



Institut zur Qualitätsentwicklung
im Bildungswesen

Forschungsdatenzentrum

WISSENSCHAFTLICHE EINRICHTUNG DER LÄNDER
AN DER HUMBOLDT-UNIVERSITÄT ZU BERLIN E.V.

Dokumentation der Eingangskontrolle / Ingest Checklist

Stand: 20.10.2023

Bibliographische Informationen / Bitte zitieren als:

Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen (FDZ am IQB) (2023). *Dokumentation der Eingangskontrolle / Ingest Checklist*. Berlin: IQB - Institut zur Qualitätsentwicklung im Bildungswesen. Verfügbar unter: https://www.iqb.hu-berlin.de/fdz/Datenuebergabe/Template_Ingest_.pdf

Das Werk steht unter der Creative-Commons-Lizenz CC0 1.0 Universell (CC0 1.0) Public Domain Dedication (<https://creativecommons.org/publicdomain/zero/1.0/deed.de>).



Themen	Antwortoptionen	
	nein	ja
FDM-Beratung		
Eine Beratung hat VOR der Datenübergabe stattgefunden.		
Beratungsprotokoll liegt vor.		
Ressourcen zur Aufbereitung der Daten/Dokumentationsmaterialien sind im Projekt vorhanden.		
rechtliche Aspekte		
Hier wird geprüft, ob informierte Einwilligungserklärungen, Ethikvoten, Genehmigungsschreiben vorliegen und ob diese eine Weitergabe an Dritte sowie eine Bereitstellung der Daten zur Nachnutzung erlauben. Nur wenn diese Prüfung (durch den VerbundFDB) positiv ausfällt, können die Daten bereitgestellt werden.		
Es liegen KEINE datenschutzrechtlichen Hindernisse vor, die einer Archivierung oder Datenweitergabe im Wege stehen.		
Es liegen KEINE urheberrechtlichen Hindernisse vor, die einer Archivierung oder Datenweitergabe im Wege stehen.		
Es liegen KEINE vertraglichen Hindernisse vor, die einer Archivierung oder Datenweitergabe im Wege stehen.		
inhaltlicher Scope		
Nur wenn der Scope dem des FDZ am IQB entspricht, werden die Daten aufgenommen.		
Es sind Kompetenzdaten enthalten.		
Es sind mehrere Kompetenzbereiche (domänenübergreifend oder domänenspezifisch) untersucht worden.		
untersuchter Bildungsbereich	<input type="checkbox"/> Berufsbildung <input type="checkbox"/> Hochschulbildung <input type="checkbox"/> sonstiges	<input type="checkbox"/> Vorschule/Kindergarten <input type="checkbox"/> Primarbereich <input type="checkbox"/> Sekundarbereich
Nachnutzungspotenzial		
Ist die Datenerhebung schwierig zu reproduzieren?		
Werden schwierig erreichbare Populationen befragt?		
War die Datenerhebung besonders aufwendig und kostenintensiv für Forschende oder die Befragten?		
Werden besonders sensible Informationen erhoben?		
Ist der Datenbestand so umfassend, dass die Analysepotentiale noch nicht ausgeschöpft sind?		
Untersucht die Studie KEINE sehr spezifische Fragestellung (z.B. Evaluation eines Trainings)?		
Untersucht die Studie KEIN einmaliges oder seltenes historisches Phänomen (z.B. Corona-Pandemie)?		
Anzahl von Variablen (Summe aus allen Datensätzen)	<input type="checkbox"/> < 100 <input type="checkbox"/> 100 - 500	<input type="checkbox"/> 501 - 1.000 <input type="checkbox"/> 1.001 - 5.000 <input type="checkbox"/> > 5.000
Stichprobengröße	<input type="checkbox"/> < 100 <input type="checkbox"/> 100 - 500	<input type="checkbox"/> 500 - 1.000 <input type="checkbox"/> 1.000 - 5.000 <input type="checkbox"/> 5.000 - 10.000 <input type="checkbox"/> > 10.000
Design	<input type="checkbox"/> Querschnitt	<input type="checkbox"/> Trend/Wiederholter Querschnitt <input type="checkbox"/> experimentelles Design <input type="checkbox"/> quasi-experimentelles Design <input type="checkbox"/> Längsschnitt
geographische Verteilung der Erhebung	<input type="checkbox"/> 1 Bundesland	<input type="checkbox"/> mehrere Bundesländer <input type="checkbox"/> gesamtdeutsch <input type="checkbox"/> international
Erhebungszeitraum	<input type="checkbox"/> > t - 5 Jahre	<input type="checkbox"/> < t - 5 Jahre
Datenqualität		
Es sind alle Datensätze übergeben worden (z. B. Schüler-, Eltern- und Lehrkräfteangaben), die benötigt werden, um zentrale Publikationen zu reproduzieren.		
Es sind alle Fälle vorhanden, die benötigt werden, um zentrale Publikationen zu reproduzieren.		
Die Namen der Teildatensätze sind eindeutig.		
Variablen- und Wertebenenennung		
Variablen und Werte sind nach einem konsistenten Schema verständlich und eindeutig bezeichnet.		
Die Namen sind möglichst kurz gewählt.		
Sonderzeichen, Umlaute oder Leerzeichen wurden nicht verwendet.		
Variablen- und Wertelabels		
Variablenlabels wurden für alle Variablen verwendet.		
Wertelabels wurden für alle Variablen verwendet.		
Variablen- und Wertelabels sind eine kurze und möglichst aussagekräftige Beschreibung der Variableninhalte bzw. -ausprägungen.		
Variablen- und Wertelabels enthalten KEINE Sonderzeichen, Umlaute und Leerzeichen.		
Es sind KEINE Wild Codes vorhanden.		
Variablenwerte		
Allen Ausprägungen numerischer Variablen sind jeweils einem eindeutigen numerischen Wert zugewiesen.		
Die zugewiesenen numerischen Werte folgen, soweit möglich, einem einheitlichen Schema.		
Offene Antworten/Freitext sind codiert, d. h. mit einem numerischen Wert versehen worden.		
Offene Antworten/Freitext sind auf datenschutzrechtliche Probleme untersucht worden.		

Fehlende Werte		
Fehlende Werte sind definiert.		
Fehlende Werte haben Wertelabel.		
Existieren verschiedene Arten von fehlenden Werten, sind diese konsistent kodiert und gelabelt.		
Es gibt keine fehlenden Werte auf ID-Variablen.		
Qualität der Dokumentationsmaterialien Ohne Dokumentationsmaterial ist eine Bereitstellung NICHT möglich!		
Es gibt eine Website mit Informationen zur Studie/zum Teilprojekt.		
Es ist eine Projektbeschreibung/Abstract vorhanden.		
Es sind die Fragebögen enthalten.		
Es ist ein Technical Report/Methodenbericht enthalten.		
Es ist ein Abschlussbericht/eine zentrale Publikation enthalten.		
Es liegt eine Publikationsliste des Projekts vor/kann erstellt werden.		
Alle Variablen im Datensatz/den Datensätzen sind im Skalenhandbuch aufgeführt.		
Die Bildung von Skalen, Indizes, Hilfsvariablen, Gewichten, Matching/Linking/Tracking-Variablen etc. ist im Code-/Skalenhandbuch, tech. Report beschrieben.		
Angaben zum Untersuchungsdesign und der Feldphase sind in einem Methodenhandbuch/Feldbericht festgehalten.		
Es gibt Variablen zur Verlinkung von Teildatensätzen.		
Datenschutz Möglichkeiten des Umgangs mit sensiblen Variablen: (1) Variablenunterdrückung: Variablen werden geleert (und beinhalten keine Werte) (2) Variablenrekodierung: - Lokale Suppression: Personen, die seltene Merkmalsausprägungen aufweisen, erhalten auf der betroffenen Variable fehlende Werte. - Vergrößerung: Zusammenfassung seltener Kategorien zu größeren Merkmalsgruppen, dies kann u. a. wie folgt geschehen: - für betroffene Variablen werden Kategorien mit Fallzahlen von $n \leq 5$ zu einer Kategorie „Sonstiges“ zusammengefasst - diskrete Merkmalsausprägungen, die selten vorkommen, werden mit ähnlichen Ausprägungen zusammengefasst und bilden gemeinsam eine neue Kategorie - Bildung von Grenzwerten: Ausprägungen auf einem Merkmal werden durch eine Ober- und Untergrenze abgeschnitten - Skalenbildung: Einzelne Items werden zu Skalen zusammengefasst bzw. wenn Skalen vorliegen, werden die Einzelitems aus den Daten entfernt		
Eigennamen von Personen wurden gelöscht und durch nicht sprechende Identifikatoren ersetzt.		
Weitere personenbezogene Daten (u. a. Eigennamen von Orten und Organisationen) liegen nicht vor oder wurden bereits anonymisiert (formale Anonymisierung).		
Regionale, berufliche und ähnliche Angaben sind nicht so kleinteilig, dass die Informationen die Identifizierung der Teilnehmer/-innen ermöglichen (andernfalls sind Hinweise an das Archiv erfolgt).		
Es liegen KEINE öffentlich zugänglichen (im Internet) detaillierten Informationen zum Projekt (inkl. Regionaler Angaben) vor.		
Die Untersuchungseinheiten stammen NICHT aus einer sehr speziellen Population.		
Es gibt KEINE/WENIG detaillierte Angaben auf Level 2 (z. B. Lehrkräfte, Einrichtungen).		
Es gibt KEINE Sperrfristen.		
Es liegen KEINE Ergebnisse aus Untersuchungen mit urheberrechtlich geschützten Tests vor.		
Es liegen KEINE urheberrechtlich geschützte Dokumente (z. B. in Verlagen publizierte Skalenhandbücher, Berichte) bei.		
technische Aspekte technisch-formale Prüfung		
Auflistung der Dateien liegt vor		
Virenscan erfolgreich durchgeführt		
Dateien auf Funktionsfähigkeit überprüft		
Dateinamen		
Dateinamen enthalten keine Sonderzeichen, Umlaute oder Leerzeichen.		
Dateinamen sind möglichst kurz gewählt.		
Dateien sind so bezeichnet, dass aus dem Namen Rückschlüsse auf Studie, Daten-/Materialtyp und ggf. Versionsnummer gezogen werden können.		
Datenformat		
Die Dateien liegen in einem empfohlenen Format vor.	<input type="checkbox"/> Stata <input type="checkbox"/> Textdatei <input type="checkbox"/> .xls	<input type="checkbox"/> SPSS <input type="checkbox"/> R <input type="checkbox"/> .csv
Es gibt Daten in anderen Formaten.		
Hinweise		