

Leitfaden zur Datenspeicherung

Wo und wie wird gespeichert?

Die Dokumentationsstelle des KFM baut unter www.centre-plurilinguisme.ch ein digitales Archiv für im Rahmen der Forschungsprojekte des IFM und KFM anfallende Daten auf. Das Archiv basiert auf der Software STAR Knowledge Center for Archives (SKCA). Diese digitale Plattform dient zugleich als Katalog für das Archiv des IFM in der Rue de Morat 24, in welchem Akten in Papierform und physische Datenträger aufbewahrt werden.

Was wird gespeichert?

Prinzipiell alle Daten, die mit dem Forschungsprojekt in Zusammenhang stehen und nach dessen Abschluss zur Verifizierung oder Weiternutzung zur Verfügung stehen sollten. Das betrifft neben den tatsächlichen Rohdaten z.B. auch Daten, die den Forschungsprozess und den Aufbau oder die näheren Umstände von Experimenten oder Studien transparent und nachvollziehbar machen (Bibliographien, Arbeitspläne, Verwaltungsdaten, Informationen über Workflows...).

Warum wird gespeichert?

- Erschliessung: die Daten im Archiv werden in gängigen Dateiformaten gespeichert und mit Schlagwörtern und einer Systematik erschlossen und logisch geordnet, was sie langfristig verwendbar macht.
- Datensicherheit: automatisierte Backups und mehrere Speicherorte sorgen dafür, dass uns alle Daten an einem zentralen Speicherort erhalten bleiben
- Sichtbarkeit: Teile der Daten werden der Öffentlichkeit zugänglich gemacht und erhöhen so die Sichtbarkeit unserer Forschung in der Fachwelt und darüber hinaus.
- Nachvollziehbarkeit und Weiterverwertung: Andere Forscher können unsere freigegebenen Daten verifizieren oder unter anderen Fragestellungen bearbeiten.

Was wird nicht gespeichert?

Nicht anonymisierte oder pseudonymisierte personenbezogene Daten werden nicht dauerhaft gespeichert. Dateien, an denen das Institut keine Rechte hat, sollen ebenfalls nicht im Archiv gespeichert werden. Das betrifft z.B. Sekundärliteratur, die im Projekt genutzt wurde und dazu temporär auf dem Server gespeichert wurde. Bei potentiell sensiblen Daten entscheiden im Zweifelsfall die Projektleiter/innen, was gespeichert wird und was nicht (→ Data Management Plan).

Bevorzugte Dateiformate

Textdokumente und Scans

→ PDF/A-1b-Standard (ISO 19005-1:2005)

Tabellen und Datenbanken

→ CSV-Format

Video- und Audiodateien

→ OGG (Audio: Opus bei einer Bitrate von 160 kb/s bei 48 kHz; Video: Theora)

Bilder und Grafiken

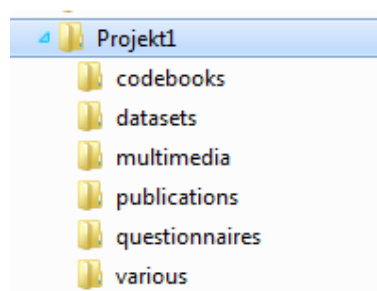
→ JPEG-Format (verlustfreie Kompression)

Proprietäre Formate gängiger Programme etwa zur Statistikerfassung und -auswertung (R, SPSS...) können im Zweifelsfall so belassen werden. Für alle anderen Dateiformate gelten folgende Leitfragen:

- können andere Forschende meines Fachgebiets die Daten nutzen, so wie ich sie abgespeichert habe?
- können sie die Daten auch nutzen, ohne teure Softwarelizenzen erstehen zu müssen?
- ist das gewählte Format voraussichtlich langfristig und plattformübergreifend nutzbar?

Bevorzugte Speicherstruktur

Grundsätzlich sollten alle Daten im Konsolidierungsordner auf dem Instituts-Server abgelegt werden. Jedes Forschungsprojekt sollte einen eigenen Unterordner innerhalb des Konsolidierungsordners erhalten. Innerhalb dieses Unterordners soll dann folgende Struktur existieren:

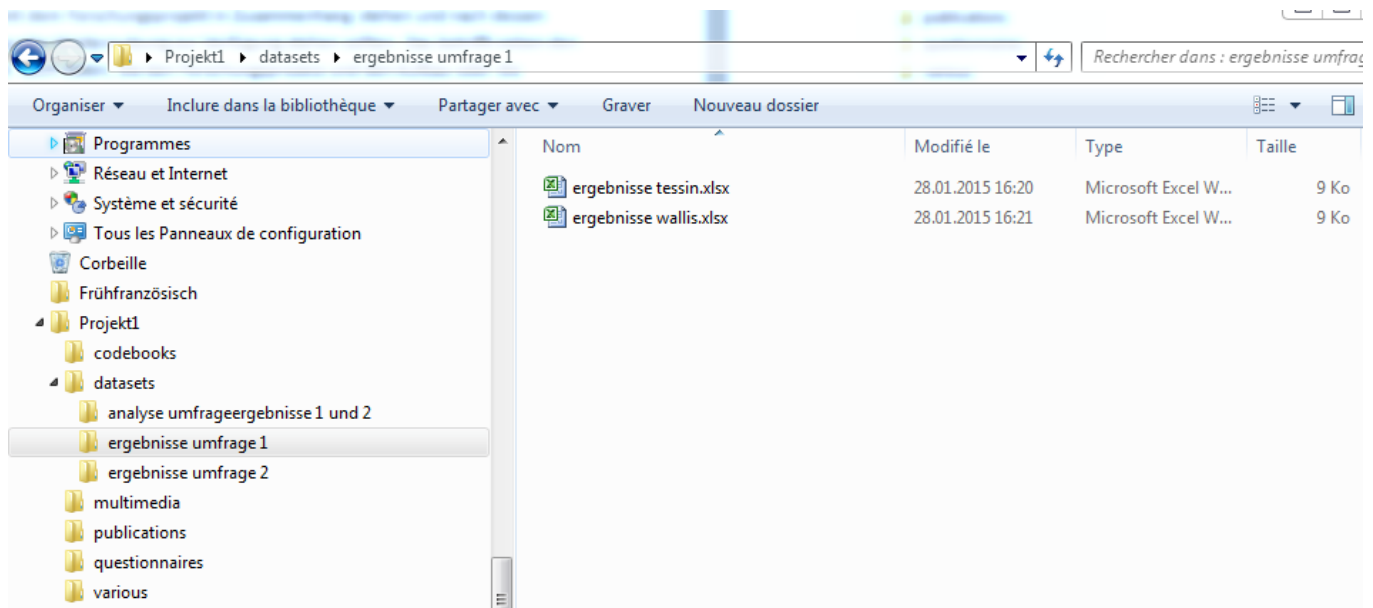


Auf diese Kategorien / Ordner sollten die Forschungsdaten nunmehr verteilt werden:

- **Codebooks** umfasst neben klassischen Codebüchern alle weiteren Dateien, die Schlüssel oder Beschreibungen zu Variablen enthalten, die in im Rahmen dieses Projekt durchgeführten Umfragen/Tests/Tabellen zum Einsatz kamen
- **Datasets** umfasst alle „Datensätze“ im weitesten Sinne: Rohdaten aus Umfragen, Studien oder Tests; Analyseskripte; (anonymisierte bzw. pseudonymisierte) Teilnehmerlisten; Transkriptionen von qualitativen Interviews; etc.
- **Multimedia** umfasst alle Audio-, Bild- und Videodateien, die als Resultat dieses Projekts angefallen sind
- **Publications** umfasst Publikationen oder auch unveröffentlichte Abschlussberichte etc., die im Zusammenhang mit dem Projekt stehen, soweit sie schon in einer veröffentlichbaren Form vorhanden sind
- **Questionnaires** umfasst alle Fragebögen, Testformulare, Aufgabenblätter etc., die in diesem Projekt zum Einsatz kamen. Das umfasst auch die Vereinbarungen zu Datenschutz und Datennutzung, die mit Studienteilnehmenden geschlossen wurden.

- **Various** umfasst alle anderen Daten, die nicht sinnvoll einer der anderen Kategorien zugeordnet werden können. Das betrifft vor allem auch Verwaltungsdaten, Arbeitspläne, Organisation, Korrespondenz, etc.

Innerhalb dieser Ordner sollten die Dateien zu Sinneinheiten gebündelt werden, für die jeweils ein Unterordner angelegt wird. Zum Beispiel können unter „datasets“ alle Dateien mit den Ergebnissen einer Umfrage im Unterordner „ergebnisse umfrage 1“ abgelegt werden, die Dateien mit Ergebnissen zu einer anderen, im Rahmen des selben Projekts durchgeführten Umfrage derweil im Unterordner „ergebnisse umfrage 2“, eine statistische Analyse aller dieser Ergebnisse im Unterordner „analyse umfrageergebnisse 1 und 2“ etc.



Zu beachten:

- Nach Möglichkeit aussagekräftige Namen für die Unterordner und für die einzelnen Dateien wählen!
- Es sollen keine weiteren Unterordner unterhalb dieser Ebene eingeführt werden. Jeder dieser Unterordner ist eine „Akte“ und stellt die kleinstmögliche Sinneinheit auf Ordnerstrukturebene dar.
- Es steht allen Forschenden frei, ob sie im Rahmen ihres Projekts diese Ordnerstruktur von Anfang an verfolgen oder eine eigene Logik verwenden und erst gegen Schluss des Projekts an der Sortierung der Daten auf die empfohlene Struktur durch die Dokumentationsstelle mitwirken. Erfahrungsgemäss ist es insgesamt weniger aufwendig, sich von Anfang an zumindest grob an der empfohlenen Struktur zu orientieren.

Beschreibung der Daten

In jedem Unterordner wird eine kleine Word-Datei mit dem Titel **metadaten.docx** hinterlegt, welche die in diesem Unterordner enthaltenen Daten knapp beschreibt. Diese Datei ist wie folgt aufgebaut:

Name	NameXYZ [Name der Akte]
Dateien	datei1.xyz
	datei2.xyz
	datei3.xyz
	...
Beschreibung	[Kurze Freitextbeschreibung des Inhalts des Akte]
Verknüpft mit	[Hier logische Verknüpfungen <i>dieses Unterordners</i> zu anderen relevanten <i>Unterordnern</i> innerhalb desselben Projekts angeben. Also z.B. den Unterordner mit dem passende Codebook zu den im aktuellen ordner gespeicherten Tabellen angeben, oder eine Transkription zu einer Audiodatei]
Veröffentlichung OK?	[Hier kann kurz und knapp angegeben werden, ob die Daten veröffentlicht werden dürfen, oder ob Einschränkungen aufgrund von Datenschutz, Coypright oder individuellen Vereinbarungen bestehen]

(Fiktives) Beispiel für eine solche Kurzbeschreibung:

Name	Ergebnisse Umfrage 1
Dateien	ergebnisse tessin.xlsx
	ergebnisse wallis.xlsx
Beschreibung	Ergebnisse der ersten, im Mai 2014 an je zwei Schulen im Tessin und Wallis durchgeführten Umfrage zur Lehrerfortbildung.
Verknüpft mit	questionnaires\fragebogen umfrage 1 datasets\analyse umfrageergebnisse 1 und 2
Veröffentlichung OK?	Ja

Fragen?

Die Dokumentationsstelle hilft euch gerne weiter: moritz.sommet@unifr.ch