

Versionierungskonzept

Stand: 10/2025

Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen (FDZ am IQB)
(2025). *Versionierungskonzept*. Berlin: IQB - Institut zur Qualitätsentwicklung im Bildungswesen.
http://doi.org/10.5159/IQB_Versionierungskonzept_v1

Das Werk steht unter der Creative-Commons-Lizenz Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International (<https://creativecommons.org/licenses/by-sa/4.0/legalcode.de>). Ausgenommen von der oben genannten Lizenz sind Teile, Abbildungen und sonstiges Drittmaterial, wenn anders gekennzeichnet.



Inhalt

1. Einführung	4
2. Neue externe Version eines Datenproduktes	5
2.1 Kriterien, wann eine neue DOI-Version erforderlich wird	5
2.2 Benennungskonvention	5
2.3 Ablage	6
3. Neue interne Version	6
3.1 Kriterien, wann eine neue interne Version erstellt wird	6
3.2 Benennungskonvention	7
3.3 Ablage	7
4. In Bezug auf die Langzeitarchivierung	7

1. Einführung

Der Zugang zu Daten einer Studie kann aufgrund der unterschiedlichen Granularität in der Sensibilität der Daten unterschiedlich sein. Folgende Zugangswege gibt es am FDZ am IQB:

- ♦ SUF via Off-site (Datenprodukt ohne sensible Daten)
- ♦ SUF via Remote (Datenprodukt mit sensiblen Daten)
- ♦ Campus Use File (CUF) via Download

Als Ergebnis der Kuratierung der Daten (= Aufbereitung) entstehen pro Studie mindestens ein, maximal drei Datenprodukte. Das FDZ am IQB vergibt auf Ebene des Datenproduktes einen persistenten Identifikator (= DOI). D. h.: Es gibt eine DOI pro Datenprodukt.

Die entstehenden Datenprodukte sind auch abhängig vom Kuratierungslevel:

- ♦ Beim Kuratierungslevel *Standard Plus* ist das einzige Datenprodukt ein SUF Off-site.
- ♦ Beim Kuratierungslevel *Fokus* kann das Datenprodukt SUF Remote hinzukommen.
- ♦ Beim Kuratierungslevel *Fokus Plus* kommt immer das Datenprodukt SUF Remote hinzu.
- ♦ Das Datenprodukt CUF kann bei jedem Kuratierungslevel entstehen.

Zusammenfassend bedeutet das, dass ein Datenprodukt das Ergebnis eines Kuratierungsprozesses ist. Ein Datenprodukt besteht aus mindestens einem Datensatz, kann aber auch mehrere Datensätze beinhalten. Es hat einen persistenten Indikator und wird als Ganzes an Nutzende weitergegeben. Die Datenprodukte einer Studie unterscheiden sich hinsichtlich der Sensibilität/Granularität der Daten und damit einhergehend in den Zugangsrestriktionen. Mehrere Datenprodukte (in unserem Fall immer auf die gleiche Studie bezogen) ergeben ein Datenpaket.

Das FDZ am IQB unterscheidet zwischen zwei Formen der Veränderung nach dem Ingest:

- ♦ inhaltliche Änderung an einem bereits bestehenden Datenprodukt: Daten (Werte von Variablen) in einem oder mehreren Datensätzen des Datenprodukts werden substantiell verändert bzw. dem Datenprodukt werden Daten in einem oder mehreren Datensätzen hinzugefügt → neue externe Version des Datenprodukts → neue DOI
- ♦ administrative Änderung an einem bereits bestehenden Datenprodukt: Änderung von Metadaten (inkl. Metadaten in Datensätzen, bspw. Wertebeschreibungen), beschreibenden Dokumenten oder ergänzenden Dateien → interne Version des Datenprodukts → keine neue DOI

Wir gehen am FDZ am IQB datensatzbezogen vor: Änderungen betreffen immer nur den einzelnen Datensatz und nicht die anderen Datensätze des Datenproduktes. Nur die Versionsangabe der **bearbeiteten** Datensätze ändert sich. D. h.: Es kann vorkommen, dass bei einem Datenprodukt mit mehreren Datensätzen nicht alle Datensätze die gleiche Versionsangabe haben.

Bei **jeder** Änderung im einzelnen Datensatz wird dort die interne Versionsvariable geändert und das Änderungsdatum angegeben. Hierbei handelt es sich um einen datensatzinternen Vorgang. Die

Kenntlichmachung der datensatzbezogenen Änderung geschieht auf Datensatzebene (s. dazu Benennungskonvention).

2. Neue externe Version eines Datenproduktes

Bei einer neuen externen Version eines Datenprodukts erstellt das FDZ am IQB alle beschreibenden und strukturellen Metadaten neu, es wird eine neue Versionsnummer erstellt und diese Datenversion ebenfalls langzeitarchiviert¹.

Dem neuen Datenprodukt wird eine neue DOI zugewiesen. D. h.: Jede externe Datenversion hat ihre eigene DOI. Auf diese Weise verweist die bereits vorhandene DOI weiterhin eindeutig auf die frühere Version des Datenproduktes. *Alle* Versionen eines Datenproduktes werden in ihren jeweiligen beschreibenden Metadaten mit einem Querverweis versehen. Neue Datenversionen einer Studie haben ein eigenes AIP. Die vorherige(n) Datenversion(en) und das/die dazugehörige(n) vorherige(n) Archival Information Package(s) (AIP) verbleiben in der Langzeitarchivierung. Aus Sicht des FDZ am IQB stellt das AIP ein Datenpaket dar.

2.1 Kriterien, wann eine neue DOI-Version erforderlich wird

Wenn Analysen und Reproduktionen bei Änderungen am Datensatz/an Datensätzen zu einem anderen Ergebnis kämen als unter der aktuellen/vorherigen Version, dann wird eine neue externe Version des Datenproduktes benötigt.

- ◆ Inhaltliche Veränderungen in den Variablen (z. B. Veränderung der IDs, um Verknüpfbarkeit zu anderen Datensätzen/Studien zu ermöglichen; Reskalierung von Testscores etc.)
- ◆ Löschen und/oder Hinzufügen von Variablen
- ◆ Hinzukommen eines Rekodierschlüssels
- ◆ Löschen oder Hinzufügen von Datensätzen zu einer Studie (z. B. Hinzufügen neuer Erhebungswellen in längsschnittlichen Studien wie BiKS, BiLWiss, BRISE)
- ◆ Veränderungen an der Stichprobe (Ausschluss oder Hinzufügen von Personen)

2.2 Benennungskonvention

Datenprodukt

Die Versionsangabe des **gesamten Datenproduktes** wird um 1 hochgezählt.

¹ https://fdz.iqb.hu-berlin.de/documents/39/Langzeitverfuegbarkeit_externe_Version_Stand-2025-Juli.pdf

Datensatz/-sätze

- ♦ **Nur die Versionsnummer der bearbeiteten Datensätze** wird um 1 erhöht.
- ♦ Wenn in einem Datenprodukt ein oder mehrere Datensätze inhaltlich angepasst werden müssen, dann erhalten diese geänderten Datensätze die Versionskennung v[Nr.+1 zur Vorversion]. Die Versionsnummern und -variablen der anderen Datensätze des Datenproduktes ändern sich jedoch nicht.
- ♦ Wenn einem Datenprodukt ein oder mehrere neue Datensätze hinzugefügt werden, dann erhalten die entsprechenden Datensätze dabei die Versionskennung v1. Die Versionsnummern und -variablen der anderen Datensätze des Datenproduktes ändern sich jedoch nicht.
- ♦ In dem entsprechenden Datensatz wird eine Versionsvariable eingefügt bzw. die bestehende Versionsvariable aktualisiert.

2.3 Ablage

- ♦ Versionierte Datensätze werden in der entsprechenden Studie abgelegt.
- ♦ Die vorherigen Versionen der entsprechenden Datensätze werden ins Archiv in die entsprechende Studie verschoben.

3. Neue interne Version

Alternativ dazu wird bei einer administrativen Änderung dies in den administrativen Metadaten dokumentiert. Es wird keine neue DOI vergeben. Die interne Versionsvariable dient dazu, bei Anfragen von Datennutzenden nachzuvollziehen, mit welcher Datensatzversion gerechnet wurde. Da die interne Versionsvariable datensatz- anstatt datenproduktbezogen vergeben wird, erhöht sich die Reproduzierbarkeit und Nachvollziehbarkeit bei Nutzendenanfragen.

Auch interne Versionen werden langzeitarchiviert.

3.1 Kriterien, wann eine neue interne Version erstellt wird

Administrative Änderungen in einem Datensatz erfordern für diesen eine interne Version.

- ♦ Rechtschreibungskorrekturen im Datensatz (nicht im Dateinamen)
- ♦ Ergänzung bzw. Änderung von Werte- und Variablenlabels
- ♦ Anpassungen des Skalenniveaus, des Formats etc. der Variablen im Datensatz

3.2 Benennungskonvention

Die Versionsangabe im Dateinamen eines bearbeiteten Datensatzes gestaltet sich wie folgt:
v[Nr]_[interne Vorversion+1]

3.3 Ablage

- ♦ Versionierte Datensätze werden in der entsprechenden Studie abgelegt.
- ♦ Die vorherigen Versionen der entsprechenden Datensätze werden ins Archiv in die entsprechende Studie verschoben.

4. In Bezug auf die Langzeitarchivierung

Sowohl neue externe als auch interne Versionen werden langzeitarchiviert, d. h., dass neue Versionen von Datenprodukten in einem neuen AIP abgelegt werden. Hierbei werden für den Fall, dass es pro Studie mehrere Datenprodukte gibt, diese gemeinsam in einem AIP abgelegt, da ein AIP immer auf Ebene der Studie angelegt wird.

Die Konvertierung in langzeitverfügbare Formate zielt darauf ab, den Inhalt der Daten zu erhalten, sie also über die Zeit sowohl les- als auch interpretierbar zu halten, da dies wesentliche Eigenschaften der Daten sind. Die Erhaltung anderer Aspekte, wie z. B. des Layouts des Eingabeformats (das "Look and Feel"), wird als weniger wichtig erachtet.

Für den Fall einer Datenkonvertierung in ein anderes Dateiformat zu Bewahrungs- oder Zugangszwecken werden die Originaldatei(en) erhalten.