

LEITFADEN 3: DAS DATENMANAGEMENT

Dieser Leitfaden ist eine praktische Anleitung zum Management Ihrer Forschungsdaten und behandelt insbesondere die ethischen Aspekte von Daten (z. B. Schutz der Studienteilnehmenden), die Organisation von Dateien und die Datensicherheit sowie deren Dokumentation.

Was ist Datenmanagement und warum ist es wichtig?

Datenmanagement ist ein allgemeiner Begriff, der zahlreiche Aktivitäten umfasst. Es beinhaltet alle Aspekte der Handhabung, Organisation, Dokumentation und Aufbereitung von Daten sowie alle Aktivitäten, die die Wiederverwendung und den Datenaustausch ermöglichen. Das Datenmanagement bezieht sich sowohl auf quantitative als auch auf qualitative Forschungsdaten (siehe Leitfaden 1 zur Datendefinition).

Ein gutes Datenmanagement ermöglicht Ihnen unter anderem:

- die Studienteilnehmenden zu schützen,
- die Qualität und das Potenzial Ihrer Daten zu erhöhen,
- Ihre Daten und Ihre Dateien langfristig nachzuvollziehen,
- Ihre Daten mit anderen teilen zu können, und
- Ihre Daten zu sichern und vor Verlust zu schützen.

Mehr denn je steht ein gutes Datenmanagement im Zentrum der wissenschaftlichen Forschung. Das Aufkommen neuer Technologien und steigende Speicherkapazitäten haben die Art und Weise, wie geforscht wird, und unser Verhältnis zu Daten revolutioniert.

Heutzutage ist gutes Datenmanagement ein Qualitätsbeweis, aber auch eine Voraussetzung für die Wiederverwendung und den Datenaustausch. Insbesondere die „Open Data“-Bewegung hat mit ihrer Forderung nach dem freien Zugang zu mit öffentlichen Mitteln finanzierten Daten auf der politischen Bühne an Dynamik gewonnen. Die Wiederverwendung von Daten und der Datenaustausch erfolgen häufig über Archive wie dasjenige von [FORS](#) (siehe Leitfaden Nr. 2).



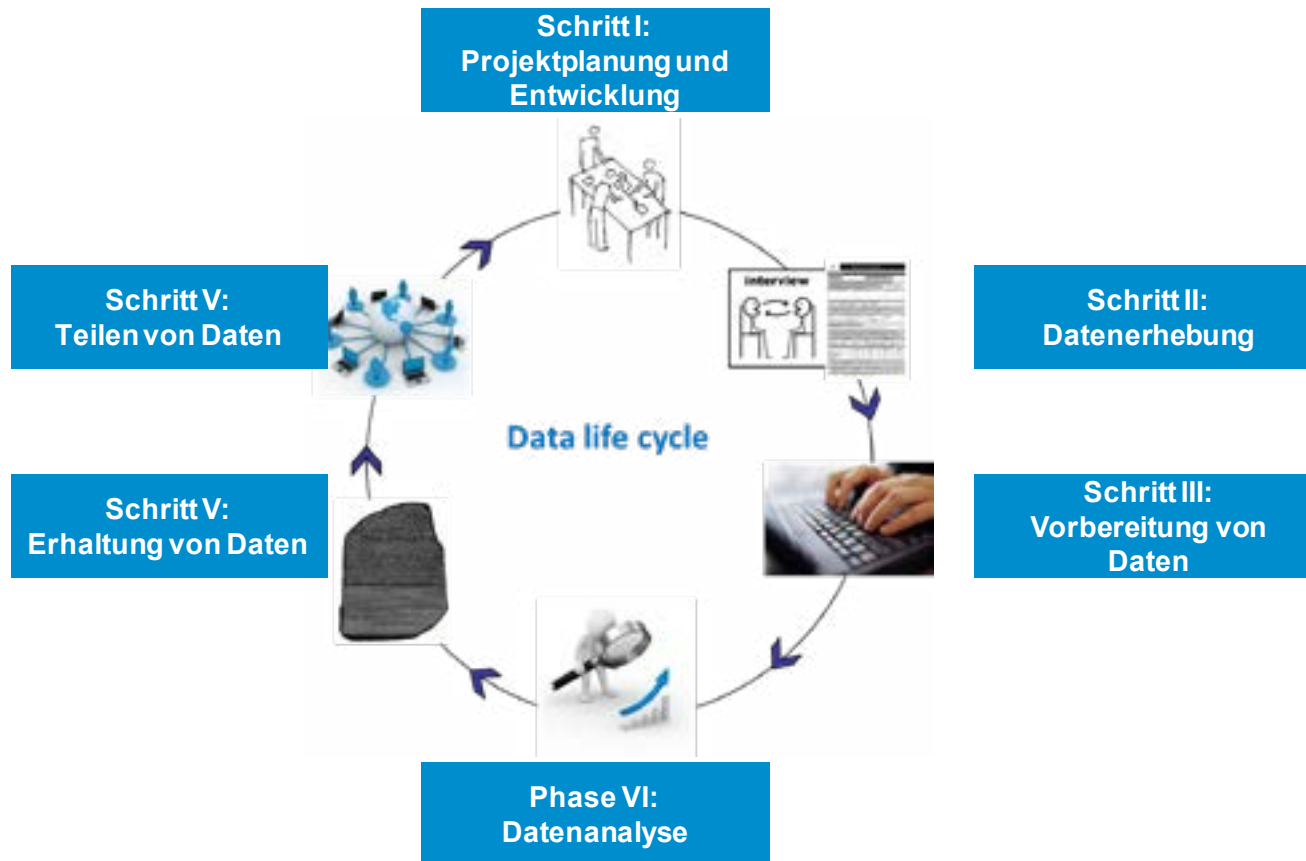
Datenmanagement in Kürze

Die wichtigsten Praktiken des Datenmanagements beziehen sich auf:

- die Sicherheit und die Speicherung der Daten
- die Organisation der Dateien
- die Dokumentation von Dateien und Daten
- die Datenbereinigung und Datenrekodierung
- die Datentranskription
- die Anonymisierung der Daten
- die Einverständniserklärung
- das Urheberrecht
- die Datenfreigabe und den Datenaustausch

DER LEBENSZYKLUS VON FORSCHUNGSDATEN

Für jedes Projekt empfehlen wir die Entwicklung eines Datenmanagementplans. Hierbei handelt es sich um eine zu Beginn des Projekts ausgearbeitete detaillierte Planung der Vorgehensweisen und Strategien, die über die verschiedenen Projektphasen hinweg angepasst werden. „Sicherheit“



Welche Praktiken in welcher Phase?

Einige Praktiken beziehen sich auf **alle Phasen**. Diese sind Aspekte im Zusammenhang mit ethischen Erwägungen und Praktiken, die sich auf die Sicherheit oder die Dokumentation der Daten beziehen.

Andere Praktiken finden nur **in bestimmten Phasen** Anwendung. Zum Beispiel sollte die Frage der Einverständniserklärung zu Beginn des Projekts geklärt werden.

In diesem Leitfaden gehen wir auf die folgenden Fragen ein:

- Wie kann der Schutz der teilnehmenden Personen gewährleistet werden?
- Wie werden die Dateien organisiert?
- Wie kann man die Datensicherheit langfristig gewährleisten?
- Wie sollen die Daten und das Projekt dokumentiert werden?

Der Schutz der teilnehmenden Personen

Der Schutz der Teilnehmenden sowie ethische Überlegungen stehen im Mittelpunkt der Forschung. Sie beeinflussen zahlreiche Verfahren während der verschiedenen Forschungsphasen und der Datenverarbeitung (Projektdesign, Sammlung, Aufbereitung, Analyse und Auswertung der Daten, oder auch Aufbewahrung, Archivierung und Datenaustausch).

Je nach Fragestellung des Forschungsprojekts kann es erforderlich sein, personenbezogene, also sehr sensible Daten zu erheben. Diese können potenziell eine Identifizierung der teilnehmenden Personen ermöglichen und diese schädigen.

Wenn Sie sensible Daten erheben, müssen der Schutz der Teilnehmenden und der Datenschutz oberste Priorität haben. Sie benötigen dazu...

- eine umfassende Aufklärung der Teilnehmenden über das Projekt und die Einholung ihrer Einwilligung zur Teilnahme an der Studie, sowie für eine mögliche Wiederverwendung (durch Dritte) der Daten (insbesondere im Rahmen von qualitativen Interviews).
- eine Anonymisierung der Daten, d. h. das Entfernen oder Modifizieren aller Elemente, die allein oder kombiniert die Identifizierung einer Person ermöglichen könnten. Anonymisierung ist für alle Daten erforderlich, die geteilt werden.
- das Anpassen der Sicherheits- und Datenzugangsregeln entsprechend ihrer Sensibilität. Im Falle von sensiblen Daten ist es wichtig, dass Sie Sicherheitsmassnahmen implementieren (siehe Abschnitt „Datensicherheit“).



Was sind persönliche und sensible Daten?

Das Bundesgesetz über den Datenschutz (DSG) vom 19. Juni 1992 Art.3 definiert **Personendaten** als „alle Angaben, die sich auf eine bestimmte oder bestimmbare Person beziehen“.

Sensible Daten sind besonders schützenswerte Personendaten und werden definiert als „Daten über:

1. die religiösen, weltanschaulichen, politischen oder gewerkschaftlichen Ansichten oder Tätigkeiten,
2. die Gesundheit, die Intimsphäre oder die Rassenzugehörigkeit,
3. Massnahmen der sozialen Hilfe,
4. administrative oder strafrechtliche Verfolgungen und Sanktionen.“

Der Prinzip der Einverständniserklärung

Die Einverständniserklärung dient dem Schutz der Teilnehmenden einer Studie. Für die Erhebung von qualitativen Daten ist meist eine explizite Zustimmung erforderlich. Sie kann jedoch auch für quantitative Daten erforderlich sein, beispielsweise in einer Fragebogenerhebung. Sie kann mündlich (vorzugsweise aufgezeichnet), schriftlich (insbesondere während eines Interviews durch Unterzeichnung einer Einverständniserklärung), oder elektronisch (beispielsweise in Form einer Checkbox bei einer Online-Umfrage) erfolgen.

Eine typische Einverständniserklärung muss verschiedene Informationen enthalten, insbesondere:

- den Titel des Projekts sowie die Namen der Projektleitenden und des Forschungsförderers
- die Projektziele
- die Stichprobenauswahl und die Art der Teilnahme
- die Vor- und Nachteile einer Teilnahme, sowie eventuelle Risiken
- Informationen über das Recht, jederzeit und ohne Angabe von Gründen die Zustimmung zur Teilnahme zu widerrufen, ohne dass den Teilnehmenden dadurch Nachteile entstehen
- Informationen zum Grundsatz der Vertraulichkeit und des Datenschutzes
- Informationen zum möglichen Datenaustausch und zu geplanten Publikationen



Wenn Ihre Daten während oder nach dem Forschungsprojekt mit anderen Forschenden ausgetauscht werden sollen, ist die Einholung der Einwilligung der Studienteilnehmenden unerlässlich. Wenn ein Datenaustausch nicht geplant, aber möglich ist, empfehlen wir die vorsorgliche Einholung der Zustimmung zu dieser Möglichkeit: Es ist sehr schwierig, die teilnehmenden Personen nachträglich zu kontaktieren. Daraus ergibt sich das Risiko, dass die Daten nicht mehr geteilt werden können.

Die Organisation der Dateien

Die strukturierte Organisation und Benennung der Ordner, Dateien und Dokumente ist essenziell, um sich gut zurechtzufinden. Der jeweilige Dokumentenname kann eine kurze Beschreibung des Inhaltes angeben (z. B. Fragebogen), das Datum (TT.MM.JJJJ), Initialen, Versionen (1., 2., ... final), Autoren usw.

Es wird dringend empfohlen...

- zu Beginn des Projekts eine Verzeichnisstruktur mit Ordnern und Unterordnern zu definieren,
- eine Regel für die Dokumentenbenennung zu definieren, und diese in einer ReadMe-Datei festzuhalten,
- „Best Practices“ anzuwenden, d. h. Dokumente in Ordnern zu klassifizieren und sie systematisch nach vordefinierten Regeln zu benennen,
- die Regeln nach Bedarf anzupassen und dabei nicht zu vergessen, die ReadMe-Datei zu aktualisieren.

Definieren Sie Ihre eigenen Regeln, halten Sie diese fest und wenden Sie sie systematisch an, dadurch sparen Sie Zeit und vermeiden Probleme.



Erstellen einer ReadMe-Datei

Eine ReadMe-Datei ist eine Textdatei (.txt), die einige Projektaspekte kurz beschreibt.

Es handelt sich um eine Datei, die wichtige Informationen über Dateien in demselben Verzeichnis enthält.

Sie kann zum Beispiel beinhalten:

Anwendungshinweise, Regeln für die Dokumentenbenennung, eine Liste mit Namen, Funktionen und Verfassern der anderen Dateien sowie den Informationen zum Datum und über die Personen, die diese erstellt haben.

Einige Beispiele für Regeln zur Benennung von Dokumenten:

- Von Ihnen erstellte Dokumente: Beschreibung_Datum_Version
- Wissenschaftliche Zeitschriften: Autor_Titel_Jahr_Zeitschrift
- Berichte: Autor_Beschreibung_Datum

Datensicherheit

Die Datensicherheit ist die Grundlage des Datenmanagements. Ohne Sicherheitsregeln ist die Lebensdauer der Daten stark eingeschränkt. Dieses gilt ebenso für den Schutz der Teilnehmenden bei personenbezogenen und sensiblen Daten. Zu den Sicherheitsfragen gehören sowohl die Datenspeicherung und das Back-up, als auch der Zugang zu den Daten und deren Weitergabe.

Speicherung und Back-up

Schlecht gespeicherte Daten können verloren gehen, beschädigt oder sogar zerstört werden. Stellen Sie zunächst sicher, dass Ihre Daten gesichert sind, d.h. unbefugte Personen dürfen nicht auf sie zugreifen können.

Anschliessend stellt sich die Frage, wo Sie Ihre Daten hinterlegen möchten. Möglich sind institutionelle Server, persönliche Computer, Online-Server und externe Geräte (DVD, USB-Stick, externer Speicher usw.). Jede Lösung hat Vor- und Nachteile. Verwenden Sie nach Möglichkeit institutionelle Server, da externe Geräte beschädigt oder verloren gehen können ebenso wie persönliche Computer. Online-Speicherplattformen (z. B. Google Drive, Dropbox) werden oft im Ausland gehostet und unterliegen anderen Datenschutzgesetzen. Wir empfehlen diese Lösungen nicht.

Kein Speicher-Tool ist vollständig sicher und es ist unerlässlich, regelmässig Sicherungskopien von wichtigen Dokumenten auf verschiedenen Medien zu erstellen.



Auswahl an Speicher- und Data-Sharing-Plattformen

Bevorzugen Sie Schweizer Lösungen wie SWITCHdrive und SWITCHfilesender.

Diese akademischen Infrastrukturen ermöglichen nicht nur die Speicherung Ihrer Dateien in einer sicheren Umgebung, sondern auch die gemeinsame Nutzung grosser Dateien und deren Synchronisation.

Zugriff auf Daten

Wenn die gesammelten Daten personenbezogen bzw. sensibel sind, stellen Sie sicher, dass nur autorisierte Personen darauf zugreifen können, d.h. Sie und möglicherweise andere Teammitglieder.

Neben der Speicherung von Daten an einem sicheren Ort, z. B. auf einem institutionellen Server, können Sie den Zugriff auf sehr sensible Dateien mit einem Kennwort schützen oder sie verschlüsseln.

Die Sicherheitsmassnahmen hängen von der Sensibilität der Daten und der Notwendigkeit des Zugriffs auf die Daten durch andere Personen ab.

Übertragung von Dateien

Das Teilen von Dateien mit anderen Personen kann auf verschiedene Arten erfolgen, zum Beispiel per E-Mail, per Übergabe eines USB-Sticks, durch Ablegen in einem gemeinsamen Verzeichnis oder durch Online-Übertragung.

Bei der Datenübertragung ist immer Vorsicht geboten, insbesondere, wenn Sie sensible Daten übertragen müssen. Seien Sie sehr vorsichtig mit Clouds und anderen Online-Servern.

Die Dokumentation

Die Dokumentation ist ein allgemeines Konzept, das bei verschiedenen Aspekten eines Forschungsprojekts Anwendung findet.

Es ist wichtig, nützliche Informationen separat festzuhalten, um sie später, wenn man seine Arbeit beschreiben oder sie anderen verständlich machen möchte, leicht wiederfinden zu können.

Die Herausforderung besteht darin, im Voraus zu bestimmen, welche Informationen später mit welcher Genauigkeit benötigt werden. Entwickeln Sie Ihren individuellen Dokumentationsmechanismus, den Sie im Verlauf des Projekts immer wieder an die Erfordernisse anpassen.

Die Dokumentation sollte sowohl auf der Projekt-, als auch auf der Datenebene erfolgen. Sie kann in Form von Notizen vorliegen, die so sortiert, gespeichert und benannt werden, dass man sie schnell wiederfinden kann. Alternativ kann die Dokumentation direkt in den Datensatz integriert werden, beispielsweise als Beschreibungen von Variablen, die Sie erstellt haben.

Einige Beispiele zu dokumentierender Aspekte

Dokumentation auf Projektebene:

- Die Entstehung des Projekts: Grund für die Wahl des Themas, Entwicklung der Ideen, bedeutsame Veränderungen usw.
- Die Entwicklung des Datenerhebungsinstruments: aufgetretene Schwierigkeiten, vorgenommene Änderungen, mögliche Probleme mit dem verwendeten Instrument usw.
- Die Stichprobenziehung: Auswahl von und Kontaktaufnahme mit den Teilnehmenden, aufgetretene Schwierigkeiten, Rücklaufquoten usw.
- Der Datenschutz: Massnahmen, Bedingungen des Datenzugangs, ...

Dokumentation auf Datenebene:

- Der Kontext: Wie, wann und wo fand die Erhebung der Daten statt (z.B. qualitative Interviews)?
- Datenaufbereitung und -bereinigung: Bereitstellung der Datenbank, Datenqualität, Codes für fehlende Werte, Anonymisierungsverfahren, Transkriptionsregeln für qualitative Daten
- Analytische Verfahren, Methoden und Tools
- Alle Informationen, die für andere nützlich sind, um Ihre Daten zu verstehen und möglicherweise wiederzuverwenden

Eine Datenmanagementstrategie entwickelt sich im Laufe eines Projekts: „Best Practices“ werden nicht nur zu Beginn eines Projekts geplant, sondern in dessen Fortgang immer wieder neu bewertet und bei Bedarf angepasst.



www.forscenter.ch, dataservice@fors.unil.ch, + 41 21 692 37 17