**DDP-Bildung**

**Mindestanforderung II.a):  
Dokumentation der Daten**

**II. Themenbereich: Dokumentation / Qualität**

# Versionierungshistorie

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | | Typ | Datum | erstellende Person |
| 1 | 20200626 | erster Entwurf | 26.06.2020 | sen |
| 2 | 20200721 | Abschnitt D ergänzt | 07.09.2020 | sen |
|  |  | Ergänzung der Spezifika 3.8 | 28.09.2020 | gf |
| 3 | 20201021 | Überarbeitung Arbeitstreffen 2020 | 21.10.2020 | aks / sen |
| 4 | 20201124 | MA überarbeitet | 24.11.2020 | sen |
| 5 | 20201125 | Version für RDMO | 25.11.2020 | sen |
|  |  |  |  |  |

# Mindestanforderung

Die Daten werden im Projektverlauf so dokumentiert, dass Dritte die gesamte Datengenese nachvollziehen, die Daten vollumfänglich verstehen und sinnvoll interpretieren sowie auffinden und zitieren können. Um die Verständlichkeit der Dokumentation zu erhöhen sowie die Auffindbarkeit und Zitation der Daten zu gewährleisten, werden zur Dokumentation bestehende Metadaten- und Dokumentationsstandards genutzt.

## Abhängigkeiten mit anderen Modulen

Diese Mindestanforderung steht in Verbindung zu folgenden weiteren DDP-Modulen:

I.a) Berücksichtigung bereits existierender Daten

II.b) Datenqualität / Nachvollziehbarkeit,

IV.a) Rechtliche Aspekte (Datenschutz)

IV.b) Urheberrechte

V.b Langfristsicherung von Daten

VI.a) Verfügbarmachung der Daten

VI.c) Datenzugang & -nachnutzung

[ggf. WEITERE MODULE ERGÄNZEN]

## Erörterungen zur Mindestanforderung

Die Mindestanforderung fordert eine für die jeweiligen Forschungsdaten und ihre Begleitmaterialien adäquate Dokumentation, die sicherstellt, dass die Daten auch für Dritte nachvollziehbar, verständlich und nutzbar (*RE-USABILITY*) sind. Dies betrifft sowohl die Inhalte der Dokumentation als auch die Nutzung von Dokumentations- und Metadatenstandards zur Steigerung der Auffindbarkeit (FINDABILITY) der Daten ebenso wie ihrer Verständlichkeit und Interoperabilität (*INTEROPERABILITY*).

* + - 1. Inventarliste (Teil III DDP-Deckblatt)

Die Inventarliste gibt den Bestand aller relevanten Daten und Begleitmaterialien im Projekt wieder. Sie bietet so eine Übersicht über die zu verarbeitenden Daten und unterstützt deren Dokumentation ebenso wie den Umgang mit den Daten im Projekt und darüber hinaus.

* + - * 1. Die Liste enthält alle relevanter Daten und Begleitmaterialien und gibt Auskunft über deren Typ, Inhalt und Herkunft.
        2. Es ist ersichtlich, welche Daten und Begleitmaterialen wie zu einander gehören.
        3. Die Softwarevoraussetzungen zum technischen Zugriff auf die Daten sind gekennzeichnet.
        4. Besonders schützenwerte Daten sind entsprechend gekennzeichnet.
        5. Die längerfristige Planung zum Umgang mit den Daten über das Projektende hinaus ist ersichtlich.
        6. Es besteht eine Arbeitsroutine zur Aktualisierung der Inventarliste.
      1. Inhalte der Dokumentation

Die Inhalte der Dokumentation hängen u. a. vom konkreten Datentyp, der Erhebungsmethode und dem Forschungsansatz ab. Je nach Datentyp etc. umfasst die Dokumentation Informationen, über

* + - * 1. das originäre Forschungsprojekt, wie etwa zu den Primärforschenden, zur Forschungsförderung oder zum Forschungsansatz und den originären Zwecken der Datenerhebung und -nutzung;
        2. die Datenerhebung, wie z. B.
    - die Untersuchungspopulation sowie der temporäre und geographische Untersuchungsraum (siehe *Anwendungsfall XY*),
    - die Methode der Stichprobenziehung, der Fallauswahl und des Feldzugangs, die Kontaktaufnahme mit den Zielpersonen, der Stichprobenfehler bzw. systematische Ausfälle von Zielpersonen,
    - die Einwilligungserklärung der Zielpersonen
    - ggf. besondere Umstände bzw. Kontextinformationen (Kontextualisierung),
    - Maßnahmen zur Sicherung der Datenqualität während der Erhebung;
      * 1. das Erhebungsinstruments, d. h. je nach Erhebungsinstrument
    - die originären Fragebögen bzw. die Beschreibung der Abfolge von Fragen, der Filterführung, der Harmonisierung von Fragen und Variablen bei Trenddaten oder von dynamischen Elementen in online-Befragungen,
    - die originären Testhefte und Itemtexten sowie die Beschreibung von Kompetenzmodellen und Normwerten, Interpretationshilfen zur Testmetrik und die Korrespondenz zwischen einzelnen Fragen bzw. Tests zur Itembildung,
    - die originären Instrumente, Leitfragen und -linien von Interviews, Experimenten oder Beobachtungen oder deren Beschreibung,
    - eventuell genutzte Stimuli, Bilder, Logos etc. bzw. deren Beschreibung und ggf. Zitation, wenn entsprechende Bilder Logos etc. den Urheberrechten Dritter unterliegen und nicht selbst Teil der Dokumentation werden können (siehe Modul IV.b) *Urheberrecht*),
    - ggf. die Nennung zur Erhebung genutzter Software;
      * 1. die Datenaufbereitung, d. h. je nach Datentyp
    - die Maßnahmen zur Sicherung der Qualität der Rohdaten,
    - die Konzepte und Vorgehensweisen zur Anonymisierung bzw. Pseudonymisierung der Daten,
    - die Beschreibung und Begründungen zur Löschung einzelner Fälle, Informationen, Variablen, Interviewteile oder Teile von (Video-)Aufnahmen,
    - die Maßnahmen zur Kontrolle auf fehlende Angaben, inkonsistente, wiederholte oder unerwartbare Antworten etc.
    - die zur Aufbereitung genutzte Software und ggf. die dabei verwendeten Syntaxdateien (siehe Modul II.b) *Datenqualität und Nachvollziehbarkeit*);
      * 1. die Daten bzw.je nach Datentyp
    - die konsistente und systematische Benennung von Werten und Variablen sowie deren Bezug zu den originären Fragen (ggf. auch über Variablen- und Wertelabel in der Datenmatrix)
    - die Beschreibung der Datenstruktur und ggf. der Zusammengehörigkeit einzelner Teildatensätze bzw. zu anderen im Projekt genutzten Daten ebenso wie zu Relationen zwischen einzelnen Variablen (etwa zur Skalenbildung)
    - die Beschreibung der Bildung und Nutzung von Gewichten
    - zur Kontextualisierung von Interviews, Videographien, Transkripten etc., d. h. zu deren Dauer und Datum der Durchführung sowie zu den interviewende und transkribierende Person
    - Angaben zu den fachlichen Hintergründen von Befragten, wie z. B. zu deren Funktion oder soziodemographische Merkmale etc.;
      * 1. die (Nach-)Nutzbarkeit der Daten durch Dritte (siehe Modul V.c) *Datenzugang und–nachnutzung*), d. h.
    - zu den die notwendigen Softwarevoraussetzungen zum Datenzugriff,
    - zu bestehenden Regelungen des Datenzugang und zu den Lizenzbedingungen,
    - zu eventuellen Nachnutzungsbeschränkungen.
      * 1. Werden die Daten durch Dritte erhoben bzw. verarbeitet, dann umfasst die Dokumentation
    - Angaben zu diesen Dritten ebenso wie zu entsprechenden Vereinbarungen, z. B. bezüglich
      * der vorgenommenen Kontrollmaßnahmen während der Datenerhebung bzw. -aufbereitung,
      * der Leitlinien zur Transkription von Interviews, Audioaufnahmen oder Videographien,
      * der Maßnahmen zur Aufbereitung und Kodierung der Daten sowie ggf. zur Konstruktionen von Gewichten,
      * der Datendokumentation und der dazu genutzten Metadaten- und Dokumentationsstandards,
      * der Administration und automatisierten Auswertung von Testbögen, z. B. Hogrefe-Tests;
    - eine klare Kennzeichnung von Arbeiten Dritter in Abgrenzung zur Eigenleistung der Forschende abgegrenzt werden.
      1. Form der Dokumentation

Zur Gewährleistung des technischen Zugriffs auf die Daten und ihre Begleitmaterialien werden die Dateien (soweit möglich) in nicht-proprietären Formaten gespeichert.

Zur Dokumentation werden bestehende Dokumentations- und Metadatenstandards ebenso wie ein kontrolliertes Vokabular berücksichtigt. Diese erhöhen nicht nur die Verständlichkeit der Dokumentation, der Daten und ihrer Begleitmaterialien, sondern sind in der Regel auch maschinenlesbar und interoperabel und unterstützen so die Auffindbarkeit der Daten durch Dritte (siehe *Anwendungsfälle*, *Standards* und *weiterführende Ressourcen* im vorliegenden Modul).

Um die Daten in (online) Datenkatalogen zu registrieren und mit einem persistenten Identifikator zu versehen (siehe Module V.a) *Verfügbarmachen der Daten* und V.c) *Datenzugang und –nachnutzung*) bedarf es zudem (einiger) standardisierter Metadaten, z. B. zum Titel der Daten, zu den Primärforschende, zum Datenzugang etc., die vom jeweiligen Registrierungsagenten vorgegeben werden (siehe *Anwendungsfälle* und *Standards* sowie *weiterführende Ressourcen* des vorliegenden DDP-Moduls).

Inhalte & Standards ggf. aber auch von Repositorien zur Datenarchivierung vorgegeben werden. im Rahmen der Dokumentation frühzeitig mit den Handreichungen und Richtlinien der entsprechenden Archivierungseinrichtung vertraut und setzen diese systematisch um