

HANDOUT
FORSCHUNGSDATENMANAGEMENT
IM SFB 1369:
GRUNDLAGEN

Inhaltsverzeichnis

Inhaltsverzeichnis.....	1
I. Einführung.....	2
A. Was bedeutet Forschungsdatenmanagement?	2
B. Grundprinzipien des Forschungsdatenmanagements	2
C. Forschungsdatenmanagement im Forschungsalltag	5
II. Tipps für Forschungsdatenmanagement im Forschungsalltag.....	6
A. Ordnerstruktur und Dateibenennung.....	6
B. Dokumentation	7
C. Daten sicher speichern	8
1. Die 3-2-1-Regel	8
2. Sicher speichern an der LMU.....	9
D. Sicherungszeitplan	12
E. Versionierung.....	12
F. Wahl von Software und Dateiformat.....	13
III. Glossar	15
Datenmanagementplan	15
Dokumentation.....	15
FAIR-Prinzipien.....	15
Forschungsdaten.....	16
Forschungsdatenmanagement.....	16
Laufwerk	16
LRZ - Leibniz Rechenzentrum	16
Metadaten.....	16
Versionierung	16
IV. FDM-Richtlinien und Einrichtungen an der LMU.....	17
V. Lizenz	18

I. Einführung

In diesem Handout möchte ich Ihnen die wichtigsten Prinzipien des FDM vorstellen und erläutern, und Ihnen anschließend Tipps für das Forschungsdatenmanagement in der Forschung geben. Am Ende des Dokuments finden Sie ein Glossar mit Definitionen gängiger Begriffe des FDM sowie Information zu den FDM-Einrichtungen an der LMU.

A. Was bedeutet Forschungsdatenmanagement?

Das Wort kann furchteinflößend wirken. Forschungsdatenmanagement (FDM) bezieht sich grundsätzlich auf Datenmanagement. Das machen die meisten aber schon in ihrem Alltag, wenn sie Urlaub- und Familienfotos sichern oder Rechnungen in einem Ordner zuhause ablegen. Das ist auch Datenmanagement.

Forschungsdatenmanagement bezeichnet konkreter, wie man mit **Forschungsdaten** umgeht: Die meisten Forschenden haben wahrscheinlich ein System um ihre Daten zu speichern und zu finden. Forschungsdatenmanagement umfasst entsprechend die Aufbereitung, die Speicherung, die Archivierung und gegebenenfalls die Veröffentlichung von Forschungsdaten.

So definiert die Deutsche Forschungsgemeinschaft (DFG) Forschungsdaten:¹

„Zu Forschungsdaten zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden.“

Forschungsdaten sind also alle Daten, die im Rahmen der Forschung verwendet oder generiert werden. Forschungsdaten können auch Notizen sein, Exzerpte von Texten, Bilder, Aufnahmen von Interviews, Zeitungsartikel usw.

Alle Forschenden im SFB generieren Forschungsdaten und müssen im Alltag mit diesen umgehen. Deshalb kann ein gutes Forschungsdatenmanagement die Arbeit in vielen Bereichen erleichtern!

B. Grundprinzipien des Forschungsdatenmanagements

Forschungsdatenmanagement ist derzeit vielen Forschungseinrichtungen ein sehr wichtiges Thema. Aufgrund des Internets und digitaler Arbeitsweisen gibt es immer mehr geteilte Daten. Zudem wird es immer einfacher Forschungsergebnisse zu veröffentlichen, zu finden, und mit anderen Forschenden auszutauschen.

¹ Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten. In: *Deutsche Forschungsgemeinschaft*, (30.09.2015), <https://www.dfg.de/resource/blob/172112/23826608514d73da82622c0a16c842db/leitlinien-forschungsdaten-data.pdf>, [letzter Zugriff: 17.06.2024].

Als Forschende an der Universität fühlen wir uns dem Humanismus verpflichtet, der sich durch folgende Werte auszeichnet: Aufgeschlossenheit für neue Ideen, Vernunft und Verstand, kritische Bewertung von Ideen und deren Austausch sowie Mündigkeit als selbstständige denkfähige Individuen.

Was bedeuten diese Werte heute, in einer immer stärker vernetzten Welt? Einerseits muss man an die gute wissenschaftliche Praxis denken.²

Dazu kommt auch Forschungsdatenmanagement ins Spiel. In vielen Kreisen, vor allem an den Universitäten, wird darüber nachgedacht. Daher sind die FAIR Prinzipien eingeführt worden, die ein Ideal des FDMs darstellen. Als Teil der Unterzeichner der Sorbonne-Erklärung für offene Forschungsdaten im Jahr 2020³ verpflichtet sich die LMU, diesen Prinzipien zu folgen.

Die **FAIR-Prinzipien** sagen, dass die Forschungsdaten in der Forschung FAIR sein sollen.⁴ Was bedeutet das konkret?

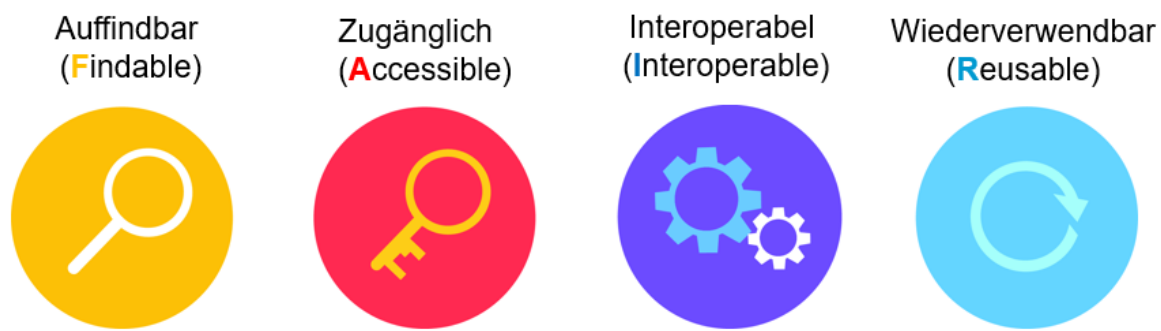


Abbildung 1:

Die FAIR-Prinzipien.

Die DFG und die LMU erwarten, dass so viele Forschungsdaten veröffentlicht werden, wie es sinnvoll und machbar ist. Wie der Grundsatz der Europäischen Kommission zu Open Data es sagt, sollte die Veröffentlichung von Forschungsdaten folgendermaßen erfolgen:⁵

"As open as possible, as closed as necessary!"

² Mehr über die gute wissenschaftliche Praxis an der LMU und die Richtlinien der LMU zur Selbstkontrolle in der Wissenschaft können Sie hier finden: Verfassen einer wissenschaftlichen Arbeit. In: *Universitätsbibliothek der Ludwig-Maximilians-Universität München* <https://www.ub.uni-muenchen.de/schreiben/wissenschaftliche-arbeit/index.html> [letzter Zugriff: 27.08.2024].

³ Sorbonne Declaration on research data rights. In: *League of European Research Universities*, <https://www.leru.org/files/Sorbonne-declaration.pdf> [letzter Zugriff: 18.06.2024].

⁴ FAIRe Daten. In: *forschungsdaten.info*, <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/> [letzter Zugriff: 15.04.2024]. Dieses Werk steht unter der Lizenz [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) und wurde von Paulina Halina Sieminska erstellt.

⁵ Open Science und Open Data. In: *Bundesministerium für Bildung und Forschung*, <https://www.horizont-europa.de/de/Open-Science-und-Open-Data-1767.html> [letzter Zugriff: 20.08.2024].

Die Veröffentlichung von Forschungsdaten erfolgt immer freiwillig und in Kooperation mit den Forschenden: Es gibt viele Möglichkeiten, den Zugang zu den Forschungsdaten nach Bedarf einzuschränken.

Forschungsdatenmanagement ist aber für Forschende immer wichtig, selbst wenn ihre Forschungsdaten nicht veröffentlicht werden.

In der Welt des Forschungsdatenmanagement wird öfter vom Lebenszyklus der Forschungsdaten gesprochen. Dieser **Forschungsdaten-Lebenszyklus** ist in dieser Grafik zu sehen:⁶

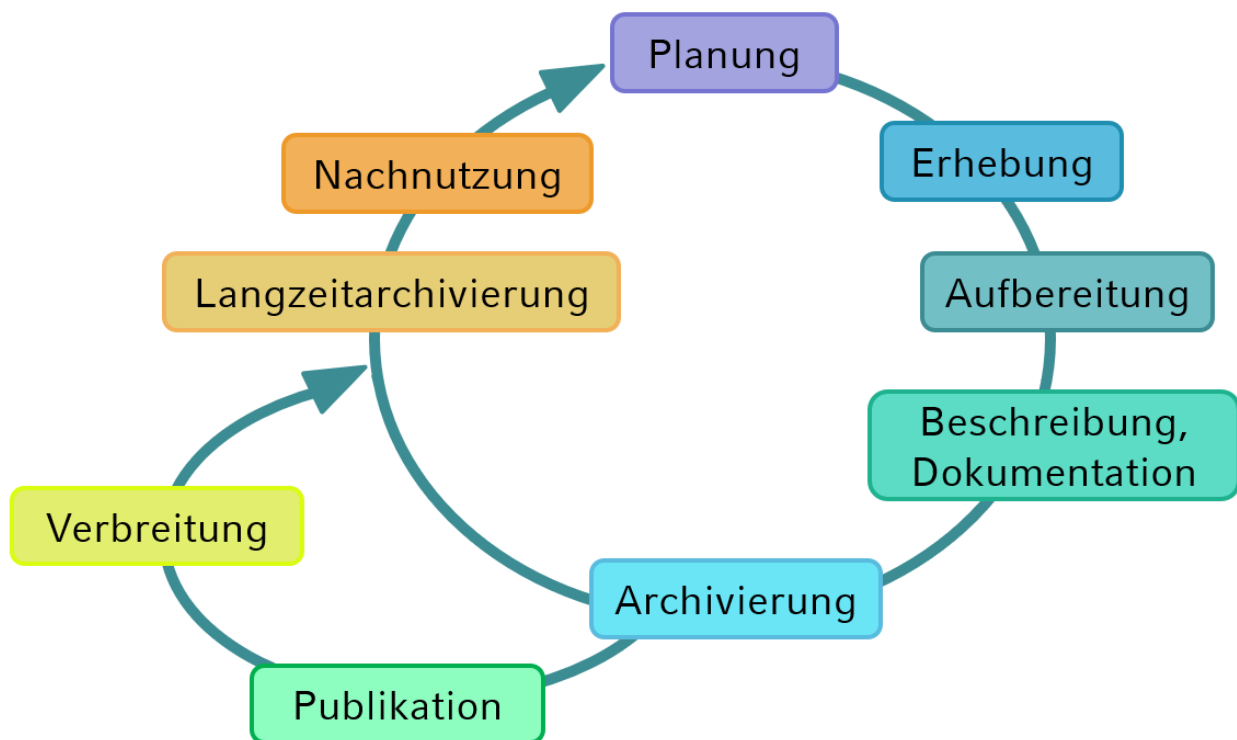


Abbildung 2:

Der Forschungsdaten-Lebenszyklus.

Forschungsdatenmanagement ist auf jeder Stufe des Lebenszyklus wichtig. Ohne ein System für die Datenspeicherung kann die Datenerhebung ein chaotischer Prozess sein. Das gilt auch für die Aufbereitung und Beschreibung der Daten. Eine Veröffentlichung oder Nachnutzung der Daten kann nur erfolgen, wenn diese langfristig und in einer Weise gesichert sind, die ihre Nachnutzung ermöglicht. Deshalb ist Forschungsdatenmanagement für alle Forschenden wichtig.

⁶Abbildung von Laure Sigalla, basierend auf: Rohrwild, Jürgen: Forschungsdaten-Lebenszyklus. In: *Forschungsdatenmanagement Bayern*, <https://www.fdm-bayern.org/forschungsdatenmanagement/#fd-lebenszyklus> [letzter Zugriff: 03.04.2024].

C. Forschungsdatenmanagement im Forschungsalltag

Zwei Prinzipien sind für Forschende wichtig, um ihre Forschungsdaten sicher zu verwalten und zu bewahren.

Das erste ist **die systematische Organisierung der Daten**. In einem Archiv oder in einer Bibliothek findet man immer, was man sucht, weil die Dokumente und Bücher dort systematisch katalogisiert werden.

Um Forschungsdaten optimal zu verwenden, brauchen Forschende ein System, das ihnen ermöglicht, die folgenden Fragen leicht zu beantworten:

- ➔ Wo liegen Ihre Daten?
- ➔ Denken Sie, dass Sie in 1, 3, 5 Jahren diese Daten finden können? Werden Sie die Daten dann immer noch bearbeiten können?
- ➔ Werden Sie immer noch verstehen, wie die Daten gespeichert sind und warum?
- ➔ Könnten Teammitglieder mit Ihren Daten klarkommen, wenn das nötig wäre?

Diese Überlegungen sind sehr wichtig, um organisiert mit Ihren Daten umgehen zu können.

Das zweite wichtige Prinzip ist **die Sicherung der Daten**. Mit der heutigen Technologie wird es immer einfacher, große Mengen an Daten zu speichern. Die DFG empfiehlt, Forschungsdaten für mindestens 10 Jahre zu archivieren.⁷ Jedoch müssen Forschende beachten, dass sie diese Daten später noch brauchen könnten, d.h. eventuell auch erst in 20 oder 30 Jahren.

Die richtige Sicherung Ihrer Daten garantiert ein langlebiges Organisationssystem Ihrer Forschungsdaten und verhindert deren Verlust.

Die Organisierung und Sicherung Ihrer Forschungsdaten sind also die wichtigsten Prinzipien für Ihren Forschungsalltag. Im Folgenden werden Empfehlungen gegeben, die Sie dabei unterstützen können, diese Prinzipien in Ihren Forschungsalltag zu implementieren.

⁷ Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten. In: *Deutsche Forschungsgemeinschaft*, (30.09.2015), <https://www.dfg.de/resource/blob/172112/23826608514d73da82622c0a16c842db/leitlinien-forschungsdaten-data.pdf>, [letzter Zugriff: 20.08.2024].

II. Tipps für Forschungsdatenmanagement im Forschungsalltag

A. Ordnerstruktur und Dateibenennung

Ein wichtiges Element der Datenorganisation auf einem Computer ist die **Ordnerstruktur** und die **Dateibenennung**.

→ Ordnerstruktur

Heutzutage arbeiten die meisten Forscherinnen und Forscher digital. Daher werden Sie den Großteil Ihrer Daten auf Ihrem Computer, auf einem Laufwerk oder auf einem anderen digitalen Speicherplatz ablegen.

Generell werden Sie diese in Ordner anlegen. Idealerweise sollten Sie Ihre Ordner so anlegen, dass die Daten gut auffindbar sind. Wenn diese Ordner oder Daten von mehreren Personen innerhalb Ihres Teilprojektes verwendet werden sollen, sollten sich alle (neuen) Teilprojekt-Mitglieder problemlos darin zurechtfinden können.

Tipps für die Ordnerstruktur:

- Daten in Ordnern ablegen.
- Systematische, inhaltsbezogene Ordnerbenennung verwenden.
- Nicht mehr als drei Unterordner-Ebenen verwenden.
- Nach Projektende prüfen, was noch gebraucht wird.

→ Dateibenennung

Innerhalb Ihres Ordners werden Dateien gespeichert sein. Wenn Dateien klar benannt werden, gibt das dem Inhalt des Ordners Struktur.

Die Struktur, die Sie in ihren Ordner einführen können, hängt sehr von der Art des Projekts ab. Deshalb sind allgemeingültige Aussagen schwierig. Das Wichtigste ist, dass Sie ein System haben, und dass Sie wissen, wie Sie dieses gestaltet haben.

Tipps für die Dateibenennung:

- Sprechende Dateinamen (mit aussagekräftigen Inhaltsschlagwörtern) verwenden.
- Eindeutigkeit herstellen, z.B. über:
 - Datierung (JJJJMMTT) am Ende oder am Anfang der Dateiname;
 - ggf. Versionskennzeichnung und/oder Bearbeiterkürzel;
 - Namen mit _ separieren;
 - Vorangestellte Nummerierung (bei einstelligen Zahlen mit 0 beginnend: 01, 02).
- Dateiname sollte objektiv, intuitiv und personenunabhängig nachvollziehbar sein.
- Leerzeichen und Sonderzeichen können in verschiedenen Programmen zu Fehlern bei der Verarbeitung führen. Sonderzeichen sind Symbole und nicht übliche Buchstaben, wie das ß, das * usw.

B. Dokumentation

Die Webseite forschungsdaten.info sagt es am besten:⁸

„Ohne Dokumentation laufen Sie Gefahr

- *Daten nicht wiederzufinden,*
- *die Entstehung von Daten nicht mehr nachvollziehen zu können,*
- *Daten wegen fehlender Kontextinformationen nicht mehr interpretieren zu können,*
- *Dateien zu verwechseln (veraltete oder konkurrierende Versionen),*
- *Daten nicht mit anderen Personen austauschen oder mit Daten aus anderen Quellen zusammenführen zu können.“*

Sie können Ihre Daten dokumentieren mithilfe:

- eines **Datenmanagementplans**,
- eines persönlichen Arbeitstagebuchs,
- einer Erklärungsdatei, wo die Daten gespeichert sind.

Der SFB 1369 muss die Vorgabe der Deutschen Forschungsgemeinschaft, dass jedes Teilprojekt sein Forschungsdatenmanagementverfahren dokumentieren muss, erfüllen. Das wird in unserem SFB durch die Erstellung eines Datenmanagementplans für jedes Teilprojekt erfolgen. Dieser wird mithilfe des Research Data Management Organizer (RDMO) der Universitätsbibliothek der LMU erstellt. Mit ihm wird Folgendes dokumentiert: Projekttitel, das Forschungsvorhaben, wie die Daten gesammelt oder generiert werden, mit welcher Software, usw.

⁸ Datendokumentation In: forschungsdaten.info, <https://forschungsdaten.info/themen/beschreiben-und-dokumentieren/datendokumentation/> [letzter Zugriff: 17.06.2024].

Auf einer kleineren Ebene kann es aber für Sie nützlich sein, einiges selbst zu dokumentieren. Nehmen wir das Beispiel von digitalisierten Archivmaterialien. Es kann sehr wichtig sein zu wissen: Wer hat die Bilder/Scans gemacht? Wann? Wie? Oder für ein aufgenommenes Interview: Wer hat das Interview geführt? Wann, wo? Wer hat die Transkription gemacht?

Diese Daten über Daten nennt man **Metadaten**.

Daher sollte es Teil Ihres Arbeitsprozesses sein, diese relevanten Elemente aufzunehmen. Es sollte immer so gemacht werden, dass beispielsweise auch neue Teammitglieder in 5 Jahren noch wissen, was gemeint war. Zum Beispiel sollten Sie Kürzungen und fach- oder ortsspezifisches Vokabular immer irgendwo dokumentieren.

Auf einer persönlichen Ebene kann das durch ein Logbuch, ein Arbeitstagebuch, erfolgen.

Auf der Ebene des Teilprojekts kann es hilfreich sein, erklärende Dateien zusammen mit den Forschungsdaten zu speichern. Zum Beispiel könnte innerhalb eines Ordners, in dem digitalisierte Quellen gespeichert sind, auch eine Datei stehen, in der – bei der Einfügung eines Scans in den Ordner – immer zusammengefasst wird, welche Teammitglieder die digitalisierten Quellen digitalisiert haben, wo, wann, und wie. An dieser Stelle könnten auch Abkürzungen gespeichert werden, die z.B. in der Dateibenennung verwendet wurden.

Wenn Sie solche Dateien in Ihre Ordnerstruktur integrieren und die Gewohnheit entwickeln, diese auszufüllen, können diese Informationen für Sie in mehreren Jahren sehr wertvoll sein.

Versuchen Sie, daran zu denken, wie Sie das in Ihren Forschungsalltag integrieren könnten, und wie Sie Ihren Arbeitsverlauf gestalten könnten, so dass die Dokumentation ein Teil davon und minimal aufwändig wird.

C. Daten sicher speichern

1. Die 3-2-1-Regel

Es ist der Alptraum aller Forschenden, ihre Forschungsdaten zu verlieren. Deshalb ist es wichtig, sie so zu sichern, dass die Gefahr eines Datenverlustes durch ein Unglück viel geringer wird.

Die Empfehlung der FDM-Welt ist es, der **3-2-1-Backupregel** zu folgen, die Sie hier illustriert sehen können:⁹

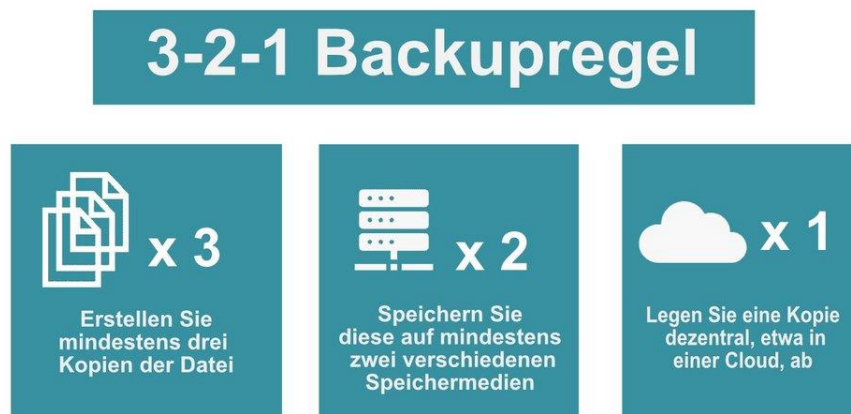


Abbildung 3:

Die 3-2-1-Backupregel

Mit 3 Kopien, auf verschiedenen Speichermedien und einer dezentralen Kopie, ist man sehr sicher. Sollten Sie Ihren Laptop und Ihre Festplatte gleichzeitig verlieren, auf denen Sie zwei von den drei Kopien haben: Kein Problem, Sie haben dann noch eine Kopie in einer Cloud.

Im SFB 1369 könnte die 3-2-1-Regel so verwendet werden: Sie haben eine Kopie lokal auf Ihrem Arbeitscomputer, eine Kopie auf einer Festplatte und eine Kopie dezentral auf Ihrem individuellen LMU-Laufwerk.

2. Sicher speichern an der LMU

Das bringt uns zu unserem nächsten Punkt: wie Sie an der LMU sicher speichern können. Zunächst ist es wichtig, zu verstehen, wie die Laufwerke funktionieren, die wir an der LMU verwenden.

Die wichtigste Frage lautet zuerst: Was ist ein Laufwerk?

„Ein Laufwerk [...] ist ein Gerät für den schreibenden und/oder lesenden Zugriff auf ein Speichermedium für digitale Daten.“¹⁰

Das bedeutet, dass, wenn Sie einen Cloud Service nutzen, die Daten immer irgendwo auf einem physischen Objekt gespeichert sind. Wenn Sie beispielsweise

⁹ 3-2-1 Backupregel. In: *forschungsdaten.info*, <https://forschungsdaten.info/themen/speichern-und-rechnen/datensicherheit-und-backup/> [letzter Zugriff: 17.06.2024].

¹⁰ Wikipedia Mitwirkenden: Laufwerk (Computer). In: *Wikipedia – Die freie Enzyklopädie*, [https://de.wikipedia.org/w/index.php?title=Laufwerk_\(Computer\)&oldid=2432612](https://de.wikipedia.org/w/index.php?title=Laufwerk_(Computer)&oldid=2432612) [letzter Zugriff: 17.06.2024].

Google Cloud in Europa nutzen, liegen ihre Daten entweder in Frankfurt, Belgien oder den Niederlanden.¹¹

Die Sicherheit der Daten, die in einer Cloud gespeichert sind, hängt von der Sicherheit der Computer ab, auf denen diese Daten liegen. Das gilt auch für die LMU-Laufwerke.

Zur Erinnerung haben alle SFB-Mitglieder Zugriff auf:

- ein individuelles Laufwerk, wie alle LMU-Mitglieder und Studierenden;¹²
- das SFB-Laufwerk, auf dem jedes Teilprojekt einen Ordner hat, in dem Daten gespeichert werden können;
- LRZ Sync+Share, vor allem empfohlen, wenn Sie häufig mit mobilen Geräten arbeiten, viel unterwegs sind und oft keinen Internetzugang haben oder Daten externen Personen außerhalb der LMU zugänglich machen möchten.

Alle diese Daten werden physisch auf einem Computer gespeichert. Da kommt das LRZ ins Spiel. Was ist dann das LRZ?

Das **Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften (LRZ)** unterstützt die LMU und andere Münchner akademische Einrichtungen mit einem IT-Angebot: mit Geräten, WLAN und Speicherplatz in Form von Laufwerken. Es befindet sich in Garching, im Norden von München, ca. 16km vom Hauptgebäude der LMU entfernt und ist mit der U6 gut erreichbar.

Das Gebäude des LRZs ist hier zu sehen:¹³



Abbildung 4:

Das Gebäude des Leibniz-Rechen-Zentrums in Garching, Bayern.

¹¹ Wikipedia Mitwirkenden: Google data centers. In: *Wikipedia, The Free Encyclopedia*, https://en.wikipedia.org/w/index.php?title=Google_data_centers&oldid=1220741486 [letzter Zugriff: 17.06.2024].

¹² Jede Fakultät der LMU bestimmt selbst, wie viel Speicherplatz ihren Angehörigen auf ihren persönlichen Laufwerken zur Verfügung steht.

¹³ Ghost writ0r: Im Vordergrund das Institutsgebäude, hinten das Rechnergebäude. In: *Wikipedia – Die freie Enzyklopädie* (28.08.2016), https://de.wikipedia.org/wiki/Leibniz-Rechenzentrum#/media/Datei:Leibniz-Rechenzentrum_2016-08-15.jpg [letzter Zugriff: 17.06.2024]. Dieses Werk steht unter der Lizenz [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/).

Ein besonderer Teil des Gebäudes ist der Rechnerwürfel, der hier zu sehen ist:¹⁴



Abbildung 5:

Rechnerwürfel des Leibniz-Rechenzentrums.

Dort sind alle Daten aufbewahrt, die auf LMU und TUM-Laufwerken gespeichert sind. Die Webseite des LRZ betont die Sicherheit dieser Einrichtungen:¹⁵

„Das LRZ betreibt als Cloud Storage hochverfügbare, weitgehend skalierbare, leistungsstarke Speichersysteme in zugriffsgesicherten, voll klimatisierten Räumen des Rechenzentrums (Private Cloud). Die Datensicherheit wird durch mehrstufige Sicherungsverfahren gewährleistet.“

sowie:¹⁶

„Der LRZ Cloud Storage befindet sich im Rechnerwürfel des Leibniz Rechenzentrums. Dieses Rechnergebäude ist als ‚dark center‘ ausgelegt, d.h. normalerweise haben Menschen dort keinen Zugang.“

Wenn Sie auf Ihrem individuellen Laufwerk oder auf dem SFB-Laufwerk etwas speichern, sind die Daten sehr sicher in Garching in diesem Würfel gesichert. Es werden regelmäßig Kopien von den Daten gemacht, so dass man Dateien einfacher restaurieren kann, sollte man etwas gelöscht haben (wenn das passiert, sollte man sich aber so schnell wie möglich bei dem LRZ melden!).

Es ist auch klar, wie die Daten gesichert sind, und damit sind wir von kommerziellen Nutzungsbedingungen unabhängig.

Deshalb empfehlen wir im SFB die Nutzung des Serverlaufwerkes um Dateien zu speichern und zu teilen! Und wenn Sie lieber einen anderen Speicherservice nutzen wollen, wird Ihnen empfohlen, das LRZ-Laufwerk zumindest für regelmäßige Sicherungen zu nutzen. Beachten Sie, dass der Zugang zu diesem Speicherservice des LRZs durch Ihre LMU-Kennung stattfindet: Sollten Sie die LMU verlassen, dann haben Sie 3 Monate Zeit, um Daten zu exportieren, die Sie weiter brauchen werden.

¹⁴ Leibniz Rechenzentrum. In: Haver & Boecker, <https://www.haverboecker.com/de/produktloesungen/architekturgewebe/referenzen/detail/leibniz-rechenzentrum/#reference-33-6> [letzter Zugriff: 17.06.2024].

¹⁵ In: Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, <https://doku.lrz.de/cloud-storage-schnelleinstieg-11476150.html> [letzter Zugriff: 17.06.2024].

¹⁶ In: Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, <https://doku.lrz.de/106-was-heisst-denn-sicherer-speicher-10746753.html> [letzter Zugriff: 17.06.2024].

D. Sicherungszeitplan

Was Ihnen helfen kann, Ihre Forschungsdaten regelmäßig zu sichern, ist ein **Sicherungszeitplan**. Überlegen Sie sich:

- ➔ Was sollten Sie sichern und wie oft?
- ➔ Wie oft werden Sie die Kopien sichern können?
- ➔ Welche Frequenz von Sicherung ist für Sie machbar?

Tipps für den Sicherungszeitplan:

- einen Sicherungszeitplan aufstellen, der für Ihren Arbeitsrhythmus sinnvoll ist: Wenn Sie jede Woche eine riesen Datenmenge erzeugen, würde ich empfehlen, jede Woche eine Sicherung zu machen. Wenn Sie weniger Daten erzeugen, kann eine Sicherung pro Monat ausreichend sein.
- einen regelmäßigen Rhythmus bestimmen: z.B. eine Sicherung pro Woche auf dem Laptop, eine pro Woche auf dem Laufwerk, eine pro Monat auf der Festplatte oder jeden Montag oder jeden 1. des Monats.
- Erinnerungen im Kalender und auf dem Handy einstellen!
- Vor dem Urlaub usw. immer sichern: Denken Sie extra daran, zu sichern, wenn Sie eine Weile nicht an Ihrem Arbeitsplatz oder zuhause sein werden. Das erlaubt Ihnen, eine Kopie des letzten Standes Ihrer Arbeit zu haben, und es ist sicherer: Auf Reisen usw. kann mehr passieren.

Sie können zusätzlich zu den Erinnerungen an die Datensicherung in Ihrem Zeitplan auch einbauen, wann Sie z.B. alte Dateien sortieren, wann Sie Ihren Computer aufräumen... Diese Forschungsdatenmanagementaufgaben sollten Teil Ihres Arbeitsablaufes werden.

E. Versionierung

Was bedeutet **Versionierung**? Es bedeutet, dass Sie mit mehreren Versionen einer Datei arbeiten und diese Versionen verwalten.

Für manche Datenbestände kann es nützlich sein, mehrere Versionen der Daten zu behalten, z.B. um sehen zu können, wann welche Daten gesammelt wurden, oder um frühere Versionen von Codes für ein Programm verfügbar halten zu können. Es kann klarer machen, was wann gemacht wurde.

Aber das kann auch zur Verwirrung führen, wenn Sie nicht mehr sicher sind, welche Version die aktuellste ist, wenn das beispielsweise in der Dateibenennung nicht deutlich wird. Um das Problem zu vermeiden, habe ich ein paar Tipps für Sie.

Tipps zur Versionierung:

- Dateien systematisch mit Datum oder mit Nummer einordnen, so dass es immer klar wird, welche Version die aktuellste ist. Z.B. V01 für Version 1 usw.
- Sollten externe Personen Ihre Datei kommentiert haben: sofort umbenennen. Datei_V03_mit_kommentaren kann sofort als Datei_V04 gespeichert werden.
- Speichern Sie nicht die Arbeitsversion Ihrer Dateien zusammen mit den anderen Versionen der Datei in einem gemeinsamen Ordner. Wenn Sie eine neue Version anlegen, können Sie die alten Versionen in einem Unterordner speichern, so dass Sie die beiden Versionen nicht verwechseln.

F. Wahl von Software und Dateiformat

In der Informatik ist ein **Dateiformat** eine besondere Anordnung von Daten in einer Datei, so dass diese Datei bestimmte Eigenschaften hat, z.B. von bestimmten Softwares verarbeitet werden können.

Gängige Dateiformate sind:

- .pdf, .docx, .txt für Textdateien;
- .jpeg, .png für Bilder;
- .mp3 für Tonaufnahmen;
- .xlsx, .csv für Tabellen;
- und viele mehr!

Es gibt tausende Dateiformate, die spezifisch für eine Art von Daten verwendet werden können, oder die für eine bestimmte Software wichtig sind.

Wieso ist das für die Sicherung der Daten wichtig?

Wenn Sie Daten sammeln, werden Sie die meiste Zeit mit einer oder mehreren Softwares arbeiten. Die Wahl einer Software für die Verarbeitung der Daten bestimmt oft das Dateiformat meiner Daten:

- ➔ Für quantitative Daten kann ich Excel verwenden → die Ausgabe ist eine Datei im .xlsx-Format.
- ➔ Meine Notizen kann ich mit Word tippen → die Ausgabe ist eine .docx-Datei.

In der täglichen Anwendung ist das kein Problem. Aber was passiert, wenn Sie alle Ihre Daten in einem Dateiformat sichern, das man nur mit einer bestimmten Software verarbeiten kann? Wenn die Software irgendwann nicht mehr existiert, ist das für Sie ein großes Problem.

In Ihren Sicherungen sollte Ihr Dateiformat für die Nachhaltigkeit Ihrer Daten so sein:

- weit verbreitet und standardisiert;
- nicht proprietär, also nicht von einer Anwendung oder einem Hersteller abhängig, und mit unterschiedlichen Programmen verwendbar;
- offen dokumentiert mit frei verfügbaren technischen Spezifikationen;
- verlustfreie oder keine Kompression;
- einfach dekodierbar oder sogar unmittelbar lesbar, also nicht durch Kodierung versteckt.

Diese Empfehlungen sind eher technisch, und ich erwarte von Ihnen nicht, dass Sie bei jedem Dateiformat wissen, ob es diese Kriterien erfüllt.

Ein paar Beispiele von Dateiformaten, die für Sicherungen gut sind, für:

- statischen Text → .pdf;
- Text → .rtf, .txt;
- Bilder → .tiff, .jpeg, .png;
- Tabellen → .csv.

Es gibt zu viele, um sie alle aufzulisten. Ich empfehle die unten verlinkte Liste des Forschungsdatendienstes OstData.¹⁷

Denken Sie daran, wenn Sie eine Software zur Datenverarbeitung aussuchen. Wichtig ist, dass die Software nicht nur ein proprietäres Format nutzt (z.B. .xlsx für Excel-Dateien), sondern es auch die Option gibt, die Dateien in geeignetere Formate zu exportieren, wie .csv.

So können Sie in Ihren Sicherungen eine Version der Daten speichern, in dem für diese Art von Dateien bzw. Daten bestmöglichen Format. Sie sind dann nicht von einer Software abhängig, sollte diese aufhören zu existieren.

¹⁷ Forschungsdatendienst OstData: Empfohlene Dateiformate für Forschungsdaten zur Langzeitarchivierung. In: *osmikon.de* (20.11.2019), https://www.osmikon.de/fileadmin/uploadMountall/user_upload/2_Dateien_nach_Rubriken_geordnet/4_Forschungsdaten/empfohlene_Dateiformate_fuer_Forschungsdaten_OstData_v1.pdf [letzter Zugriff: 27.01.2025].

III. Glossar

3-2-1-Backupregel

Die 3-2-1-Backupregel gilt als goldene Regel der Datensicherung. Sie empfiehlt die Erstellung von 3 Kopien der Daten auf mindestens 2 verschiedenen Speichermedien und 1 der Kopien soll dezentral gespeichert werden. Deshalb heißt sie die 3-2-1 Regel.

Diese Regel soll garantieren, dass immer eine Kopie der Daten existiert, auch wenn man Pech hat und gleichzeitig zwei Kopien verliert.

Dateiformat

In der Informatik ist ein Dateiformat eine besondere Anordnung von Daten in einer Datei, so dass diese bestimmten Eigenschaften hat, z.B. von bestimmten Softwares verarbeitet werden kann.

Sowohl PDF- als auch .docx-Dateien können Text enthalten. Trotzdem sind sie so unterschiedlich wie eine VHS-Kassette und eine DVD, auch wenn diese beiden Objekte Videos erhalten können.

Datenmanagementplan

Ein Datenmanagementplan (DMP) ist ein Dokument, in dem Forschende oder eine Forschungsgruppe ihren Umgang mit Forschungsdatenmanagement dokumentieren können. Ein DMP kann vielen Formen haben und soll vor allem Metadaten der Forschungsdaten der Forschenden oder der Forschungsgruppe dokumentieren. Zum Beispiel kann man in einem DMP dokumentieren, welche Daten im Rahmen der Forschung generiert und verwendet werden, wo sie gesichert sind, unter welchem Dateiformat, wer Zugriff darauf hat, ob die Forschungsdaten sensible Daten erhalten, welche möglichen rechtlichen Einschränkungen sich daraus ergeben und noch vieles mehr.

Manche Forschungseinrichtungen erwarten, dass ihre Forschenden einen DMP ausfüllen und dafür eine bestimmte Vorlage verwenden.

Dokumentation

Dokumentation bedeutet so viel wie die Zusammenfassung von Informationen über die Forschungsdaten (d.h. Metadaten). Dokumentation kann mithilfe eines Datenmanagementplans, eines Arbeitstagebuchs oder einer erklärenden Datei, die zusammen mit den Forschungsdaten gespeichert werden kann, erfolgen.

FAIR-Prinzipien

Die FAIR-Prinzipien sind Leitlinien für das Forschungsdatenmanagement in der akademischen Welt. Laut den FAIR-Prinzipien sollten Forschungsdaten *Findable* (auffindbar), *Accessible* (zugänglich), *Interoperable* (interoperabel) und *Reusable* (wiederwendbar) sein.

Forschungsdaten

Forschungsdaten sind alle Daten, die im Rahmen der Forschung verwendet oder generiert werden. Forschungsdaten können Notizen, Exzerpte von Texten, Bilder, Aufnahmen von Interviews, Zeitungsartikel, Messdaten, Laborwerte, audiovisuelle Informationen, Erhebungsdaten, Objekte aus Sammlungen oder Proben sein, die im Rahmen einer wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden.

Forschungsdatenmanagement

Forschungsdatenmanagement (FDM) bezieht sich auf Datenmanagement. Das machen alle schon im Alltag, wenn sie Familienbilder sichern oder Rechnungen in einem Ordner zuhause ablegen.

Forschungsdatenmanagement ist die Form, wie man mit Forschungsdaten umgeht: Alle Forschenden haben wahrscheinlich ein System um ihre Daten zu speichern und zu finden. Forschungsdatenmanagement umfasst also die Aufbereitung, die Speicherung, die Archivierung und gegebenenfalls die Veröffentlichung von Forschungsdaten.

Laufwerk

Ein Laufwerk ist ein Gerät für den schreibenden und/oder lesenden Zugriff auf ein Speichermedium für digitale Daten. Daten, die in einer Cloud gespeichert werden, können ebenso auf einem Laufwerk oder auf einem anderen Speichergerät gespeichert werden.

LRZ - Leibniz Rechenzentrum

LRZ steht für das Leibniz Rechenzentrum. Das LRZ unterstützt die LMU, die TU und andere Münchner wissenschaftliche Einrichtungen mit einem IT-Angebot sowie Supercomputing Services.

Metadaten

Als Metadaten bezeichnet man alle zusätzlichen Informationen, die zur Interpretation der eigentlichen Daten, z. B. Forschungsdaten, notwendig oder sinnvoll sind und die eine (automatische) Verarbeitung der Forschungsdaten durch technische Systeme ermöglichen. Kürzer kann man sie als „Daten über Daten“ beschreiben.

Zu den Metadaten eines Fotos zählt beispielsweise: Wann wurde das Foto gemacht, wo, vom wem, mit welchem Gerät usw. Ähnliche Informationen kann man auch über andere Arten von Daten speichern. Metadaten für eine Excel-Tabelle könnten beispielsweise sein: Wer hat die Tabelle ausgefüllt, wann und mit welcher Version von Excel.

Versionierung

Versionierung bezeichnet die Verwaltung von mehreren Versionen einer Datei oder eines Datensatzes. Um Verwirrung durch konkurrierende Versionen einer Datei zu vermeiden, ist es wichtig, die Versionierung konsequent zu führen.

IV. FDM-Richtlinien und Einrichtungen an der LMU

Im SFB 1369 ist die Forschungsdatenmanagementbeauftragte ihre Ansprechpartnerin, wenn Sie Fragen bezüglich des Forschungsdatenmanagement haben. Darüber hinaus gibt es an der LMU Richtlinien und Einrichtungen für das Forschungsdatenmanagement.

Wenn Sie mehr über Forschungsdatenmanagement in den digitalen Geisteswissenschaften an der LMU erfahren wollen, können Sie die Leitfaden für den Umgang mit Forschungsdaten in den digitalen Geisteswissenschaften an der Ludwig-Maximilians-Universität München (LMU)¹⁸ lesen oder sich über die Webseite des FDM-Portals der Universitätsbibliothek der LMU¹⁹ informieren, die Angebote für die Geistes- und Kulturwissenschaften zusammenfasst.

¹⁸ IT-Gruppe Geisteswissenschaften, Universitätsbibliothek der LMU München (Hrsg.): Leitfaden für den Umgang mit Forschungsdaten in den digitalen Geisteswissenschaften an der Ludwig-Maximilians-Universität München (LMU). In: *Open Access LMU* (29. Mai 2024), <https://doi.org/10.5282/ubm/epub.116800> [letzter Zugriff: 26.08.2024].

¹⁹ Geistes- und Kulturwissenschaften. In: *Research Data LMU*, <https://fdm.ub.lmu.de/wb/geistes-kulturwissenschaften/> [letzter Zugriff: 27.08.2024].

V. Lizenz

Dieser Text steht unter der Lizenz Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0), <https://creativecommons.org/licenses/by-sa/4.0/>.

Dieses Handout wurde von Laure Sigalla, Forschungsdatenmanagement-Beauftragte im SFB 1369-Vigilanzkulturen, erstellt. Es liegt nun in der Version 1.0 vom 04.02.2025 vor.