

Stamp

- Standardisierter Datenmanagementplan: Fallbeispiele

Version 1.0 // Februar 2024



Zitationsvorschlag: DDP-Bildung & VerbundFDB (2024). *Stamp – Standardisierter Datenmanagementplan. Fallbeispiele*. Version 1.0. Frankfurt am Main: DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation.

Kontakt: Bei Problemen, Fragen, Anmerkungen oder Anregungen zum Stamp schreiben Sie uns unter: verbund@forschungsdaten-bildung.de.



This publication is licensed under a Creative Commons Attribution 4.0 (CC-BY 4.0):
<https://creativecommons.org/licenses/by/4.0>

Inhaltsverzeichnis

1. Planungshilfe	5
1.1. Planung von Aufwendungen	5
2. Datenschutz	7
2.1. Datenschutzkonzept.....	7
2.2. Aufklärung der Betroffenen	12
2.3. Kontaktdaten über Dritte einholen (Ankerpersonen).....	13
2.4. Einwilligung Dritter einholen	14
2.5. Einwilligung nach Datenerhebung einholen.....	15
2.6. Einwilligung bei verschiedenen Erhebungsmethoden	16
2.7. Der Begriff der <i>Daten</i> in der Einwilligung	17
2.8. Einsatz eines externen Dienstleisters	18
2.9. Wahrnehmung von Betroffenenrechten: Widerspruch und Widerruf.....	19
2.10. Einschränkung von Betroffenenrechten.....	20
2.11. Kontaktdaten über mehrere Wellen hinweg managen	21
2.12. Schutzverletzung durch Hackerangriff	22
3. Urheberrecht.....	23
3.1. Nutzungsrechte an eigenen Materialien.....	23
3.1. Daten in der Cloud speichern.....	25
3.2. Nutzungsrechte an Materialien Dritter.....	26
3.3. Urheberrecht und Testschutz bei der Verwendung von Tests.....	28
3.4. Urheberrecht bei der Verwendung von Unterrichtsmaterialien	29
4. Datenorganisation	30
4.1. Konzept zur Datenorganisation.....	30

4.2.	Eigene Back-Up-Strategie	36
5.	<i>Nachvollziehbarkeit</i>	38
5.1.	Stichprobenverfahrens und Fallauswahl dokumentieren	38
5.2.	Untersuchungspopulation und Ausschöpfungsquote dokumentieren	40
5.3.	Harmonisierung von Variablen dokumentieren	41
5.4.	Bildung von Gewichten dokumentieren	43
5.5.	Bildung von Skalen und Indizes dokumentieren	43
5.6.	Integration zusammengehöriger (Teil-)Datensätze dokumentieren.....	44
5.7.	Variablen, Fragen und Filterführung dokumentieren	45
5.8.	Dynamische Elemente in Online-Fragebögen dokumentieren	48
5.9.	Leitfadengestützte Interviews dokumentieren	51
5.10.	Feldzugang dokumentieren.....	53
5.11.	Verlaufsprotokoll einer Unterrichtsstunde dokumentieren	54
6.	<i>Datenteilen</i>	55
6.1.	Konzept zum Datenteilen	55
7.	<i>Langfristsicherung</i>	58
7.1.	Konzept zur Langfristsicherung	58
8.	<i>Bestandsliste</i>	61
8.1.	Bestandsliste.....	61

Hinweise

In den folgenden Fallbeispielen werden verschiedene Herausforderungen des Datenmanagements beschrieben und potenzielle Lösungen skizziert. Jedes Fallbeispiel beginnt mit einer kurzen Einführung in ein fiktives Projekt und die konkret bestehende Herausforderung im Datenmanagement. Zur Förderung der Verständlichkeit führen eine Vielzahl von Fallbeispielen in ein identisches fiktives Projekt ein.

1. Planungshilfe

1.1. Planung von Aufwendungen

Im Rahmen eines durch das *Bundesministerium für Bildung und Forschung* (BMBF) geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) arbeiten zwei Lehrstühle zweier Hochschulen (A & B) mit einer außeruniversitären Forschungseinrichtung des Bundes (C) zusammen. Das Projekt begann im Mai 2022 und endet im April 2025. Der Förderbescheid des BMBF setzt für entsprechende Projekte das weitestgehende Teilen der Forschungsergebnisse, wie Daten und Begleitmaterialien, im Sinne des *Open Science* voraus.

Im Projekt wird die Lesekompetenz von Grundschulkindern in der dritten Klasse vor den Sommerferien 2023 getestet (Daten der 1. Welle: LesLer 2023-1). Dieser Test wird nach den Sommerferien 2023 bei den gleichen Grundschulkindern wiederholt (Daten der 2. Welle: LesLer 2023-2). Der Test wird an insgesamt 34 zufällig ausgewählten Grundschulklassen mit dem Förderschwerpunkt *Sprache* im gesamten Bundesgebiet durchgeführt.

Das entwickelte Messinstrument basiert auf einem urheberrechtlich geschützten Testinstrument Dritter (Lesekompass 2018). Für die Nachnutzung dieses Testinstruments im Rahmen des Forschungsvorhabens liegt die Genehmigung der Urhebenden vor. Die im Projekt zu erhebenden Daten werden durch vorhandene Daten einer vergleichbaren Studie ergänzt (FördLes 2019).

Im Anschluss an eine erste Analyse der Daten nach der 2. Welle sind zusätzlich zehn Interviews mit Grundschullehrkräften geplant (LesInt 2023). Diese Interviews werden per Video aufgezeichnet. Die Transkription der Videoaufzeichnungen erfolgt durch eine*n externe*n Dienstleistende*n (Auftragsverarbeitende*r D). Im Rahmen des Projekts sollen zudem zwei empirische Dissertationen entstehen.

Im Rahmen der Projektplanungsphase (Herbst 2021) wurde der Stamp zur Planung der Aufwendungen genutzt. Entsprechend der Vorgaben im Abschnitt *Planungshilfe* wurden folgende Kostenpositionen für die Aufbereitung der Leistungsdaten identifiziert (und in einem separaten Dokument als Vorlage für die Antragstellung skizziert):

Aufwendungen für die Aufbereitung und Dokumentation der Leistungsdaten

Maßnahmen	materielle Aufw.	zeitliche Aufw.	Anmerkungen
Forschungsethik			
...	
Nachvollziehbarkeit			
Rohdatenkontrolle und Aufbereitung der Daten, u. a. abhängig vom Datentypus, der Erhebungsmethode und der Komplexität der Daten		40 Stunden	eine Woche Vollzeitäquivalent (VZÄ) zur Kontrolle der Rohdaten (im Feld und nach Dateneingang)
Digitalisieren und/oder Transkribieren der Daten		0 Stunden/PM	Leistungsdaten werden digital erfasst, keine weitere Digitalisierung notwendig
Pseudonymisierung und Anonymisierung, u. a. abhängig vom Datentyp, dem Inhalt der Daten und den betroffenen Personen(-gruppen)		120 Stunden	drei Wochen VZÄ zur Aufbereitung und Pseudonymisierung der Leistungsdaten (inkl. ggf. weitere Aufbereitung und Pseudonymisierung zum Datenteilen)
Kosten für spezielle Aufbereitungssoftware	€ 0,0		Standardsoftware ausreichend, keine weitere Software zur Aufbereitung der Leistungsdaten erforderlich
Dokumentation erstellen, u. a. abhängig vom Datentypus und der Dokumentation der Datenerhebung und -aufbereitung		80 Stunden	zwei Wochen VZÄ zur Dokumentation der Leistungsdaten (inkl. ggf. weitere Dokumentation zum Datenteilen)
Kosten für spezielle Dokumentationssoftware	€ 0,0		Standardsoftware ausreichend, keine weitere Software zur Aufbereitung der Leistungsdaten erforderlich
ggf. Aufwendungen für Auftragsverarbeitung (Kosten und Zeitaufwand zur Kommunikation und Prüfung erbrachter Leistungen)	€ 0,0	0 Stunden/PM	keine externen Dienstleistenden für die Erhebung, Aufbereitung etc. der Leistungsdaten
Datenteilen			
...	

Zur Aufbereitung und Kontrolle der Leistungsdaten fallen keine Sachkosten, z. B. für Software oder externe Dienstleistende, an (€0,00). Für die Kontrolle der Rohdaten (40Std.), zur Aufbereitung und Pseudonymisierung (120 Std.) ebenso wie für die Dokumentation der Daten (80 Std.), auch in Bezug auf das Teilen der Daten zur Nachnutzung durch Dritte, werden insgesamt 240 Wochenstunden bzw. 1,5 Personenmonaten veranschlagt. Diese werden in der weiteren Kalkulation der Aufwendungen mit den Personalkosten (gemäß Tarifvertrag) der für die Bearbeitung eingeplanten Person gewichtet.

2. Datenschutz

2.1. Datenschutzkonzept

Im Rahmen eines durch das *Bundesministerium für Bildung und Forschung* (BMBF) geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) arbeiten zwei Lehrstühle zweier Hochschulen (A & B) mit einer außeruniversitären Forschungseinrichtung des Bundes (C) zusammen. Das Projekt begann im Mai 2022 und endet im April 2025. Der Förderbescheid des BMBF setzt für entsprechende Projekte das weitestgehende Teilen der Forschungsergebnisse, wie Daten und Begleitmaterialien, im Sinne von Open Science voraus.

Im Projekt wird die Lesekompetenz von Grundschulkindern in der dritten Klasse vor den Sommerferien 2023 getestet (Daten der 1. Welle: LesLer 2023-1). Dieser Test wird nach den Sommerferien 2023 bei den gleichen Grundschulkindern wiederholt (Daten der 2. Welle: LesLer 2023-2). Der Test wird an insgesamt 34 zufällig ausgewählten Grundschulklassen mit dem Förderschwerpunkt *Sprache* im gesamten Bundesgebiet durchgeführt.

Das entwickelte Messinstrument basiert auf einem urheberrechtlich geschützten Testinstrument Dritter (Lesekompass 2018). Für die Nachnutzung dieses Testinstruments im Rahmen des Forschungsvorhabens liegt die Genehmigung der Urhebenden vor. Die im Projekt zu erhebenden Daten werden durch vorhandene Daten einer vergleichbaren Studie ergänzt (FördLes 2019).

Im Anschluss an eine erste Analyse der Daten nach der 2. Welle sind zusätzlich zehn Interviews mit Grundschullehrkräften geplant (LesInt 2023). Diese Interviews werden per Video aufgezeichnet. Die Transkription der Videoaufzeichnungen erfolgt durch eine*n externe*n Dienstleistende*n (Auftragsverarbeitende*r D). Im Rahmen des Projekts sollen zudem zwei empirische Dissertationen entstehen.

Zu Beginn der Projektlaufzeit (Mai 2022) entwerfen die Mitarbeitenden der außeruniversitären Forschungseinrichtung (C) ein **Datenschutzkonzept**. Im Folgenden sind einige zentrale Passagen des Konzepts dargestellt:

1. Grundlagen und Begriffsdefinitionen

... NICHT ABGEBILDETE PASSAGEN ...

2. Verantwortlichkeit für den Datenschutz

Allgemein verantwortlich für das Datenschutzkonzept und dessen systematische Umsetzung ist die außeruniversitäre Einrichtung (C). Diese erhebt die Daten, bereitet sie auf, dokumentiert sie und übernimmt deren Archivierung zum Teilen der Daten durch Dritte über das hauseigene Forschungsdatenzentrum ebenso wie die Langfristsicherung weiterer Daten und Begleitmaterialien.

Als Einrichtung des Bundes erfolgt die Verarbeitung gemäß den Vorgaben der EU-Datenschutzgrundverordnung (DSGVO) und des Bundesdatenschutzgesetzes (BDSG). Vor dem Aufsetzen der Einwilligungserklärung (zum Einholen der informierten Einwilligungen) wird durch die Datenschutzbeauftragten der außeruniversitären Einrichtung (C) geprüft, ob und ggf. welche weiteren Regelungen der Landesdatenschutzgesetze (LDSG) neben der DSGVO und dem BDSG berücksichtigt werden müssen, insbesondere im Kontext der Datenerhebung an Grundschulen der Bundesländer.

Die Speicherung der Daten im Projekt erfolgt im Cloudsystem der Hochschule A. Die Hochschule ist verantwortlich für das rechtskonforme Sichern der Daten und Begleitmaterialien. Neben den Regelungen der DSGVO gelten für die zu ergreifende Maßnahmen zum gesicherten Verarbeiten der Daten die Regelungen des LDSGs der Hochschule A.

3. Rechtsgrundlagen der Datenverarbeitung

Die Verarbeitung personenbezogener Daten in den LesLer-2023-Daten und LesInt-2023-Daten erfolgt auf Grundlage einer informierten Einwilligung der betroffenen Personen bzw. ihrer Erziehungsberechtigten.

... NICHT ABGEBILDETE PASSAGEN ...

Eine Prüfung der Datenschutzbeauftragten der außeruniversitären Einrichtung (C) kommt zu dem Schluss, dass die Konkretisierung der DSGVO durch die LDSG hinsichtlich der Altersgrenzen der Einwilligungsfähigkeit nicht zum Tragen kommen. Bei den Zielpersonen der LesLer-2023-Daten handelt es sich um Grundschüler*innen, deren Einwilligung generell nur über die Einwilligung der Erziehungsberechtigten hergestellt werden kann. Bei den Zielpersonen der LesInt-2023-Daten handelt es sich um Lehrkräfte, deren alleinige Einwilligungsfähigkeit gegeben ist.

... NICHT ABGEBILDETE PASSAGEN ...

Beim Einholen der Einwilligung zum Erheben und Verarbeiten der Videodaten (LesInt-2023-Daten) ist auch die Einwilligung gemäß §22 des Gesetzes betreffend das Urheberrecht an Werken der bildenden Künste und der Photographie (KunstUrhG) zur Verarbeitung der Bilder der Lehrkräfte einzuholen.

... NICHT ABGEBILDETE PASSAGEN ...

4. Weitere Genehmigungen zur Datenverarbeitung

Notwendige Genehmigungen zum Zugang zu sowie zur Datenerhebung an den Grundschulen werden entsprechend der landesspezifischen Schulgesetze in den betroffenen Bundesländern eingeholt.

... NICHT ABGEBILDETE PASSAGEN ...

5. Risikobewertung der Verarbeitung

Das Risiko der Datenverarbeitung wird nach aktuellem Planungsstand wie folgt eingeschätzt:

- LesLer 2023: Mittleres Risiko potenzieller Folgen für die Schüler*innen. *Begründung:* Neben der Erfassung der Lesekompetenzen werden weitere soziodemographische Daten, wie Alter, Geschlecht, primäre Sprache im Haushalt der Schülerin bzw. des Schülers erfasst. *Schutzmaßnahmen:* Die Rohdaten werden verschlüsselt gespeichert und verarbeitet sowie pseudonymisiert.
- LesInt 2023: Mittleres bis hohes Risiko negativer Folgen für Lehrkräfte und Schüler*innen. *Begründung:* In den Interviews (Videoaufzeichnungen) mit den Lehrkräften kann nicht ausgeschlossen werden, dass sensible Informationen über die Interviewten, andere Lehrkräfte bzw. Schüler*innen offengelegt werden. *Schutzmaßnahmen:* Die Videoaufzeichnungen ebenso wie die erstellten Transkripte werden verschlüsselt gespeichert und verarbeitet. Für die Langfristsicherung werden die Videoaufzeichnungen und Transkripte teil-pseudonymisiert; für das Teilen der Daten werden die Transkripte umfassend pseudonymisiert.

... NICHT ABGEBILDETE PASSAGEN ...

6. Geschützes Verarbeiten der Daten

Verantwortlich für den Schutz der Daten in der Cloud ist die Hochschule A; die ergriffenen Maßnahmen umfassen:

- Zugangskontrolle zur Cloud: Die Zugangskontrolle zu den Servern (Cloudsystem) obliegt dem Rechenzentrum der Hochschule A. Die Regelungen zur Nutzung der Cloud sind allen Projektmitarbeitenden bekannt. Die Verantwortlichkeiten auf Seiten des Projekts ebenso wie im Rechenzentrum sind klar definiert. Mögliche Sicherheitslücken im Zugangssystem werden (bei Auftreten) umgehend an die Verbundpartner gemeldet. Dies gilt auch für etwaige Änderungen der Zugangsregelungen.
- Verschlüsselung in der Cloud: Die Verschlüsselung von Dateien in der Cloud erfolgt mittels *VeraCrypt*¹. Die Datenschutzbeauftragten der außeruniversitären Einrichtung (C) und die verantwortliche Stelle des Rechenzentrums der Hochschule A besitzen Kenntnis über die in der Cloud gespeicherten Daten und deren Inhalte sowie über die genutzte Verschlüsselungssoftware. Die Datenschutzbeauftragten bzw. die Verantwortlichen des Rechenzentrums melden bekanntwerdende Schwachstellen in der Verschlüsselungssoftware umgehend den Verbundpartnern. Die Projektkoordination prüft zudem monatlich die Betreiber-Homepage der Verschlüsselungssoftware auf Updates bzw. Sicherheitswarnungen.
- Passwörter der Mitarbeitenden: Die Passwörter der Mitarbeitenden zu ihren Rechnern ebenso wie zum Cloudsystem werden halbjährlich erneuert. Veraltete oder unsichere Passwörter werden umgehend ersetzt. Das Erstellen von und der Umgang mit Passwörtern folgt den Empfehlungen des *Bundesamts für Sicherheit in der Informationstechnik*².

... NICHT ABGEBILDETE PASSAGEN ...

7. Einwilligungsmangement

Die unterzeichneten Einwilligungserklärungen (Papierdokumente) werden an geschützter Stelle innerhalb der außeruniversitären Einrichtung (C) aufbewahrt. Die Einwilligungserklärungen der Erziehungsberechtigten werden (zusammen mit den Rohdaten) nach der Pseudonymisierung der LesLer-2023-Daten, spätestens aber zum Projektende (April 2025), vernichtet. Die Einwilligungserklärungen der Lehrkräfte werden nach dem Teilen der pseudonymisierten Transkripte im Forschungsdatenzentrum der außeruniversitären Einrichtung (C) vernichtet.

Die Einwilligungserklärungen sind mit einer eindeutigen Kennung versehen, die über eine Schlüsseldatei mit den einzelnen Schüler*innen bzw. Lehrkräften verknüpft werden kann.

Die in den Einwilligungserklärungen genannten Zwecke der Datenverarbeitung sowie potenzielle weitere Empfänger*innen der Daten und Angaben zur Speicherdauer bzw. zu den Kriterien der Löschung der Daten sind systematisch im Einwilligungsmanagement erfasst.

Im Rahmen des Einwilligungsmanagements wird zudem die Wahrnehmung einzelner Rechte der Schüler*innen bzw. ihrer Erziehungsberechtigten sowie der Lehrkräfte und die damit einhergehenden Konsequenzen für die weitere Verarbeitung der Daten dokumentiert, d. h.

- a) welche betroffene Person hat wann, welches Recht gegenüber welchem Verbundpartner geltend gemacht,
- b) welche*r Projektmitarbeitende hat wann, wie mit der betroffenen Person die Wahrnehmung des Rechts kommuniziert,
- c) welche Konsequenzen ergeben sich durch die Wahrnehmung des Rechts für die weitere Verarbeitung der Daten, z. B. Löschung von Daten der betroffenen Person,
- d) wie wird gewährleistet, dass das wahrgenommene Recht in der weiteren Verarbeitung der Daten berücksichtigt wird, z. B. durch das tatsächliche Löschen der Daten,
- e) welche weiteren Konsequenzen ergeben sich durch das wahrgenommene Recht im Projekt, z. B. die Dokumentation gelöschter Informationen,
- f) wie und mit welchem Resultat wurden die Datenschutzbeauftragten der außeruniversitären Einrichtung (C) in die Wahrnehmung der Rechte und die Kommunikation mit der betroffenen Person einbezogen.

... NICHT ABGEBILDETE PASSAGEN ...

8. Transkription der Videoaufnahmen durch den Dienstleistenden D

Die Transkription der Videoaufzeichnungen der Interviews mit den Lehrkräften (LesInt 2023) erfolgt durch den externen Dienstleistenden D. Aus dem Vertrag mit den Dienstleistenden (D) geht u. a. Folgendes hervor:

1. Grundlagen des Vertrags

... NICHT ABGEBILDETE PASSAGEN ...

IV. Datenaustausch

Die Übermittlung der Videoaufzeichnungen sowie der (nicht-pseudonymisierten) Transkripte erfolgt über einen passwortgeschützten Austauschserver des Dienstleistenden (D). Beim Verbundpartner C (außeruniversitäre Einrichtung) sind zwei verantwortliche Personen namentlich bestimmt, die vom Dienstleistenden einen Link sowie ein individualisiertes Passwort zum Austauschserver erhalten. Die nicht-pseudonymisierten Transkripte können von den verantwortlichen Personen, nach Fertigstellung durch den Dienstleistenden D, vom Austauschserver heruntergeladen werden.

Der Dienstleistende D trägt die Verantwortung dafür, dass die Videoaufnahmen ebenso wie die erstellten Transkripte im korrekten Verzeichnis des Austauschservers abgelegt werden und dass nur die zum Projekt gehörenden Videoaufzeichnungen, Transkripte und das Übermittlungsprotokoll (siehe Abschnitt III.) auf dem Austauschserver gespeichert werden.

V. Vernichtung der Daten

Nach Download der Transkripte führen die Mitarbeitenden der außeruniversitären Einrichtung C innerhalb von 14 Tagen eine Kontrolle der Transkripte durch und kommunizieren eventuelle Rückfragen zu einzelnen Transkripten mit dem Dienstleistenden (D):

- bestehen keine offenen Fragen, werden die Videoaufzeichnungen und zugehörigen Transkripte beim Dienstleistenden D nach Ablauf einer 14-Tagefrist (beginnend mit dem Download der Transkripte) vernichtet;
- bestehen offene Rückfragen, erfolgt die Vernichtung der entsprechenden Videoaufzeichnungen und der zugehörigen Transkripte unmittelbar nach Klärung der offenen Fragen, frühestens aber 14 Tage nach Download der Transkripte; eine Rückfrage gilt erst dann als geklärt, wenn die Mitarbeitenden der außeruniversitären Einrichtung C dies explizit gegenüber dem Dienstleistenden D kommunizieren.

Die Vernichtung der Videoaufzeichnungen, der zugehörigen Transkripte sowie ggf. weiterer personenbezogener Daten beim Dienstleistenden D erfolgt durch ein Neuaufsetzen der entsprechenden Bereiche des Austauschservers beim Dienstleistenden D.

VII. Schlussbestimmungen

... NICHT ABGEBILDETE PASSAGEN ...

-
1. VeraCrypt, Version 1.25.9 unter: <https://www.veracrypt.fr/code/VeraCrypt/>.
 2. Bundesamts für Sicherheit in der Informationstechnik: Sichere Passwörter erstellen. unter: https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Cyber-Sicherheitsempfehlungen/Accountschutz/Sichere-Passwoerter-erstellen/sichere-passwoerter-erstellen_node.html
Bundesamts für Sicherheit in der Informationstechnik: Sicherer Umgang mit Passwörtern Schritt-für-Schritt erklärt. unter: https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Cyber-Sicherheitsempfehlungen/Accountschutz/Sichere-Passwoerter-erstellen/Umgang-mit-Passwoertern/umgang-mit-passwoertern_node.html

2.2. Aufklärung der Betroffenen

Im Rahmen des BMBF-geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) werden die Lesekompetenzen von Grundschüler*innen vor und nach den Sommerferien 2023 an insgesamt 34 zufällig ausgewählten Grundschulklassen im gesamten Bundesgebiet untersucht. Zum Einholen der Einwilligung reichen die Projektmitarbeitenden zunächst schriftliche Informationen zu ihrem Forschungsvorhaben über die Lehrkräfte an die Erziehungsberechtigten der Grundschüler*innen weiter. Anschließend werden eine Reihe von Informationsveranstaltungen an den ausgewählten Grundschulen durch die Projektmitarbeitenden durchgeführt.

Die Projektmitarbeitenden gehen am Ende der Vorstellung ihres Vorhabens explizit darauf ein, dass

1. mit einer Verweigerung zur Teilnahme am Verbundprojekt keinerlei (negative) schulische Konsequenzen für die Grundschüler*innen verbunden sind,
2. mit einer Verweigerung von Teilen der Datenverarbeitung, wie z. B. zum Teilen der Daten mit Dritten, keinerlei (negative) schulische Konsequenzen für die Grundschüler*innen verbunden sind,
3. mit der Teilnahme am Verbundprojekt keinerlei (negative bzw. positive) schulische Konsequenzen für die Grundschüler*innen verbunden sind,
4. die nicht-pseudonymisierten Ergebnisse der Lesetests nicht an die Lehrkräfte bzw. Grundschulen weitergegeben werden.

In den anschließenden Fragerunden werden die Erziehungsberechtigten noch einmal explizit dazu ermutigt, bestehende Fragen zur (Verweigerung der) Teilnahme und daraus resultierender, potenzieller Konsequenzen zu stellen.

2.3. Kontaktdaten über Dritte einholen (Ankerpersonen)

In einem Forschungsprojekt werden soziale Netzwerke von Schüler*innen der Oberstufe an Gymnasien untersucht. Mittels eines standardisierten Fragebogens werden sowohl die (schulinternen) sozialen Netzwerke der Schüler*innen innerhalb der jeweiligen Schule analysiert als auch soziale Netzwerke der Schüler*innen außerhalb ihrer Gymnasien (schulextern).

Zur Ermittlung der externen Netzwerke werden von den Schüler*innen auch Angaben zu Personen aus ihren externen Netzwerken erfragt. Hierzu zählen beispielsweise Angaben zum Geschlecht und Alter jeder Person im externen Netzwerk der Schüler*innen sowie deren Kontaktdaten (Klarname und E-Mail-Adresse). Kontaktdaten kontaktieren die Forschenden die Personen in den externen Netzwerken zur weiteren Befragung. Für das Forschungsvorhaben und das Einholen von Kontaktinformationen Dritter liegt ein positives Ethikvotum vor.

Nach Eingang der Rohdaten zu den schulinternen Netzwerken werden die Kontaktinformationen der Personen der externen Netzwerke von den Umfragedaten getrennt und zusammen mit der Herkunft der Informationen (ID der Schüler*in, die bzw. der den Kontakt im Fragebogen angegeben hat) in einer separaten Kontaktdatei gespeichert. Unvollständige bzw. unbrauchbare Angaben werden dabei gelöscht. Nach Aufbereitung der Kontaktdatei werden alle darin befindlichen Personen per E-Mail kontaktiert.

In der E-Mail informieren die Forschenden zunächst über sich selbst und ihr Forschungsanliegen, den Grund des Kontaktversuches, die Herkunft der Kontaktinformationen, die bereits erhaltenen personenbezogenen Informationen (Klarname, Emailadresse, Alter und Geschlecht) sowie die Zwecke der Verarbeitung dieser Angaben. Im Anhang der E-Mail findet sich zudem eine Einwilligungserklärung zur a) Nutzung der Kontaktdaten zur Befragung im Rahmen des Forschungsprojekts sowie b) zur Teilnahme an der Befragung mit der Bitte, diese unterschrieben an die Forschenden zurückzuschicken.

Alle per E-Mail angeschriebenen Kontaktpersonen werden in das Einwilligungsmanagement des Projekts aufgenommen. Dort werden alle Rückmeldungen der Personen zur Teilnahme bzw. Teilnahmeverweigerung vermerkt. Kontaktinformationen von Personen, die eine Nutzung ihrer personenbezogenen Daten im Rahmen des Projekts ablehnen, werden vernichtet und entsprechend als „vernichtet / fehlende Einwilligung“ im Rahmen des Einwilligungsmanagements gekennzeichnet. Analog werden personenbezogene Informationen über diese Personen aus den (Roh-)Daten der schulinternen Befragung entfernt.

2.4. Einwilligung Dritter einholen

In einem Forschungsprojekt werden die sprachlichen Kompetenzen von 11.-Klässler*innen an Gymnasien untersucht. Von besonderem Interesse ist dabei die Sprachkompetenz von Schüler*innen aus Haushalten mit Migrationshintergrund. Die Forschenden werten dazu Klassenarbeiten im Fach Deutsch aus.

Ergänzt werden die Daten durch einen standardisierten Fragebogen, den die Schüler*innen zu Hause selbst ausfüllen. Im Rahmen dieses Selbstausfüllerbogens (PAPI) werden u. a. Angaben zum Haushalt und dem familiären Hintergrund der Schüler*innen, wie etwa das Alter von Geschwistern und Großeltern, ebenso wie zum sozialen Hintergrund der Eltern, wie z. B. zur ethnischen Herkunft und Religionszugehörigkeit, zum Schulabschluss und zum Erwerbsstatus, erhoben.

Zur Auswertung der Klassenarbeiten und zur Verarbeitung der Daten des Selbstausfüllerbogens werden von allen Schüler*innen vor Beginn der Datenerhebung die notwendigen Einwilligungen eingeholt. Da bereits in der Fragebogenkonstruktion ersichtlich ist, dass neben Informationen über die Schüler*innen auch personenbezogene Daten über die Eltern der Schüler*innen erhoben werden, bedarf es zur Verarbeitung dieser Angaben der informierten Einwilligung der Eltern. Diese wird zusammen mit dem Selbstausfüllerbogen an die Schüler*innen ausgeteilt.

Die Einwilligungserklärung für die Eltern ist dabei analog zu jener der Schüler*innen formuliert. Sie informiert über die Forschenden und das Forschungsvorhaben, die Zwecke der Verarbeitung der Daten und die Rechte der Eltern. Darüber hinaus werden die besonderen Kategorien personenbezogener Daten (Art. 9 Abs. 1 DSGVO), wie ethnische Herkunft oder Religionszugehörigkeit, sowohl im Informationsteil als auch in der eigentlichen Einwilligungserklärung erörtert.

Wird eine Einwilligung durch die Eltern verweigert, liegt diese aber von der betreffenden Schülerin bzw. dem betreffenden Schüler vor, so werden die Informationen über die Eltern zunächst aus dem Selbstausfüllerbogen entfernt. Der Fragebogen ist dazu so konzipiert, dass Angaben zu beiden Elternteilen auf jeweils getrennten Seiten erfasst werden. Diese werden bei einer Teilnahmeverweigerung (des entsprechenden Elternteils) aus den Papierfragebogen entfernt und vernichtet, bevor die (verbleibenden) Informationen digitalisiert werden. Die Vernichtung von Teilen des Selbstausfüllerbogens wird entsprechend im Einwilligungsmanagement zur*m jeweiligen Schüler*in vermerkt.

Einwilligungserklärungen der übrigen Familienangehörigen, d. h. von Geschwister und Großeltern, werden nicht eingeholt. Die Forschenden argumentieren hier, dass als einziges relevantes personenbezogenes Merkmal das Alter dieser Personen erhoben wird. Das Einholen einer Einwilligung und das entsprechende Einwilligungsmanagement würde demgegenüber die Verarbeitung weiterer personenbezogener Daten erfordern, wie etwa die Klarnamen der Familienmitglieder, und somit dem Prinzip der Datenminimierung zuwiderlaufen.

2.5. Einwilligung nach Datenerhebung einholen

Ein Forschungsprojekt untersucht das Verhalten von Lehramtsstudierenden in Konfliktsituationen mittels eines experimentellen Settings. In den Diskussionsrunden zu Vorträgen der Lehramtsstudierenden über Schulpraktika, die Teil des regulären Studiums sind, werden durch die Dozierenden künstliche Konfliktsituationen herbeigeführt. Die Forschenden beobachten das Verhalten der Studierenden im Verlauf dieser Konfliktsituationen und zeichnen es per Video auf.

Im Rahmen der Projektplanung holen die Forschenden ein Ethikvotum zu ihrem experimentellen Setting und die damit einhergehende Täuschung der Lehramtsstudierenden bei ihrer institutionellen Ethikkommission ein. Parallel werden die Dozierenden auf ihre Rollen als Provokateure bzw. Provozierte vorbereitet, die verschiedenen Rollen in der Konfliktsituation klar definiert und durch Rollenspiel eingeübt. Im Rahmen dieser Vorbereitung wird von den Dozierenden auch die informierte Einwilligung zur Aufzeichnung der Diskussion und des Konflikts eingeholt.

Um die Studierenden in ihrem Verhalten möglichst wenig zu beeinflussen, werden diese im Vorfeld der Beobachtung nicht über die tatsächlichen Forschungsinteressen und Zwecke der Verarbeitung der Daten informiert. Stattdessen wird ihnen ein fiktives Forschungsprojekt zur Untersuchung didaktischer Fähigkeiten angehender Lehrkräfte vorgestellt. Hierzu wird die Einwilligung zur Aufzeichnung der Vorträge und Diskussionen ebenso wie zu den Zwecken der Verarbeitung der Daten eingeholt. Diese vorläufige Einwilligungserklärung entspricht, so weit als möglich, der tatsächlichen (noch einzuholenden) Einwilligung, z. B. in Bezug auf die Informationen zu den Forschenden oder hinsichtlich verschiedener Zwecke der Datenverarbeitung, wie etwa zum Teilen der Daten nach Projektende.

Erst nach Abschluss der Gruppendiskussion und dem dabei ausgetragenen Konflikt werden die Studierenden über die eigentlichen Interessen der Forschenden, das experimentelle Setting und die künstlich herbeigeführte Konfliktsituation informiert. Dabei werden die im Vorfeld der Vorträge eingeholten (vorläufigen) Einwilligungserklärungen der Studierenden wieder an diese zurückgegeben. Die Studierenden erhalten schließlich die tatsächliche Einwilligungserklärung, inklusive der eigentlichen Forschungsinteressen und Zwecke der Verarbeitung der Daten. Verweigern Studierende ihre Einwilligung zur Aufnahme bzw. Datenverarbeitung, löschen die Forschenden die entsprechenden Videoaufnahmen umgehend.

2.6. Einwilligung bei verschiedenen Erhebungsmethoden

In einer Studie sollen personenbezogene Daten rechtskonform auf Basis der informierten Einwilligung der Teilnehmenden verarbeitet werden. Teil der Einwilligungserklärung ist die eindeutige Willensbekundung zur Teilnahme an der Datenerhebung, -speicherung, -aufbereitung, -analyse, -teilen etc.

Dabei kann die eindeutige Willensbekundung, je nach Befragungssituation, in unterschiedlicher Art und Weise abgegeben werden, solange

- a) die betroffene Person einwilligungsfähig ist,
- b) die betroffene Person aktiv und informiert einwilligt, z. B. durch Unterschreiben der Einwilligungserklärung, und
- c) die eindeutige Willensbekundung durch die Projektbeteiligten belegbar ist, z. B. durch eine unterschriebene Einwilligungserklärung.

Beispiel für eine eindeutige Willensbekundung in Paper and Pencil Interviews (PAPI)

Eine Studie soll mittels papierbasierter Fragebögen durchgeführt werden. Die betroffenen Personen werden im Vorfeld der Erhebung der Daten über deren Verarbeitung aufgeklärt. Von einer eindeutigen Willensbekundung kann ausgegangen werden, wenn die betroffene Person den ausgefüllten Fragebogen (inklusive des Informationsteils) zurückschickt. Der ausgefüllte Fragebogen dient dann als Nachweis für die Einwilligung.

Beispiel für eine eindeutige Willensbekundung in Computer Assisted Telephone Interview (CATI)

Eine Studie soll mittels telefonischer Befragung durchgeführt werden. Die betroffenen Personen werden zunächst am Telefon über die Erhebung und Verarbeitung der Daten aufgeklärt. Von einer eindeutigen Willensbekundung kann ausgegangen werden, wenn die betroffene Person zustimmt, mit der Befragung zu beginnen. Der Nachweis für die Einwilligung wird von der interviewenden Person (meist) elektronisch im Rahmen des Einwilligungsmanagement dokumentiert bzw. aufgezeichnet.

Beispiel für eine eindeutige Willensbekundung in Computer Assisted Web Interview (CAWI)

Eine Studie soll mittels Online-Befragung durchgeführt werden. Die betroffenen Personen werden vor der Freischaltung des eigentlichen Fragebogens über die Erhebung (auch von Logdaten) und die Verarbeitung der Daten aufgeklärt. Von einer eindeutigen Willensbekundung kann ausgegangen werden, wenn die betroffene Person zum Freischalten des eigentlichen Fragebogens z. B. einen „Beginnen“-Button anklicken oder ein Häkchen setzen muss. Der Nachweis über die Einwilligung wird elektronisch in der Datenbank (Logdaten) dokumentiert.

Beispiel für eine eindeutige Willensbekundung bei Aufzeichnungen von Videokonferenzen

Eine Studie soll mittels einer Konferenzsoftware durchgeführt und aufgezeichnet werden. Die betroffenen Personen erhalten die datenschutzrechtlich relevanten Informationen für eine informierte Einwilligung vor der Veranstaltung per Email. Darin wird die aktive Einwahl zur Teilnahme an der Studie als aktive Willensbekundung zur Nutzung der Konferenzsoftware sowie zur Aufzeichnung erörtert. Zu Beginn der digitalen Veranstaltung werden die Teilnehmenden nochmals auf die Aufzeichnung hingewiesen und gebeten, bei fehlender Einwilligungsbereitschaft zur Aufzeichnung den virtuellen Raum wieder zu verlassen. Als Nachweis zur Einwilligung dient die tatsächliche Teilnahme (virtueller Zutritt zur Veranstaltung) an der Veranstaltung (Aufzeichnung).

2.7. Der Begriff der Daten in der Einwilligung

Im Rahmen des BMBF-geförderte Verbundprojekt *Lesen Lernen* (LesLer 2022) werden die Lesekompetenzen von Grundschüler*innen vor und nach den Sommerferien 2023 (2 Erhebungswellen) an insgesamt 34 zufällig ausgewählten Grundschulklassen im gesamten Bundesgebiet untersucht. Zusätzlich werden zehn Interviews mit Grundschullehrkräften geführt.

Bei der Erstellung der Einwilligungserklärung orientieren sich die Projektmitarbeitenden an der Stamp-Checkliste zur *Informierten Einwilligung* und den dortigen Empfehlungen. In der Einwilligungserklärung erörtern sie die verschiedenen Arten von Daten, deren Verarbeitung, die Zwecke der Verarbeitung sowie die Sicherungsdauer bzw. deren Löschfristen und -kriterien.

- a) Kontaktdaten beinhalten u. a. den Namen der entsprechenden Schüler*in sowie Name und Anschrift der Erziehungsberechtigten. Die Kontaktdaten auf den Einwilligungserklärungen (Papierdokumente) werden so bald als möglich, d. h. mit Abschluss der Datenaufbereitung und deren Pseudonymisierung, spätestens aber zum Projektende (April 2015), vernichtet.
- b) Leistungsdaten umfassen die Ergebnisse der Lesetests und sind die eigentlich interessierenden Forschungsdaten des Projekts. Sie werden im Rahmen der Datenaufbereitung (weitestgehend) pseudonymisiert und mit Dritten zur Nachnutzung geteilt.
- c) Soziodemographische Daten der Schüler*innen beinhalten z. B. das Alter und die primäre Sprache im Haushalt und sind Teil der Forschungsdaten, die gemeinsam mit den Leistungsdaten in pseudonymisierter Form mit Dritten geteilt werden.

Analog erörtern die Projektbeteiligten in der Einwilligungserklärung für die Lehrkräfte die Verarbeitung der Videoaufzeichnungen und der daraus erstellten Transkripte. Dabei gehen sie auch auf das Teilen der pseudonymisierten Transkripte ebenso wie die langfristige Sicherung von (Teilen der) Videodateien und nicht-pseudonymisierten Transkripte ein.

2.8. Einsatz eines externen Dienstleisters

In einem Forschungsprojekt werden in drei Wellen wiederholt Daten bei den gleichen Schüler*innen erhoben (Paneldesign). Die Datenerhebung erfolgt unmittelbar im Unterricht. Im Rahmen der zweiten Welle wird zudem allen Erziehungsberechtigten ein Selbstausfüllerbogen (PAPI) postalisch zugesandt. Die Adressinformationen für diese Erhebung entstammen der Schüler*innen-Befragung der ersten Welle. Zur Durchführung und Digitalisierung der postalischen Befragung wird ein Erhebungsinstitut beauftragt, welches für den Postversand wiederum einen Dienstleistenden einsetzt.

Im Rahmen der Einwilligungserklärung werden die Teilnehmenden über die Verarbeitung ihrer personenbezogenen Daten informiert. Dies betrifft auch die Weitergabe der Kontaktdaten an Dritte, wie etwa die Angestellten des Erhebungsinstituts aber auch Mitarbeitende des vom Erhebungsinstitut beauftragten Postdienstleistenden. Dieser erhält vom Erhebungsinstitut vorbereitete Briefe mit eingedruckten Adressen.

Für derartige Verarbeitungszwecke ist eine Löschfrist für den jeweiligen Datenempfängenden in der Einwilligungserklärung anzugeben. Im konkreten Fall ist die Verarbeitung der personenbezogenen Daten (Speicherung und Nutzung) beim Postdienstleistenden nur bis zum Versand des Selbstausfüllerbogens gestattet. Das Erhebungsinstitut verarbeitet hingegen neben den Adressdaten auch die Befragungsdaten (Digitalisierung). Es kann die personenbezogenen Daten folglich so lange verarbeiten, wie es für den vertraglich festgelegten Zweck der Verarbeitung notwendig ist. Dies kann z. B. der Abschluss der Datenerhebung und Übergabe an die Forschenden sein.

Wie in der Einwilligungserklärung spezifiziert, werden die pseudonymisierten Befragungsdaten über ein Repositorium anderen Forschenden zur Nachnutzung zugänglich gemacht. Diese anderen Forschenden und das Repositorium müssen ihrerseits aber nicht (notwendigerweise) als mögliche Empfangende der Daten in der Einwilligungserklärung genannt werden, da pseudonymisierte Daten nicht mehr dem Datenschutz unterliegen.

2.9. Wahrnehmung von Betroffenenrechten: Widerspruch und Widerruf

Im Rahmen des BMBF-geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) werden die Lesekompetenzen von Grundschüler*innen vor und nach den Sommerferien 2023 (zwei Erhebungswellen) an insgesamt 34 zufällig ausgewählten Grundschulklassen im gesamten Bundesgebiet untersucht. Zusätzlich werden zehn Interviews mit Grundschullehrkräften geführt.

Unmittelbar nach Erhebung der Lesekompetenz wurden die Leistungsdaten ebenso wie verschiedene soziodemographische Merkmale der Schüler*innen digitalisiert und eine Datenmatrix zur Auswertung erstellt. Nach verschiedenen Maßnahmen zur Kontrolle der digitalisierten Daten wurden die originären Leistungstests und der Fragebogen zur Soziodemographie (Papierdokumente) vernichtet. Anschließend wurde die erstellte Datenmatrix pseudonymisiert, so dass eine Re-Identifikation einzelner Schüler*innen in den Daten faktisch nicht mehr möglich ist. Die einzelnen Schüler*innen können über eine eindeutige ID in der Datenmatrix jedoch weiterhin mit den Einwilligungserklärungen (der Erziehungsberechtigten) in Verbindung gebracht werden und ermöglicht so u. a. die Verknüpfung der Leistungsdaten aus beiden Erhebungswellen.

Kurz nach Abschluss der Pseudonymisierung der Daten **widersprechen** die Erziehungsberechtigten eines Schülers der weiteren Verarbeitung der Daten und verlangen deren Löschung. Die Projektbeteiligten prüfen zunächst, ob der Schüler in den Daten identifiziert und die Daten des Schülers entfernt werden können. Dies ist nicht mehr unmittelbar in den Daten möglich. Jedoch erlaubt die Verknüpfung mit der Einwilligungserklärung eine Re-Identifikation des Schülers. Seine Daten werden daraufhin aus der Datenmatrix entfernt und dies den Erziehungsberechtigten entsprechend kommuniziert.

Aufgrund des Widerspruchs und der Prüfung möglicher Re-Identifikationen einzelner Schüler*innen in den Daten, entscheiden die Projektbeteiligten, die noch existierenden Einwilligungserklärungen zu vernichten. Dadurch geht auch die letzte Möglichkeit zur Re-Identifikation einzelner Schüler*innen in den Daten verloren. Die Projektbeteiligten erreichen so, dass zukünftige Widersprüche zur Verarbeitung der Daten mit dem Hinweis auf die faktische Anonymität bzw. die Unmöglichkeit, einzelne Schüler*innen in den Daten zu Re-Identifizieren, abgewendet werden können.

Während der Videoaufzeichnung eines Interviews kommt es zudem zu einem **Widerruf** der Einwilligung zur Erhebung der Daten. Die interviewte Lehrkraft bricht das Interview ab. Neben dem Abbruch des Interviews kommt es durch die Lehrkraft jedoch nicht zum Widerspruch gegen die weitere Verarbeitung der Daten. Die Projektbeteiligten transkribieren den bereits aufgezeichneten Teil des Interviews und nutzen das Transkript im weiteren Projektverlauf entsprechend den in der Einwilligungserklärung dargelegten Verarbeitungszwecken.

2.10. Einschränkung von Betroffenenrechten

In einer Studie werden dieselben Personen wiederholt befragt (Paneldesign). Die Forschenden holen hierzu eine informierte Einwilligung der Teilnehmenden ein, die u. a. die Nutzung der Befragungsdaten zum Zweck der wissenschaftlichen Auswertung umfasst.

Da es sich bei den erhobenen Daten um personenbezogene Informationen handelt, stehen den Teilnehmenden grundsätzlich verschiedene Rechte zu, über die sie im Rahmen des Einholens der Einwilligung aufgeklärt werden müssen. Allerdings können Teile dieser Rechte für bestimmte Daten unter bestimmten Umständen durch die Forschenden eingeschränkt werden.

So gilt beispielsweise im Kontext der Befragungsdaten, dass die Wahrnehmung der Rechte der betroffenen Personen zu keiner ernsthaften Beeinträchtigung des Forschungsinteresses oder gar zu dessen Unmöglichkeit führen darf. Dies betrifft beispielsweise das Recht auf Berichtigung von personenbezogenen Informationen. Die Forschenden können dies ablehnen, wenn die Forschungsziele eine unverfälschte wissenschaftliche Nutzung von ad hoc gegebenen Informationen der Teilnehmenden voraussetzen. Eine nachträgliche Berichtigung von Informationen würde dem zuwiderlaufen.

Analog kann auch das Recht auf Einsichtnahme in die Daten, zumindest vorübergehend, eingeschränkt werden. Im vorliegenden Projekt wird ein Paneldesign genutzt, um wiederholt gegebene Antworten zu vergleichen. Aus wissenschaftlicher Sicht ist es dabei unabdingbar, dass die Teilnehmenden die Fragen (in den wiederholten Erhebungen) unverfälscht beantworten und nicht durch ein vorausgegangenes Abfragen ihrer Antworten aus vorherigen Erhebungen beeinflusst werden. Eine (zwischenzeitliche) Einsichtnahme in die Daten könnte die Datenqualität folglich unkontrolliert kompromittieren und die wissenschaftliche Nutzung unmöglich machen. Diese Argumentation ist jedoch an den Zeitraum der Datenerhebung gebunden und begründet keine Einschränkung der Einsichtnahme in die Daten im weiteren Projektverlauf.

Letztendlich können die Forschenden im Kontext der Umfragedaten die Rechte der betroffenen Personen auch durch die faktische Anonymisierung der Daten einschränken. Diese fallen dann mangels Personenbezug nicht mehr unter die rechtlichen Regelungen der Datenschutzgesetze.

2.11. Kontakt Daten über mehrere Wellen hinweg managen

In einem Forschungsprojekt werden in drei Wellen wiederholt Daten bei den gleichen Schüler*innen erhoben (Paneldesign). Die Erhebung der Daten erfolgt unmittelbar im Unterricht. Im Rahmen der zweiten Welle wird zudem allen Erziehungsberechtigten ein Selbstausfüllerbogen (PAPI) postalisch zugesandt. Die Adressinformationen für diese Erhebung entstammen der Schüler*innen-Befragung der ersten Welle.

Um die Verknüpfbarkeit der Daten über die einzelnen Wellen hinweg zu gewährleisten, existiert eine **Korrespondenztabelle**, die über eine eindeutige Kennziffer die Identifikation jeder Schüler*innen ermöglicht. Darüber hinaus enthält die Tabelle personenbezogene Informationen zu den Schüler*innen, wie Klarnamen und Angaben zur Klassen- bzw. Schulzugehörigkeit.

Gemäß den Einwilligungserklärungen werden personenbezogenen Daten, wie etwa Kontaktinformationen, nur so lange gespeichert, wie es für den weiteren Projektverlauf notwendig ist. Nicht mehr benötigte personenbezogene Daten werden vernichtet.

Die in der ersten Welle erhobenen Kontaktinformationen zu den Erziehungsberechtigten werden nach Eingang der Rohdaten unverzüglich von den (übrigen) Daten getrennt, um die Kennziffern der entsprechenden Schüler*innen ergänzt und zur weiteren Verarbeitung geschützt gespeichert. Nach Ende der Feldphase zur Befragung der Erziehungsberechtigten werden die eingegangenen Rohdaten zunächst kontrolliert und (teilweise) pseudonymisiert. Anschließend werden die Kontaktinformationen der Erziehungsberechtigten (Adressdaten) vernichtet.

Nach dem Zusammenspielen aller drei Erhebungswellen sowie der Umfragedaten der Erziehungsberechtigten in einen integrierten Datensatz erfolgt eine abschließende Datenkontrolle. Daraufhin wird auch die Korrespondenztabelle der Schüler*innen vernichtet.

Die Kriterien, wann genau welche Kontaktinformationen zu vernichten sind und welche Dateien bereits tatsächlich wie vernichtet wurden, wird im Rahmen des Vernichtungsmanagements dokumentiert.

2.12. Schutzverletzung durch Hackerangriff

Im Rahmen des BMBF-geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) kooperieren die beiden Hochschulen A und B mit der außeruniversitären Einrichtung C. Im Projekt wird die Lesekompetenz von Grundschüler*innen vor und nach den Sommerferien 2023 (zwei Erhebungswellen) an insgesamt 34 zufällig ausgewählten Grundschulklassen im gesamten Bundesgebiet untersucht. Zum kollaborativen Arbeiten nutzen die Mitarbeitenden des Projekts einen institutseigenen Clouddienst der Hochschule A. Daten mit sensiblen oder personenbezogenen Informationen werden dort in einem verschlüsselten Container mittels VeraCrypt¹ verarbeitet.

Im Projektverlauf wurden die Server der Hochschule A gehackt, wobei in bislang unbekanntem Umfang Dritte auf Daten zugreifen, diese kopieren, verändern und löschen konnten. Durch die Datenschutzbeauftragten und das Rechenzentrum der Hochschule A wurde bereits überprüft, welche Server von dem Hack betroffen sind. Hierzu zählen nach ersten Erkenntnissen auch die Server des im LesLer-Projekt genutzten Clouddienstes.

Die Projektmitarbeitenden prüfen zunächst ihre in der Cloud gesicherten Dateien (stichpunktartig) auf Funktionalität und Vollständigkeit ebenso wie auf Verschlüsselung aller Dateien mit sensiblen oder personenbezogenen Informationen. Zunächst können keine Veränderungen an den Dateien festgestellt werden. Alle zu verschlüsselnden Dateien waren zum Zeitpunkt des Hacks verschlüsselt. Die Projektbeteiligten erstellen eine Liste der Dateien mit sensiblen oder personenbezogenen Informationen und beschreiben kurz deren Inhalt, wie z. B. Kontaktdaten zu den Erziehungsberechtigten, Videoaufzeichnung mit Lehrkräften oder Verträge mit Dritten (siehe [Bestandsliste](#)). Analog dokumentieren sie die Kontrolle der Verschlüsselung und die zur Verschlüsselung genutzte Technologie. Sie übermitteln diese Liste an die Datenschutzbeauftragten und beachten im weiteren Verlauf Anfragen und Vorgaben der Datenschutzbeauftragten.

¹ VeraCrypt, Version 1.25.9 unter: <https://www.veracrypt.fr/code/VeraCrypt/>.

3. Urheberrecht

3.1. Nutzungsrechte an eigenen Materialien

Im Rahmen eines durch das *Bundesministerium für Bildung und Forschung* (BMBF) geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) arbeiten zwei Lehrstühle zweier Hochschulen (A & B) mit einer außeruniversitären Forschungseinrichtung des Bundes (C) zusammen. Das Projekt begann im Mai 2022 und endet im April 2025. Der Förderbescheid des BMBF setzt für entsprechende Projekte das weitestgehende Teilen der Forschungsergebnisse, wie Daten und Begleitmaterialien, voraus.

Im Projekt wird die Lesekompetenz von Grundschulkindern in der dritten Klasse vor den Sommerferien 2023 getestet (Daten der 1. Welle: LesLer 2023-1). Dieser Test wird nach den Sommerferien 2023 bei den gleichen Grundschulkindern wiederholt (Daten der 2. Welle: LesLer 2023-2). Der Test wird an insgesamt 34 zufällig ausgewählten Grundschulklassen mit dem Förderschwerpunkt *Sprache* im gesamten Bundesgebiet durchgeführt.

Das entwickelte Messinstrument basiert auf einem urheberrechtlich geschützten Testinstrument Dritter (Lesekompass 2018). Für die Nachnutzung dieses Testinstruments im Rahmen des Forschungsvorhabens liegt die Genehmigung der Urhebenden vor. Die im Projekt zu erhebenden Daten werden durch vorhandene Daten einer vergleichbaren Studie ergänzt (FördLes 2019).

Im Anschluss an eine erste Analyse der Daten nach der 2. Welle sind zusätzlich zehn Interviews mit Grundschullehrkräften geplant (LesInt 2023). Diese Interviews werden per Video aufgezeichnet. Die Transkription der Videoaufzeichnungen erfolgt durch eine*n externe*n Dienstleistende*n (Auftragsverarbeitende*r D). Im Rahmen des Projekts sollen zudem zwei empirische Dissertationen entstehen.

Zu Beginn der Projektlaufzeit (Mai 2022) entwirft die Projektleitung ein erstes Konzept zum Managen von Urheberrechten im Projekt. Im Rahmen der allgemeinen Projektorganisation betrifft dies a) alle im Projekt erstellten Forschungsergebnisse, wie Fachbeiträge oder Daten, sowie b) alle im Projekt zur Generierung von Forschungsergebnissen nachgenutzten Materialien anderer Personen. Aus dem Konzept zum Managen von Urheberrechten im Projekt geht für Daten und Begleitmaterialien Folgendes hervor:

Nutzungsrechte an Daten(-sammlungen) und Begleitmaterialien

Im Projekt und allen damit einhergehenden Publikationen von Forschungsergebnissen gelten die Regeln der guten wissenschaftlichen Praxis (Zitation, inkl. Quellangabe).

Verantwortlichkeiten für das Managen der Nutzungsrechte an Daten und Begleitmaterialien

Die Verantwortlichkeit für das Managen der Urheberrechte an projekteigenen Daten und Begleitmaterialien sowie an Materialien Dritter im Kontext der Datenerhebung, Aufbereitung und dem Datenteilen liegt bei der außeruniversitären Forschungseinrichtung (C).

Nutzungsrechte an projekteigenen Daten und Begleitmaterialien

Gemäß Förderbescheid des BMBF sind die Hochschulen (A & B) und die außeruniversitäre Forschungseinrichtung (C) Zuwendungsempfänger und besitzen die Verwertungsrechte an den im Projekt erstellten Daten und Begleitmaterialien.

Die Verteilung der Nutzungsrechte an den projekteigenen Daten und Begleitmaterialien sind gemäß Kooperationsvertrag wie folgt geregelt:

- Die Verbundpartner stimmen darin überein, gemeinsam Schöpfer der Datensammlung LesLer 2023 und LesInt 2023 sowie damit einhergehender Begleitmaterialien zu sein.
- Die Nutzungsrechte an projekteigenen Datensammlung und Begleitmaterialien stehen im Rahmen des Vorhabens allen Verbundpartnern in gleicher Weise zu.
- Die Nutzung der Datensammlung und Begleitmaterialien für die beiden Dissertationen (ggf. auch über April 2025 hinaus) wird von den Verbundpartnern gewährleistet.
- Die Verbundpartner übertragen, gemäß Förderbescheid, im Rahmen des Teilens ihrer Datensammlung einfache Verwertungsrechte zur Sicherung, Vervielfältigung, technischen Anpassung und Bereitstellung der Daten an das Forschungsdatenzentrum der außeruniversitären Forschungseinrichtung (C).

3.1. Daten in der Cloud speichern

Das Verbundprojekt *Lesen Lernen* (LesLer 2022) bedient sich zur Sicherung und Übermittlung der vollständig anonymisierten Daten und ihrer Begleitmaterialien zwischen den Standorten der Verbundpartner eines (kommerziellen) Clouddienstes.

Aus den Nutzungsbedingungen des Clouddienstes geht hervor, dass Informationen, die über das Cloudsystem „gespeichert, verarbeitet und übertragen werden“, vom Anbietenden des Clouddienstes genutzt werden dürfen, um „die Dienste anbieten, verbessern, sichern und vermarkten zu können“. Entsprechend dieser Nutzungsbedingungen übertragen die Forschenden mit dem Upload der Daten und Begleitmaterialien einfache Nutzungsrechte an die Anbietenden des Cloudsystems, z. B. zur Sicherung und Vervielfältigung der Dateien. Darüber hinaus sichern sich die Anbietenden aber auch weiterreichende Nutzungsrechte, etwa um ihre „Dienste zu ,verbessern“, ohne dass die damit einhergehende Verwendung in den Nutzungsbedingungen konkretisiert werden.

Eine Miturheberschaft der Anbietenden des Clouddienstes an den Daten und Begleitmaterialien, die z. B. bei der Verarbeitung oder dem Teilen der Daten berücksichtigt werden müsste, entsteht durch die Nutzung des Dienstes jedoch nicht. Dennoch sollten es Forschende nach Möglichkeit vermeiden, derart unkonkrete Bedingungen zur Datennutzung zu akzeptieren.

3.2. Nutzungsrechte an Materialien Dritter

Im Rahmen eines durch das *Bundesministerium für Bildung und Forschung* (BMBF) geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) arbeiten zwei Lehrstühle zweier Hochschulen (A & B) mit einer außeruniversitären Forschungseinrichtung des Bundes (C) zusammen. Das Projekt begann im Mai 2022 und endet im April 2025. Der Förderbescheid des BMBF setzt für entsprechende Projekte das weitestgehende Teilen der Forschungsergebnisse, wie Daten und Begleitmaterialien, voraus.

Im Projekt wird die Lesekompetenz von Grundschulkindern in der dritten Klasse vor den Sommerferien 2023 getestet (Daten der 1. Welle: LesLer 2023-1). Dieser Test wird nach den Sommerferien 2023 bei den gleichen Grundschulkindern wiederholt (Daten der 2. Welle: LesLer 2023-2). Der Test wird an insgesamt 34 zufällig ausgewählten Grundschulklassen mit dem Förderschwerpunkt *Sprache* im gesamten Bundesgebiet durchgeführt.

Das entwickelte Messinstrument basiert auf einem urheberrechtlich geschützten Testinstrument Dritter (Lesekompass 2018). Für die Nachnutzung dieses Testinstruments im Rahmen des Forschungsvorhabens liegt die Genehmigung der Urhebenden vor. Die im Projekt zu erhebenden Daten werden durch vorhandene Daten einer vergleichbaren Studie ergänzt (FördLes 2019).

Im Anschluss an eine erste Analyse der Daten nach der 2. Welle sind zusätzlich zehn Interviews mit Grundschullehrkräften geplant (LesInt 2023). Diese Interviews werden per Video aufgezeichnet. Die Transkription der Videoaufzeichnungen erfolgt durch eine*n externe*n Dienstleistende*n (Auftragsverarbeitende*r D). Im Rahmen des Projekts sollen zudem zwei empirische Dissertationen entstehen.

Zu Beginn der Projektlaufzeit (Mai 2022) entwirft die Projektleitung ein erstes Konzept zum Managen von Urheberrechten im Projekt. Im Rahmen der allgemeinen Projektorganisation betrifft dies a) alle im Projekt erstellten Forschungsergebnisse, wie Fachbeiträge oder Daten, sowie b) alle im Projekt zur Generierung von Forschungsergebnissen nachgenutzten Materialien anderer Personen. Aus dem Konzept zum Managen von Urheberrechten im Projekt geht für Daten und Begleitmaterialien Folgendes hervor:

Nutzungsrechte an Materialien Dritter im Kontext der Datenverarbeitung:

Im Projekt werden zwei Quellen von urheberrechtlich geschützten Materialien anderer Personen nachgenutzt:

1. FördLes-2019-Daten: Die Daten sind zur Nutzung für wissenschaftliche Zwecke über das Forschungsdatenzentrum der außeruniversitären Forschungseinrichtung (C) verfügbar. Die Nutzungsbedingungen erlauben die Verwendung der Daten gemäß dem Vorhaben ebenso wie für die beiden Dissertationen (auch über das Projekt hinaus).

Gemäß Nutzungsbedingungen müssen die Daten nach Abschluss der Nutzung, d. h. (spätestens) zum Projektende (April 2025) vernichtet werden. Zu Replikations- bzw. Verifikationszwecken von Projektergebnissen, können die Daten auch nach Projektende weiter über das Forschungsdatenzentrum bezogen werden.

Die Zitation der Daten, ihrer Schöpfenden und ihrer Herkunft (Quelle) folgt den Vorgaben des Forschungsdatenzentrums.

2. LeseKompass 2016: Die Nutzung des Messinstruments zur Erhebung der LesLer-2023-Daten liegt in Form einer individuellen Nutzungsvereinbarung mit den Urheberrechtsinhabenden vor. Gemäß dieser darf das Messinstrument im Rahmen des Projekts zur Datenerhebung und -auswertung genutzt werden.

Eine Veröffentlichung des Messinstruments ist nicht gestattet, weder in Publikationen von Forschungsergebnissen des Projekts noch in der Datendokumentation oder den beiden Dissertationen.

Das Messinstrument darf zu Replikations- bzw. Verifikationszwecken langfristig über das Projektende hinaus intern gesichert werden.

Die Zitation des Messinstruments, der Schöpfenden und der Quelle ergibt sich aus der Vereinbarung mit den Urheberrechtsinhabenden.

3.3. Urheberrecht und Testschutz bei der Verwendung von Tests

Ein Forschungsprojekt untersucht die Zusammenhänge zwischen Persönlichkeitseigenschaften von Lehrkräften und den schulischen Leistungen der von ihnen unterrichteten Schüler*innen im Deutschunterricht. Von besonderem Interesse ist dabei der Einstieg in den Lehrberuf.

In drei Erhebungswellen werden bei neu eingestiegenen Lehrkräften Persönlichkeitseigenschaften erfasst. Zur Erhebung der Persönlichkeitseigenschaften wird das urheberrechtlich geschützte Messinstrument *NEO-Persönlichkeitsinventar* (NEO-PI-R)² genutzt. Um in der Analyse auf mögliche Wechselwirkungen mit der Intelligenz der Schüler*innen zu kontrollieren, wird zusätzlich bei den Schüler*innen der ebenfalls urheberrechtlich geschützte *Grundintelligenztest Skala 2* (CFT-20-R Test)³ durchgeführt.

Zur größtmöglichen Nachvollziehbarkeit der geteilten Forschungsdaten, ist die Bereitstellung des Originalwortlauts von Fragebogenitems, z. B. im Rahmen eines Codebuchs, ein wesentlicher Bestandteil der Datendokumentation. Dem stehen im konkreten Fall allerdings (1) urheberrechtliche Aspekte sowie (2) Bedenken zum Testschutz im Sinne der Wahrung der Validität von Testverfahren entgegen. Zudem ist aufgrund der wiederholten Datenerhebung (drei Wellen) (3) eine wiederholte Prüfung etwaiger Änderungen der Nutzungsbedingungen für beide Messinstrumente erforderlich.

Urheberrechtliche Aspekte

Die Nutzungsrechte am Wortlaut der Items des Messinstrument *NEO-Persönlichkeitsinventar* (NEO-PI-R) liegen beim herausgebenden Verlag. Die Nutzungsbedingungen des Verlags schließen eine Verbreitung der Items (im Wortlaut) explizit aus, was auch die Dokumentation des Items zur Nachvollziehbarkeit der Daten betrifft. Durch die adäquate Zitation (Quellangabe) des zugehörigen Testmanuals erhalten Nachnutzende der Daten jedoch die Möglichkeit das Messinstrument (bei Bedarf) selbst zu beziehen. Die Forschenden schaffen so (größtmögliche) Transparenz und Nachvollziehbarkeit (bezüglich Erhebung und Aufbereitung) ihrer Daten.

Testschutz zur Wahrung der Validität von Testverfahren

Besonders bei Items von Intelligenztests, kann deren freie Verfügbarkeit die Validität des Tests einschränken, da sich Proband*innen auf etwaige Tests vorbereiten und im Vorhinein richtige Antworten einüben könnten. Für das Messinstrument *Grundintelligenztest Skala 2* (CFT-20-R Test) bedeutet dies, dass im Rahmen des Tests präsentierte Stimuli nicht veröffentlicht werden dürfen (unabhängig etwaiger urheberrechtlicher Auflagen), um die Validität des Testverfahrens zu sichern.

Die Forschenden beschränken sich bei der Dokumentation der zu teilenden Daten auf die Angabe, ob eine Aufgabe richtig oder falsch gelöst wurde; die tatsächlich vorliegenden bzw. gewählten Antwortoptionen werden zur Wahrung des Testschutzes hingegen nicht dokumentiert und mit Dritten geteilt. Analog zur Dokumentation des Messinstruments *NEO-Persönlichkeitsinventar* (NEO-PI-R) wird es Nachnutzenden der Daten durch eine adäquate Zitation (Quellangabe) des Messinstruments *Grundintelligenztest Skala 2* (CFT-20-R Test) jedoch ermöglicht, dieses selbst zu verwenden.

² Beim *NEO-Persönlichkeitsinventar* (NEO-PI-R) handelt es sich um einen Fragebogen zur Messung von Persönlichkeiten (240 Items) anhand eines Fünf-Faktoren-Modells (Neurotizismus, Extraversion, Offenheit für Erfahrung, Verträglichkeit und Gewissenhaftigkeit).

³ Durch das Messinstrument *Grundintelligenztest Skala 2* (101 Items) lassen sich das allgemeine intellektuelle Niveau erfassen.

Prüfung etwaiger Änderungen der Lizenzbedingungen bzw. der Nutzungserlaubnis

Aus Gründen der Nachvollziehbarkeit und zur Dokumentation der Datenerhebung halten die Forschenden sorgfältig fest, welche Version des Testverfahrens in welcher Welle genutzt wurde. Zudem kontrollieren die Forschenden vor jeder Erhebungswelle auf etwaige Änderungen der Nutzungsbedingungen durch den Verlag und prüfen die sich aus potenziellen Nutzungsänderungen ergebenden Auswirkungen auf die Nutzung der Testverfahren im Projekt. Auftretende Änderungen der Lizenzbedingungen werden ebenso wie damit einhergehende Konsequenzen für die Verarbeitung der Daten im Rahmen der Datendokumentation festgehalten.

3.4. Urheberrecht bei der Verwendung von Unterrichtsmaterialien

In einem Forschungsprojekt werden Videoaufzeichnungen von Unterrichtssituationen erstellt. Dabei werden auch Unterrichtsmaterialien von Lehrkräften und Schüler*innen (wie z. B. Tafelbilder oder selbst erstellte Arbeitsblätter, schriftliche Arbeiten, Bilder, Referate etc.) sowie Abbildungen aus Schulbüchern und im Unterricht eingebundene Apps, Streams oder Videos (DVD, VHS) aufgezeichnet.

Die Forschenden lassen sich im Vorfeld der Datenerhebung (Videoaufzeichnung) durch die Lehrkräfte und Schüler*innen die Verarbeitung (Erheben, Sichern, Auswerten etc.) urheberrechtlich geschützter Materialien dieser Personen, die Teil der Aufnahmen werden, schriftlich genehmigen.

Nach der Datenerhebung prüfen die Forschenden zunächst die Urheberschaft an aufgezeichneten Materialien. Besonders berücksichtigt werden dabei solche Materialien, bei denen die Urheberschaft nicht (ausschließlich) bei den Lehrkräften bzw. Schüler*innen liegt, wie z. B. bei Abbildungen aus Schulbüchern.

Unterrichtsmaterialien von Lehrkräften und Schüler*innen

Die Forschenden entfernen nach Abschluss der Datenerhebung zunächst alle urheberrechtlich geschützten Materialien von Lehrkräften und Schüler*innen, für die keine Genehmigung zur Verarbeitung vorliegt. Entsprechende Passagen in den Videoaufzeichnungen werden nach Einzelfallprüfung entweder verpixelt (Erhalt der jeweiligen Videopassage) oder ganz aus den Aufnahmen geschnitten (Löschung der jeweiligen Videopassage) und die Eingriffe in die Aufzeichnungen dokumentiert.

Für das Teilen der Daten werden enthaltene Materialien von Lehrkräften und Schüler*innen entsprechend pseudonymisiert (teilweise entfernt bzw. verpixelt) und die Veränderung in den Aufzeichnungen entsprechend dokumentiert.

Materialien aus (kommerziellen) Schulbüchern, Streams und Apps

Im Kontext der Aufzeichnung urheberrechtlich geschützter Materialien Dritter, wie etwa Abbildungen aus Schulbüchern, verzichten die Forschenden mit Verweis auf §60c UrhG auf das Einholen entsprechender Nutzungsrechte bei den Urheberrechtsinhabenden. Gemäß §60c UrhG stellen die Forschenden sicher, dass sie die entsprechenden Materialien im Rahmen des Forschungsvorhabens nutzen dürfen, indem beispielsweise weniger als 15 Prozent eines Werkes aufgezeichnet wurde.

Beim Teilen der Daten zur Nachnutzung durch Dritte werden entsprechende urheberrechtliche Materialien Dritter aus den Videoaufzeichnungen entfernt bzw. verpixelt. Die veränderten Passagen in den Aufzeichnungen werden im Rahmen der Datendokumentation gekennzeichnet, Gründe für die Löschung bzw. Veränderung erörtert und die originären Materialien adäquat zitiert (Quellangabe), so dass interessierte Nachnutzende der Daten diese Materialien eindeutig identifizieren und wiederfinden können.

4. Datenorganisation

4.1. Konzept zur Datenorganisation

Im Rahmen eines durch das *Bundesministerium für Bildung und Forschung* (BMBF) geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) arbeiten zwei Lehrstühle zweier Hochschulen (A & B) mit einer außeruniversitären Forschungseinrichtung des Bundes (C) zusammen. Das Projekt begann im Mai 2022 und endet im April 2025. Der Förderbescheid des BMBF setzt für entsprechende Projekte das weitestgehende Teilen der Forschungsergebnisse, wie Daten und Begleitmaterialien, im Sinne des *Open Science* voraus.

Im Projekt wird die Lesekompetenz von Grundschulkindern in der dritten Klasse vor den Sommerferien 2023 getestet (Daten der 1. Welle: LesLer 2023-1). Dieser Test wird nach den Sommerferien 2023 bei den gleichen Grundschulkindern wiederholt (Daten der 2. Welle: LesLer 2023-2). Der Test wird an insgesamt 34 zufällig ausgewählten Grundschulklassen mit dem Förderschwerpunkt *Sprache* im gesamten Bundesgebiet durchgeführt.

Das entwickelte Messinstrument basiert auf einem urheberrechtlich geschützten Testinstrument Dritter (Lesekompass 2018). Für die Nachnutzung dieses Testinstruments im Rahmen des Forschungsvorhabens liegt die Genehmigung der Urhebenden vor. Die im Projekt zu erhebenden Daten werden durch vorhandene Daten einer vergleichbaren Studie ergänzt (FördLes 2019).

Im Anschluss an eine erste Analyse der Daten nach der 2. Welle sind zusätzlich zehn Interviews mit Grundschullehrkräften geplant (LesInt 2023). Diese Interviews werden per Video aufgezeichnet. Die Transkription der Videoaufzeichnungen erfolgt durch eine*n externe*n Dienstleistende*n (Auftragsverarbeitende*r D). Im Rahmen des Projekts sollen zudem zwei empirische Dissertationen entstehen.

Zu Beginn der Projektlaufzeit (Mai/Juni 2022) wurde ein erster Konzeptentwurf zum Managen der Daten erstellt. Aus dem Entwurf geht u. a. Folgendes hervor:

Speichermöglichkeiten und Ordnerstruktur

Um das kollaborative Arbeiten im Verbundprojekt zu erleichtern, wird das Cloud-System der Hochschule A genutzt (online Bearbeitung und Speicherung der Dateien). Die zu managenden Dateien sind in der Bestandsliste des Projekts zusammengefasst (siehe Fallbeispiel *Bestandsliste*).

Verantwortlichkeiten für den Betrieb und die Nutzung der Cloud: Die Verantwortung für den Betrieb des Cloud-Systems liegt bei Hochschule A (IT).

Speicherkapazität: Der zur Verfügung gestellte Speicherplatz von derzeit 50 Gigabyte wird als für das Projekt ausreichend erachtet. Eine weitere Aufstockung der Speicherkapazität ist auf Antrag an der Hochschule A möglich (siehe Bestandsliste).

Datenschutz in der Cloud: Zur Gewährleistung datenschutzrechtlicher Vorschriften bei der Nutzung der Cloud, wurde dort ein verschlüsselter Container mittels VeraCrypt (Version 1.25.9)¹ angelegt, in dem alle gespeicherten Dateien (end-to-end) verschlüsselt gespeichert und verarbeitet werden.

Nutzung der Cloud: In der Cloud

- besteht eine Ordnerstruktur (aufbauend auf den Arbeitspaketen des Projekts), auf deren oberster Eben sich eine *readme*-Datei zur Nutzung der Cloud sowie u. a. zur Ordnerstruktur und zum gesicherten Speichern von Dateien mit sensiblen und/oder personenbezogenen Inhalten findet;
- besteht ein automatisiertes Versionierungssystem (wie in der *readme*-Datei erklärt);
- können beschädigte, versehentlich gelöschte oder überschriebene Dateien 90 Tage lang wiederhergestellt werden (kurzzeitiges Back-Up).

Dateibenennung, Dateiformate und genutzte Software zur Datenverarbeitung

Im Rahmen der Projektkoordination bestehen für alle Mitarbeitenden im Projekt frei zugängliche, verschriftlichte Regeln (siehe *readme*-Datei) zu

- in der Cloud genutzter Software (Version) zur Datenverarbeitung;
- Dateiformaten in der Cloud, zum Upload von Dateien ebenso wie zum Download und zur Formattransformationen;
- der Benennung von Dateien in der Cloud.

Back-Ups (langfristig)

Da die Möglichkeiten der Wiederherstellung von Dateien in der Cloud auf 90 Tage begrenzt sind, wird monatlich ein zusätzliches, verschlüsseltes Back-Up aller Dateien in der Cloud durch die Projektmitarbeitenden der Hochschule A erstellt.

Verantwortlichkeiten für das händische Erstellen der Back-Ups: Die Verantwortung für das Erstellen und die Übermittlung der zusätzlichen Back-Ups liegt in der Verantwortung der Hochschule A. Die Sicherung und die Kontrolle des Zugangs zu den händischen Back-Ups an den jeweiligen Einrichtungen liegt bei den Projektmitarbeitenden der entsprechenden Einrichtung.

Erstellen händischer Back-Ups: Die verantwortlichen Mitarbeitenden der Hochschule A erstellen zum Ende jedes Monats zusätzlich ein verschlüsseltes Back-Up aller Dateien in der Cloud und übermitteln diese verschlüsselte Datei (über die Cloud) an die Verbundpartner (B & C). Die drei Verbundpartner speichern die verschlüsselten Back-Ups in einem gesicherten, zugangskontrollierten Serverbereich ihrer jeweiligen Einrichtung.

Zugriff auf die Back-Ups: Ein Zugriff auf die Back-Ups ist jeder*m Projektmitarbeitenden über die an der jeweiligen Einrichtung gesicherten Dateien möglich. Die Regelungen für den Zugriff auf die Back-Up-Dateien folgen den institutionellen Vorgaben der jeweiligen Verbundpartner. Die Nutzung der händischen Back-Ups durch die Verbundpartner ist den übrigen Verbundpartnern unter Nennung der Gründe für den Rückgriff auf die Back-Ups schriftlich mitzuteilen.

Vernichtungsmanagement:

Die im Projekt nachgenutzten urheberrechtlich geschützten Daten und Materialien (Messinstrument) werden entsprechend den Nutzungsbedingungen bzw. den individuellen Vereinbarungen mit den Urheberrechtsinhabenden vernichtet.

Verantwortlichkeiten für das Vernichtungsmanagement: Die Verantwortung für das Vernichtungsmanagement liegt bei der außeruniversitären Einrichtung (C), in deren Verantwortung auch die Aufbereitung der (Roh-)Daten fällt.

Dokumentation zu vernichtender Dateien: Für alle zu löschenden Dateien dokumentieren die verantwortlichen Projektmitarbeitenden im Projektverlauf den Umfang und die Inhalte der Daten, ihre Herkunft (Zitation mit Quellangaben und (soweit möglich) persistentem Identifikator) sowie die Nutzungsbedingungen im Projekt und die Löschauflagen.

Vernichtungsmanagement: Um diesen Löschaufgaben nachzukommen, führen die verantwortlichen Projektmitarbeitenden im Rahmen der Projektorganisation ein geeignetes Vernichtungsmanagement und protokollieren dieses fortlaufend in einer Tabelle, die allen Projektbeteiligten zugänglich ist. Zum aktuellen Zeitpunkt (Mai 2022) betrifft das Vernichtungsmanagement u. a.:

1. FördLes 2019 (im Projekt nachgenutzte Daten Dritter):

- die Daten werden spätestens zum Projektende vernichtet; zu prüfen ist, ob die Daten auch nach Projektabschluss zur Fertigstellung/Finalisierung der beiden Dissertationen benötigt und folglich auch über das eigentliche Projektende (April 2025) erhalten werden müssen;
- die FördLes-2019-Daten können auch nach der Vernichtung im Projekt, z. B. zu Replikationszwecken, über das Forschungsdatenzentrum der außeruniversitären Forschungseinrichtung (C) bezogen werden.

2. LeseKompass 2018 (im Projekt nachgenutztes Messinstrument):

- entsprechend der Vereinbarung mit den Urheberrechtsinhabenden, kann das Messinstrument intern für 15 Jahre über das Projektende hinaus gesichert werden.

3. LesLer 2023 (beide Wellen):

- die Vernichtung bzw. die Notwendigkeit eines längerfristigen Erhalts der Rohdaten beider Wellen wird noch geprüft;
- die beiden aufbereiteten Daten (LesLer 2023-1 und LesLer2023-2) werden nach Erstellen, Prüfen und Teilen des integrierten Datensatzes (LesLer 2023, beide Wellen) vernichtet; mit dem Archivieren und Teilen des integrierten Datensatzes ist ein weiterer Erhalt der beiden Einzeldatensätze nicht mehr nötig;
- Kontaktdaten (Schüler*innen und Erziehungsberechtigten) werden unmittelbar nach Kontrolle und Aufbereitung der zweiten Welle (LesLer 2023-2) durch die außeruniversitäre Forschungseinrichtung (C) vernichtet.

4. LesInt 2023:

- die Vernichtung der Daten LesInt 2023 bzw. die Notwendigkeit eines längerfristigen Erhalts der originalen Videoaufnahmen ebenso wie der nicht-pseudonymisierten Transkripte (Rohversion) wird noch geprüft.

1. VeraCrypt, Version 1.25.9 unter: <https://www.veracrypt.fr/code/VeraCrypt/>.

Auszug aus der Tabelle zum Vernichtungsmanagement (Stand: Juni 2022)

Dateiname	Kurzbeschreibung (Inhalt)	Speicherorte	(vertraglicher) Löschzeitpunkt	Dokumentation (Inhalt & Grund)	vernichtet, am	vernichtet, durch	verantwortliche Person	Anmerkungen
20220304_LesLer-2023-1_R-Daten_v00.Rdata	Roh-Leistungsdaten & Sozio-demographie, Welle 1	Arbeitsverzeichnis (Server)	30.04.2025	wird erstellt			Partner A (IT)	
		3 Back-Ups an den Einrichtungen der Partner (A, B & C)	30.04.2025				Partner A, B & C (IT)	
202202XX_LesLer-2023-1_F-Daten_v00.Rdata	aufbereitete Leistungsdaten & Soziodemographie, Welle 1	Arbeitsverzeichnis (Server)	30.04.2025				Partner A (IT)	
		3 Back-Ups an den Einrichtungen der Partner (A, B & C)	30.04.2025				Partner A, B & C (IT)	als Teil des integrierten Datensatzes teilen
		Arbeitsrechner der Projektbeteiligten	30.04.2025				Mitarbeitende	
...
202202XX_LesLer-2023_Kontaktdaten_v00.txt	Kontaktdaten zu Schüler*innen & Erziehungsberechtigte	Arbeitsverzeichnis (Server)	09.2023	wird erstellt			Partner A (IT)	zu vernichten nach Abschluss der zweiten Erhebung
		3 Back-Ups an den Einrichtungen der Partner (A, B & C)	09.2023				Partner A, B & C (IT)	
... WEITERE DATEIEN IN AKTUELLEM FALLBEISPIEL AUSGELASSEN ...								
20220207_Lesekompass-2018_v1.0.rft	Messinstrument zu Leistungstests	Arbeitsverzeichnis (Server)	30.04.2025	liegt vor			Partner A (IT)	
		3 Back-Ups an den Einrichtungen der Partner (A, B & C)	30.04.2025				Partner A, B & C (IT)	
		Langfristsicherung	30.04.2040				Partner C	
202212XX_LesInt_Dok-Int-1_v0.0.rft	Dokumentationsbogen zum Interview mit Lehrkraft 01 (nicht pseudonymisiert)	Arbeitsverzeichnis (Server)	30.04.2025	wird erstellt			Partner A (IT)	
		3 Back-Ups an den Einrichtungen der Partner (A, B & C)	30.04.2025				Partner A, B & C (IT)	
202212XX_LesInt_Vid-Int-1_v0.0.rft	Videoaufzeichnung des Interviews mit Lehrkraft 01 (nicht pseudonymisiert)	Arbeitsverzeichnis (Server)	30.04.2025	wird erstellt			Partner A (IT)	
		3 Back-Ups an den Einrichtungen der Partner (A, B & C)	30.04.2025				Partner A, B & C (IT)	
202212XX_LesInt_Trans-Int-1_v0.0.rft	Roh-Transkript Videoaufzeichnung	zur des Arbeitsverzeichnis (Server)	30.04.2025	Anonymisierungsprotokoll wird erstellt				Partner A (IT)

Interviews mit Lehrkraft 01
(nicht pseudonymisiert)

... WEITERE DATEIEN IN AKTUELLEM FALLBEISPIEL AUSGELASSEN ...

4.2. Eigene Back-Up-Strategie

Zur Umsetzung eines gemeinsamen Promotionsprojekts erhalten zwei Stipendiat*innen (S & T) eine Förderung durch eine private Stiftung. Ausgaben für Infrastrukturmaßnahmen, wie z. B. Computer, Speichermedien etc., sind von dieser finanziellen Förderung explizit ausgenommen. Darüber hinaus sind die Stipendiat*innen nicht in eine bestehende institutionelle Infrastruktur eingebunden, so dass ihnen beispielsweise keine Büroräume oder Serverbereiche zum Speichern von Daten und Dateien zur Verfügung gestellt werden. Zwar erlaubt die Stiftung die Nutzung eines stiftungseigenen Cloud-Systems zum Austauschen von Dateien zwischen Stipendiat*innen, für das Cloud-System bestehen jedoch keine Back-Up-Verfahren.

Aufgrund mangelnder Alternativen entwickeln die Stipendiat*innen entsprechend ihrer Bedarfe ein eigenes Back-Up-Verfahren. Aus den verschriftlichten Vereinbarungen der beiden geht hervor:

Back-Up Verfahren

Das gesamte Arbeitsverzeichnis des Projekts wird im Cloud-System der Stiftung gesichert und bearbeitet. Pro Woche werden insgesamt vier Sicherungskopien dieses Arbeitsverzeichnisses an zwei unterschiedlichen Wochentagen erstellt:

1. Die ersten beiden Sicherungskopien werden jeden Mittwoch durch die Stipendiatin bzw. den Stipendiaten S erstellt; eine der beiden Sicherungskopien erfolgt auf dem privaten Rechner von S, die zweite Sicherungskopie wird extern auf einem USB-Stick gespeichert, der bei S zu Hause gelagert wird.
2. Die beiden verbleibenden wöchentlichen Sicherungskopien erfolgen jeden Freitag durch die Stipendiatin bzw. den Stipendiaten T und werden ebenfalls zum einem auf dem Privatrechner von T sowie zum anderen auf einem USB-Stick gespeichert, der zu Hause bei T gelagert wird.

Vorgehen

1. Zum Schutz personenbezogener und sensibler Informationen im Projekt wird zunächst ein verschlüsselter Zip-Ordner des Arbeitsverzeichnisses mittels *7-Zip*¹ erstellt.
2. Der verschlüsselte Zip-Ordner wird auf dem Privatrechner und dem USB-Stick der*des erstellenden Stipendiat*innen gespeichert und mittels *Checksum*² mit der Ausgangsdatei (Zip-Ordner im Cloud-System der Stiftung) verglichen (Kontrolle der Datenintegrität); der Zip-Ordner im Cloud-System wird daraufhin gelöscht.
3. Die Stipendiat*innen bewahren die beiden letzten (von ihnen erstellten) Sicherungskopien auf ihrem Privatrechner bzw. dem USB-Stick auf (geschätztes notwendiges Speichervolumen auf den Privatrechnern und den USB-Sticks jeweils ca. 10 bis 15 GB). Um die notwendigen Speicherkapazitäten für die Back-Ups gering zu halten, werden ältere Sicherungskopien auf
 - a) den Privatrechnern vernichtet (siehe hierzu auch Punkt 5 dieser Vereinbarung),
 - b) dem USB-Sticks gelöscht (siehe hierzu auch Punkt 6 dieser Vereinbarung).
4. Das Erstellen der Sicherungskopien liegt in der Verantwortung der*des beteiligten Stipendiat*innen S (mittwochs) bzw. der*des Stipendiat*innen T (freitags); im Falle der Abwesenheit einer*eines Stipendiat*innen übernimmt die*der jeweils andere Stipendiat*in die anfallenden Sicherungskopien der abwesenden Person.
5. Zum Schutz personenbezogener und sensibler Informationen im Projekt werden die Daten auf den Festplatten der Privatrechner am Projektende durch Überschreiben der Dateien mittels *Eraser*³ vernichtet.
6. Um technischen Defekten vorzubeugen, werden die zur Sicherung genutzten USB-Sticks alle sechs Monate durch neue USB-Sticks (32GB) ersetzt. Die veralteten USB-Sticks werden neu formatiert, um die darauf befindlichen Dateien endgültig zu vernichten.

Zum Ende des ersten Projektjahres vereinbaren die beiden Stipendiat*innen eine Evaluierung und ggf. Anpassung des vereinbarten Back-Up-Verfahrens.

-
1. <https://www.7-zip.org/>.
 2. https://www.chip.de/downloads/Checksum_71303943.html.
 3. https://www.chip.de/downloads/Eraser_12994923.html.

5. Nachvollziehbarkeit

5.1. Stichprobenverfahrens und Fallauswahl dokumentieren

Ein Forschungsprojekt führte eine Studie zur Untersuchung zum Einfluss von Familienverhältnissen auf die Sprachentwicklung von Viertklässler*innen an Grundschulen im Berliner Bezirk Spandau durch. Zentrale Interessen lagen dabei einerseits auf Unterschieden in der sprachlichen Entwicklung von Schüler*innen in privaten und öffentlichen Grundschulen. Im Bezirk Spandau existierten zum Zeitpunkt der Stichprobenziehung laut Verzeichnis der Berliner Schulen vier private und 29 öffentliche Grundschulen. Von diesen insgesamt 33 Grundschulen wurden vier Schulen (eine private und drei öffentliche) per Zufallsauswahl ausgewählt und alle Viertklässler*innen im Rahmen des Unterrichts mittels eines standardisierten Fragebogens befragt (Schulinterviews).

Andererseits interessierte die Forschenden der Einfluss des familiären Hintergrunds auf die Entwicklung der Sprachkompetenz, wobei im Projekt zwischen Schüler*innen mit und ohne familiären Migrationshintergrund unterschieden wurde. Hierzu wurden im Anschluss an die Schulinterviews Interviews in den Haushalten der Schüler*innen mit einem Elternteil / Erziehungsberechtigten (Elterninterview) sowie (falls gegeben) mit einem Geschwisterkind (Geschwisterinterview) durchgeführt. Dazu zogen die Primärforschenden auf Basis der an den Schulinterviews teilnehmenden Schüler*innen Subsample im Umfang von 70 % der Haushalte von Schüler*innen mit familiärem Migrationshintergrund sowie 50 % der Haushalte von Schüler*innen ohne familiären Migrationshintergrund.

Aus dem Methodenbericht des Projektes geht unter dem Punkt Fallauswahl und Stichprobenziehung Folgendes hervor:

Stichprobenziehung: Schulen und Schüler*innen

1. Entsprechend des Schulverzeichnisses der Berliner Schulen wurde eine Schulliste aller 33 Grundschulen (4 private und 29 öffentliche Grundschulen) in Spandau (mit insgesamt 2671 Schüler*innen der 4. Klassen) angelegt.
2. Zur Ziehung der geschichteten Zufallsstichprobe wurden die Grundschulen bezüglich des interessierenden Merkmals (private versus öffentliche Grundschulen) entsprechend ihrer Verteilung im Schulverzeichnis in zwei Cluster eingeteilt und die Schulen in beiden Clustern nach der Anzahl der Schüler*innen sortiert, wobei:
 - a) bei den privaten Grundschulen die Schulen nach absteigender Schüler*innenzahl sortiert (von vielen zu wenigen Schüler*innen) und,
 - b) bei den öffentlichen Grundschulen in umgekehrter Reihenfolge, d. h. von wenigen zu vielen Schüler*innen sortiert wurde.

Dadurch wurde die Wahrscheinlichkeit, eine große öffentliche Schule zu ziehen, minimiert und die Wahrscheinlichkeit, eine große private Grundschule zu ziehen, maximiert.

3. Mithilfe einer einfachen Zufallsauswahl (Simple Random Sampling mit Zufallsstart und gleichen Intervallen) wurden dann 4 Schulen gezogen, an denen alle Schüler*innen der 4. Klassen im Zeitraum vom 13. Oktober bis 25. November 2021 (Schulinterviews) befragt wurden.

Stichprobenziehung: Eltern und Geschwister

Nach Abschluss der Schulinterviews wurden Haushalte zur weiteren Befragung (der Eltern / Erziehungsberechtigten sowie Geschwisterkinder) ausgewählt.

1. Auswahlgesamtheit für die Ziehung der Haushalte bildeten alle Schüler*innen, die an den Schulinterviews teilnahmen und für die eine Einwilligung (der Eltern / Erziehungsberechtigten) zur weiteren Befragung im Haushalt vorlag.

Bei Doppelnennungen von Haushalten, z. B. aufgrund von befragten Schüler*innen, die im gleichen Haushalt leben (gemeinsame Befragung von Geschwisterkindern an einer Schule kam in insgesamt vier Fällen bzw. bei acht befragten Schüler*innen vor; Befragungen von Schüler*innen eines Haushalts in unterschiedlichen Schulen kamen nicht vor), wurde eine der beiden Schüler*innen aus der Grundgesamtheit entfernt, so dass jeder Haushalt die gleiche Wahrscheinlichkeit der Ziehung hatte und kein Haushalt mehrfach gezogen werden konnte.

2. Jede der vier Schulen wurde für die Zufallsauswahl der Haushalte zunächst als einzelner Cluster betrachtet, der wiederum in zwei Subgruppen unterteilt wurde:
 - a) Schüler*innen mit familiärem Migrationshintergrund im Haushalt (Subgruppe 1),
 - b) Schüler*innen ohne familiären Migrationshintergrund im Haushalt (Subgruppe 2).
3. Aus jeder der acht Gruppen (vier Schulcluster mit je zwei Subgruppen) wurden dann per einfacher Zufallsauswahl (Simple Random Sampling mit Zufallsstart und gleichen Intervallen)
 - a) 70 % der Schüler*innen in Subgruppe 1 (Haushalte mit familiärem Migrationshintergrund) sowie
 - b) 50 % der Schüler*innen in Subgruppe 2 (Haushalte ohne familiären Migrationshintergrund) gezogen.
4. In den Haushalten der betroffenen Schüler*innen wurde dann der erste verfügbare Elternteil bzw. Erziehungsberechtigte, der dauerhaft im Haushalt lebt, befragt (Elterninterview).
5. Analog wurde (sofern dauerhaft im Haushalt lebend) ein Geschwisterkind befragt (Geschwisterinterview), wobei bei mehreren Geschwisterkindern zur Auswahl die Methode des nächsten Geburtstags (next birthday method, d. h. befragt wurde das Geschwisterkind, das vom Tag der Befragung ausgesehen, als nächstes Geburtstag feiert) angewendet wurde.
6. War das ausgewählte Geschwisterkind
 - a) sechs Jahre oder jünger, wurde anstelle des Geschwisterkindes ein Auskunftsgespräch mit jenem Elternteil / Erziehungsberechtigten geführt, das auch im Elterninterview befragt wurde;
 - b) zwischen sechs und zwölf Jahren, so beantwortete das Geschwisterkind selbstständig den Fragebogen zum Geschwisterinterview; zusätzlich wurde ein Auskunftsgespräch mit jenem Elternteil / Erziehungsberechtigten geführt, welches/welcher auch im Elterninterview befragt wurde;
 - c) über 12 Jahre alt, so wurde nur ein Geschwisterinterview mit dem ausgewählten Geschwisterkind geführt.

Mit Geschwisterkindern über 16 Jahren wurde kein Geschwisterinterview geführt; die Daten der*des betroffenen Schüler*in wurde dann als „ohne zu befragendes Geschwisterkind“ gekennzeichnet.

5.2. Untersuchungspopulation und Ausschöpfungsquote dokumentieren

Im Rahmen des BMBF-geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) wird die Lesekompetenz von Grundschulkindern vor (Welle 1) und nach (Welle 2) den Sommerferien 2023 an insgesamt 34 Grundschulklassen im gesamten Bundesgebiet untersucht (Paneldesign). Die Auswahl der Grundschulen erfolgte mittels einer einfacher Zufallsstichprobe (Random Sampling). Ausgehend von einer Bruttostichprobe von 1000 Grundschüler*innen wurden in Welle 1 die Lesekompetenz von insgesamt 605 Schüler*innen (Nettostichprobe) mittels eines standardisierten Leistungstests erfasst. Aus dem Methodenbericht geht für die erste Erhebungswelle Folgendes hervor:

Geographischer Raum:	Deutschland (DE)
Grundgesamtheit:	alle Schüler*innen an Grundschulen Auswahlverfahren
Zufallsauswahl:	1. Auf Basis aller staatlichen Grundschulen in Deutschland wurden zufällig zehn Schulen ausgewählt; Grundschulen privater bzw. freier Träger wurden nicht berücksichtigt.
Erhebungsverfahren:	2. Auf Basis des Schüler*innen-Verzeichnisses der ausgewählten Grundschulen wurden zufällig 100 Schüler*innen an jeder Schule ausgewählt. standardisierter Fragebogen Erhebungsperiode
Erhebungsperiode:	Lesetests und soziodemographischer Fragebogen unmittelbar im Unterricht 2023-05-27 bis 2023-07-30
Ausschöpfungsquote	
Bruttostichprobe:	1000
Schüler*innen nicht angetroffen:	89
Nettostichprobe:	911
Teilnahmeverweigerungen:	234
unvollständige Beantwortung:	72
	(>30 % unbeantwortete Fragen)
Realisierte Stichprobe:	605
Ausschöpfungsquote:	60,5%

In Welle 2 konnten, ausgehend von einer Bruttostichprobe von 605 Grundschüler*innen, in Welle 1 445 Grundschüler*innen erneut getestet werden. Aus dem integrierten Methodenbericht für beide Wellen geht hervor:

Geographischer Raum:	Deutschland (DE)
Grundgesamtheit:	alle Schüler*innen an Grundschulen Auswahlverfahren (Zweistufige Zufallsauswahl)
Zufallsauswahl:	1. Auf Basis aller staatlichen Grundschulen in Deutschland wurden zufällig zehn Schulen ausgewählt; Grundschulen privater bzw. freier Träger wurden nicht berücksichtigt. 2. Auf Basis des Schüler*innen-Verzeichnisses der ausgewählten Grundschulen wurden zufällig 100 Schüler*innen an jeder Schule ausgewählt.
Erhebungsverfahren:	standardisierter Fragebogen Erhebungsperiode
Erhebungsperiode:	Lesetests unmittelbar im Anschluss an den Unterricht Welle 1: 2023-05-27 bis 2023-07-30 Welle 2: 2023-08-023 bis 2023-10-03

Ausschöpfungsquote beider Wellen

	Welle 1	Welle 2
Bruttostichprobe	1000	605
Schüler*innen nicht angetroffen:	89	71
Nettostichprobe	911	534
Teilnahmeverweigerungen	234	48
Unvollständige Beantwortung	72	41
Realisierte Stichprobe	605	445
Ausschöpfungsquote	60,5 %	74,2 %
Anteil Brutto-SP W2/W3 an Brutto-SP W1		60 %
Anteil Netto-SP W2/W3 an Netto-SP W1		58,1 %
Anteil realisierte SP W2/W3 an realisierter SP W1		73,6 %
Anteil realisierte SP W2/W3 an Brutto-SP W1		44,5 %
Anzahl der Befragten, die keine weitere Teilnahme wünschen (= Reduktion der Stichprobe in der Folgewelle um diese Anzahl)	5	22

5.3. Harmonisierung von Variablen dokumentieren

Im Rahmen des BMBF-geförderte Verbundprojekt *Lesen Lernen* (LesLer 2022) wird die Lesekompetenz von Grundschulkindern vor und nach den Sommerferien 2023 an insgesamt 34 zufällig ausgewählten Grundschulklassen im gesamten Bundesgebiet untersucht (Paneldesign mit zwei Wellen). Darüber hinaus wurden einige soziodemographische Merkmale der Grundschüler*innen, wie Geschlecht, Alter oder primäre

genutzte Sprache im Haushalt der Schüler*innen, erfasst. Die Daten des Projekts (LesLer 2023) wurden nach der Erhebung beider Wellen um bereits existierende Daten einer vergleichbaren Studie (FördLes 2019) erweitert (Trenddesign).

Zur Harmonisierung der LesLer-2023-Daten mit den FörderLes-2019-Daten müssen zunächst einige der soziodemographischen Angaben der Schüler*innen harmonisiert werden. Die Mitarbeitenden des Projekts legten dazu zunächst eine Korrespondenztabelle der Variablen in beiden Datensätzen an:

	LesLer-2023	FördLes-2019	Trendfile
...			
Geburtsjahr	V01	–	V01
Alter	V01_1	age	V01_1
Geschlecht	V02	sex	V02
Deutsch im HH	V03	–	V03
Deutsch im HH (abgeleitet)	V03_1	german	V03_1
...			

Daran anschließend wurden einzelne Variablen bzw. deren Ausprägungen harmonisiert. So fragt die LesLer- Studie die Kinder nach ihrem Geburtsjahr (V01) während in den FördLes-2019-Daten nur Informationen zum Alter (in Jahren) vorliegen. Folglich wird aus der Angabe zum Geburtsjahr in LesLer-2023-Daten zunächst das Alter in einer neuen Variable (V01_1) abgeleitet und entsprechend dokumentiert.

Bei den Angaben zur Verwendung der Sprache Deutsch in den Haushalten der Schüler*innen (V03) divergieren die Antwortoptionen zwischen den LesLer-2023- und den FördLes-2019-Daten. Die Daten der FördLes-Studie differenzieren lediglich zwischen „immer oder meistens Deutsch gesprochen“ und „manch- mal oder nie Deutsch gesprochen“. In den LesLer-2023-Daten wird hingegen unterschieden, zwischen „immer Deutsch gesprochen“, „meistens Deutsch gesprochen“, „manchmal Deutsch gesprochen“ und „nie Deutsch gesprochen“. Zur Harmonisierung der beiden Datensätze wird zunächst eine abgeleitete Variable (V03_1) in den LesLer-2023-Daten erstellt. Aus dem Codebuch geht hierzu folgende Übersicht zur Harmonisierung hervor:

	LesLer-2023	FördLes-2019
Name der Quellvariable	V03	V03_1
Name der Zielvariable	V03	V03_1
Ausprägungen	1 immer Deutsch gesprochen	1 immer oder meistens Deutsch gesprochen
	2 meistens Deutsch gesprochen	
	3 manchmal Deutsch gesprochen	2 manchmal oder nie Deutsch gesprochen
	4 nie Deutsch gesprochen	2 manchmal oder nie Deutsch gesprochen

5.4. Bildung von Gewichten dokumentieren

Ein Forschungsprojekt untersuchte die Mathematikleistungen von Grundschüler*innen in Berlin, im Vergleich zwischen privaten und öffentlichen Grundschulen. Aus einer Schulliste mit 33 Grundschulen wurden eine private sowie vier öffentliche Grundschulen gezogen, wobei eine der gezogenen öffentlichen Schulen die Teilnahme verweigerte. An den drei öffentlichen und der einen privaten Grundschule wurden anschließend jeweils 10 Schüler*innen zur Untersuchung ausgewählt.

Um potenzielle Verzerrungen der Schätzwerte infolge ungleicher Auswahlwahrscheinlichkeiten zu reduzieren und die für die Forschungsfrage relevanten Merkmale der Grundschulen (privat/öffentlich) an die entsprechende Verteilung in der Grundgesamtheit im Rahmen einer Poststratifizierung anzupassen, wurden nach der Datenerhebung Gewichte gebildet. Aus dem Methodenbericht des Forschungsprojekts:

Zur Reduktion potenzieller Verzerrungen wurden folgende Basisgewichte für die Schulen, Schüler*innen sowie Schul-Adjustierungsfaktoren und ein finales Gesamtgewicht für die Datenanalyse gebildet:

1. Das Schul-Basisgewicht gibt die Auswahlwahrscheinlichkeit einer Schule wieder. Es wurde als Quotient aus „Gesamtzahl der Schulen“ ($N = 33$) dividiert durch „Anzahl der gezogenen Schulen insgesamt“ ($n = 4$) gebildet.
2. Der Schul-Adjustierungsfaktor wurde pro Schicht, d. h. getrennt für private und öffentliche Schulen, gebildet. Er berechnete sich aus dem Quotienten „Anzahl der gezogenen Schulen“ dividiert durch „Anzahl der teilgenommenen Schulen“, also z. B. $3/4$ für die öffentlichen Schulen.
3. Schließlich wurde ein Basisgewicht für die Schüler*innen berechnet. Dies wurde aus dem Quotienten „Anzahl der Schüler*innen pro Schule, insgesamt“ dividiert durch „Anzahl der zu ziehenden Schüler*innen pro Schule“ gebildet. Das Basisgewicht für die private Schule mit insgesamt 98 Schüler*innen, von denen zehn Personen befragt werden sollten, beträgt z. B. $98/10$.
4. Abschließend wurden die drei Faktoren, Schul-Basisgewicht, Schul-Adjustierungsfaktor und Basisgewicht für Schüler*innen, multipliziert und so das finale Gesamtgewicht für die weitere Analyse gebildet.

5.5. Bildung von Skalen und Indizes dokumentieren

Im Rahmen des BMBF-geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) wird die Lesekompetenzen von Grundschulkindern vor und nach den Sommerferien 2023 (2 Erhebungswellen) an insgesamt 34 zufällig ausgewählten Grundschulklassen im gesamten Bundesgebiet untersucht. Dazu wurde ein bereits existierendes Messinstrument (LeseKompass 2017) genutzt und daraus eine Skala zur Lesekompetenz der Grundschüler*innen abgeleitet. Aus dem Codebuch geht zur Dokumentation der Skala und ihrer Items Folgendes hervor:

<i>Itemlabel</i>	
stmot1_MZP	Das Lesen von Geschichten macht mir Spaß.
stmot2_MZP	Eine meiner Lieblingsbeschäftigungen ist das Lesen.
stmot3_MZP	Zuhause lese ich gerne.
stmot5_MZP	Beim Lesen empfinde ich Freude.

stmot6_MZP Ich freue mich auf jede freie Zeit, in der ich lesen kann.						
Vierstufiges Antwortformat: trifft nicht zu (1) - trifft eher nicht zu (2) - trifft eher zu (3) - trifft genau zu (4)						
Itemkennwerte						
Itemlabel	MZP1			MZP2		
	M	SD	r_{it-i}	M	SD	r_{it-i}
stmot1_MZP	3,07	.80	.62	2,97	.71	.48
stmot2_MZP	2,66	.76	.49	2,68	.84	.46
stmot3_MZP	2,72	.80	.47	2,77	.82	.47
stmot5_MZP	2,97	.70	.42	2,86	.72	.46
stmot6_MZP	4,12	.80	.42	2,96	.81	.45
Skalenkennwerte						
	MZP1		MZP2			
	stmot_1s		stmot_2s			
M	2,92		2,72			
SD	0,64		0,65			
Mo	4,00		4,00			
Md	4,00		2,70			
a	.69		.69			
Skewness	-0,29		-0,29			
N	605		445			

5.6. Integration zusammengehöriger (Teil-)Datensätze dokumentieren

Im Rahmen des BMBF-geförderten Verbundprojekts *Lesen Lernen* (LesLer-2022) wird die Lesekompetenz von Grundschulkindern vor und nach den Sommerferien 2023 an insgesamt 34 zufällig ausgewählten Grundschulklassen im gesamten Bundesgebiet untersucht (Paneldesign mit zwei Wellen). Darüber hinaus werden in der ersten Welle einige soziodemographische Merkmale der Grundschüler*innen, wie Alter und Geschlecht der Schüler*innen, erfasst.

Die Daten der beiden Wellen und die soziodemographischen Angaben werden nach Aufbereitung der Daten in einem integrierten Datensatz zusammengefasst. Dazu verfügen alle drei Teildatensätze über eine ID-Variable mit einer eindeutigen Kennung für jede*n Schüler*in. Das Zusammenspielen der drei Teildatensätze erfolgt dann über eben diese ID-Variable. Der integrierte Datensatz wird unter einem neuen Namen gespeichert.

Anschließend wird der integrierte Datensatz (LesLer-2023) um einen bereits existierenden Datensatz (FördLes-2019) erweitert (Trendfile). Dazu werden zunächst die Ausprägungen der interessierenden Variablen harmonisiert und weitere abgeleitete Variablen gebildet. Schließlich werden die Variablen in beiden Datensätzen gleich gelabelt und die beiden zu integrierenden Datensätze im gleichen Dateiformat gespeichert. Das Zusammenspielen erfolgt dann über das Anhängen der FördLes-2019-Daten an die

LesLer-2023-Daten. Abschließend werden die fehlenden Werte in Variablen, die ausschließlich in den LesLer-2023-Daten vorliegen, für die Fälle der FördLes-2019-Daten rekodiert. Schließlich wird der integrierte Datensatz unter einem neuen Namen gespeichert.

5.7. Variablen, Fragen und Filterführung dokumentieren

Im Rahmen des BMBF-geförderten Verbundprojekts Lesen Lernen (LesLer 2022) wird die Lesekompetenz von Grundschulkindern vor und nach den Sommerferien 2023 an insgesamt 34 zufällig ausgewählten Grundschulklassen mit dem Förderschwerpunkt „Sprache“ im gesamten Bundesgebiet untersucht. Neben einem Test zur Lesekompetenz werden in der ersten Erhebungswelle zudem einige soziodemographische Merkmale der Schüler*innen erfasst, wie Alter und Geschlecht oder primäre Sprache im Haushalt der Kinder.

Der nachfolgende Auszug aus dem Codebuch des LesLer-Projekts gibt einen Einblick in die Dokumentation einzelner Variablen (V) und ihrer zugehörigen Fragen (F) sowie die mit den Fragen einhergehende Filterführung im originären Messinstrument (Fragebogen).

V01	<p>Geburtsjahr der Schülerin / des Schülers F01: Weißt du, wann du geboren wurdest? Interviewendenanweisung: Nur Jahr (JJJJ) erfassen</p> <p>Jahreszahl (numerischer Wert), z. B. 2012 2013 2014 2015</p> <p>9998 weiß ich nicht 9999 fehlende Angabe (missing)</p> <p>Anmerkungen:</p> <ul style="list-style-type: none"> • Ausgangsvariable für V01a (Alter der Schülerin / des Schülers)
V01_1	<p>Alter der Schülerin / des Schülers F01: aus V01 abgeleitet Variable</p> <p>Jahre (numerischer Wert), z. B. 8 9 10 11</p> <p>99 fehlende Angabe (missing) in V01</p> <p>Anmerkungen</p> <ul style="list-style-type: none"> • aus V01 (Geburtsjahr) abgeleitete Variable V01_1 = 2023 – V01

V02 Geschlecht der Schülerin / des Schülers
F02: Bist du ...

- | | |
|---|---------------------------|
| 1 | ein Mädchen (weiblich) |
| 2 | ein Junge (männlich) |
| 3 | weder noch |
| 7 | sag ich nicht |
| 8 | weiß ich nicht |
| 9 | fehlende Angabe (missing) |

V03 Primäre Sprache im Haushalt der Schülerin / des Schülers: Deutsch
F03: Bei euch zu Hause wird ...

- | | | |
|---|-----------------------------|-----------------|
| 1 | Deutsch gesprochen | weiter mit F04 |
| 2 | meistens Deutsch gesprochen | weiter mit F03b |
| | manchmal Deutsch gesprochen | weiter mit F03a |
| 4 | nie Deutsch gesprochen | weiter mit F03a |
| 8 | weiß ich nicht | |
| 9 | fehlende Angabe (missing) | |

V03a Primäre Sprache im Haushalt der Schülerin / des Schülers:
F03a: Welche Sprache spricht ihr hauptsächlich zu Hause?

- Freitext (re-codiert)
- | | |
|----|--------------------|
| 01 | Albanisch |
| 02 | Arabisch |
| 03 | Aserbaidtschanisch |
| 04 | Bosnisch |
| 05 | Estnisch |
| 06 | Finnisch |
| 07 | Französisch |
| 08 | Hebräisch |
| 09 | Italienisch |

WEITERE KATEGORIEN NICHT DARGESTELLT

- | | |
|----|---------------------------|
| 97 | sag ich nicht |
| 98 | weiß ich nicht |
| 99 | fehlende Angabe (missing) |

Anmerkungen:

- Freitextnennung der Schuler*innen

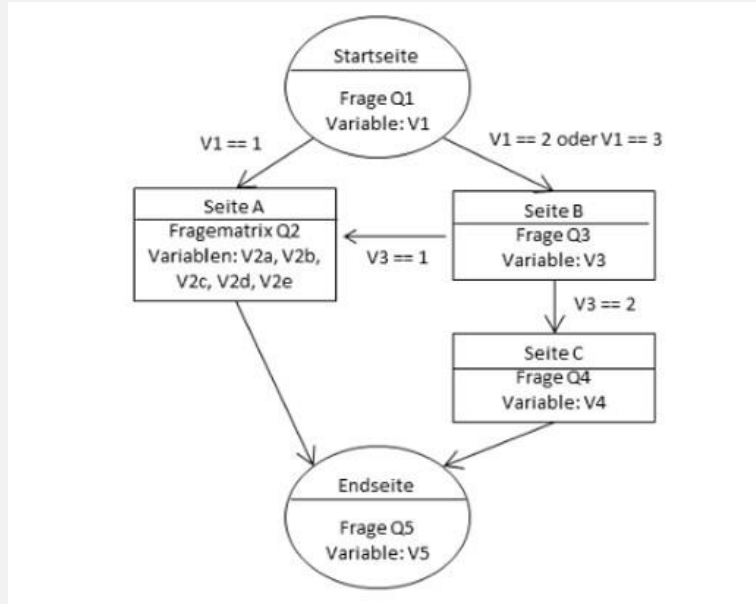
- Codierung nach ISO-639-1-Code
- codiert wurden nur tatsächlich genannte Sprachen

... WEITERES CODEBUCH NICHT DARGESTELLT ...

5.8. Dynamische Elemente in Online-Fragebögen dokumentieren

In einem Forschungsprojekt werden Daten von Studierenden mittels Web-Survey erhoben. Das digitale Erhebungsinstrument beinhaltet dabei zwei dynamische Elemente. Zum einen wird durch eine Filterführung (question flow) festgelegt, welchen Personengruppen jeweils welche Fragen zur Beantwortung angezeigt werden. Diese Personengruppen sind BAföG-Empfänger*innen (V1==1), Stipendiaten*innen (V1==2) und Personen, die keine Förderung erhalten (V1==3). Dabei wird per Randomisierung die Reihenfolge der Items der Fragematrix Q2 für jede befragte Person zufällig festgelegt. Dies dient der Verminderung von Reihenfolgeeffekten. Aus der Dokumentation des Projekts geht zur Filterführung Folgendes hervor:

Die Nutzung der Eingangsfilter sowie der daran anschließenden Filterführung folgt der nachstehenden Abbildung. Darin sind die einzelnen Seiten des Ergebnisinstrumentes, die Bezeichnung der Fragen und Variablen, die auf den einzelnen Seiten erhoben werden, die Übergänge einer Seite zu möglichen Folgeseiten sowie die jeweilige Bedingung, die für einen Übergang zur Folgeseite erfüllt sein muss, abgebildet.



Ein Eingangsfilter beschreibt dabei die Gruppe von Personen, der die jeweilige Frage angezeigt wird:

Frage	Eingangsfilter
Q1	–
Q2	$V1 == 1 \ \ ((V1 == 2 \ \ V1 == 3) \ \&\& \ V3 == 1)$
Q3	$V1 == 2 \ \ V1 == 3$
Q4	$(V1 == 2 \ \ V1 == 3) \ \&\& \ V3 == 2$
Q5	–
Q1	–

Die Randomisierung der Reihenfolge der Aussagen in einer Fragematrix basiert auf folgendem Schema:

	stimme voll und ganz zu	stimme eher zu	teils / teils	stimme eher nicht zu	stimme überhaupt nicht zu
Aussage A	()	()	()	()	()
Aussage B	()	()	()	()	()
Aussage C	()	()	()	()	()
Aussage D	()	()	()	()	()
Aussage E	()	()	()	()	()

Die Aussagen A und E der Fragematrix Q2 werden immer als erste bzw. als letzte Aussage der Matrix angezeigt. Die Aussagen B, C und D der Matrix werden für jede befragte Person in einer zufälligen

Reihenfolge angezeigt. In der Variablen Q2_item-order wird für jede befragte Person die jeweils realisierte Reihenfolge der Aussagen gespeichert.

5.9. Leitfadengestützte Interviews dokumentieren

Im Rahmen des BMBF-geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) wird die Lesekompetenz von Grundschulkindern vor und nach den Sommerferien 2023 (2 Erhebungswellen) an 34 Grundschulklassen im gesamten Bundesgebiet untersucht. Im Anschluss an eine erste Analyse der Daten nach der zweiten Erhebungswelle werden zusätzlich zehn Interviews mit Grundschullehrkräften geführt (LesInt 2023). Diese Gespräche werden per Video aufgezeichnet. Als Stimulus verwenden die Forschenden dabei ein dreiminütiges Video, in dem den Lehrkräften zwei gängige online-Lernplattformen zur Förderung der Lesekompetenz vorgestellt wurden. Nach Durchführung der ersten beiden Interviews (Pre-Test) nehmen die Projektbeteiligten einige Änderungen an ihrem Leitfaden vor.

Alle während der Interviews zum Einsatz kommenden Materialien, Erhebungsinstrumente, Stimuli und dabei gewonnene Dokumente wurden von den Forschenden dokumentiert:

Dokumentation des Leitfadens

Im Projekt wurde ein Leitfaden zur Erhebung von Interviews entwickelt, wie im Methodenbericht beschrieben. Dazu zählen auch Angaben und Resultate aus Pre-Tests und die sich daraus ergebenden Anpassungen des Leitfadens. Die Projektbeteiligten dokumentieren den Leitfaden vollständig und in leicht zugänglicher, maschinenlesbar Form und führen ihn in der Bestandsliste des Projekts auf. Die Dokumentation des Leitfadens umfasst u. a. die Leitfragen, Interviewer*innen-Anweisungen, Vorgaben zum geplanten und tatsächlichen Ablauf der Interviews sowie Informationen zu den Interviewern*innen, wie z. B. ihrer Disziplinzugehörigkeit, Vorkennnisse sowie Angaben zu Interviewer*innen-Schulungen.

Dokumentation des Stimulus

Die Projektbeteiligten setzen in den Interviews ein Video als Stimulus ein. Die Gründe zur Auswahl der im Video vorgestellten online-Lernplattformen zur Förderung der Lesekompetenz werden im Methodenbericht erörtert. Darin beschrieben sind sowohl die Inhalte des Videos als auch dessen konkrete Einsatzweise im Interview, z. B. wann das Video gezeigt wurde. Die Projektbeteiligten dokumentieren darüber hinaus den technischen Einsatz von Hardware und Software zur Vorführung des Videos, das in einer gesicherten Cloud gespeichert und den Lehrkräften per Tablet zu Beginn der Interviews mithilfe der Software Media Player Classic (MPC-HC) vorgespielt wurde. Mit Zustimmung der Urheberrechtsinhabenden wird das Video in einem nicht-proprietären Dateiformat gespeichert und im Methodenbericht adäquat, d. h. inklusive eines persistenten Identifikators, zitiert.

Verwendung eines Forschungstagebuchs (Feldnotizen)

Während der Durchführung der Interviews (Feldphase) erstellen die Projektbeteiligten ein prozessgeleitetes Forschungstagebuch in Form einer Textdatei. Darin werden Erfahrungen und Probleme in den einzelnen Interviews, z. B. zur Interviewsituation, den Interviewpartnern*innen, der Methode und dem Instrument, dokumentiert. Das Forschungstagebuch wird in pseudonymisierter Form Dritten zur Nachnutzung als Teil der Daten bereitgestellt.

Verwendung von Dokumentationsbögen in der Erhebungssituation

Für jedes Interview wurde ein Dokumentations- bzw. Kontextualisierungsbogen erstellt, der Hinweise zum Ablauf des Interviews enthält und durch die Interviewenden direkt im Anschluss an das Interview verfasst wurde, wie nachfolgenden beispielhaft dargestellt:

Interviewkennung	S05_int03_v1
Datum	21.06.2021
Ort	Frankfurt a. Main
Zeitraum / Dauer	9:45 – 11:15 Uhr / 90min
Interviewer*in / Institution	Peter Meier / Universität Frankfurt
Institution	Grundschule am See
interviewte Person	Martina Schulz
Alter der interviewten Person	44
Berufserfahrung (Jahre)	17
verwendete Stimuli	3-minütiges Video zu Online- Lernplattformen zur Förderung der Lesekompetenz
verwendete Hardware / Software	<ul style="list-style-type: none"> • Tablet / VLC-Player • digitales Aufnahmegerät, Typ X
Hinweise zum Ablauf des Interviews	<ul style="list-style-type: none"> • Beginn mit fünfzehn-minütiger Verspätung • zunächst technische Probleme aufgrund instabiler WLAN-Verbindung (Zugang wurde für Vorführung des Videos benötigt, Cloud) • Interview ab Minute 19 einmal für etwa zwei Minuten durch Kollegen der Befragten unterbrochen

5.10. Feldzugang dokumentieren

In einem Forschungsprojekt wurden Lehrkräfte an neun Schulen in Hessen und Rheinland-Pfalz untersucht. Zur Fallauswahl wurde dabei zunächst jede Schule als einzelner Fall betrachtet. Alle neun Schulen erklärten sich grundsätzlich bereit, an der Untersuchung teilzunehmen, verwiesen aber zunächst auf die zuständigen Schulministerien und notwendige Genehmigungen. Nach Einholen dieser behördlichen Genehmigungen kontaktierten die Projektbeteiligten zu interviewende Lehrkräfte auf unterschiedliche Weise. So wurde ein Teil der Lehrkräfte bereits durch die zuständigen Behörden informiert, während der Zugang zu anderen Lehrkräften über persönliche Kontakte der Projektbeteiligten zustande kam.

Zur Gewährleistung der Transparenz und zur späteren Nachvollziehbarkeit wurden die unterschiedlichen Zugangswege und -versuche (unabhängig von ihrem Erfolg) im Methodenbericht in pseudonymisierter Form dokumentiert:

Fall Name, Nachname	Institution	Rolle (Position des Kontaktes)	Kontaktaten (E-Mail oder Telefon-Nr.)	Feldzugangsnotizen (Eintragen, welche Schwierigkeiten oder Möglichkeiten sich für den Feldzugang durch die Person ergeben).
Fall Name, 01 Nachname 01	Gymnasium	Schulleitung	Schulleitung@Schule_I.de	Die Anfrage wurde an einer höheren Stelle weitergeleitet. Es wurde ein Telefongespräch mit der Schulleitung vereinbart. Die Schulleitung schlägt vor, nach der Untersuchung die Ergebnisse in der Schule bei der nächsten Fachbesprechung vorzustellen mit der Bitte aus den Untersuchungsergebnissen anwendungsorientierte Rückschlüsse auf die Unterrichtspraxis zu ziehen.
Fall Name, 02 Nachname 02	IGS	Fachlehrer*in Philosophie	Fachlehrer*in@Schule_II.de	Der Kontakt zur Schule II (Fall 02) hat sich durch eine zufällige Begegnung mit einem*r ehemaligen Kommilitonen*in ergeben, die die Anfrage an die Fachleiter*in für Philosophie weitergeben wird.
... WEITERE FÄLLE NICHT DARGESTELLT ...				

5.11. Verlaufsprotokoll einer Unterrichtsstunde dokumentieren

Im Rahmen einer Studie an weiterführenden Schulen in Nordrhein-Westfalen werden Unterrichtseinheiten beobachtet. Die Forschenden fertigen dazu Videoaufnahmen sowie Verlaufsprotokolle der teilnehmenden Beobachtungen an. Die Verlaufsprotokolle werden mit einer Kennung versehen, in die [Bestandsliste](#) aufgenommen sowie anschließend pseudonymisiert und ausgewertet. Die im Rahmen der Beobachtung gewonnenen Feldmaterialien werden ebenfalls mit Kennungen versehen, die eine Verknüpfung zu den entsprechenden Unterrichtsstunden ermöglichen.⁴

Datum:	17.05.2022		
Schulform:	Gymnasium Klasse: 8		
Unterrichtsstunde:	Fach: Mathematik		
Beginn der Stunde:	12:25 Uhr Ende der Stunde: 13:10 Uhr		
Protokollant/In:	Petra Meier Protokollkennung: S03_UB04_P_v2		
Zeit / Szenenwechsel	Didaktische Einheit / Lehrkraftaktivität	Sozialform	eingesetzte Medien
Begrüßung und Hausaufgabenkontrolle (15 Min)	Die Lehrkraft reaktiviert das Wissen der letzten Unterrichtsstunde	Klassengespräch	Tafelanschrieb
Erarbeitung (10 Min)	Reaktivierung des Wissens zum bisherigen Lösungsverfahren quadratischer Gleichungen	Einzelarbeit	Arbeitsbuch
... WEITERE BEOBACHTUNGEN DER STUNDE NICHT DARGESTELLT ...			

⁴ Die folgende Darstellung basiert auf Tiedtke, Michael (2008). *Erläuterungen zum Verlaufsprotokoll. SPS Leitfaden Verlaufsprotokoll*. <https://www.apaek.uni-frankfurt.de/58404529/verlaufsprotokoll.pdf>.

6. Datenteilen

6.1. Konzept zum Datenteilen

Im Rahmen eines durch das *Bundesministerium für Bildung und Forschung* (BMBF) geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) arbeiten zwei Lehrstühle zweier Hochschulen (A & B) mit einer außeruniversitären Forschungseinrichtung des Bundes (C) zusammen. Das Projekt begann im Mai 2022 und endet im April 2025. Der Förderbescheid des BMBF setzt für entsprechende Projekte das weitestgehende Teilen der Forschungsergebnisse, wie Daten und Begleitmaterialien, im Sinne des *Open Science* voraus.

Im Projekt wird die Lesekompetenz von Grundschulkindern in der dritten Klasse vor den Sommerferien 2023 getestet (Daten der 1. Welle: LesLer 2023-1). Dieser Test wird nach den Sommerferien 2023 bei den gleichen Grundschulkindern wiederholt (Daten der 2. Welle: LesLer 2023-2). Der Test wird an insgesamt 34 zufällig ausgewählten Grundschulklassen mit dem Förderschwerpunkt *Sprache* im gesamten Bundesgebiet durchgeführt.

Das entwickelte Messinstrument basiert auf einem urheberrechtlich geschützten Testinstrument Dritter (Lesekompass 2018). Für die Nachnutzung dieses Testinstruments im Rahmen des Forschungsvorhabens liegt die Genehmigung der Urhebenden vor. Die im Projekt zu erhebenden Daten werden durch vorhandene Daten einer vergleichbaren Studie ergänzt (FördLes 2019).

Im Anschluss an eine erste Analyse der Daten nach der 2. Welle sind zusätzlich zehn Interviews mit Grundschullehrkräften geplant (LesInt 2023). Diese Interviews werden per Video aufgezeichnet. Die Transkription der Videoaufzeichnungen erfolgt durch eine*n externe*n Dienstleistende*n (Auftragsverarbeitende*r D). Im Rahmen des Projekts sollen zudem zwei empirische Dissertationen entstehen.

Bereits in der Projektplanungsphase (Herbst 2021) wurde ein erster Entwurf zum Teilen der Daten und Begleitmaterialien entworfen und in Teilen in den Projektantrag integriert, wonach Folgendes vorgesehen ist:

Teilen der Daten mit Dritten im Sinne von Open Science

Dem Gedanken von Open Science folgend, werden die Daten und Begleitmaterialien des Projekts (LesLer 2023 & LesInt2023):

- a) inhaltlich so umfangreich wie möglich, d. h. mit dem größtmöglichen Analysepotential zur Nachnutzung, und
- b) so offen wie möglich, d. h. mit den geringstmöglichen Nutzungsbeschränkungen, mit Dritten geteilt.

Im Vorfeld des Datenteilens werden relevante Regelungen des Datenschutzes (z. B. die in der Einwilligungserklärung definierten Zwecke der Verarbeitung) und des Urheberrechtes (z. B. in Bezug auf die Übertragbarkeit von Verwertungsrechten) sowie forschungsethische Grundsätze geprüft.

1. Verantwortlichkeiten für das Datenteilen

Die LesLer-2023-Daten ebenso wie die LesInt-2023-Daten werden über das Forschungsdatenzentrum (FDZ) der außeruniversitären Forschungseinrichtung (C) bereitgestellt. Der Verbundpartner C (außeruniversitäre Forschungseinrichtung) ist verantwortlich für das Teilen der Daten und Begleitmaterialien im FDZ.

2. Teilen der Daten und Begleitmaterialien

Nach derzeitigem Planungsstand (Herbst 2021) werden folgende Daten und Begleitmaterialien zum Projektende mit Dritten geteilt:

1. LesLer-2023-Daten: Die Daten werden als kumulierter (beide Wellen) und pseudonymisierter Datensatz zusammen mit einem Methodenbericht und einem Codebuch archiviert und mit Dritten geteilt. Im weiteren Projektverlauf wird geprüft, inwiefern das Nachnutzungspotential der Daten durch die Pseudonymisierung (unverhältnismäßig) eingeschränkt wird und ob die Daten ggf. sowohl als Public als auch als Scientific Use File geteilt werden können. Das im LesLer-Projekt nachgenutzte Messinstrument LeseKompass 2023 darf aufgrund urheberrechtlicher Beschränkungen nicht im Rahmen der Datendokumentation geteilt werden und wird in der Datendokumentation (Codebuch) entsprechend zitiert.

2. LesInt 2023: Die pseudonymisierten Transkripte können je nach Grad (möglicher) Pseudonymisierung nur unter restriktiven Nutzungsbedingungen (Nachnutzung auf Antrag) Dritten zugänglich gemacht werden. Im weiteren Projektverlauf werden der Zugang zu den Transkripten und notwendige Einschränkungen ihrer Nachnutzung geprüft und mit dem FDZ abgestimmt.

2. Nachnutzungspotential und geeignetes Forschungsdatenzentrum

Die LesLer-2023- und LesInt-2023-Daten entsprechen dem Portfolio des FDZs der außeruniversitären Forschungseinrichtung (C). Sie bieten durch ihren zeitlichen Vergleich (vier Wellen, in Kombination mit FördLes-2019-Daten), die Abdeckung des gesamten Bundesgebiets und die ergänzenden Interviews (Transkripte) ein hohes Nachnutzungspotential für die Erforschung der Sprachentwicklung bei Grundschulkindern.

Das Teilen der Daten über das FDZ der außeruniversitären Forschungseinrichtung (C) ermöglicht die Zugänglichkeit und Nachnutzbarkeit der Daten. Durch den Erhalt eines persistenten Identifikators und die Registrierung der Daten in digitalen Datenkatalogen wird zudem die Auffindbarkeit der Daten und Begleitmaterialien für Dritte gesteigert.

Die unterschiedlichen Lizenzen des FDZs ermöglichen die Spezifikation unterschiedlicher Nachnutzungszwecke sowie Nutzungsbedingungen. Je nach Grad der Pseudonymisierung der Daten kann die Nachnutzung sowohl auf bestimmte Personengruppen als auch auf bestimmte Nachnutzungszwecke eingeschränkt werden. Die Notwendigkeit konkreter Einschränkungen für die Nachnutzung wird im weiteren Projektverlauf, getrennt für die LesLer-2023- und die LesInt-2023-Daten geprüft. Im Kontext der Fertigstellung der beiden Dissertationen auch nach Projektende (April 2025) wird ggf. ein zeitliches Embargo zur Nachnutzung der Daten durch Dritte im weiteren Projektverlauf geprüft. Das Embargo beträgt maximal ein Jahr und endet spätestens im April 2026.

3. Vorbereitung und Teilen der Daten und Begleitmaterialien

Die Übertragung einfacher Verwertungsrechte zum Datenteilen an das FDZ ist im Kooperationsvertrag geregelt. Im Kontext der Datenerhebung wird sichergestellt, dass alle notwendigen Genehmigungen (der Schulbehörden) und Einwilligungen (der betroffenen Personen) zum Teilen der Daten vorliegen.

Die Daten werden nach Absprache mit dem FDZ ca. sechs Monate vor Projektende (ab November 2024) an das FDZ übermittelt und archiviert. Mit Projektende (April 2025) werden die Daten mit Dritten zur Nachnutzung (unter entsprechenden Nutzungsbedingungen) geteilt. Die Projektmitarbeitenden der außeruniversitären Einrichtung (C) stimmen die Vorbereitung der Daten und Begleitmaterialien (wie Dateiformate, Dateibenennungen etc.) direkt mit den Datenkuratierenden des FDZs ab. Die Checklisten des Stamps zum Datenteilen finden daher im Folgenden keine weitere Anwendung.

7. Langfristsicherung

7.1. Konzept zur Langfristsicherung

Im Rahmen eines durch das *Bundesministerium für Bildung und Forschung* (BMBF) geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) arbeiten zwei Lehrstühle zweier Hochschulen (A & B) mit einer außeruniversitären Forschungseinrichtung des Bundes (C) zusammen. Das Projekt begann im Mai 2022 und endet im April 2025. Der Förderbescheid des BMBF setzt für entsprechende Projekte das weitestgehende Teilen der Forschungsergebnisse, wie Daten und Begleitmaterialien, im Sinne des *Open Science* voraus.

Im Projekt wird die Lesekompetenz von Grundschulkindern in der dritten Klasse vor den Sommerferien 2023 getestet (Daten der 1. Welle: LesLer 2023-1). Dieser Test wird nach den Sommerferien 2023 bei den gleichen Grundschulkindern wiederholt (Daten der 2. Welle: LesLer 2023-2). Der Test wird an insgesamt 34 zufällig ausgewählten Grundschulklassen mit dem Förderschwerpunkt *Sprache* im gesamten Bundesgebiet durchgeführt.

Das entwickelte Messinstrument basiert auf einem urheberrechtlich geschützten Testinstrument Dritter (Lesekompass 2018). Für die Nachnutzung dieses Testinstruments im Rahmen des Forschungsvorhabens liegt die Genehmigung der Urhebenden vor. Die im Projekt zu erhebenden Daten werden durch vorhandene Daten einer vergleichbaren Studie ergänzt (FördLes 2019).

Im Anschluss an eine erste Analyse der Daten nach der 2. Welle sind zusätzlich zehn Interviews mit Grundschullehrkräften geplant (LesInt 2023). Diese Interviews werden per Video aufgezeichnet. Die Transkription der Videoaufzeichnungen erfolgt durch eine*n externe*n Dienstleistende*n (Auftragsverarbeitende*r D). Im Rahmen des Projekts sollen zudem zwei empirische Dissertationen entstehen.

Bereits in der Projektplanungsphase (Herbst 2021) wurde ein erster Konzeptentwurf zur internen Langfristsicherung von Daten und Begleitmaterialien über einen Zeitraum von 15 Jahren nach Projektende (April 2025) erstellt:

Langfristig zu sichernde Daten und Begleitmaterialien

Alle Daten und Begleitmaterialien werden für 15 Jahre über das Projektende hinaus, d. h. bis April 2040, langfristig, intern gesichert, wenn

- a) diese nicht mit Dritte zur Nachnutzung geteilt wurden,
- b) diese auf Basis rechtlicher Vorgaben, Genehmigungen oder Vereinbarungen gesichert werden dürfen,
- c) auf diesen publizierte Forschungsergebnisse des Projekts beruhen (Replikation & Verifikation der Ergebnisse).

Dies betrifft zum aktuellen Planungszeitpunkt:

1. das Messinstrument LeseKompass 2018,
2. die teil-pseudonymisierten Rohdaten der Leistungstests (LesLer 2023), inklusive der Aufbereitungssyntax (für die teil-pseudonymisierten Rohdaten und die geteilten Forschungsdaten), und
3. die teil-pseudonymisierten Videoaufnahmen der Interviews (LesInt 2023) sowie die zugehörigen teil-pseudonymisierten Transkripte, die Transkriptionsregeln und die Pseudonymisierungsprotokolle (für die teil-pseudonymisierten Videoaufnahmen sowie für die geteilten Transkripte).

Die Notwendigkeit, die Rohdaten der beiden Wellen (LesLer 2023-1 & 2023-2) sowie die nicht-pseudonymisierten Videoaufnahmen und die zugehörigen nicht-pseudonymisierten Transkripte ebenfalls langfristig zu sichern, wird im weiteren Projektverlauf geprüft.

Verknüpfbarkeit langfristig gesicherter und geteilter Daten

Eine Verknüpfung der langfristig gesicherten Daten mit den geteilten Daten wird über eindeutige Identifikatoren auch über das Projektende hinaus gewährleistet:

1. bei zusätzlich langfristig gesicherten LesLer-2023-Daten und geteilten LesLer-2023-Daten über die ID der Befragten und
2. bei zusätzlich langfristig gesicherten LesInt-2023-Daten (Transkripte und Videodateien) und geteilten LesInt-2023-Daten (Transkripte) über IDs der Interviews.

Verantwortlichkeiten und Aufwendungen für die Langfristsicherung

Verantwortlich für die Langfristsicherung ist die außeruniversitäre Forschungseinrichtung (C). Diese trägt auch etwaige Kosten (Aufwendungen) im Kontext der Langfristsicherung und gewährleistet die Vernichtung der Dateien zum Ende der Langfristsicherung (April 2040).

Durchführung der Langfristsicherung

Spätestens drei Monate vor Projektende (Januar 2025) beginnen die Mitarbeitenden der außeruniversitären Forschungseinrichtung (C) mit der Vorbereitung der zu sichernden Daten und Begleitmaterialien entsprechend den Vorgaben des Stamps.

Die Langfristsicherung folgt den institutionellen Verfahren der außeruniversitären Einrichtung (C). Die Systeme zur Langfristsicherung der Einrichtung gewährleisten u. a.:

- den physischen Erhalt der Daten und Begleitmaterialien bis April 2040,
- die kontinuierliche, rechtskonforme Zugangskontrolle zu den Daten,
- die Sicherung der langfristig gesicherten Dateien gegen versehentlichen Verlust, Zerstörung etc.

Die langfristige Sicherung sensibler und personenbezogener Daten erfolgt:

1. teil-pseudonymisiert, d. h. in den langfristig gesicherten Daten der Teilstudie
 - a) LesLer 2023 werden ausschließlich jene Variablen erhalten, auf denen publizierte Forschungsergebnisse beruhen, alle weiteren Variablen werden entfernt,

- b) LesInt 2023 werden alle Informationen aus den Videoaufnahmen und Transkripten entfernt, die ein Risiko der Re-Identifizierung und potenzielle negative Konsequenzen für die betroffenen Personen mit sich bringen, erhalten werden nur solche Risiko-Passagen, die unmittelbar in wissenschaftliche Publikationen einfließen;
2. verschlüsselt, wobei die Verantwortung für den Schlüssel über die gesamte Sicherungsdauer hinweg bei der Projektleitung der außeruniversitären Forschungseinrichtung (C) bzw. (bei Personalwechseln) bei der zugehörigen Leitung der Forschungseinrichtung (C) liegt;
3. zugriffsgeschützt, durch das IT-Sicherheitssystem der Einrichtung (C).

Zugriff auf die Daten und Begleitmaterialien in der Langfristsicherung

Der Zugriff der Projektbeteiligten auf die Daten und Begleitmaterialien in der Langfristsicherung ist, wie folgt, im Kooperationsvertrag geregelt:

Die Daten und Begleitmaterialien in der Langfristsicherung können von allen Projektmitarbeitenden über die gesamte Sicherungsdauer bis April 2040 zur Verifikation publizierter Ergebnisse des Projekts genutzt werden. Eine solche Nutzung ist den übrigen Projektmitarbeitenden bzw. den beteiligten Lehrstühlen der beiden Hochschulen (A & B) und der Leitungen der außeruniversitären Forschungseinrichtungen (C) schriftlich (per E-Mail) anzuzeigen.

Zur Nachnutzung der langfristig gesicherten Daten und Begleitmaterialien durch die originären Projektmitarbeitenden, z. B. in weiteren Forschungsprojekten oder zur Lehre, bedarf es eines positiven Votums aller Urheberrechtsinhabenden. Dieses Votum kann, z. B. aufgrund von Personalwechseln, durch ein Votum der Lehrstuhlinhabenden (A & B) sowie der Leitungen der außeruniversitären Forschungseinrichtung (C) ersetzt werden. Den Lehrstuhlinhabenden und Einrichtungsleitungen werden hierzu die notwendigen Verwertungsrechte übertragen.

Der Zugriff auf die Daten und Begleitmaterialien ist schriftlich (per Email) bei der Leitung der außeruniversitären Forschungseinrichtung (C) zu beantragen. Der Nachweis zur Nutzung zu Replikationszwecken bzw. das Vorliegen von weiteren Nutzungsrechten (positives Votum zur Nachnutzung) obliegt den Antragsstellenden. Der Zugriff auf die Daten erfolgt, nach Freigabe durch die Leitung, durch das Rechenzentrum der außeruniversitären Forschungseinrichtung (C).

8. Bestandsliste

8.1. Bestandsliste

Im Rahmen eines durch das *Bundesministerium für Bildung und Forschung* (BMBF) geförderten Verbundprojekts *Lesen Lernen* (LesLer 2022) arbeiten zwei Lehrstühle zweier Hochschulen (A & B) mit einer außeruniversitären Forschungseinrichtung des Bundes (C) zusammen. Das Projekt begann im Mai 2022 und endet im April 2025. Der Förderbescheid des BMBF setzt für entsprechende Projekte das weitestgehende Teilen der Forschungsergebnisse, wie Daten und Begleitmaterialien, im Sinne des *Open Science* voraus.

Im Projekt wird die Lesekompetenz von Grundschulkindern in der dritten Klasse vor den Sommerferien 2023 getestet (Daten der 1. Welle: LesLer 2023-1). Dieser Test wird nach den Sommerferien 2023 bei den gleichen Grundschulkindern wiederholt (Daten der 2. Welle: LesLer 2023-2). Der Test wird an insgesamt 34 zufällig ausgewählten Grundschulklassen mit dem Förderschwerpunkt *Sprache* im gesamten Bundesgebiet durchgeführt.

Das entwickelte Messinstrument basiert auf einem urheberrechtlich geschützten Testinstrument Dritter (Lesekompass 2018). Für die Nachnutzung dieses Testinstruments im Rahmen des Forschungsvorhabens liegt die Genehmigung der Urhebenden vor. Die im Projekt zu erhebenden Daten werden durch vorhandene Daten einer vergleichbaren Studie ergänzt (FördLes 2019).

Im Anschluss an eine erste Analyse der Daten nach der 2. Welle sind zusätzlich zehn Interviews mit Grundschullehrkräften geplant (LesInt 2023). Diese Interviews werden per Video aufgezeichnet. Die Transkription der Videoaufzeichnungen erfolgt durch eine*n externe*n Dienstleistende*n (Auftragsverarbeitende*r D). Im Rahmen des Projekts sollen zudem zwei empirische Dissertationen entstehen.

Im Projektverlauf wird eine [Bestandsliste](#) aller Daten und Begleitmaterialien angelegt und über den gesamten Projektverlauf aktualisiert. Aus der Bestandsliste (Stand: Juni 2022) geht u. a. Folgendes hervor:

Auszug aus der Bestandsliste (Stand: Juni 2022)

Name der Datei(-gruppe), inkl. Version	Speicherformat der Datei(-gruppe)	verwendete Software	Speicher- volumen	Beschreibung der Datei(-gruppe)	Zusammen- gehörigkeit von Dateien	datenschutzrechtliche / sensible Inhalte	Status der Datei(-gruppe) nach Projektende	Herkunft der Datei(-gruppe)
20220304_LesLer-2023-1_ R-Daten_v00.Rdata	.RData	R (4.0.2)	~10MB	Roh-Leistungsdaten & Soziodemographie, Welle 1 (je 50-80 Schüler*innen an 34 Schulen) [TEMPLATE]	1.a	ja, nicht-pseudonymisiert, inkl. Schulbezug & Kontaktdaten	Langfristsicherung???	eigene Datei
20220304_LesLer-2023-2_ R-Daten_v00.Rdata	.RData	R (4.0.2)	~8 MB	Roh-Leistungsdaten, Welle 2 (je 50-80 Schüler*innen an 34 Schulen) [TEMPLATE]	1.b	ja, nicht-pseudonymisiert, inkl. Schulbezug & Kontaktdaten	Langfristsicherung???	eigene Datei
202302XX_LesLer-2023-1_ F-Daten_v00.Rdata	.RData	R (4.0.2)	~10 MB	aufbereitete Leistungsdaten & Soziodemographie, Welle 1 [GEPLANT]	1.a	ja, ggf. schwach pseudonymisiert (Schulbezug)	Vernichtung	eigene Datei
202302XX_LesLer-2023-2_ F-Daten_v00.Rdata	.RData	R (4.0.2)	~8 MB	aufbereitete Leistungsdaten, Welle 2 [GEPLANT]	1.b	ja, ggf. schwach pseudonymisiert (Schulbezug)	Vernichtung	eigene Datei
202312XX_LesLer-2023_ F-Daten_v00.Rdata	.RData	R (4.0.2)	~20MB	integrierter Datensatz, Wellen 1 & 2 [GEPLANT]	1	ja, ggf. schwach pseudonymisiert (Schulbezug)	archivieren & teilen	eigene Datei
2022XXXX_LesLer-2023_ Methodenbericht_v00.rtf	.rtf	Microsoft Word (14.0)	~350 KB	integrierter Methodenbericht der Wellen 1 & 2 zu den aufbereiteten LesLer-2023-Daten	1	nein	archivieren & teilen	eigene Datei
2022XXXX_LesLer-2023_ Codebuch_v00.rtf	.rtf	Microsoft Word (14.0)	~150 KB	integriertes Codebuch der Wellen 1 & 2 zu den LesLer-2023-Daten (inkl. Schulmerkmalen)	1	nein	archivieren & teilen	eigene Datei
20200713_FördLes-2019_ F-Daten_v3.1.Rdata	.RData	R (4.0.2)	16,3MB	Leistungsdaten & Soziodemographie zu 2581 Schüler*innen aus 41 Grundschulen zu zwei Messzeitpunkten	2	ja, schwach pseudonymisiert mit Schulbezug	Vernichtung	doi:10.782671/ST_78537-1, Zitation: Literaturverwaltung
20200713_FördLes-2019_Doku_v2.7.pdf	.pdf	Adobe Acrobat (XII)	647 KB	Dokumentation (Codebuch & Methodenbericht) zu FördLes-2019-Daten	2	nein	Vernichtung	doi:10.782671/ ST_78537-2 Zitation: Literaturverwaltung
20220207_Lesekompass-2018_v1.0.rft	.rtf	Microsoft Word (14.0)	276 KB	Messinstrument zu Leistungstests	1 / 2	nein	Langfristsicherung	Zitation: Literaturverwaltung
... WEITERE DATEIEN IN AKTUELLEM FALLBEISPIEL AUSGELASSEN ...								
20220412_LesInt_Leitfaden_v0.2.rft	.rtf	Microsoft Word (14.0)	~300KB	Leitfaden für alle Interviews mit Lehrkräften	3	nein	archivieren & teilen	eigene Datei
202212XX_LesInt_Dok-Int-1_v0.0.rft	.rtf	Microsoft Word (14.0)	~250KB	Dokumentationsbogen zum Interview mit Lehrkraft 01	3a	ja, Angaben zu interviewter und interviewender Person, Schulbezug	Langfristsicherung	eigene Datei
202212XX_LesInt_pDok-Int-1_v0.0.rft	.rtf	Microsoft Word (14.0)	~150KB	pseudonymisierter Dokumentationsbogen zum Interview mit Lehrkraft 01	3a	ja, ggf. schwach pseudonymisiert (Schulbezug)	archivieren & teilen	eigene Datei
202212XX_LesInt_Vid-Int-1_v0.0.rft	.mp4	MPEG-4	~500MB	Videoaufzeichnung des Interviews mit Lehrkraft 01	3a	Ja, nicht-pseudonymisierte Videoaufnahme	Langfristsicherung???	eigene Datei

202212XX_LesInt_Trans-Int-1_v0.0.rft	Microsoft Word (14.0)	~800KB	wörtliches Transkript zur Videoaufzeichnung des Interviews mit Lehrkraft 01	3a	Ja, ggf. Nennung von Klarnamen etc.	Langfristsicherung	eigene Datei
202212XX_LesInt_pTrans-Int-1_v0.0.rft	Microsoft Word (14.0)	~500KB	pseudonymisiertes Transkript zur Videoaufzeichnung des Interviews mit Lehrkraft 01	3a	ja, ggf. schwach pseudonymisiert (Schulbezug)	archivieren & teilen	eigene Datei
202212XX_LesInt_Dok-Int-2_v0.0.rft	Microsoft Word (14.0)	~250KB	Dokumentationsbogen zum Interview mit Lehrkraft 02	3b	ja	Langfristsicherung	eigene Datei
... WEITERE DATEIEN IN AKTUELLEM FALLBEISPIEL NICHT DARGESTELLT ...							