



HANDREICHUNGEN AUS DER AG FORSCHUNGSDATENMANAGEMENT SPEICHERSTANDARDS ZUR DATENORGANISATION

Exposé

Datenqualität | Datenorganisation | Ordnerstruktur | Dateikonventionen | Versionskontrolle | Dokumentation | Datenformate | Metadatenschemata

Erstellt | bearbeitet

17. Dezember 2025 (V 1.7) | Richter Anja,
Keyvan Djahangiri

Kontakt

forschungsdaten@uni-passau.de

[Webseite](#)

Hintergrund

Ein fachgerechter, qualitätssichernder Umgang mit Forschungsdaten ist entscheidend für die Anschlussfähigkeit der Forschung wie auch die Überprüfbarkeit wissenschaftlicher Erkenntnisse.

Forschungsdaten sollen **FAIR** (Findable, Accessible, Interoperable und Reusable) sein – d.h. sie sollen auffindbar, zugänglich, interoperabel und wiederverwendbar sein. Für die langfristige Nachnutzbarkeit Ihrer Forschungsdaten spielen deshalb vor allem der Erhalt der **Authentizität**, der **Integrität**, der **Zugänglichkeit** und der **Verständlichkeit** eine zentrale Rolle. Wesentlich hierzu beitragen können eine durchdachte Planung, die präzise Dokumentation und Beschreibung Ihrer Daten, die Berücksichtigung fachspezifischer (Metadaten-)Standards und die Bereitstellung der technischen Infrastruktur.

Vor der Datenarchivierung bzw. Publikation kann es nützlich sein, durch den Prozess der **Datenvalidierung** eine genaue Selektion darüber zu treffen, welche Daten dauerhaft abgelegt werden sollen. Ebenso sollten Sie Ihre Daten spätestens an dieser Stelle in offene Dateiformate übertragen, um so die langfristige Lesbarkeit und Nutzbarkeit zu erhöhen.

Speicherstandards zur Daten- und Dateiorganisation

Datenqualität

Eine planvolle Daten- und Dateiorganisation ist essentiell für die Datenqualität, die Aspekte wie die Vollständigkeit und die Richtigkeit der Daten, aber auch Kriterien wie das Datenformat umfasst.

Intrinsische Datenqualität

- *Objektivität und Genauigkeit:* Daten sollen unvoreingenommen und neutral sein.
- *Reputation und Glaubwürdigkeit:* Die Quelle der Daten beeinflusst ihre Qualität, demnach müssen z.B. befragte Personen echt und authentisch sein, während Feldforschungsanbietende Vertrauenswürdigkeit und Seriosität vorweisen müssen.

Kontextuelle Datenqualität:

Manche Aspekte der Datenqualität können nur vor dem Hintergrund der jeweiligen Aufgabe beurteilt werden und erfordern deshalb eine besonders gründliche Planung vor der Durchführung eines Forschungsprojekts.

- *Vollständigkeit:* Daten sollen alle relevanten Informationen enthalten, um eine umfassende Analyse zu ermöglichen.
- *Aktualität:* Daten müssen als Grundlage genauer Schlussfolgerungen stets aktuell sein.
- *Relevanz:* Daten müssen im spezifischen Forschungskontext relevant sein.

Datenorganisation

Um die Daten nachvollziehbar zu halten, sie leichter wiederzufinden und auf diese Weise zugleich Verzögerungen im Forschungsprozess und Mehrarbeit zu vermeiden, sollten Sie von Beginn an eine konsistente Datenorganisation zu grunde legen, die Regeln enthält und Verantwortlichkeiten benennt. Dazu gehört eine klare Ordnerstruktur, Ontologien und Konventionen zur Ordner- und Dateibenennung, die von allen Projektbeteiligten eingehalten werden, wie auch die Berücksichtigung besonderer Anforderungen an die Datenorganisation bspw. im Falle sehr großer Datenmengen, gesondert geltenden Datenschutzrichtlinien im Umgang mit personenbezogenen Daten und Regelungen der Zugriffsrechte innerhalb einer Projektgruppe.

- Ordnerstruktur: Im Sinne einer klaren und konsistenten Ordnerstruktur ist es ideal, eine nachvollziehbare, hierarchische Datengliederung mit maximal drei Unterordner-Ebenen vorzunehmen. Daten, die strukturell oder inhaltlich zusammengehören, sollten gemeinsam abgelegt und der Ordner dementsprechend systematisch und inhaltsbezogenen benannt werden. Überprüfen Sie nach dem Projektende, welche Daten, und damit welche Ordner, noch gebraucht und welche gelöscht werden können, damit die Ordnerstruktur möglichst einfach und verständlich bleibt.
- Dateikonventionen: Konventionen zur Datei- und Ordnerbenennung sowie die Bedeutung benutzter Abkürzungen sollten sorgfältig gewählt, plausibel festgelegt und ebenso gut dokumentiert werden. Verwenden Sie aussagekräftige Komponenten wie z.B. Erstellernamen, Versionen und Bearbeitungsstadium und achten Sie darauf, die Benennung konsistent zu halten.

Datumsformate sind in der Form JJJJMMTT anzugeben, Nummern immer zweistellig oder dreistellig mit 0 als Platzhalter und verschiedene Versionen einer Datei als V mit Nummerierung (V01, V02 etc.). Leerzeichen, Punkte, Sonderzeichen, Umlaute, zu lange Dateinamen sind nicht empfehlenswert, stattdessen sollten Großbuchstaben und Unterstriche genutzt werden. Vermeiden Sie Wiederholungen von Informationen aus Ordnernamen in den Dateinamen.

Projektname
Musikkultur_Deutschland
1. Ebene: Kategorie
/[Ordnernummer]_[Datenstatus]
/01_Rohdaten
/02_Prozessierte Daten
...
2. Ebene: Unterkategorie
/01_Rohdaten/[Zeitraum]_[Daten]_[Ersteller]
/01_Rohdaten/2025_Interviews_Schmitt
/01_Rohdaten/2025_Interviews_Meier
...
3. Ebene: Datei
/01_Rohdaten/2025_Interviews_Schmitt/[YYYYMMDD]_[Region]_[Interviewpartner]_[Thema].[Dateiformat]
/01_Rohdaten/2025_Interviews_Schmitt/20250326_Bayern_Müller_TraditionelleMusik.mp3

- **Versionskontrolle:** Wenn mit Dateiversionen gearbeitet wird, um Entwicklungsschritte besser verfolgen zu können, führen Sie regelmäßig eine Versionskontrolle durch, legen Sie obsolete Versionen nach einem Backup separat ab oder löschen Sie diese bei Bedarf. So verringern Sie das Risiko, mit veralteten Dateien weiterzuarbeiten. Eine klare Kennzeichnung der Versionen im Dateinamen in Form von V01, V02 etc. oder in einer standardisierten Kopfzeile innerhalb der Datei schafft eine einfache und schnelle Übersicht. Sie können unterstützend auch mit Versionskontroll-Software arbeiten wie z.B. Git und Apache Subversion.

Dokumentation

Eine gut geführte, genaue und stetig aktualisierte Dokumentation hilft dabei, Daten wiederzufinden, ihre Entstehung nachzuvollziehen, Kontextinformationen für die Interpretation zu liefern und Verwechslungen mit alternativen Versionen zu vermeiden. Darüber hinaus gehört eine akkurate Dokumentation zur [guten wissenschaftlichen Praxis](#) (GWP, §9, S.6).

Halten Sie in Ihrer Dokumentation neben einer dauerhaften Kontaktangabe und allgemeinen Informationen zu Ihrem Projekt auch Angaben zu Erhebungsmethoden, Struktur und Beziehung der Daten, Qualitätsmaßnahmen, ggf. ein Codebook, Dateiversionen sowie Informationen zum Zugang, Vertraulichkeit und Nutzungsbedingungen fest.

Zur Dokumentation kann beispielsweise ein Datenmanagementplan (DMP), eine ReadMe-Datei, eine gut strukturierte Metadatenbank oder auch ein projektinternes Wiki genutzt werden. ReadMe-Dateien sind dabei eine besonders einfache Form der Dokumentation, die Details zu Daten, Ordnern und ggf. Änderungen bereitstellen. Im Unterschied zu Metadaten, die für die maschinelle Analyse gedacht sind, ist das Ziel von ReadMe-Dateien, anderen Forschenden eine klare Übersicht und wichtige Kontextinformationen zu den Daten zu bieten.

Unabhängig davon, ob die (Meta-)Daten automatisiert oder manuell erfasst wurden, ist beim Speichervorgang die Verwendung standardisierter und offener Formate, fachspezifischer Vokabulare und gebräuchlicher Programmiersprachen wichtig. Verwenden Sie zudem strukturierte Metadaten, damit Ihre Forschungsdaten maschinenlesbar sind, und achten Sie dabei auf eine eindeutige Verknüpfung der Metadaten mit dem jeweiligen Datensatz über persistente Identifikatoren (IDs wie bspw. DOI).

Dateiformate

Für die langfristige Nutzbarkeit und die Langzeitarchivierung ist es von außerordentlicher Wichtigkeit, bei der Formatauswahl auf die Kompatibilität (Interoperability) sowie auf die verlustfreie Konvertierung in alternative Formate zu achten. Daher werden insbesondere offene, nicht-proprietary Formate empfohlen.

Die Eignung bestimmter Dateiformate ist immer abhängig von der verwendeten Software und den spezifischen Zielen der Forschungsprojekte, allerdings können bereits zahlreiche Dateiformate problemlos in offene Formate umgewandelt werden. Ferner geben Dateiformate Auskunft über die Struktur der in der Datei gespeicherten Daten sowie über deren Zweck und Zugehörigkeit.

This README.txt file was generated on 2024-02-19 by Naiara Korta Martiartu (naiara.korta@unibe.ch)

----- GENERAL INFORMATION-----

Contributor information:

Name: Naiara Korta Martiartu
Role/Function: Principal Investigator, Data Collector (main contact person)
Institution: Institute of Applied Physics, University of Bern
Address: Siderstrasse 5, 3012 Bern, Switzerland
Email: naiara.korta@unibe.ch

Name: Parisa Salemi Yolgunlu
Role/Function: Data Collector
Institution: Institute of Applied Physics, University of Bern
Address: Siderstrasse 5, 3012 Bern, Switzerland

Name: Martin Frenz
Role/Function: Sponsor
Institution: Institute of Applied Physics, University of Bern
Address: Siderstrasse 5, 3012 Bern, Switzerland

Name: Michael Jaeger
Role/Function: Principal Investigator (alternative contact person)
Institution: Institute of Applied Physics, University of Bern
Address: Siderstrasse 5, 3012 Bern, Switzerland
Email: michael.jaeger@unibe.ch

Date of data collection: December 2022 - May 2023

Location of data collection: Institute of Applied Physics, University of Bern, Bern, Switzerland

Keywords: pulse-echo ultrasound, attenuation imaging, tissue characterization, ultrasound tomography,

Abb.1: Auszug aus einer [ReadMe-Datei](#)

Dateityp	empfehlenswerte Formate	weniger empfehlenswerte Formate
Tabellendaten	CSV, TSV, SPSS portable	Excel (.xlsx/.xls)
Text	TXT, HTML, RTF, PDF/A	Word (.docx/.doc)
Präsentation	PDF/A	PowerPoint (.pptx/.ppt)
Multimedia	MP4, WAV, AVI	WMV, QuickTime, H264
Bilder	TIFF, JPEG2000, PNG	GIF, JPG
strukturierte Daten, Datenaustausch	XML, RDF, JSON	SDXF, RDBMS

Insbesondere das CSV-Format ist aufgrund seiner Offenheit und Kompatibilität eine gute Wahl. Im Gegensatz dazu ist das XLS-Format von Microsoft proprietär und dadurch weniger zugänglich.

Metadatenschemata

Metadaten sind „Daten über Daten“, d.h. sie bieten strukturierte Informationen über die vorliegenden Daten, damit Mensch und Maschine die Daten besser verstehen und verarbeiten können.

Unterschiedliche Arten von Metadaten (bspw. bibliografische, administrative oder deskriptive) erfüllen dabei verschiedene Funktionen. Sie dokumentieren u. a., wer die Daten erhoben, bearbeitet und publiziert hat, mit welchen Methoden die Daten generiert wurden oder unter welchen rechtlichen Bedingungen sie verwendet werden können. Metadatenstandards erleichtern die Einhaltung von Vorgaben, wie etwa der FAIR-Prinzipien, und stellen sicher, dass zentrale Elemente beschrieben werden.

Metadaten können sowohl fachübergreifende als auch fachspezifische Eigenschaften haben. Vorteilhaft ist es, sich für die Datenstrukturierung an Metadatenschemata der eigenen Fachcommunity zu orientieren.

Zusätzliche Hinweise & Wissenswertes

- Das bloße Speichern der Daten auf Ihrem eigenen Rechner oder einem externen Speichermedium stellt noch keine fachgerechte Archivierung nach Projektende dar. Empfehlenswerter ist dagegen die Ablage in einem fachspezifischen oder generischen Repository.
- Zur Unterstützung der Planung, Organisation und Verwaltung Ihrer Daten können Sie die [RDMO-Instanz](#) (Research Data Management Organiser) von [forschungsdaten.info](#) oder das [GF BIO DMP Tool](#) verwenden.
- Der Forschungsdatendienst [OstData](#) (in Deutsch) und die [Cornell University](#) (in Englisch, inkl. Anleitung) bieten zur Erstellung einer ReadMe-Datei auch ein Template an.
- Bei der Auswahl der zu veröffentlichten bzw. zu archivierenden Forschungsdaten können Sie die Checkliste „[Five steps to decide what data to keep](#)“ des Digital Curation Center (DCC) nutzen.
- Für die Umbenennung großer Mengen an Dateien und Ordnern können Sie unterstützende Software nutzen, z.B. [Ant Renamer Portable](#), [ReNamer](#) (für Mac), [GPRename](#) (für Linux).
- Ein Standard für bibliografische Metadaten ist z.B. das [DataCite Metadata Schema](#); für administrative Metadaten in der Langzeitarchivierung gibt es z.B. [PREMIS](#) und für fachspezifische Metadaten existiert eine Vielzahl von Standards. Eine Übersicht über existierende Metadatenstandards bieten: [Metadata Standards Catalog der Research Data Alliance](#), [FairSharing.org](#) und [DDC](#) (Digital Curation Centre).



Links

- Übersicht über [freie Versionsverwaltungssoftware](#)
- Weitere Beispiele für die [Dateibenennung](#) und Hinweise zur [Dokumentation](#) bei VerbundFDB
- Beispiel für einen [Daten- und Methodenbericht](#) und Beispiel für einen [Datensatz](#) mit kürzerer ReadMe-Datei
- Informationen über die langfristige Nutzbarkeit von Dateiformaten bei [KOST](#) (Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen) und [Landesinitiative Langzeitverfügbarkeit NRW](#)
- Siehe auch die übrigen Handreichungen aus der AG Forschungsdatenmanagement