem die Daten in Form einer csv-Datei gespeichert sind
04_PrimaryData	<ul style="list-style-type: none"> Gemessene oder simulierte Daten; jedes Unterverzeichnis mit einer "readme.json"-Datei zur Beschreibung der Daten ein *.pdf aus dem ELN mit QR-code (siehe unten 5)
05_ProgramSources	<ul style="list-style-type: none"> Z.B. OpenFOAM-Solver, Python/Matlab code, etc. Fallbeispiele (Netze) for OpenFOAM, Fluent, ABAQUS etc.
06_References	<ul style="list-style-type: none"> Die Referenzen *.bib oder *.enl zusammen mit *.ris Zitierte Arbeiten/Bücher (wenn es die Veröffentlichungsrechte erlauben)
07_Supplement	<ul style="list-style-type: none"> Ergänzende Dateien
08_Misc	<ul style="list-style-type: none"> Kommentare der Gutachter*innen / Widerlegungen

Tabelle 1: Die Daten jeder Arbeit müssen in acht Unterverzeichnissen gespeichert werden; weitere Hinweise:

*Wenn die Arbeit druckreif ist, lädt der korrespondierende Autor die letzte und überarbeitete Version (d.h. nur die tatsächlich verwendeten, aber absolut vollständigen Daten) in das geschützte Verzeichnis

**Wenn die in einer Abbildung/Tabelle wiedergegebenen Daten aus verteilten Primärdatenverzeichnissen stammen, genügt es, hier nur das erzeugende Programm zu speichern

Um Ihre Veröffentlichung zu vervollständigen, sollte jede zitierte Arbeit in das Verzeichnis "06_References" für internen Gebrauch aufgenommen werden. Allzu oft können die Ideen einer Arbeit nicht nachvollzogen und reproduziert werden, weil es Referenzen gibt, die Sie nicht bekommen können, weil Ihre Institution keinen Zugang zu dieser speziellen Zeitschrift bietet, oder die Referenzen stammen aus einer Masterarbeit, die nicht verfügbar ist. Speichern Sie daher alle Referenzen mit geeigneten Dateinamen für einen direkten Bezug - zusammen

mit allen zuvor genannten Daten und Informationen - das stellt Ihre Arbeit in ihrer Gesamtheit dar. Darüber hinaus ermöglicht EndNote eine komfortable Direktverlinkung zu den zitierten Publikationen, d.h. Sie sollten diese Links in Ihrer '*.enl'-Datei hinzufügen. Das Gleiche ist auch in BibTeX möglich. Wenn Ihre Arbeit auf ein öffentliches Repository übertragen werden soll, können jedoch nur frei zugängliche Publikationen hochgeladen werden, da viele Verlage, wie Springer und Elsevier, es nicht erlauben, Veröffentlichungen aus ihren Zeitschriften oder Büchern so zu veröffentlichen.

Das Verfassen einer Arbeit ist ein mehrstufiger Prozess. Mit der Entscheidung für eine bestimmte Zeitschrift kommt der Prozess an einen entscheidenden Punkt, an dem alle Daten in der oben und im folgenden Screenshot angegebenen Struktur gespeichert werden müssen und der Begutachtungsprozess beginnt. Wird der Beitrag angenommen, werden alle Änderungen integriert und das endgültige pdf mit Band, Jahr und Seitenzahlen freigegeben; die letzten Aktionen sind dann:

- Prüfen Sie, ob die wirklich benötigten Daten gespeichert sind, nicht mehr und nicht weniger.
- Überprüfen Sie die Übereinstimmung der Nummerierung der Abbildungen und Tabellen im Manuskript mit den entsprechenden Verzeichnissen und dem Erzeugungsprogramm der einzelnen Abbildungen (sehen Sie sich das Manuskript im DummyPaper an). Tabellen ohne Bezug zu Mess- oder Simulationsdaten müssen nicht gesondert im Verzeichnis "03_Tables" abgelegt werden.
- Prüfen Sie, ob alle ELN-Dateien, die das Experiment/die Simulation beschreiben, abgelegt wurden.
- Allerletzter Schritt: Legen Sie eine Kopie der kompletten Daten des Papers in das Verzeichnis 'Published' auf dem Fileserver. Das ist ein spezielles Verzeichnis: Sie können etwas dorthin kopieren, aber danach können Sie nichts mehr ändern. Es ist wie ein Briefkasten: Wenn Sie etwas hineingeworfen haben, ist es weg!

Bei Änderungen der Reihenfolge von Abbildungen oder Tabellen, was während des Review-Prozess häufig vorkommt, müssen diese entsprechend neu durchnummeriert werden. Die dafür notwendigen Umbenennungen werden mittels eines Python-Programms (abgelegt auf dem IWT-Fileserver unter →Austausch →Forschungsdatenmanagement →DataManagement-Guidelines-ServiceFiles →ImageSorter) sowohl für in Word als auch in LaTeX geschriebene Manuskripte erleichtert.

Poster und Präsentationen

Poster oder Präsentationen werden z.B. gespeichert unter:

/VT-Server/<Abteilung>/Projekte/<Förderung>/<Projektname>/Poster/

/VT-Server/<Abteilung>/Projekte/<Förderung>/<Projektname>/Präsentationen/

Die Verzeichnisstruktur ist ähnlich wie bei den Papers (siehe Tabelle 2), d.h. Sie müssen hier

Tabelle 2: Die Daten jedes Beitrags müssen in sieben Unterverzeichnissen gespeichert werden; weitere Hinweise:

* Wenn das Poster präsentiert wurde, lädt der jeweilige Präsentator die letzte und überarbeitete Version (d.h. nur die wirklich verwendeten, aber absolut vollständigen Daten) in das geschützte Verzeichnis

**Wenn die in einer Abbildung/Tabelle wiedergegebenen Daten aus verteilten Rohdatenverzeichnissen stammt, genügt es, hier nur das erzeugende Programm zu speichern

00_FinalPoster*	<ul style="list-style-type: none"> Die allerletzte Version Ihres Posters als pdf mit Band, Jahr und Seitenzahlen auf dem VT-Server in 'VT-Publikationen' EndNote-Eintrag für IWT-WiKo mit DOI-Nummer, und ein EndNote *.ris ASCII-Export, der den FAIR-Prinzipien entspricht
01_Poster+Abstract	<ul style="list-style-type: none"> das *.pdf und *.pptx or *.tex
02_Figures**	<ul style="list-style-type: none"> ein Unterverzeichnis für jede Abbildung (figure_01, ...) mit Datenpunkten als *.csv or *.txt, oder die erzeugende Datei/ das erzeugende Programm für die Abbildungspunkte (Matlab, Python, Origin,...); die Ergebnisse stammen aus Daten, die im Verzeichnis 04_PrimaryData gespeichert oder daraus extrahiert wurden
03_Tables**	<ul style="list-style-type: none"> ein Unterverzeichnis für jede Tabelle (Tabelle_01, ...) in dem die Daten in Form einer csv-Datei gespeichert
04_PrimaryData	<ul style="list-style-type: none"> gemessene oder simulierte Daten, jedes Unterverzeichnis mit einer "readme.json" Datei zur Beschreibung der Daten ELN pdf mit QR-code (siehe unten)
05_ProgramSources	<ul style="list-style-type: none"> e.g. OpenFOAM-Solver, Python/Matlab code, etc. Fallbeispiele (Netze) for OpenFOAM, Fluent, ABAQUS etc. "readme.json" Datei mit Metadaten
06_References	<ul style="list-style-type: none"> die Referenzen *.bib oder *.enl zusammen mit *.ris Zitierte Arbeiten/Bücher (wenn Veröffentlichungsrechte erlaubt sind)
07_Supplement	<ul style="list-style-type: none"> Ergänzende Dateien
08_Misc	<ul style="list-style-type: none"> Konferenzprogramm

wieder alle verwendeten Mess- oder Simulationsdaten speichern. Einige Verzeichnisse haben abweichende Namen und einen anderen Inhalt, wie z.B. "01_Poster+Abstrakt", das Poster und Abstrakt der Konferenz enthält. Falls Sie eine mündliche Präsentation gehalten haben, heißen die ersten beiden Verzeichnisse "00_FinalPresentation" und "01_Presentation+Abstrakt".

Bei Änderungen der Reihenfolge von Abbildungen oder Tabellen müssen diese entsprechend neu durchnummeriert werden.

4 Dokumentation und Metadaten

Metadaten sind Daten über Daten. Metadaten sind dazu da, die eigentlichen Daten genauer zu beschreiben. Ein Druckwert alleine sagt nichts aus, es muss immer ein Kontext mit angegeben werden, an welchem System (z.B. in welchem Fluid) und an welcher Stelle gemessen wurde, mit welchem Drucksensor (Firma, Modell) und an welchem Messgerät war der Drucksensor angeschlossen, usw.

Die Dokumentation von Daten sowie die Erstellung von Metadaten sind unerlässlich, um Ihre Daten im Detail zu verstehen und auch anderen Forschenden zu helfen, Ihre Daten zu finden und weiter zu nutzen bzw. richtig zu zitieren. Damit werden Daten nachvollziehbar, ein wichtiger Aspekt für FAIR Data (siehe Kapitel 7).

Für die Erstellung von Metadaten gibt es bereits viele Vorlagen bzw. Metadaten-Standards bzw. -Schematas. Jede Disziplin hat eigene, fachspezifische Schemata, die Sie selber erfragen oder nachschlagen müssen. Falls es noch kein Schema für Ihr Gebiet oder Ihre Anwendung gibt, können selber Metadatenschema entworfen werden. Weitere allgemeine Leitlinien finden sie im folgenden Text.

4.1 Wichtige Information beim Sammeln oder Erstellen von Daten

- Notieren Sie alle Dateinamen und -formate, die mit dem Projekt in Verbindung stehen, wie die Daten organisiert sind, wie die Daten erzeugt wurden (einschließlich aller verwendeten Geräte samt Software), und Informationen darüber, wie die Daten verändert oder verarbeitet wurden.
- Version und Hersteller von Messgeräten und Computerprogrammen (auch von den Steuerprogrammen der Messgeräte)
- Abkürzungen oder Variablen, die in den Daten oder in der Dateibenennungsstruktur verwendet werden.

4.2 Dokumentation allgemeiner Daten im DataCite-Schema

Am L-IWT wird ein weltweit verwendetes Metadatenschema verwendet zur Angabe von allgemeinen Metadaten. Unser Schema folgt den Empfehlungen des weltweiten DataCite-Konsortiums (TIB Hannover, Britische Bibliothek, Denmark Technical Information Center, ETH Zürich, CalTech, Australian National Data Service, ...). Diese Metadaten können mithilfe unseres 'Readme File Creator' (siehe 1) erzeugt und in einer "readme.json"-Datei in **JEDEM VERZEICHNIS** Ihrer Daten gespeichert werden, um diese zu beschreiben. Alternativ kann dieselbe DataCite-Metadaten-Struktur auch im JSON-Editor des ELNs (siehe 5) abgebildet bzw. eingegeben werden.

Im Folgenden werden einige der extrahierten Eigenschaften erläutert; weitere Informationen finden Sie in Tabelle 3 in [4]:

ID 1: Kennung

Nummer, die zur Identifizierung der Daten verwendet wird. Dies kann ein DOI (Digital Object Identifier) im Falle einer Veröffentlichung oder eine PID (Persistent Identifier) aus dem ELN sein (eLabFTW generiert eine lange Nummer für jedes Experiment, siehe 5 Electronic Lab Notebook).

ID 2: Erstellerin oder Ersteller

Name der Person, die für die beschriebenen Daten verantwortlich ist (ORCID, Open Researcher Contributor Identification).

ID 3: Titel

Kann der Titel eines Datensatzes, der Name einer Software oder der Titel eines Artikels sein.

ID 4: Herausgeber

Der Name der Einrichtung, die die Daten produziert, aufbewahrt, archiviert, veröffentlicht, druckt, vertreibt, freigibt oder herausgibt (z.B. "Leibniz-Institut für Werkstofforientierte Technologien - IWT").

ID 5: Jahr der Veröffentlichung

Das Jahr, in dem die Daten öffentlich zugänglich gemacht wurden oder werden.

ID 8: Datum

Entstehungsdatum der Daten, einschließlich des Start- und Enddatums des Projekts, des Datums der Datenänderung und des von den Daten abgedeckten.

ID 9: Sprache

Sprache(n) des Inhalts der Quelle.

ID 10: Ressourcentyp

Eine Beschreibung der Ressource; meist "DataSet", kann aber auch "Audiovisuell", "High-Speed Images", "DataPaper", etc. sein.

ID 19: Förderkennzeichen

Organisationen oder Behörden, die die Forschung finanziert haben.

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

ID 16: Rechte

Alle bekannten Rechte am geistigen Eigentum an den Daten. Die beste Wahl ist normalerweise die Creative Commons Lizenz 'CC-BY' (BY attribution) oder 'CC-0' für Daten sowie die 'CC Version 4.0' für schriftliche Dokumente.

ID 17: Beschreibung

Wie die Daten erzeugt wurden, einschließlich der verwendeten Geräte oder Software, des Versuchsprotokolls und anderer Dinge, die Sie in ein Laborjournal aufnehmen könnten.

ID 18: Geolokalisierung

Wenn sich die Daten auf einen physischen Ort beziehen, halten Sie Informationen über die geografische Abdeckung fest.

Das oben in Figure 1 abgebildete Tool kann vom IWT-File-Server (→Austausch →Forschungsdatenmanagement →DataManagementGuidelines-ServiceFiles →ReadmeFileCreator) für jedes Betriebssystem heruntergeladen werden, um die Eingabe von Metadaten zu vereinfachen. Sie können automatisch eine Datei namens 'readme.json' mit Ihren Metadaten, eingebettet in ein JSON (Java Script Object Notation)-Format, erstellen, die sowohl als Dokumentation als auch als Metadaten für Data Science Methods verwendet wird. Diese Datei wird in jedem Verzeichnis gespeichert, das Daten oder Programmcode enthält. Felder können einfach leer bleiben, wenn sie unzutreffend sind.

Dieselbe DataCite-Metadaten-Struktur kann auch im JSON-Editor des ELNs (siehe Kapitel 5) abgebildet bzw. eingegeben werden.

4.3 Erzeugen eines Metadaten-Schema


Ein Schema ist eine logische Struktur, die sowohl die Beziehungen zwischen den Metadaten-Elementen aufzeigt als auch Regeln für die Anwendung und Verwaltung der Metadaten liefert. Ein Beispiel für die Entwicklung eines Metadaten-Schema ist in der Arbeit von Eberskirch et al. [2] beschrieben, in der alle relevanten Parametern gesammelt wurden zur physikalischen als auch biologischen (toxikologischen) Charakterisierung von Nanopartikeln. Die Entwicklung eines für die eigene Fachdisziplin passenden Schemas kann sehr hilfreich sein für die Strukturierung der Daten und des eigenen Wissens.

5 Electronic Lab Notebook (ELN)

5.1 Allgemeine Informationen

Electronic Lab Notebooks ermöglichen es Forschende, experimentelle Verfahren, Protokolle, Notizen und Daten über ihren Computer oder ein mobiles Gerät zu organisieren und zu speichern. ELNs bieten gegenüber dem traditionellen Papiernotizbuch mehrere Vorteile bei der Dokumentation von Forschungsarbeiten in der aktiven Phase eines Projekts. Dazu gehören die Durchsuchbarkeit innerhalb und zwischen den Notizbüchern, die sichere Speicherung

mit mehreren Redundanzen, der Fernzugriff auf die Notizbücher und die Möglichkeit, die Beschreibungen von Experimenten oder Simulationen problemlos mit anderen Teammitgliedern und Mitarbeitenden zu teilen.

eLabFTW  ist ein OpenSource, generisches, browserbasiertes ELN, entwickelt/initialisiert von Nicolas Carpi (Institut Pasteur, Paris) im Jahr 2012. Die Daten werden in MySQL/MariaDB gespeichert. Es wird von vielen Entwickler*innen gepflegt und ist weltweit im Einsatz (Berkley, Indian Institute of Technology, KIT, ...).

- Zugang: <https://elabftw.iwt.zz/>
- Frei definierbarer Status der Experimente ("beendet", "läuft", ...)
- Definition von Vorlagen für eine schrittweise Ablaufbeschreibung von Messungen
- Jedes Experiment erhält eine eindeutige ID und kann leicht als pdf-Datei zusammengefasst werden
- Frei definierbare Kategorien/Tags
- Grafischer Texteditor zur Beschreibung Ihrer Experimente und Simulationen
- Dateianhänge von Daten mit Vorschau der gängigen Formate (pdf, tiff, png, ...)
- Verknüpfung von Experimenten
- Zeitstempel-Dienst (RFC 3161, z.B. DFN)
- Datenimport/-export (csv, zip, json, ...)
- Zugriff mit z.B. Python über eine Anwendungsprogrammierschnittstelle (API)

5.2 IWT-Vorgaben zur Dokumentation und Struktur

Um die wissenschaftlichen Experimente übersichtlich, klar und verständlich für andere zu dokumentieren und auch noch Jahre später schnell auffinden zu können, werden die Titel in eLabFTW einheitlich mit dieser Systematik aus sechs Komponenten vorgegeben:

{Versuchsnummer}, {Probenname/Kürzel}, Material, {Gerät/Anlage}, Methode, Projektkürzel

Die Angaben in eckigen Klammern sind optional, aber die drei Angaben für "Material", "Methode" und das "Projekt" bzw. "Projektkürzel" sind zwingend erforderlich, es können aber auch mehr als die oben genannten sein. Mit genau derselben Systematik werden Tags (d.h. Marker bzw. Schlagwörter) vergeben. Der Grund für diese Redundanz liegt in dem SQL basierten Datenbank-System von eLabFTW, das mithilfe von Tags wesentlich schnellere Antworten auf Suchanfragen liefert.

Zur Dokumentation der Ergebnisse kann für ein Experiment (was auch eine Simulation sein kann) eine PDF-Datei erstellt werden, indem Sie im Bearbeitungsmodus einfach auf 'PDF erstellen' klicken (siehe '04_PrimaryData' in Tabelle 1). Diese Art der Dokumentation enthält einen eindeutigen QR-Code und ist für Daten, die veröffentlicht werden, erforderlich! Der QR-Code ermöglicht einen direkten Link zu den Daten zusammen mit einer eindeutigen ID.

Die **maximale Größe** der hochgeladenen Daten sollte **100Mbyte pro Datei** nicht überschreiten. Weitere Tipps und Tricks finden Sie im eLab-Dokument (bitte fügen Sie dort Kommentare hinzu, wenn Sie neue Möglichkeiten der Nutzung entdecken) auf dem IWT file server (→Austausch →Forschungsdatenmanagement →eLabFTW)

Eine Vorlage für den in jedem Experiment vorhandenen JSON-Feld mit Angaben zu den Metadaten (siehe Kapitel 4) ist ebenfalls auf dem IWT-Fileserver abgelegt.

6 Git Software Repository

Git ist ein webbasiertes Tool, das ein Repository für die Versionskontrolle bereitstellt. Es wird verwendet, um Ihre Softwareentwicklungen zu verwalten, zu planen, zu erstellen, zu verifizieren und zu überwachen. Außerdem kann es als System zur Versionierung von Dokumenten verwendet werden. Bei der Erstellung eines größeren Textes (Dissertation, Referate, Anträge, ...) kann zum Beispiel der aktuelle Stand des Dokuments eingegeben werden, und alle vorherigen Versionen werden mitgespeichert. Damit lässt sich jeder geänderte Satz nachvollziehen und rekonstruieren.

Die Git-Instanz am IWT ist hier zugänglich: <https://github.com/Leibniz-IWT/>. Anmeldung mittels Email und Betreff "IWT Organization access request" an: github@iwt.uni-bremen.de

7 FAIR Data und Open Science

FAIR (Findable, Accessible, Interoperable, Reusable) werden Daten dann genannt, wenn sie im Internet gefunden, herunterladbar und tatsächlich auch genutzt werden können. Insbesondere für das in Kapitel 3.4.2 geforderte Verzeichnisstruktur der Datenablage gilt die Anforderung, Daten FAIR abzulegen, d.h. Messdaten müssen klar nachvollziehbar sein (Dokumentation des verwendeten Messgerät mit allen Einstellungen, Resultate mit Einheiten als ASCII-File, etc.) und Simulationsdaten müssen nachgerechnet werden können (Quellcode, Rechengitter, Randbedingungen etc.; keine kommerziellen Programmpakete, aber case-Dateien schon). Mit diesem Vorgehen wird die Gute Wissenschaftliche Praxis (GWP) realisiert.

Für mit öffentlichen Geldern geförderten Projekte (und das sind die meisten am L-IWT) gilt zusätzlich der Grundsatz, dass alle im Projekt gewonnen Daten und entwickelten Methoden der Öffentlichkeit zugänglich gemacht werden müssen. Dazu kann die oben beschriebene Verzeichnisstruktur auf zenodo <https://zenodo.org/>, einem Repository für Wissenschaftliche Daten, hochgeladen werden, wofür dann eine DOI bereitgestellt wird. Allerdings dürfen die

PDFs bei Veröffentlichung in Fachzeitschriften nur wenn dies explizit erlaubt ist (was genau geprüft werden muss) mit abgelegt werden, oder nach einer gewissen Embargo-Zeit. In der Regel darf aber eine selbst formatierte Version des Textes samt Grafiken veröffentlicht werden.

Daten aus anderen Forschungseinrichtungen, die z.B. in einer Kooperation mit in eine Veröffentlichung einfließen, müssen entsprechend unserer Vorgaben ebenfalls FAIR abgelegt werden, also entweder uns überlassen und dann nach Kapitel 3.4.2 abgespeichert, oder mittels Verlinkung auf ein Repositorium zugänglich gemacht werden.

Damit die abgelegten Daten im Sinne von FAIR von anderen Personen weitergenutzt werden können, müssen sie mit entsprechenden Lizenzen versehen werden (ohne Angabe ist keine Weiternutzung und nur das Betrachten erlaubt). Dazu hat es sich bewährt, Texte unter der 'CC-BY-Lizenz' und Software unter der '3-clause BSD license' (oder 'BSD-3') abzulegen.

Im Rahmen von Open Science ist FAIR ein Teil von Open Data, das zusammen mit Open Access von Fachartikeln und Open Source von Software drei wesentliche Prinzipien für eine offene Wissenschaft bilden. Weitere Informationen, u.a. mit dem Leibniz Open Science Leitbild, sind auf der IWT Webseite zu finden².

8 EndNote

Wir haben EndNote als Literaturverwaltungssoftware für alle (Einzelplatzsystem). EndNote kann in Word integriert werden, um zitierte Referenzen in Publikationen zu verwalten. Es dient auch als Literaturdatenbank für Aufsätze, Lehrbücher, Präsentationen usw. Und es wird als Literaturdatenbank für alle Publikationen des IWTs verwendet.

9 Weitere Tools für Datenmanagement

- Suche nach Daten-Repositorien: <https://www.re3data.org/>
- <https://zenodo.org/> → wiss. Daten, Veröffentlichungen, Berichte, Präsentationen, Videos usw.
- <https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>
→ Liste der empfohlenen Datenformaten
- <http://rd-alliance.github.io/metadata-directory/standards/> → disziplinspezifische Metadaten und Standardformate
- Umbenennungs-Tools: PSRenamer, ExifToolGUI

²<https://www.iwt-bremen.de/en/institute/about-us/open-science>

Literatur

- [1] Deutsche Forschungsgemeinschaft (DFG). *Umgang mit Forschungsdaten*. 2021. URL: https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/index.html.
- [2] Linda Elberskirch u. a. „Digital research data: from analysis of existing standards to a scientific foundation for a modular metadata schema in nanosafety“. In: *Part. Fibre Toxicol.* 19 (2022), S. 1–1-19.
- [3] Leibniz Gemeinschaft. *Leitlinie zum Umgang mit Forschungsdaten in der Leibniz-Gemeinschaft*. 2018. URL: https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/Bilder_und_Downloads/Forschung/Open_Science/Leitlinie_Forschungsdaten_2018.pdf.
- [4] 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>.
- [5] Kristin Briney. *Data Management for Researchers*. Pelagic Publishing, 2015. ISBN: 9781784270117.
- [6] Leibniz Universität Hannover. *Leitfaden zur Erstellung eines Datenmanagementplans*. Version 2.3. 18. Aug. 2020. URL: https://www.fdm.uni-hannover.de/fileadmin/fdm/Dokumente/Leitfaden_DMP_LUH_v2.3.pdf.
- [7] Lisa Gitelman. *“Raw Data” Is an Oxymoron*. MIT Press, 2013.
- [8] Synology®. *Download-Zentrum*. URL: <https://www.synology.com/de-de/support/download/RS3618xs?version=7.1#utilities>.

A Anhang

A.1 Speicherung der Daten einer Veröffentlichung - „Dummy-Paper“

Der folgende Screenshot zeigt die komplette Datenstruktur eines Dummy-Papers. Sie finden dieses Dummy Paper auf unserem IWT-Server IWT file server under →Austausch →Forschungsdatenmanagement →DataManagementGuidelines-ServiceFiles →Dummy-Paper. Grundsätzlich geht es darum, dass alle Daten im Paper, egal ob es sich um eine Abbildung oder eine Tabelle handelt, mit der ursprünglichen Quelle der Messung oder der Simulation in Verbindung gebracht werden können. Dies wird im Dummy Paper durch Matlab-Skripte realisiert, kann aber auch mittels Python oder Excel erfolgen. Entscheidend ist: Egal wie, aber es muss ein Bezug zwischen den im Verzeichnis '04_PrimaryData' gespeicherten Daten und ihrer Darstellung in der Publikation hergestellt werden. Alle Abbildungen und die dazugehörigen Skripte - hier Matlab-Dateien - werden zusammen mit den Messdaten angegeben, so dass die Ausführung eines Matlab-Skripts in diesem Verzeichnis die entsprechende Abbildung erzeugt, damit schließlich Ihre Ergebnisse in Ihrer Veröffentlichung nachvollzogen werden können. Im Falle von anderen Daten, wie z.B. REM-Bildern, können Sie auf diese durch ein vom ELN generiertes PDF verweisen, das einen Link zu dem dort gespeicherten Experiment enthält.

Das "Dummy Paper" enthält eine Word- und eine LaTeX-Vorlage in '01_Manuscript'. Die EndNote- oder BibTeX-Literaturreferenzen befinden sich in '06_References'. Beide, die EndNote-Datei '*.enl' oder die BibTeX-Datei '*.bib', sollten Links zwischen den Referenzen und den freizugänglichen PDFs der Papiere und Bücher enthalten, die ebenfalls in '06_References' gespeichert sind.

Name	Date Modified
00_FinalPublishedPaper	2020-11-19 16:35:53
Riefler-ImpedanceCharacterization+CoupledPiezoTubeFluidSystem+...	2020-12-16 15:06:19
01_Manuscript	2023-10-10 12:53:26
01_Submission	2023-10-10 12:53:26
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
DummyPaperWord-Export.pdf	2020-12-21 15:21:24
02_Revision	2023-10-10 12:53:16
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_SpherePacking_StreamLines.eps	2020-12-16 15:06:26
Figure5_FluidFlow_SpherePacking_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	2022-04-06 09:36:18
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_SpherePacking_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
Link to eLabFTW-json-template.json	2022-02-25 10:29:04
05_ProgramSources	2020-11-27 09:48:19
Figure5_FluidFlow_SpherePacking_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

Abbildung 2: Complete data structure of a dummy paper.

A.2 Schreiben eines Datenmanagementplans

A.2.1 Motivation

Ein Datenmanagementplan (DMP) dokumentiert die Überlegungen eines Antragstellers zu den Daten, die im Rahmen eines laufenden Forschungsprojekts anfallen. Forschungsförderer (DFG etc.) erwarten mittlerweile, dass öffentlich geförderte Projekte ihre Daten der Öffentlichkeit zugänglich machen, wofür in der Vergangenheit eine einfache Veröffentlichung ausreichte.

Aufgrund der Vielzahl unterschiedlicher Forschungsaufgaben gibt es keine einheitliche Struktur für DMPs. Im Folgenden sind die wesentlichen Elemente eines DMP der TIB Hannover aufgeführt, die sich so auch bei den Empfehlungen der DFG wiederfinden.

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

A.2.2 Elemente eines DMP

0) Administrative Informationen

- Projektname
- Projektteilnehmende
- Projektbeschreibung
- Projektgrund (Promotion, Drittmittel, etc.)
- Projektdauer
- Version des DMP

1) Methoden und Arten der Datenerhebung

- Werden Primärdaten generiert oder werden Sekundärdaten verwendet?
- Welche Datentypen/-formate werden erzeugt und verarbeitet?
- Wie groß sind die Daten?
- Welche Ausrüstung (Instrumente, Hardware, Software usw.) wird verwendet?
- Wie sind die Daten organisiert? (Verzeichnis oder dateiorientierte Struktur? Versionskontrolle?)
- Wie werden der Forschungsprozess und die Daten dokumentiert?
- Welche (technischen) Standards werden für die Beschreibung/Dokumentation verwendet (Metadaten, Klassifizierung)?
- Wie werden die Metadaten erzeugt (z.B. automatisch, manuell, nach einem Leitfaden, selbst definiert)?

2) Sicherung und Datensicherheit

- Wo werden die Daten gespeichert?
- Welche Kapazität wird benötigt?
- Intervall der Datensicherheit?
- Sind irgendwelche Schutzmaßnahmen für sensible Daten erforderlich?
- Gibt es Dritte unter den Projektpartnern (z. B. bei gemeinsamen Projekten), die Zugang zu den Daten benötigen?

3) Archivierung

- Welche Daten werden archiviert?
- Auf welchem Datenträger?
- Gibt es Anforderungen an die Betreiberinnen und Betreiber der Infrastruktur? Z.B. Datenkuratierung?
- Welche Metadaten müssen bereitgestellt werden, um die archivierten Daten zu finden?
- Welche Informationen werden zusätzlich benötigt, um den Kontext der Daten zu verstehen?
- Wie lange werden die Daten archiviert?
- Gibt es rechtliche Besonderheiten bei der Datenarchivierung?
- Wie hoch sind die Kosten für welche Dienstleistung?

4) Gemeinsame Nutzung und Veröffentlichung von Daten

- Gibt es Daten, die mit anderen geteilt werden müssen?
- Welche Systeme sind für die gemeinsame Nutzung der Daten möglich?
- Welche Metadaten und Dokumentationen werden zusätzlich benötigt, damit Dritte die Daten nutzen können?
- Wo (Datenrepositorium, Datenzeitschrift) werden die Daten veröffentlicht und wie (z.B. Open Access, mit Sperrfrist, beschränkter Zugang)?
- Wie lauten die Lizenzbedingungen für die veröffentlichten Daten? (z. B. "CC BY 4.0")

5) Ressourcen und Verantwortlichkeiten

- Wie ist die Verteilung der Verantwortung in diesem Projekt geregelt?
- Wer ist für das Datenmanagement verantwortlich (Prozesse, IT, Richtlinien, Formate, Monitoring)?
- Erforderliche personelle Ressourcen für eine erfolgreiche Implementierung/Realisierung?
- Wie hoch sind die Kosten innerhalb der Projektphase und ggf. danach?
- Welche Infrastruktur-Ressourcen werden benötigt, fallen zusätzlichen Kosten an?

A.3 Glossar

AiF	...	Arbeitsgemeinschaft industrieller Forschungsvereinigungen
ASCII	...	American Standard Code for Information Interchange
BMBF	...	Bundesministerium für Bildung und Forschung
BSD	...	Berkeley Software Distribution
CC-BY	...	Creative Commons - BY attribution
DFG	...	Deutsche Forschungs-Gemeinschaft
DMP	...	Data Management Plan
ELN	...	Elektronic Laboratory Notebook
ERC	...	European Research Council
FAIR	...	Findable Accessible Interoperable Reusable
GWP	...	Gute Wissenschaftliche Praxis
ISO	...	International Organization for Standardization
JSON	...	Java Script Object Notifier
NAS	...	Network Access Storage
XML	...	eXtended Markup Language

III