

Das erweiterte Metadatenschema der VFU soeb 3

Uwe Jensen, Stefan Schweers

GESIS-Technical Reports 2014|13

Das erweiterte Metadatenschema der VFU soeb 3

Uwe Jensen, Stefan Schweers

GESIS-Technical Reports

GESIS – Leibniz-Institut für Sozialwissenschaften

Unter Sachsenhausen 6-8

50667 Köln

Telefon: (0221) 476 94 - 0

Telefax: (0221) 476 94 - 199

E-Mail: uwe.jensen@gesis.org

ISSN: 1868-9043 (Print)

ISSN: 1868-9051 (Online)

Herausgeber,

Druck und Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln

Inhaltsverzeichnis

1	Einführung	5
2	Das erweiterte Metadatenschema der VFU soeb 3 im Überblick	7
2.1	Darstellung der Metadatenelemente in den Schematabellen.....	8
2.2	Erforderliche, empfohlene und optionale Elemente.....	10
3	Das Metadatenschema Studie	13
3.1	Einführung.....	13
3.2	Die Metadatenelemente des Schemas Studie.....	13
3.3	Metadatenelemente zum Upload von Dateien zum Schema Studie.....	24
4	Das Metadatenschema Datennutzung	26
4.1	Einführung.....	26
4.2	Die Metadatenelemente des Schemas Datennutzung.....	26
4.3	Metadatenelemente zum Speichern von Dateien zu einer Datennutzung.....	29
5	Das Metadatenschema Syntax	31
5.1	Einführung.....	31
5.2	Die Metadatenelemente des Schemas Syntax.....	31
5.3	Metadatenelemente zum Speichern von Dateien zu einer Syntax.....	35
	Referenzen	37

1 Einführung

Ziel des Projekts „Virtuelle Forschungsumgebung“ (VFU) ¹ war die Entwicklung einer Arbeitsplattform, die die Zusammenarbeit räumlich verteilter wissenschaftlicher Einrichtungen in Verbundprojekten unterstützt. Virtuelle Forschungsumgebungen können nur forschungsnah und projekt- oder community-spezifisch entwickelt werden.

Dazu wurden Infrastrukturentwicklung und fachwissenschaftliche Forschung projektförmig verknüpft, wobei das vom BMBF geförderte Verbundprojekt „Berichterstattung zur sozioökonomischen Entwicklung Deutschlands“ (soeb 3) ² (2013 bis 2016) als Anwendungsfall für die projektbezogenen, forschungsnahen Entwicklung der VFU dient.

Das VFU Projekt stützte sich auf konzeptionelle Vorarbeiten aus dem Teilprojekt „Kollaborative Datenauswertung und Virtuelle Arbeitsumgebung“ (VirtAug) im Rahmen der Konzeptphase ³ zu soeb 3. Im Zuge der Konzeptentwicklung sollte das Modellprojekt am praktischen Fall

„der sozioökonomischen Berichterstattung untersuchen und dokumentieren, wie der gemeinsame Datenzugang und die datenbezogene Kooperation von Sozialwissenschaftler/innen an verschiedenen quantitativ-empirisch orientierten Forschungseinrichtungen und insbesondere eine kollaborative Auswertung der Mikrodaten von Forschungsdatenzentren künftig besser organisiert und technisch unterstützt werden könnten.“ (vgl. Bartelheimer; Schmidt. 2011)

Die vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Umsetzungsphase ⁴ begann im Juni 2012 und dauerte bis Ende Juni 2014. Das in dieser Zeit entstandene Portal VFU soeb 3 und die modular implementierten Werkzeuge unterstützen im Projektbetrieb insbesondere die kollaborative Nutzung sozialwissenschaftlicher Mikrodaten über die gesamte wissenschaftliche Leistungskette, die sich vom Datenzugang bis hin zur Archivierung von Forschungsergebnissen erstreckt. Darüber hinaus werden durch weitere Komponenten die VFU Kommunikation und Koordination im Forschungsverbund soeb 3 erleichtert.

Unter einer gemeinsamen Oberfläche integriert das Portal die entwickelten der Metadateneditoren (siehe unten) und weitere Werkzeuge wie Forum, Wiki, Dateiverwaltung, Terminkalender, Ankündigungs- und Absprachenkomponente, sowie – vermittelt der Middleware – die Speicherung und das Durchsuchen von Metadaten (Archivfunktion).

Die Datenverwaltung soll schwerpunktmäßig die Archivierung, Dokumentation und Nachnutzung bzw. kollaborative Nutzung von Auswertungssyntax (Syntax-Sharing) unterstützen. Neben Syntaxdateien und (freigegebenen) Forschungsdaten werden über sie auch alle anderen Dateiformate, etwa Outputdateien, Tabellen- oder Textdokumente, formatunabhängig abgelegt und mittels Inhaltssuche sowie über Metadaten auffindbar gemacht werden.

Es soll möglich sein, im Rahmen der bestehenden Nutzungsberechtigungen logische Verknüpfungen zwischen den verschiedenen Datenarten (z.B. zwischen Syntax und genutzten Forschungsdaten), herzustellen, um den Forschungsprozess zusammenhängend und transparent nachvollziehen zu können. Eine entscheidende Voraussetzung für die Nutzung einer gemeinsamen Arbeitsumgebung besteht weiterhin darin, dass die Daten anhand von Metadaten gut beschrieben werden (vgl. SOFI. 2013).

¹ <http://www.soeb.de/vfu-soeb-3/>

² <http://www.soeb.de/ueber-soeb-3/>

³ <http://www.soeb.de/vfu-soeb-3/konzeptphase/>

⁴ <http://www.soeb.de/vfu-soeb-3/umsetzungsphase/>

Ziel des Projekts ist es auch, über den Forschungsverbund Sozioökonomische Berichterstattung hinaus Lösungen von allgemeiner Bedeutung für die quantitativ-empirisch forschenden Sozialwissenschaften zu erarbeiten. Am 24.01.2014 präsentierte der Projektverbund "Virtuelle Forschungsumgebung für die sozioökonomische Berichterstattung" (VFU soeb 3) in Berlin die erste operativ genutzte Version der VFU einer breiteren Fachöffentlichkeit und stellte sie im Rahmen von fünf Panel zur Diskussion.⁵

Das für den Produktivbetrieb eingerichtete Portal „Virtuelle Forschungsumgebung für die sozioökonomische Berichterstattung“ ist für die registrierten Mitglieder des Forschungsverbundes Portal zugänglich⁶. Eine Vorstellung, wie die einzelnen Bereiche und Funktionalitäten der VFU zusammenhängen, vermittelt Schmidt (2014). Die Einführung in die Nutzung der VFU stellt den Aufbau und die Zusammenhänge der wichtigsten Komponenten einschließlich beispielhafter Bearbeitungsabläufe als Einstieg für die Nutzerinnen und Nutzer des Portals dar.

Die Virtuelle Forschungsumgebung wurde in einem interdisziplinären Projektverbund von folgenden Partnern entwickelt:

- Soziologisches Forschungsinstitut Göttingen (SOFI) e.V. an der Georg-August-Universität Göttingen (Projektleitung)
- GESIS – Leibniz-Institut für Sozialwissenschaften e.V., Mannheim
- Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG), Göttingen
- Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg
- Niedersächsische Staats- und Universitätsbibliothek (SUB) der Georg-August-Universität Göttingen

in Kooperation mit

- D-Grid-gGmbH, Dortmund
- Forschungsdatenzentrum der Rentenversicherung Bund, Berlin und Würzburg
- Forschungsdatenzentrum des Sozio-ökonomischen Panels, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

GESIS definierte die für die VFU benötigten Metadaten und spezifizierte diese in zusammenhängenden Metadatenschemata, die in diesem Bericht beschrieben werden. Darauf aufbauend wurden im Zuge fachwissenschaftlicher Spezifikationen mit den Projektpartnern weitere kollaborative IT-Werkzeuge, insbesondere drei Metadateneditoren von GESIS entwickelt und implementiert. Die Editoren und die implementierten Metadatenelemente werden im Report „Die Metadateneditoren der VFU soeb 3“ dokumentiert (Jensen, Schweers, Carevic. 2014).

In dem hier vorliegenden Report werden die Metadatenelemente des erweiterten Metadatensets der VFU soeb 3 beschrieben. Zur Dokumentation der für Datenanalysen relevanten Objekte Studie, Datennutzung und Syntax wurde jeweils ein Metadatenschema entwickelt. Diesen Hauptobjekten können außerdem Dateien zugeordnet werden, die im Archiv der VFU soeb 3 gespeichert werden können.

Kapitel 2 gibt einen Überblick über die einzelnen Schemata und beschreibt wie die Metadatenelemente in den Schematabellen spezifiziert werden. Weiterhin werden erforderliche, empfohlene und optionale Elemente sowie automatische erzeugte Metadaten schemäübergreifend beschrieben.

Die Kapitel 3 bis Kapitel 5 spezifizieren alle Elemente der Metadaten-Schemata Studie, Datennutzung und Syntax und die mit ihnen verbundenen Subschemata zur Dokumentation der in im VFU Archiv gespeicherten Dateien.

⁵ <http://www.soeb.de/vfu-soeb-3/abschlussstagung/>

⁶ <https://vfu.sofi.gwdg.de/>

2 Das erweiterte Metadatenchema der VFU soeb 3 im Überblick

Das erweiterte Metadatenchema der VFU soeb 3 besteht aus den drei Teilschemen Studie, Datennutzung und Syntax. Die drei Schemata beschreiben ausschließlich Objekte und die mit ihnen verknüpften Dateien, die Gegenstand üblicher Arbeitsabläufe bei der Durchführung von Datenanalysen sind. Im soeb 3 Forschungsverbund werden diese Analysen in achtzehn wissenschaftlichen Arbeitspaketen von über fünfunddreißig Wissenschaftlerinnen und Wissenschaftlern aus siebzehn Einrichtungen durchgeführt. Die Erfassung und Erschließung von analysebezogenen Metadaten in der VFU soeb 3 unterstützt damit übergreifend die kollaborative Nutzung sozial- und wirtschaftswissenschaftlicher Mikrodaten. Die dazu im Produktivsystem VFU soeb 3 implementierten Werkzeuge und Metadatenelemente sind im Bericht „Die Metadateneditoren der VFU soeb 3“ (Jensen, Schweers, Carevic. 2014) dokumentiert.

Im Folgenden werden die Metadatenelemente des erweiterten Metadatenchemas der VFU soeb 3 vorgestellt. Mit Hilfe dieser Metadaten können die Forschenden u. a. in Analysen genutzte Ausgangsdaten, Untersuchungsziele, methodische Strategien und Analyseparameter der Datennutzung, angewendete statistische Methoden und konkret benötigten Syntaxen und deren Entwicklung systematisch dokumentieren. Gleichzeitig fördern sie die informationelle Vernetzung zwischen den wissenschaftlichen Arbeitspaketen, indem z. B. alle von den Arbeitspaketen dokumentierten Datensätze im VFU Archiv durchsucht werden können. Insbesondere fördern die Mehrwertdienste des Archivs (vgl. ebenda, Kapitel 8) auf Grundlage der erfassten Metadaten die gemeinsame Entwicklung, den Austausch und die Nachnutzung von Auswertungssyntaxen. Im Sinne der Sicherung guter wissenschaftlicher Praxis dokumentieren die Metadaten der VFU soeb 3 relevante Information des Forschungsprozesses von den Ausgangsdaten bis zu den Ergebnissen der Datenanalysen. Diese Angaben in den Metadaten werden nachprüfbar, nachnutzbar und dauerhaft im VFU Archiv gesichert und stehen damit auch über das Projektende hinaus für weitere Infrastrukturzwecke zur Verfügung.

Die in drei Schemata organisierten Metadatenelemente definieren jeweils eine Liste von strukturierten Angaben, mit deren Hilfe die folgenden Objekte und ihnen zugeordnete Dateien beschrieben werden.

Das Metadatenchema **Studie (STB)** beschreibt die Ausgangsdaten und die Dokumentationen der Datensätze (einer Studie), die für Datenanalysen im soeb 3 Verbundprojekt genutzt werden. Dokumente und / oder Datenfiles des Datengebers können der Studienbeschreibung zugeordnet und (soweit sinnvoll und rechtlich zulässig) im Archiv gespeichert werden. Durch Upload im Archiv gespeicherte Dateien werden durch die Subschemen „STB-Upload“ beschrieben. Die Inhalte der einzelnen Metadatenelemente werden mit dem Metadateneditor Studie bearbeitet.

Das Metadatenchema **Datennutzung (DN)** enthält die Elemente, die methodische Konzepte und Parameter der Datenanalysen beschreiben. Diesem Schema können im Archiv gespeicherte Datensätze zugeordnet werden, die von soeb 3 Arbeitspaketen aus Ausgangsdaten generiert wurden. Solche Datendateien und Textdokumente zur Datennutzung werden durch die Elemente im Subschema „DN-Upload“ beschrieben. Die Inhalte werden mit dem gleichnamigen Metadateneditor Datennutzung erfasst.

Im Metadatenchema **Syntax (SX)** werden Syntaxdateien in verschiedenen Entwicklungsstadien sowie dazugehörige Outputdateien beschrieben. Diesem Schema können im Archiv gespeicherte Syntaxdateien und insbesondere die zu einer Syntaxdatei gehörende Ergebnisdateien zugeordnet werden. Durch Upload im Archiv gespeicherte Syntaxdateien und Outputdateien werden mit dem Subschema „SX-Upload“ beschrieben. Die Metadaten werden mit dem Editor Syntax verarbeitet.

Bei der Entwicklung der Metadaten schemata Studienbeschreibung und Datennutzung wurde auf die Elemente des Metadatenstandards DDI 2 (DDI Alliance. 2014) zurückgegriffen. Demgegenüber existiert aktuell noch keine systematisch zusammenhängende Metadatenstruktur zur ausführlichen Dokumentation von Syntaxdateien und Outputdateien und des projektbezogenen Kontextes von kollaborativ durchgeführten Datenanalysen (vgl. dazu Friedhoff u.a. 2013: 12 ff.). Aus diesem Grund wurde für die Bedürfnisse der VFU soeb 3 das Metadaten schema „Syntax“ neu entwickelt.

Die implementierten Elemente werden seit Ende 2013 im Produktivsystem der VFU soeb 3 genutzt (Jensen, Schweers, Carevic. 2014). Schlussfolgerungen zur praktischen Nutzung der Metadaten und welche „Veränderungen in den Arbeitsabläufen durch die technischen Möglichkeiten tatsächlich stattfinden, können nur zukünftige Nutzungsstudien, die sich auf die VFU im Echtbetrieb beziehen, zeigen“ (Schmidt. 2014b).

Die in diesem Report dokumentierten Schemata und Elemente werden in die laufenden Entwicklungen der DDI Alliance eingebracht (vgl. Jensen. EDDI 2014). Die Kapitel 3 bis 5, in denen die drei Metadaten schemata Studie, Datennutzung und Syntax spezifiziert werden, beschreiben weitere konzeptuelle Aspekte der Schemata.

Den Zusammenhang zwischen den drei Metadaten schemata und ihren Dateien sowie deren Erfassung und Bearbeitung durch entsprechende Editoren und der Bezug zum VFU Archiv zeigt die folgende Abbildung.

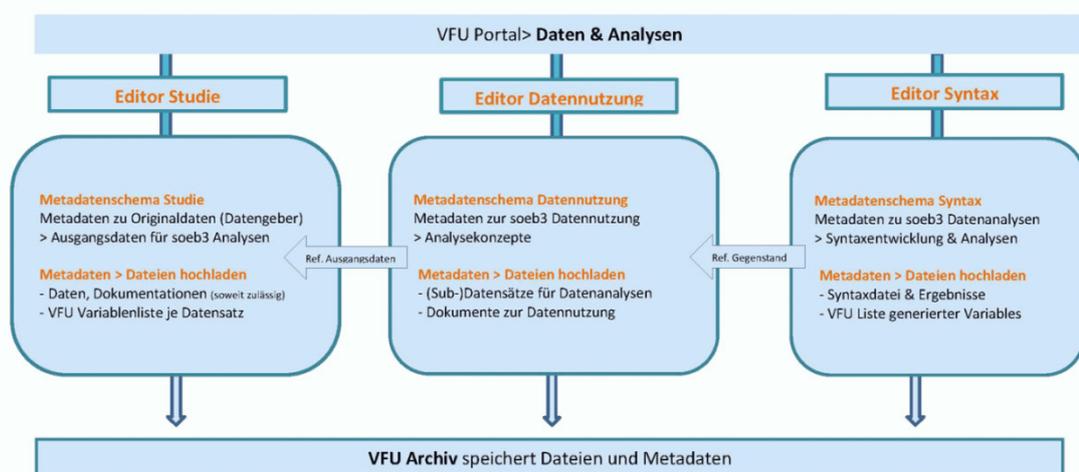


Abbildung 1: Editoren und Metadaten schemata im Überblick (Jensen. 2014)

2.1 Darstellung der Metadatenelemente in den Schematabellen

Die Tabellen in den Kapiteln 3 bis 5 beschreiben alle Elemente des jeweiligen Metadaten schemas. Zusätzlich enthält jedes einzelne Schema der Objekte Studie, Datennutzung und Syntax zwei weitere Tabellen (Subschemen) mit Elementen, die Dateitypen dokumentieren, die beim Speichern im Archiv dem jeweiligen Objekt zugeordnet werden.

Jede Tabelle enthält Information entsprechend der Einträge im Tabellenkopf, die im Folgenden näher erläutert werden.

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
-----	-------------	-------------------------------------	----

Abbildung 2: Spaltennamen der Metadaten schematabellen

- **Spalte Nr.**

Hier ist die laufende Nr. des nebenstehenden Elementnamens angegeben.

Ist einem Zeileneintrag keine Nummer zugeordnet, handelt es sich um eine einleitende Anmerkung einer logischen Gruppe, die den folgenden Elementen zur Erläuterung vorangestellt wird.

Ist einem Zeileneintrag eine ganzzahlige Nummer (z. B. 5.) zugeordnet, handelt es sich i.d.R. um einen Oberbegriff, ohne das zu diesem Element ein Inhalt erfasst wird. Die letzte Spalte ist in diesem Fall leer. Oberbegriffe dienen als Kontextinformation, die den folgenden Sub-Elementen (z. B. 5.1 / 5.2) vorangestellt werden, welche die Inhalte enthalten.

- **Spalte Elementnamen**

Name des Metadatenelementes. Zur Erleichterung der Zuordnung werden z. T. übergeordnete Begriffe vorausgehender Elemente im Namen des aktuellen Elementes mitgeführt.

- **Spalte „Definition und Anmerkungen (Schema Abkürzung)“**

Das Kürzel in der Klammer benennt das aktuelle Schema, z. B. steht MD-STB für Metadaten Studienbeschreibung.

Der Inhalt in dieser Spalte hat eine doppelte Funktion. Einleitend wird das Element definiert. Zusätzliche Anmerkungen erläutern soweit notwendig u.a. inhaltliche oder formale Besonderheiten des Elements, zeigen logische Abhängigkeiten zu anderen Elementen, verweisen auf die Wiederholbarkeit von Elementen und geben Beispiele.

Aus praktischen Gründen wird in diesem Bericht darauf verzichtet, die logischen Beziehungen zwischen den Elementen (A/C; Child/Attribut) sowie zum Attribut „Occurance“ (Anzahl möglicher oder notwendiger Instanzen eines Elements) gesondert in Spalten darzustellen. Die Angabe 1-n im Text bedeutet in diesem Zusammenhang, dass das Element einmal (1) vorkommen muss, und mehrmals (n) vorkommen kann. 0-n bedeutet, dass das Element nicht vorkommen muss (0), aber mehrfach vorkommen kann.

Weiterhin werden in dieser Spalte die Kontrollierten Vokabulare aufgeführt, die als standardisierte Listeneinträge in den Editoren genutzt werden. Soweit Vokabulare anderen Schemata entstammen, werden sie hier als Quelle aufgeführt und in der Referenzliste nachgewiesen.

- **Spalte „FA“ (Funktionale Aspekte)**

In der letzten Spalte werden die Elemente hinsichtlich funktionaler Aspekte näher gezeichnet.

Dazu gehört insbesondere die Angabe zu erforderlichen (M) und empfohlene (E) bzw. optionalen (ohne Kodierung) Einträgen.

Bei den erforderlichen Angaben wird zwischen Metadaten unterschieden, die von einer Person erfasst werden müssen (M), und solchen, die von einem technischen System der VFU (Archiv oder Portal) automatisch erzeugt werden.

Aus Gründen der Übersichtlichkeit werden die erforderlichen, die empfohlen und die optionalen Elemente nicht gesondert in Kurzform als Tabelle ausgewiesen. Stattdessen werden sie inhaltlich und schemaübergreifend in Kapitel 2.2 zusammenfassend erläutert.

Die entsprechenden Angaben werden in allen Schemata außerdem einheitlich in der letzten Spalte codiert. Alle verwendeten Abkürzungen zeigen die nächste Übersicht.

Übersicht der Kürzel zur Kodierung von funktionalen Angaben in Spalte „FA“

- AA Erforderliche Angabe, die automatisch vom Archivsystem erzeugt wird
- AP Erforderliche Angabe, die automatisch vom Portalsystem erzeugt wird
- M Erforderliche Angabe, die von einer Person erstellt wird
- E Empfohlene Angabe
- FT Freitextfeld
- KL Kalenderfunktion
- KV Kontrolliertes Vokabular
- URI Uniform Resource Identifier. Einheitlicher Bezeichner zur Identifizierung von Quellen. Die im Schema verwendete Form der URL verweist auf eine Quelle im Web.

2.2 Erforderliche, empfohlene und optionale Elemente

Die erforderlichen, empfohlenen und optionalen Elemente sowie automatisch erzeugte Metadaten werden schemaübergreifend in den folgenden Abschnitten beschrieben.

Erforderliche Elemente (M)

Beim Anlegen eines Objektes Studie, Datennutzung oder Syntax sind zwei Angaben zu den folgenden Elementen erforderlich,

- Titel (des Objektes)

Der Titel bezeichnet das jeweilige Objekt durch einen aussagekräftigen Inhalt. Dieser ist zur sinnvollen Erschließung der Objekte erforderlich und wird als zentrales Metadatum bei der Archivsuche verwendet. In der VFU soeb 3 wurden dazu Konventionen zur Vergabe von Dateinamen und Titeln vorgeschlagen (Schmidt 2014a: 5).

- Arbeitspaket

Ein VFU Mitglied kann Mitglied in einem oder mehreren wissenschaftlichen bzw. administrativen Arbeitspaketen sein. Um die Bearbeitung eines Objektes dem jeweiligen Arbeitskontext zuzuordnen, wählt das Mitglied das jeweilige Arbeitspaket aus einer vordefinierten Liste aus.

Werden einem der Objekte Studie, Datennutzung oder Syntax Dateien durch Speichern im VFU Archiv zugeordnet, sind Angaben zu diesen Elementen erforderlich:

- Arbeitspaket (siehe oben)
- Referenz Titel < Studie, Datennutzung oder Syntax >

Die Referenzierung mit dem jeweiligen Objekt erfolgt durch die Auswahl aus einer Liste mit allen angelegten Titeln in einem Objektbereich.

- Titel (der Datei) (siehe oben)
- Dateinamen (siehe oben Titel des Objektes)

Angabe des vollständigen Namens (Dateibezeichner, Dateiformat) der im VFU Archiv gespeicherten Datei. Er kann beim Upload in die VFU übernommen oder vor dem Speichern editiert werden.

Automatisch erzeugte administrative Metadaten

Weiterhin werden einige Elemente automatisch vom System Portal oder vom Archivspeicher generiert, die erforderlich sind, um das Objekt zu verwalten. Gleichzeitig sind diese Metadaten auch erforderlich, weil sie als Kontextinformation bei der Bearbeitung eines Objektes in einem der Editoren automatisch angezeigt werden sollen (vgl. Jensen, Schweers, Carevic. 2014: Kapitel 6).

Zu den automatisch erzeugten (technisch-administrativen) Metadaten zählen im Besonderen die folgenden Elemente, die in allen drei Schemata und ihren dateibezogenen Subschemen benutzt werden.

Vom VFU Portal erzeugte administrative Angaben (AP)

- Vorname, Nachname der Person, die das Objekt (und zugeordnete Dateien) angelegt bzw. zuletzt bearbeitet hat.

Diese Metadaten werden aus Angaben der im Portal angemeldeten Person entnommen, die einen der Metadateneditoren zur Erstellung oder Bearbeitung des Objektes benutzt (hat).

- Arbeitspaket

Angaben zu einem Arbeitspaket sind über eingetragene Mitgliedschaften der jeweiligen Person ableitbar. Bei Mitgliedschaft in mehreren APs soll eine Auswahlliste angeboten werden, über die das zum Arbeitskontext passende AP zugeordnet wird.

- Platzhalter für einen VFU Projekttitel, z. B. VFU soeb 3

Vom Archiv erzeugte administrative Angaben (AA)

- Speicherdatum (Objekt Studie, Datennutzung oder Syntax)
- System-ID (dito)
- Angaben im Kontext der Editorfunktion „Publizieren“ (dito)

Eine Editorfunktion mit Schalter Ja / Nein. Durch diese Funktion soll vermieden werden, dass unvollständige bzw. unverständliche Metadaten veröffentlicht werden, z. B. weil die Bearbeitung unterbrochen wurde.

Ein entsprechendes Flag zeigt der Person an, ob der aktuelle Objektzustand bereits in der VFU publiziert wurde.

- Angaben im Kontext der Editorfunktion „Duplizieren“ (dito)

Eine Editorfunktion zum Duplizieren des kompletten Metadatensatz eines Objektes unter einem neuem Objektitel, z. B. um (einzelne) Inhalte im neuen Objekt weiterzuverwenden, während andere Inhalte verändert werden sollen.

Das neue Objekt ist über folgende Metadaten mit dem Ursprungsobjekt verknüpft:

- Titel des ursprünglichen Objekts Studie, Datennutzung oder Syntax
- System-ID des ursprünglichen Objekts
- Ref Objekt-ID.

Durch den Verweis wird das neue Objekt mit dem Ursprungsobjekt navigierbar verknüpft. Über den Link werden alle Metadaten des Ursprungsobjektes angezeigt.

- Angaben im Kontext der Editorfunktion „Berechtigungen“ (dito)
 - Freigabedatum
Datum, an dem die Freigabe erteilt wurde

- Angaben im Kontext der Verknüpfung einer Datei mit dem Objekt Studie, Datennutzung oder Syntax
 - Objekt <STB, DN oder SX> Titel (Auswahlliste)
Zuordnung einer im VFU Archiv gespeicherten Datei zum einem im VFU Archiv gespeicherten Titel eines Referenzobjektes Studienbeschreibung (STB), Datennutzung (DN) oder Syntax (SX).
 - Ref Parent-ID
Referenz, mit der die gespeicherte Datei mit der System-ID des Referenzobjektes verknüpft wird.

Die referenzierten Dateien werden zusammen dem Referenzobjekt im jeweiligen Bearbeitungskontext der Editoren als „zugehörige Dateien“ angezeigt.
- Weitere Angaben zur verknüpften Datei
 - Dateigröße
 - Speicherdatum
 - Fingerprint
 - Freigabedatum

Empfohlene (E) und optionale Elemente

Diese Unterscheidung betrifft insbesondere das Metadatenschema Studie. Um die genutzten Ausgangsdaten transparent und nachvollziehbar auf Studienebene zu dokumentieren, werden bestimmte Einzelelemente oder Gruppen von Einzelelementen zur Erfassung empfohlen. Je nach Bedarf, Kontext und Nachvollziehbarkeit der Daten können außerdem weitere Elemente dokumentiert werden, z. B. durch Verweise auf vom Datengeber bereitgestellte Materialien (Datendokumentationen, Methodenreports, Literatur zu den Daten).

Demgegenüber wird als Regel empfohlen, zu allen Metadatenelemente der Objekte (und Dateien) Datennutzung und Syntax (schrittweise) Angaben zu machen, um die Erschließung und kollaborative Nutzung dieser Objekte und ihrer Bestandteile im Zusammenhang mit den Datenauswertungen durch die wissenschaftlichen Arbeitspakete zu unterstützen.

3 Das Metadatenschema Studie

3.1 Einführung

Das Schema Studie enthält die Metadatenelemente, mit denen die sog. Ausgangsdaten (Dickman u.a. 2010) dokumentiert werden. Dies sind die Daten, die für Datenanalysen im Projekt VFU soeb 3 genutzt werden und von verschiedensten Datengebern (Datenarchive, Forschungsdatenzentren, Statistische Ämter, Bundesbankdaten, OECD usw.) bereitgestellt werden. Für die soeb 3 Datenanalysen werden insgesamt 50 Datensätze und Datenbanken von über 35 Datengebern genutzt. Zusätzlich sind Updates und neue Daten zu den jeweiligen Datensätzen zu berücksichtigen. Die Metadaten für alle in den 18 wissenschaftlichen Arbeitspaketen genutzten Datensätze sollen in der VFU durchsucht werden können.

Darüber hinaus sollen im Sinne der Sicherung guter wissenschaftlicher Praxis nicht nur Analyseprozeduren und Syntaxen sondern auch die zugrunde liegenden (Ausgangs)Daten transparent und nachvollziehbar dokumentiert werden. Im Interesse einer nachhaltigen Sicherung der VFU soeb 3 Ergebnisse sollen deshalb Datenanalysen und Informationen über Ausgangsdaten zur Nachprüfung und Nachnutzung dokumentiert und dauerhaft gesichert werden.

Der Entstehungskontext von sozialwissenschaftlichen Forschungsdaten sowie deren fachliche Inhalte und methodologische Eigenarten werden in Archivkontexten seit langem durch die sog. Studienbeschreibung des DDI Dokumentationsstandards (DDI Alliance. 2014) beschrieben. Deshalb wird das Metadatenschema der Studienbeschreibung auch als generischer Ansatz zur Dokumentation der soeb 3 Ausgangsdaten eingesetzt. Das DDI 2 kompatible Schema des GESIS Datenbestandskatalogs (Zenk-Möltgen; Habel. 2012) dient als Referenzschema und wurde an den Bedarf der VFU soeb 3 angepasst. Zusätzlich wurden Elemente des da|ra Metadatenschemas (Hausstein u.a. 2012, 2014), z. B. zur Dokumentation von Persistent Identifiern, einbezogen.

Mit Blick auf die Rechercheinteressen in einer VFU und Anforderungen an eine nachhaltige Sicherung datenbezogener Kontextinformationen ist abzuwägen, wie umfangreich und genau Ausgangsdaten dokumentiert werden sollen. So sind einerseits Brüche beim Zugang zu Ausgangsdaten (soweit sie nicht bereits vertraglichen Beschränkungen unterliegen) und / oder deren Dokumentationen zu berücksichtigen, wenn z. B. URLs zu entsprechenden Internetseiten nicht mehr verfügbar sind. Andererseits reduziert sich der Dokumentationsaufwand erheblich, wenn der Datengeber die Ausgangsdaten mit einem Persistent Identifier (PID) versieht, die über die entsprechende URL auf Dauer erreichbar sind.

3.2 Die Metadatenelemente des Schemas Studie

Erforderlich sind die Angaben zu 1.3 Arbeitspaket und 7.1 Titel (M in Spalte FA).

Empfohlen werden Mindestangaben zu den Elementen (vgl. FT / KV in Spalte FA) der Gruppen, die nachvollziehbar Zugang, Herkunft und Kontext des Datensatzes dokumentieren, der die Ausgangsdaten für soeb 3 Analysen enthält. Die betrifft insbesondere die Elemente der folgenden Gruppen:

- Nr. 5 Datengeber: 5.1 Namen, 5.2 Spracheversion Daten, 5.3 URL;
- Nr. 6 Datenzugang: 6.1 Nutzungsbedingung (+ URL), 6.2 Zugangsart;
- Nr. 7 Studie: 7.3 Studie ID (+ Typ), 7.4 Inhalt (+ URL);
- Nr. 9 Datenpublikation: 9.1 Datum, 9.2 Version in Verbindung mit
- Nr. 10 Persistent Identifier: 10.1 PID, 10.3 PID URL, 10.4 Datenzitation

- Nr. 11-14 Methodologie: 11 Referenzzeitraum, zeitliche (12) und räumliche (13) Dimension der Daten, 14 Grundlagen der Datenerhebung (u. a. Grundgesamtheit, Auswahl- und Erhebungsverfahren);
- Nr. 16-18 Datensatz: 16 Struktur, 17 Anonymisierung, 18 Formale Kennzeichen (u. a. Variablenanzahl, Analyseeinheit);
- Nr. 19-22 Datenfile: Struktur der Dateien (+ URL), Dateiformat, Dateinamen, Fingerprint

Um den Erfassungsaufwand insbesondere bei komplexen Datensätzen wie dem Sozio-oekonomischen Panel (SOEP) zu minimieren, soll zusätzlich zum jeweiligen Element Beschreibung soweit wie möglich auf die entsprechende Informationsseite des Datengebers durch das jeweilige URL Element verwiesen werden. Ist ein Persistent Identifier (PID) für den Datensatz vorhanden, ist dieser immer anzugeben und mittels URL auf die Seite beim Datenanbieter (Landing Page) zu verweisen. Damit ist gewährleistet, dass die Ausgangsdaten einfach zitiert werden können und leicht auffindbar sind. Optional können weitere Elemente aus der Gruppe Nr. 8 (Primärforscher / Institution), sowie Angaben zu weiterem Material Nr. 23-26 (Nutzungsbedingung; Dokumente, Literatur) bei Bedarf genutzt werden. So erleichtern Verweise auf vom Datengeber bereitgestellte Datendokumentationen oder Methodenreports die Nachvollziehbarkeit der Daten und fördern ihre Nachnutzung im Zusammenhang mit Re-Analysen von soeb 3 Ergebnissen nach Projektende.

Tabelle 1: Die Metadatenelemente des Schemas Studie (MD-STB)

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
	Administration VFU STB	Angaben zu VFU Mitglied und Arbeitspaket, in dessen Kontext die Studienbeschreibung in der VFU bearbeitet wird und weitere administrative Elemente	
1.	Person	Person, die die Beschreibung der Studie anlegt und bearbeitet	
1.1	Vornamen	Vorname der Person	AP
1.2	Nachnamen	Nachname der Person	AP
1.3	Arbeitspaket	Kontext in dem das Objekt bearbeitet wird <ul style="list-style-type: none"> • Auswahlliste: Namen der Arbeitspakete 	M KV
1.4	Projekttitel	Platzhalter für VFU Projektnamen, z. B. „VFU soeb 3“	AP
2.	Objekt	Angaben zum Objekt „VFU Studienbeschreibung“	
2.1	Titel	Der Titel des gespeicherten Objekts „VFU-STB“ <ul style="list-style-type: none"> • Wird automatisch aus 7.1 Studientitel erzeugt 	AA
2.2	ID	Eindeutige menschenlesbare ID der VFU-Studienbeschreibung <ul style="list-style-type: none"> • Format z. B. sb001, sb002, etc. • Auswahl aus ID Liste Beim duplizieren der VFU STB wird der Studientitel geändert und eine neue STB-ID zugeordnet	KV
2.3	Datum Aktualisierung	Logische Kennzeichnung des Metadatenstatus als (neue) Version durch den User. Er gibt ein Datum, an dem die Angaben zuletzt inhaltlich aktualisiert wurden.	KL

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
		<ul style="list-style-type: none"> Kalenderfunktion: Format JJJJ.MM.TT 	
2.4	Datum Beschreibung	<p>Logische Kennzeichnung des Metadatenstatus als (neue) Version durch den User. Er beschreibt den Stand der Version bzw. die inhaltlichen Änderungen einer neuen Version, z. B.</p> <ul style="list-style-type: none"> Stand der Metadaten beim Datengeber laut Datum der Erfassung der VFU STB neue Version des Datensatzes des Datengebers, die in der VFU benutzt wird 	FT
2.5	Versionsnummer	<p>Angabe einer numerischen Version</p> <ul style="list-style-type: none"> Format je nach Versionierungskonzept 2 oder 3 stellig 	FT
2.6	Speicherdatum	Systemdatum bei Anwendung der Funktion Speichern	AA
2.7	Publizieren	<p>Steuerung der Veröffentlichung der VFU-STB zur Vermeidung bruchstückhafter Informationen z. B. Arbeitsunterbrechungen</p> <ul style="list-style-type: none"> Auswahl: Ja / Nein 	KV
2.8	Objekt-ID	System-ID der angelegten VFU-STB	AA
3.	Duplizieren VFU-STB Ursprung	Angaben zur ursprünglichen VFU-STB, die beim Duplizieren automatisch in das Duplikat übernommen werden.	
3.1	Titel	<p>Titel der ursprünglichen VFU-STB</p> <ul style="list-style-type: none"> Wird automatisch aus 2.1 übernommen 	AA
3.2	Objekt ID	<p>System-ID der ursprünglichen Studienbeschreibung</p> <ul style="list-style-type: none"> Wird automatisch aus 2.8 übernommen 	AA
3.3	Ref Objekt-ID	<p>Verweis auf die Objekt-ID der ursprünglichen VFU-STB. Das Duplikat ist durch die ID mit dem Ursprungsobjekt verknüpft und navigierbar. Es werden alle Metadaten des Ursprungsobjektes angezeigt.</p>	AA
4.	VFU Nutzung Daten	Angaben zur Nutzung der Daten dieser Studie durch die wissenschaftlichen VFU Arbeitspakete und deren Mitglied(er)	
4.1	Arbeitspaket Name	<p>Angabe, welches Arbeitspaket den Datensatz nutzt</p> <ul style="list-style-type: none"> Auswahl aus Liste aller AP (1-n) 	KV
4.2	Arbeitspaket Person	<p>Angabe, welche Person(en) des AP den Datensatz nutzen</p> <ul style="list-style-type: none"> Angabe: „Nachname, Vorname“ (1-n) je AP 	FT
4.3	Vertrag Daten	<p>Angabe, ob ein Vertrag zur Nutzung des genannten Datensatzes notwendig ist</p> <ul style="list-style-type: none"> Filter ja / nein 	KV
4.4	Vertrag Anmerkung	<p>Anmerkungen zu Besonderheiten der Bereitstellung und Nutzung der Daten, z. B. durch Projektleitung bereitgestellt.</p> <ul style="list-style-type: none"> Wenn ein Vertrag zur Datennutzung notwendig ist („ja“ in Feld 4.3), müssen folgende Angaben erfasst werden 	FT

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
4.5	Vertrag Unterzeichner	VFU Mitglied, das den Vertrag (für ein AP) abgeschlossen hat <ul style="list-style-type: none"> • Angabe: „Nachname, Vorname“ (1-n) je Vertrag / Datensatz 	FT
4.6	Vertrag Dauer	Angaben zur Laufzeit des Vertrages	
4.6.1	Vertrag Dauer von	Dauer des Vertrages je Nutzer / AP <ul style="list-style-type: none"> • Datumsfeld von (TT.MM.JJJJ) 	KL
4.6.2	Vertrag Dauer bis	Dauer des Vertrages je Nutzer / AP <ul style="list-style-type: none"> • Datumsfeld bis (TT.MM.JJJJ) 	KL
	Datengeber	Angaben zur Quelle der Studie, der Daten und Bedingungen der Datennutzung	
5.	Datengeber	Administrative Angaben zum Datengeber	
5.1	Namen	Namen der Einrichtung, die die Studie / den Datensatz und Material zur Nutzung im Projekt bereitstellt	FT
5.2	Sprache	Sprachversion der Studie (Datensatz, Materialien), die in VFU genutzt wird: <ul style="list-style-type: none"> • Deutsch: DEU (Alphanumerischer Kode nach ISO 639-2) • Englisch: ENG (Alphanumerisch Kode nach ISO 639-2) • DEU + ENG (bei gemischt sprachlicher Dokumentation) 	KV
5.3	URL	Link zur Website des Datengebers	URI
6.	Datenzugang	Angaben des Datengebers zu Nutzungsbedingungen und Bereitstellung des Datensatzes	
6.1	Nutzungsbedingung	Rechtliche bzw. formale Nutzungsbedingungen des Datensatzes (ohne Kosten) <ul style="list-style-type: none"> • Frei (Download; eventuell unter Berücksichtigung von Copyright o.ä.) • Kontrolliert (Registrieren und i.d.R. explizite Zustimmung zu Nutzungsbedingungen des Datengebers) • Vertrag (vertragliche Regelung in Schriftform) • Andere 	KV
6.1.1	URL	Webseite zu Nutzungsbedingung und Datenbereitstellung	URI
6.2	Zugang	Zugang nach Grad der Anonymisierung und Zugangstyp <ul style="list-style-type: none"> • Absolut (PUF: Download, CD/ DVD, o. ä.) • Faktisch (SUF: Off-Site) • Faktisch (Gastaufenthalt On-Site) • Projektbezogen (Gastaufenthalt On-Site) • Formal (Kontrolliertes Fernrechnen On-Site) Quelle KV: FDZ Statistische Ämter Bund und Länder	KV
6.2.1	URL	Webseite mit speziellen Informationen zum Datenzugang	URI

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
	Bibliographie	Angaben des Datengebers zur Bibliographie	
7.	Studie	Studie (Datensatz), die in einem soeb 3 für Datenanalysen benutzt wird	
7.1	Titel	Titel der Studie / des Datensatzes (laut Datengeber)	M FT
7.2	weitere Titel	Angabe zusätzlicher Titel der Studie	FT
7.2.1	weitere Titel Typ	Art des zusätzlichen Titels, z. B. englischer Originaltitel: <ul style="list-style-type: none"> • Originaltitel • Alternativer Titel • Paralleltitel • Untertitel • Projekttitel Quelle KV: Metadatenschema DBK (Zenk-Möltgen 2012)	KV
7.3	Studie ID	Eindeutige Kennung der Studie	FT
7.3.1	Studie ID Art	Typ der Studien ID: <ul style="list-style-type: none"> • Studiennummer • Datensatznummer • Sonstige 	KV
7.4	Inhalt	Abstract, Kurzbeschreibung der Studie, kodierte Variablen, Demographie, etc. Angaben zur Studie sollten mit den Angaben zum Datensatz abgeglichen werden	FT
7.4.1	URL	Link zur Seite des Datengebers mit weiteren Informationen zum Inhalt der Studie	URI
8.	Primärforscher/in bzw. Institution	Name der Person und / oder Institution, die als Autor für die Studie bzw. die Daten verantwortlich ist (sind) <ul style="list-style-type: none"> • 8.1 und/oder 8.2 sollten angegeben werden • Zur Datenerhebung vgl. Elemente 14 und weitere 	
8.1	PF Name	Name der Person (Primärforscher/in) <ul style="list-style-type: none"> • Je nach Konzept 1-n Felder für 8.1.1 und 8.1.2 oder • Freitextfeld mit Eingabe Nachname, Vorname je Zeile 	
8.1.1	PF Vorname	Karl	FT
8.1.2	PF Nachnamen	Maier	FT
8.2	Institution	Name der Institution, der die Person zugeordnet oder die für Studie / Daten verantwortlich ist <ul style="list-style-type: none"> • Institut für vergleichende Sozialforschung; Deutsche Bundesbank, Genesis Datenbank ... 	FT

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
9.	Datenpublikation	Angaben des Datengeber zu Publikation, Version, PID, Zitation des Datensatzes	
9.1	Publikation Datum	Datum, an dem der Datenbieter den Datensatz veröffentlicht hat <ul style="list-style-type: none"> • Format JJJJ.MM.TT 	KL
9.2	Publikation Version	Versionsnummer des Datensatzes zum Zeitpunkt der Publikation durch den Datengeber	FT
9.3	Versionsgrund	Beschreibung des Grundes für die neue Version des Datensatzes	FT
10.	Persistent Identifier	Persistent Identifier des Datensatzes	
10.1	PID	z. B. DOI: 10.5684/soep.v28 für alle Daten der SOEP Welle 28.	FT
10.2	PID Art	Art des Persistent Identifier (PID): <ul style="list-style-type: none"> • ARK • DOI • EAN13 • EISSN • Handle • ISBN • ISSN • ISTC • LISSN • LSID • PURL • UPC • URL • URN <p>Quelle KV: da ra Metadatenschema 3.0 (Tabelle 3.1.10) (Hausstein u.a. 2014)</p>	KV
10.3	PID URL	Verweis auf Zielseite zur PID beim Datengebers (Landing Page) <ul style="list-style-type: none"> • z. B. SOEPv28.1: http://dx.doi.org/10.5684/soep.v28.1 	URI
10.4	Datenzitation URL	Vorgabe zur Zitation der Forschungsdaten dieser Studie (Regeln der Datenzitation) – Beispiel SOEP: http://www.diw.de/deutsch/soep/faq/32014.html#274746	URI
	Methodologie	Angaben des Datengebers zum methodischen Kontext / Design der Datenerzeugung	
11.	Referenzzeitraum	Zeitraum den die Daten inhaltlich abdecken (Feldzeit Umfragen / Referenzzeitraum Statistiken)	
11.1.1	Zeitraum Von	Angaben des Zeitraums als Spanne: Von <ul style="list-style-type: none"> • Format JJJJ.MM.TT 	KL

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
11.1.2	Zeitraum Bis	Angaben des Zeitraums als Spanne: Bis <ul style="list-style-type: none"> • Format JJJJ.MM.TT 	KL
11.2	Zeitraum (von - bis) frei	Möglichkeit, einen Zeitraum anzugeben, wenn dieser nicht im Kalendermodus angegeben werden kann. <ul style="list-style-type: none"> • Jahreszeiten (z.B. Herbst 1989) 	FT
12.	Dimension Zeit	Zeitlichen Ausdehnung und Häufigkeit der Datengewinnung	
12.1	Zeitliche Dimension (kontrolliert)	Zeitliche Ausdehnung der Datenerhebung – Typisiert <ul style="list-style-type: none"> • Längsschnitt, Trend, Kohorte, Zeitreihe • Längsschnitt: Panel • Querschnitt • Andere Quelle KV: da ra Metadatenschema 3.0 (Tabelle 3.1.5)) auf Grundlage der DDI Empfehlungen (Hausstein u.a. 2014)	KV
12.2	Zeitliche Dimension (frei)	Möglichkeit, die zeitliche Ausdehnung zu beschreiben, wenn KV keine passenden Begriffe enthält	FT
12.3	Frequenz	Häufigkeit der Datenerhebung <ul style="list-style-type: none"> • monatlich, vierteljährlich, jährlich, o. ä. 	FT
13.	Dimension Raum	Beschreibt die geographische Einheit, die der Stichprobe (Auswahl) zugrunde liegt	
13.1	Geographischer Raum (kontrolliert)	Räumliche Ausdehnung im Design der Datenerhebung (Namen der geografischen Einheiten) <ul style="list-style-type: none"> • Namen von Staaten gemäß ISO 3166-1 z. B. DE • ISO 3166-2 für deren Regionen z. B. DE-BY (Bayern) 	KV
13.2	Geographischer Raum (frei)	Angabe von Gebietseinheiten (innerhalb höherer ISO-Einheiten 12.1) oder von nicht standardisierter Einheiten: <ul style="list-style-type: none"> • räumlicher Gebietseinheiten z. B. Raumordnungsregionen (ROR), Mikrozensuskreisregionen (MZKR), • oder Norddeutschland bzw. näherer Spezifizierungen (BRD ohne Berlin/W.) 	FT
14	Datenerhebung Grundlagen	Methodische Grundlagen und Verfahren der Datenerhebung	
14.1	Grundgesamtheit	Beschreibung der (statistischen) Einheiten, die der Auswahl zu Grunde liegen <ul style="list-style-type: none"> • z.B. Privathaushalte mit Personen, die zum Befragungszeitpunkt das 18. Lebensjahr vollendet hatten 	FT
14.2	Auswahlverfahren	Beschreibung des Auswahlverfahrens zur Erstellung der Stichprobe (Stichprobenverfahren) <ul style="list-style-type: none"> • z. B. 1980 bis 1992 und 1998 mehrstufig geschichtete Zufallsauswahl aus allen Privathaushalten mit Personen, die 	FT

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
		zum Befragungszeitpunkt das 18. Lebensjahr vollendet hatten (ADM-Stichprobendesign)	
14.3	Erhebungsverfahren (kontrolliert)	Beschreibung der Methode, mit der die Daten erhoben wurde (Typen) <ul style="list-style-type: none"> • Quantitatives Interview (Fragebogen) • Qualitatives Interview (Leitfaden) • Prozessgenerierte Daten • Beobachtung • Experiment • Inhaltsanalyse Quelle KV: da ra Metadatenschema 3.0 (Tabelle 3.1.7) auf Grundlage der DDI Empfehlungen (Hausstein u.a. 2014)	KV
14.4	Erhebungsverfahren (frei)	Freie Beschreibung des Erhebungsverfahrens, wenn das KV keine passender Begriff gefunden wurde	FT
15.	Datenerhebung Durchführung	Name der Institution / Person, die die Daten erhoben hat. u. a. Feldarbeit Erhebungsinstitute Einträge sind in 15.1.x und / oder in 15.2 möglich <ul style="list-style-type: none"> • Einrichtungen wie Statistische Ämter und Deutsche Bundesbank, die die Daten aufbereiten und anbieten sollten als Institution weiter oben unter Primärforscher/in und Institution eingetragen werden 	
15.1	Person	Name der Person, die die Daten erhoben hat	FT
15.1.1	Vorname	Elke	FT
15.1.2	Nachnamen	Maier	FT
15.2	Institution	Name der Institution, die die Daten erhoben hat	FT
	Datensatz	Angaben des Datengebers zu inhaltlichen, formalen und technische Aspekte des Datensatzes Inhaltliche Angaben zum Datensatz sollten mit Angaben zu Studie Inhalt 7.4 abgeglichen werden	
16.	Datensatz Logisch	Logische Aspekte des Datensatzes (Inhalt, Anonymisierung, Struktur, etc.) – Vorgehen: <ul style="list-style-type: none"> • Komplexe Daten: Beschreibung unter „Datensatz Struktur“ mit Verweis auf die Datengeberseite und / oder • Kurzbeschreibung in den einzelnen Abschnitten zu 19.x (Datensatz Technisch) 	
16.1	Struktur	Allgemeine Beschreibung, insbesondere bei komplexen Datenstrukturen (hierarchisch, mehrdimensional), mit Hinweisen auf Dokumentationen zum Verständnis des Datensatzes. <ul style="list-style-type: none"> • Z. B. „Das SOEP verfügt über eine detaillierte Dokumentation der Erhebung und der generierten Daten“ 	FT

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
16.2	Struktur URL	z. B. SOEB.v28 - Online - Dokumentation und Neuerungen	URI
17	Anonymisierung	Angaben des Datengebers zur Anonymisierung der Daten	
17.1	Grad	Typ des Anonymisierung (und Datenzugangstyp) <ul style="list-style-type: none"> • Absolut (PUF) • Faktisch (SUF) • Faktisch (On-Site) • Formal (On-Site) 	KV
17.2	Grad Beschreibung	Beschreibung zur Anonymisierung des Datensatzes	FT
18	Datensatz Formal	Beschreibung formaler Eigenschaften des Datensatzes	
18.1	Variablen Anzahl	Anzahl der Variablen im Datensatz (unabhängig von der Aufteilung auf Einzeldateien)	FT
18.2	Analyseeinheit Typ	Typ der Einheiten, zu denen der Datensatz Aussagen trifft: <ul style="list-style-type: none"> • Individuum, • Organisation • Familie/im selben Haushalt • Haushalt/ Wohneinheit • Ereignis/ Prozess • Geographische Einheit • Zeiteinheit • Texteinheit • Gruppe • Objekt • Sonstiges Quelle KV: da ra Metadatenschema 3.0 (Tabelle 3.1.8) auf Grundlage der DDI Empfehlungen (Hausstein u.a. 2014)	KV
18.3	Analyseeinheit Anzahl Typ	Anzahl der Einheiten je Objekttyp (unabhängig von der Aufteilung auf Einzeldateien), <ul style="list-style-type: none"> • z. B. Anzahl der Befragten (Fallzahlen) 	FT
18.4	Daten Erhebungstyp	Formaler Typ der Datenerhebung: <ul style="list-style-type: none"> • Befragungsdaten • Prozessdaten • Statistiken • Indikatoren • Andere 	KV
19.	Datensatz Technisch	Beschreibung der technischen Eigenschaften des Datensatzes, der aus 1-n Datenfiles besteht. Eine große Anzahl Datenfiles kann im Element „Datenfile Struktur“ mit URL erfasst werden.	

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
19.1	Datenfile Anzahl	Anzahl der Dateien (Datenfile je Format), aus denen der Datensatz besteht	FT
20.0	Datenfile Struktur	Beschreibung zur Struktur der Datenfiles bei komplexen Datensätzen	
20.1	Beschreibung	Systematische Übersicht, z.B. zu Dateinamen, Dateiformat, Größenangabe, Sprachversion je Datenfile	FT
20.1.1	URL	Verweis auf die Quelle beim Datengeber	URI
21.	Einzeldatei	Beschreibung eines einzelnen Datenfiles (0-1)	
21.1	Datei Namen	Dateinamen je Datenfile (1-n)	
21.2	Datei Größe	Speichergröße des Objekts Nr. 21.1	
21.3	Datenfile Format	Technisches Format der Datenfile(s) eines Datensatzes <ul style="list-style-type: none"> • Dateiformat (i. d. R Quantitative Daten) <ul style="list-style-type: none"> • RData • Sav • Dta • Xls • Csv • Sas • Dateiformate (Qualitative Daten) <ul style="list-style-type: none"> • Audio • Video • Dateiformate (Text) <ul style="list-style-type: none"> • Text (txt, doc, pdf, etc.) • Dateiformate (Andere) <ul style="list-style-type: none"> • Andere 	KV
22	Datenfile Fingerprint	Beschreibung der Authentizität eines digitalen Objekts (Einzeldatei) durch eine Prüfsumme	
22.1	Fingerprint Objekt	Prüfsumme eines einzelnen Objekts Nr. 21.1 (optional)	FT
22.2	Fingerprint Liste Objekte	Übersicht der Dateien und deren Prüfsummen, z. B. zu soep.v28 Abschnitt „MD5 Fingerprints der einzelnen Dateien“	FT
22.2.1	Liste URL	Beispiel SOEP: Stata deutsch TXT, 16.16 KB	URI
22.3	Fingerprint Verfahren	Technisches Verfahren, mit dem die Prüfsumme gebildet wird (ggf. mit Angabe des Verzeichnisses): <ul style="list-style-type: none"> • UNF • MD5 • SHA • RIPEMD-160 • Tiger 	KV

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
		<ul style="list-style-type: none"> • HAVAL • Whirlpool • LM-HASH • NTLM 	
	Material	Dokumentationen des Datengebers zur Studie und den Daten	
23	Nutzungsbedingung	Angaben zur Nutzung und Zitation des Materials (ohne Daten)	
23.1	Nutzungsrecht Sonstige	<p>Nutzungsrechte der Webseiteninhalte und Zitierregeln für Materialien zu einer Studie (z. B. Methodenbericht):</p> <p>Deutsch</p> <ul style="list-style-type: none"> • Copyright • Rechtshinweise • Urheberrechte • Impressum <p>Englisch</p> <ul style="list-style-type: none"> • Disclaimer • Terms of Condition • Condition of use 	KV
23.2	Nutzungsrechte URL	Verweis auf die Quelle beim Datengeber	URI
24	Dokumente	Angaben zu Dokumenten, z. B. Datendokumentation, Methodenbericht, Messinstrument etc.	
24.1	Dokumente Übersicht	Übergreifende Beschreibung von Dokumentationen, insbesondere bei komplexen Datenstrukturen eines Datensatzes.	FT
24.1.1	Übersicht URL	Verweis auf die Quelle beim Datengeber	URI
25.	Dokumente Einzelbeschreibung	Angaben zu einzelnen Dokumenten / Dokumentationen z. B. weil die Dokumente (0-n) aus Gründen der Transparenz / Nachvollziehbarkeit erfasst werden soll	
25.1	Titel	Angabe des Dokumenttitels (laut Datengeber)	FT
25.2	Dokument Typ	<p>Typ des Dokuments - Auswahlliste (KV)</p> <ul style="list-style-type: none"> • Studienbeschreibung / Kurzbeschreibung Daten • Datensatzbeschreibung • Erhebungsdesign und -Instrumente • Fragebogen • Methodenbericht / Methodenbeschreibung • Codebuch / Datenhandbuch / Datenreport • Codeplan • Schlüsselverzeichnis • Variablenliste • Andere 	KV

Nr.	Elementname	Definition und Anmerkungen (MD-STB)	FA
25.3	Dokument Inhalt	Nähere Angaben zum Dokumentinhalt	FT
25.3.1	Dokument URL	Verweis auf die Quelle beim Datengeber	URI
	Literatur	Angaben zu Literatur, die sich auf diese Studie / Daten bezieht	
26	Veröffentlichungen	Bibliographische Literaturangabe(n)	FT
26.1	URL	Verweis auf die Quelle beim Datengeber	URI

3.3 Metadatenelemente zum Upload von Dateien zum Schema Studie

Erforderlich Angaben sind 1.3 Arbeitspaket, 2.1 Referenz, 3.1 Titel und 3.4 Dateiname (M Spalte FA). Empfohlen werden Angaben zu allen weiteren editierbaren Elementen (FT / KV in Spalte FA).

Tabelle 2: Metadatenelemente zum Speichern einer Datei im VFU Archiv (soweit rechtlich zulässig) und Referenzierung mit der Beschreibung zur Studie

Nr.	Elementname	Definition und Anmerkungen (STB-Upload)	FA
	Administration STB-Upload	Angaben zu VFU Mitglied und Arbeitspaket, in dessen Kontext der Upload von Dateien in das VFU Archiv erfolgt	
1.	Person	Mitglied der VFU, das die Datei im VFU Archiv speichert	
1.1	Vornamen	Vorname der Person	AP
1.2	Nachname	Nachname der Person	AP
1.3	Arbeitspaket	Kontext in dem das Objekt bearbeitet wird <ul style="list-style-type: none"> Auswahlliste: Namen der Arbeitspakete 	M KV
	Datei	Angaben zur Datei, die gespeichert und mit einer Studienbeschreibung (VFU STB) verknüpft wird.	
2	Datei Typ	Typ der Datei, die im Archiv gespeichert wird: Auswahlliste (KV) <ul style="list-style-type: none"> Datenfile Dokument 	KV
2.1	Referenz Titel VFU STB	Verknüpfung der Datei mit dem Titel einer Studienbeschreibung <ul style="list-style-type: none"> Auswahlliste: Titel der VFU STB (MD-STB Nr. 2.1) 	M KV
2.2	Ref Parent-ID	Verknüpfung der Datei mit der System-ID der STB	AA
3	Eigenschaft	Weitere Angaben zur gespeicherten Datei	
3.1	Titel	Angabe eines Titel zur gespeicherten Datei (laut Datengeber)	M FT
3.2	Typ	Spezifizierung des Objekttyps gemäß der Angabe zu Nr. 2 Angaben erfolgen zu entweder 3.2.1 (Datenfile) oder zu 3.2.2 (Dokument)	
3.2.1	Typ Datenfile	Typ des Datenfiles - Auswahlliste (KV)	KV

Nr.	Elementname	Definition und Anmerkungen (STB-Upload)	FA
		<ul style="list-style-type: none"> • Dateifile ist Teil eines Datensatzes • Einzelner Datenfile 	
3.2.2	Typ Dokument	Typ des Dokuments - Auswahlliste (KV) <ul style="list-style-type: none"> • Studienbeschreibung / Kurzbeschreibung Daten • Datensatzbeschreibung • Erhebungsdesign und -Instrumente • Fragebogen • Methodenbericht / Methodenbeschreibung • Codebuch / Datenhandbuch / Datenreport • Codeplan • Schlüsselverzeichnis • Variablenliste • Andere 	KV
3.3	Inhalt	Beschreibung zur gespeicherten Datei	FT
3.4	Dateiname	Vollständiger Dateiname laut Datengeber	M FT
3.4.1	Dateigröße	Angaben in GB / Mbyte/Kbyte	AA
3.4.2	Speicherdatum	Datum an dem die Datei im Archiv gespeichert wurde	AA
3.4.3	Fingerprint	Fingerprint des Datengebers (wenn verfügbar)	FT
3.5	Freigabe Nutzungsbedingung	Nutzungsbedingungen der Datei in der VFU gemäß <ul style="list-style-type: none"> • STB Nr. 23.1 Nutzungsrechte laut Datengeber 	FT
3.6	Freigabe	Zuweisung definierter VFU-Rechte je Person / je Arbeitspaket <ul style="list-style-type: none"> • Laut VFU Rollen- und Rechtmanagement 	KV
3.6.1	Freigabedatum	Datum, an dem die Freigabe erteilt wurde. Angabe erfolgt automatisch nach Speichern von 3.6	AA

4 Das Metadatenchema Datennutzung

4.1 Einführung

Mit dem Schema Datennutzung werden Metadaten zur Nutzung der Ausgangsdaten, insbesondere hinsichtlich des Untersuchungsgegenstandes und methodischer Parameter von geplanten Datenanalysen im Projekt VFU soeb 3, beschrieben. Das Schema Datennutzung wird formal vom Metadatenchema Syntax abgegrenzt. Mit der Datennutzung werden Informationen zur Planung von Analysen erfasst, während die Metadaten zur Syntax konkret ausgeführte Datenanalysen dokumentieren (vgl. Kapitel 5).

Angaben zur Datennutzung sollen es den Forschenden der VFU soeb 3 ermöglichen, mit Hilfe von Recherchen z. B. zu explorieren, welche Fragestellungen und Untersuchungsgegenstände mit welchen Daten und Parametern untersucht werden. Damit können etwa folgende Aspekte kollaborativer Datenanalysen durch Recherchen in den Metadaten systematisch untersucht und beantwortet werden:

- Für welche Untersuchungsgegenstände und -ziele, wird der Datensatz einer Studie genutzt?
- Welche Datensatzart wird genutzt z.B. Quer-, Längsschnittfiles, Jahr / Welle
- Für welchen Zeitraum? Für welches Gebiet? Für welche Analyseeinheit?
- Für welche kleinste Gebietseinheit werden die Daten analysiert, z.B. Kreisebene, ROR?
- Welche Methoden wendet der / die Nutzer/in bzw. das Arbeitspaket dabei an?
- Welche Datensätze werden aus den Ausgangsdaten generiert?

Dateien, z. B. gematchte Daten oder Planungsmemos, können mit der Datennutzung verknüpft werden. Datenfiles können z. B. aus Datenbanken wie Genesis (Statistisches Bundesamt) zusammengestellt oder aus anderen Ausgangsdaten für Analysen vorbereitet werden. Solche Daten sind logisch der Datennutzung zugeordnet, während deren Quelle durch die Studienbeschreibung dokumentiert wird.

4.2 Die Metadatenelemente des Schemas Datennutzung

Erforderlich sind Angaben zu 1.3 Arbeitspaket, 4.1 Titel und 5.1 Referenz (M in Spalte FA). Empfohlen werden Angaben zu allen weiteren editierbaren Elementen (FT / KV in Spalte FA).

Tabelle 3: Metadatenelemente des Schemas Datennutzung (MD-DN)

Nr.	Elementname	Definition und Anmerkungen (MD-DN)	FA
	Administration VFU DN	Angaben zu VFU Mitglied und Arbeitspaket, in dessen Kontext die Beschreibung der Datennutzung in der VFU bearbeitet wird und weitere administrative Elemente	
1	Person	Person, die die Datennutzung anlegt und bearbeitet	
1.1	Vornamen	Vorname der Person	AP
1.2	Nachname	Nachname der Person	AP
1.3	Arbeitspaket	Kontext in dem das Objekt bearbeitet wird	M

Nr.	Elementname	Definition und Anmerkungen (MD-DN)	FA
		<ul style="list-style-type: none"> Auswahlliste: Namen der Arbeitspakete 	KV
2	Objekt Datennutzung	Angaben zum Objekt „VFU Datennutzung“	
2.1	Titel	Der Titel des gespeicherten Objekts „VFU- Datennutzung“ <ul style="list-style-type: none"> wird automatisch aus Feld 4.1 Titel erzeugt 	AA
2.2	ID	Eindeutige menschenlesbare ID der Datennutzung <ul style="list-style-type: none"> Format z. B. dn001, dn002, etc. Auswahl aus ID Liste Beim duplizieren der Datennutzung wird der Titel geändert und eine neue ID zugeordnet	KV
2.3	Datum Aktualisierung	Logische Kennzeichnung des Metadatenstatus als (neue) Version durch den User. Er gibt ein Datum, an dem die Angaben zuletzt inhaltlich aktualisiert wurden. <ul style="list-style-type: none"> Kalenderfunktion: Format TT.MM.JJJJ 	KL
2.4	Datum Beschreibung	Logische Kennzeichnung des Metadatenstatus als (neue) Version durch den User. Er beschreibt den Stand der Version bzw. die inhaltlichen Änderungen einer neuen Version	FT
2.5	Versionsnummer	Angabe einer numerischen Version <ul style="list-style-type: none"> Format je nach Versionierungskonzept 2 oder 3 stellig 	FT
2.6	Speicherdatum	Systemdatum bei Anwendung der Funktion Speichern	AA
2.7	Publizieren	Steuerung der Veröffentlichung einer Datennutzung zur Vermeidung bruchstückhafter Informationen z. B. beim Speichern wegen Arbeitsunterbrechungen. <ul style="list-style-type: none"> Auswahl: Ja / Nein 	KV
2.8	Objekt-ID	System-ID der angelegten Beschreibung einer Datennutzung	AA
3	Duplizieren DN Ursprung	Angaben zur ursprünglichen Datennutzung, die automatisch in ein Duplikat übernommen werden, wenn im Editor der Befehl Duplizieren gewählt wird	
3.1	Titel	Titel der ursprünglichen Datennutzung <ul style="list-style-type: none"> Wird automatisch aus 2.1 übernommen 	AA
3.2	Objekt ID	System-ID der duplizierten Datennutzung (Ursprungsobjekt) <ul style="list-style-type: none"> Wird automatisch aus 2.8 übernommen 	AA
3.3	Ref Objekt-ID	Link auf Objekt-ID des duplizierten Ursprungsobjekts. Das Duplikat ist über diese ID mit dem Ursprungsobjekt verknüpft und navigierbar (Link). Es werden alle Metadaten des Ursprungsobjektes angezeigt.	AA
	Gegenstand	Angaben zu Gegenstand der Untersuchung, genutzte Studie / Datensatz und Methoden Aspekte der Untersuchung	

Nr.	Elementname	Definition und Anmerkungen (MD-DN)	FA
4	Thema	Beschreibung eines Untersuchungsgegenstand /-ziels	
4.1	Titel	Thema des Untersuchungsgegenstands	M FT
4.2	Inhalt	Beschreibung des Untersuchungsgegenstands, für die die Studie bzw. der Datensatz benutzt wird	FT
5	Ausgangsdaten	Beschreibung der spezifischen Daten einer Studie, die für Analyse genutzt werden (soll)	
5.1	Referenz Studie	Auswahl der Studie mit Beschreibung der Ausgangsdaten, die für die Analysen genutzt werden (VFU STB) <ul style="list-style-type: none"> • Auswahlliste: Titel VFU STB (MD-STB Nr. 2.1) 	M AA
5.2	Datensatz Typ	Typ des für Analysen benutzten (Sub-) Datensatzes (aus 1-n Datenfiles) der Studie (ohne temporär erzeugte Datensätze, die im Schema Syntax dokumentiert werden können) <ul style="list-style-type: none"> • Subdatensatz (aus Einzeldatenfile) • Einzeldatensatz (original Datenfile) • Selbst generierter Datensatz (Daten aus Datenbank, etc.) • Gematchter Datenfile (vorab aus mehreren Datenfiles erzeugt) 	KV
5.3	Beschreibung Datensatz	Beschreibung des gewählten Datensatzes (1-n Datenfiles) und seine spezifische Nutzung für geplante Analysen <ul style="list-style-type: none"> • z. B. Subset Beschreibung, Datensatzstruktur, Beschreibung selbstgenerierter DS, Spezifische Formatanwendung (Long-Format; Wide Format); Nutzung von Spelldaten etc. 	FT
6	Methode	Methodische Aspekte, hinsichtlich der der Gegenstand untersucht wird	
6.1	Zeitraum	Beschreibung für welchen Zeitraum (Referenzzeitraum) die Daten genutzt werden (sollen)	FT
6.2	geografischer Raum	Beschreibung für welches Gebiet (Land / Untereinheit) die Daten genutzt werden (sollen) <ul style="list-style-type: none"> • Namen (nicht Code) gemäß ISO 3166-1 für Staaten und ISO 3166-2 für deren Regionen 	FT
6.3	Gebietseinheit	Beschreibung der kleinsten Gebietseinheit (z. B. Kreisebene ROR) für die Daten genutzt werden <ul style="list-style-type: none"> • Administrative Gebietseinheit <ul style="list-style-type: none"> • Regierungsbezirk • Region • Kreis • Gemeinde • Statistische Gebietseinheit <ul style="list-style-type: none"> • ROR • BIK 	KV

Nr.	Elementname	Definition und Anmerkungen (MD-DN)	FA
		<ul style="list-style-type: none"> • Weitere Gebietseinheit <ul style="list-style-type: none"> • PLZ • Finanzamtsbezirk • Gerichtsbezirk • Arbeitsagenturbezirk • Bundestagswahlkreis • Landtagswahlkreis • Bundestagsbezirk • Kammerbezirk HWK • Kammerbezirk IHK • Andere 	
6.4	Analyseeinheit	Beschreibung für welche Analyseeinheit(en), die Daten genutzt werden	FT
6.5	Methoden	Beschreibung welche Methoden bei der Untersuchung / Datenanalyse angewendet werden	FT
7	Datenanalyse	Konzeptuelle Beschreibung der geplanten Datenanalysen bezüglich des Untersuchungsgegenstandes	
7.1	Analysekonzept	Spezifizierung der Datenanalyse auf Konzeptebene, z. B. Datenanalyseplan. Die Operationalisierung des Konzepts erfolgt durch die Syntaxen, die mit der Beschreibung dieser Datennutzung und dem dokumentierten Untersuchungsgegenstand verknüpft werden	FT

4.3 Metadatenelemente zum Speichern von Dateien zu einer Datennutzung

Erforderlich sind Angaben zu 1.3 Arbeitspaket, 2.1 Referenz, 3.1 Titel und 3.4 Dateiname (M Spalte FA).
Empfohlen werden Angaben zu allen weiteren editierbaren Elementen (FT / KV in Spalte FA).

Tabelle 4: Metadatenelemente zum Speichern einer Datei im VFU Archiv (soweit rechtlich zulässig) und Referenzierung mit der Beschreibung der Datennutzung (DN-Upload)

Nr.	Elementname	Definition und Anmerkungen (DN-Upload)	FA
	Administration DN-Upload	Angaben zu VFU Mitglied und Arbeitspaket, in dessen Kontext der Upload von Dateien in das VFU Archiv erfolgt	
1	Person	Mitglied der VFU, das die Datei im VFU Archiv speichert	
1.1	Nachnamen	Vorname der Person	AP
1.2	Vornamen	Nachname der Person	AP
1.3	Arbeitspaket	Arbeitspaket in dessen Kontext das Mitglied die Datei speichert <ul style="list-style-type: none"> • Auswahlliste: Namen der Arbeitspakete 	M KV

Nr.	Elementname	Definition und Anmerkungen (DN-Upload)	FA
	Datei	Angaben zur Datei, die gespeichert und mit einer Datennutzung (DN) verknüpft wird.	
2	Datei Typ	Typ der Datei, die im Archiv gespeichert wird: Auswahlliste (KV) <ul style="list-style-type: none"> • Datenfile • Dokument 	KV
2.1	Referenz Titel Datennutzung	Titel des beschriebenen Untersuchungsgegenstands / Ziels, mit der die gespeicherte Datei verknüpft wird <ul style="list-style-type: none"> • Auswahlliste: Titel Datennutzung (MD-DN Nr. 2.1) 	M KV
2.2	Ref Parent-ID	Verknüpfung der Datei mit der System-ID der gewählten DN	AA
3	Eigenschaft	Weitere Angaben zur gespeicherten Datei	
3.1	Titel	Angabe eines Titel der Datei	M FT
3.2	Typ	Spezifizierung des Objekttyps gemäß der Angabe zu Nr. 2 Angaben erfolgen entweder zu 3.2.1 (Datenfile) oder zu 3.2.2 (Dokument)	
3.2.1	Typ Datenfile	Typ des Datenfiles - Auswahlliste (KV) <ul style="list-style-type: none"> • Subdatensatz (aus Einzeldatenfile • Einzeldatensatz (original Datenfile) • Selbst generierter Datensatz (Daten aus Datenbank, etc.) • Gematchter Datenfile (vorab aus Datenfiles erzeugt) • Sonstiger Datenfile 	KV
3.2.2	Typ Dokument	Typ des Dokuments - Auswahlliste (KV) <ul style="list-style-type: none"> • Beschreibung eines selbstgenerierten Datensatzes • Beschreibung eines einzelnen Untersuchungsgegenstands • Beschreibung der Analyseplanung • Sonstiges Dokument 	KV
3.3	Inhalt	Beschreibung zur gespeicherten Datei	FT
3.4	Dateiname	Vollständiger Dateinamen	M FT
3.4.1	Dateigröße	Angaben in GB / Mbyte/Kbyte	AA
3.4.2	Speicherdatum	Datum an dem die Datei im Archiv gespeichert wurde	AA
3.4.3	Fingerprint	Fingerprint der gespeicherten Datei (wenn implementiert)	
3.5	Freigabe Nutzungsbedingung	Bedingung zur Nutzung der Datei nach Freigabe für weitere(s) VFU Mitglied(er) und / oder VFU AP(s)	FT
3.6	Freigabe	Zuweisung definierter VFU-Rechte je Person / je Arbeitspaket <ul style="list-style-type: none"> • Laut VFU Rollen- und Rechtemanagement 	KV
3.6.1	Freigabedatum	Datum, an dem die Freigabe erteilt wurde. Angabe erfolgt automatisch nach Speichern von 3.6	AA

5 Das Metadatenchema Syntax

5.1 Einführung

Das Metadatenchema Syntax enthält Elemente zur Beschreibung von Aufgaben und Inhalten der für die Durchführung von Datenanalysen eingesetzten Syntaxdateien. So lässt sich die Anwendung einer Syntaxdatei im Arbeitsverlauf einer Datenanalyse grob nach den Kategorien Datenselektion, Datenaufbereitung, Datenverknüpfung und Datenauswertung unterscheiden (vgl. Dickmann u.a. 2010: 9).

Weiterhin dokumentieren die Metadatenelemente in diesem Schema das Erstellen und Bearbeiten von Syntaxinhalten und deren Versionen. Die Dokumentation von Auswertungssyntax in der VFU dient damit auch sowohl der verbundinternen Qualitätssicherung als auch der Zusammenarbeit bei der Erstellung und Entwicklung von Syntaxdateien im Zuge der Datenauswertung. Durch ein entsprechendes Rechtemanagement können die Syntaxdateien mit Forschenden und / oder wissenschaftlichen Arbeitspaketen des soeb 3 Projektes geteilt werden. Beispielsweise stellt das Arbeitspaket 11 Syntaxbausteine oder beschreibende Metadaten bereit, damit Haushaltsvariablen möglichst harmonisiert generiert werden können. Zur Sicherung der Herkunft von Syntaxen und Syntaxbausteinen können Angaben zum Urheber und eventuelle Zitationsvorgaben im entsprechenden Metadatum „Nutzungsbedingungen“ dokumentiert werden.

Schließlich können die Output-Dateien, die durch die Anwendung einer Syntaxdatei softwarespezifisch produziert werden, durch entsprechende Metadaten erfasst, dokumentiert und im Zuge eines Uploads in das Archiv der VFU verwaltet und erschlossen werden. Dazu zählen z. B. Logfiles, Listing, Graphiken oder Tabellen und Tabellendokumente.

5.2 Die Metadatenelemente des Schemas Syntax

Erforderliche sind Angaben zu 1.3 Arbeitspaket, 4.1 Referenz, 7.1 Titel, 7.3 Dateiname (M Spalte FA). Empfohlen werden Angaben zu allen weiteren editierbaren Elementen (FT / KV in Spalte FA).

Tabelle 5: Die Metadatenelemente des Schemas Syntax (MD SX)

Nr.	Elementname	Definition und Anmerkungen (MD-SX)	FA
	Administration VFU SX	Angaben zu VFU Mitglied und Arbeitspaket, in dessen Kontext die Beschreibung der Syntax in der VFU bearbeitet wird und weitere administrative Elemente	
1	Person	Person, die die Syntax anlegt und bearbeitet	
1.1	Vornamen	Vorname der Person	AP
1.2	Nachnamen	Nachname der Person	AP
1.3	Arbeitspaket	Kontext in dem das Objekt bearbeitet wird <ul style="list-style-type: none"> • Auswahlliste: Namen der Arbeitspakete 	M KV
2	Objekt	Angaben zum im Archiv gespeicherten Objekt „VFU Syntax“	
2.1	Titel	Der Titel des gespeicherten Objekts „VFU- Syntax“	

Nr.	Elementname	Definition und Anmerkungen (MD-SX)	FA
		<ul style="list-style-type: none"> wird automatisch aus 7.1 Titel erzeugt 	AA
2.2	ID	<p>Eindeutige menschenlesbare ID der Syntaxbeschreibung</p> <ul style="list-style-type: none"> Format z. B. sx001, sx002, etc. Auswahl aus ID Liste <p>Beim duplizieren der Syntax wird der Titel geändert und eine neue ID zugeordnet</p>	KV
2.3	Datum Aktualisierung	<p>Logische Kennzeichnung des Metadatenstatus als (neue) Version durch den User. Er gibt ein Datum, an dem die Angaben zuletzt inhaltlich aktualisiert wurden.</p> <ul style="list-style-type: none"> Kalenderfunktion: Format JJJJ.MM.TT 	KL
2.4	Datum Beschreibung	<p>Logische Kennzeichnung des Metadatenstatus als (neue) Version durch den User. Er beschreibt den Stand der Version bzw. die inhaltlichen Änderungen einer neuen Version</p>	FT
2.5	Versionsnummer	<p>Angabe einer numerischen Version</p> <ul style="list-style-type: none"> Format je nach Versionierungskonzept 2 oder 3 stellig 	FT
2.6	Speicherdatum	Systemdatum bei Anwendung der Funktion Speichern	AA
2.7	Publizieren	<p>Steuerung der Veröffentlichung einer Syntaxdatei zur Vermeidung bruchstückhafter Informationen z. B. beim Speichern wegen Arbeitsunterbrechungen.</p> <ul style="list-style-type: none"> Auswahl: Ja / Nein 	KV
2.8	Objekt-ID	System-ID der angelegten Syntaxbeschreibung	AA
3	Duplizieren VFU-SX Ursprung	Angaben zur ursprünglichen SX, die automatisch in ein Duplikat übernommen werden,, wenn im Editor der Befehl Duplizieren gewählt wird	
3.1	Titel	<p>Titel der ursprünglichen Syntaxdatei</p> <ul style="list-style-type: none"> Wird automatisch aus 2.1 übernommen 	AA
3.2	Objekt ID	<p>System-ID der ursprünglichen Syntaxdatei</p> <ul style="list-style-type: none"> Wird automatisch aus 2.8 übernommen 	AA
3.3	Ref Objekt-ID	<p>Link auf Objekt-ID des duplizierten Ursprungsobjekts. Das Duplikat ist über diese ID mit dem Ursprungsobjekt verknüpft und navigierbar. Es werden alle Metadaten des Ursprungsobjektes angezeigt</p>	AA
	Syntax Eigenschaft	Formale Angaben zur Syntaxdatei	
4	Gegenstand	Gegenstand, Ziel, Thema der Untersuchung, der die Syntax zugeordnet wird. Die Angabe wird im Schema Datennutzung Nr. 4.1 Titel und 4.2 Inhalt gemacht.	
4.1	Referenz Datennutzung	<p>Verknüpfung der Syntax mit dem Titel der Datennutzung.</p> <ul style="list-style-type: none"> Auswahlliste: Titel Datennutzung (MD-DN Nr. 2.1) 	
5	Syntax Dateistruktur	Angabe zum Syntaxaufbau (Einzeldatei, Syntaxset, Modul)	

Nr.	Elementname	Definition und Anmerkungen (MD-SX)	FA
5.1	Dateistruktur Typ	Typ der Dateistruktur der Syntax <ul style="list-style-type: none"> • Einzelsyntax: Datei ist eine einzelne Syntax • Syntaxset: Datei ist Teil eines Sets von Syntaxdateien • Syntax-Modul: z. B. eine Haushaltstypologie 	KV
6	Syntax Statistikprogramm	Statistiksoftware mit der die Syntax lauffähig ist	
6.1	Name	Namen der Software, mit der die Syntax lauffähig ist <ul style="list-style-type: none"> • STATA • SPSS • R 	KV
6.2	Version	Software Version, mit der die Syntax lauffähig ist	FT
6.3	Pakete	Zusätzliches Paket (+ Version), das zum Ausführen benötigt werden (Packages / Ados)	FT
7	Syntax Ablage Archiv	Dokumenttitel und Ablage der Syntax im Archiv	
7.1	Titel	Titel der Syntaxdatei	M FT
7.2	Beschreibung	Allgemeine Beschreibung der Syntax (Abstrakt). Eine Spezifikation je Funktion erfolgt gesondert nach <ul style="list-style-type: none"> • Auswahl einer Kategorie in Element 9.1 und der jeweiligen Beschreibung in den Elementgruppen 10 bis 14 	E FT
7.3	Dateinamen	Name der gespeicherten Datei	M FT
7.4	Speicherdatum	Systemdatum bei Anwendung der Funktion Speichern	AA
8.	Syntax Berechtigung	Teilen von Syntax mit anderen VFU Mitgliedern bzw. Arbeitspaketen, durch Vergabe entsprechender Berechtigungen	
8.1	Nutzungsbedingungen	Bedingung zur Nutzung der Datei nach Freigabe für weitere(s) VFU Mitglied(er) und / oder VFU AP(s)	FT
8.2	Freigabe	Zuweisung von Berechtigung(en) je Person / je Arbeitspakete <ul style="list-style-type: none"> • Laut VFU Rollen- und Rechtemanagement 	KV
8.3	Freigabedatum	Datum, an dem die Freigabe erteilt wurde	AA
	Syntax Funktion Aufgabe	Grobe Kategorisierung der Funktion sowie Beschreibung der Aufgabe einer einzelnen Syntaxdatei. Genauere Angaben erfolgen in Elementgruppen 10 bis 14	
9	Syntax Funktion	Beschreibung der Syntaxfunktion und die auszuführende Aufgabe der Syntax	
9.1	Funktion Kategorie	Angabe zu Aufgabe bzw. Funktion der Syntax nach Kategorien <ul style="list-style-type: none"> • Datenselektion: Subset Variablen aus Ausgangsdaten • Datenaufbereitung: Variablen recodieren etc. 	KV

Nr.	Elementname	Definition und Anmerkungen (MD-SX)	FA
		<ul style="list-style-type: none"> • Datenverknüpfung: Analysefile erstellen • Datenauswertung: Analyse mit Syntax ausführen 	
9.2	Aufgabe Beschreibung	Beschreibung der spezifischeren Syntaxaufgabe je gewählter Funktionskategorie aus 9.1	FT
9.3	Aufgabe Entwicklungsstatus	Entwicklungsstatus der Syntax <ul style="list-style-type: none"> • Entwurf • Work in Progress • Final 	KV
	Syntax Funktion Operationalisierung	Beschreibung der Ausführung einer Aufgabe (der geplanten Operationalisierung) je gewählter Funktion / Kategorie aus 9.1	
10	Syntax Datenselektion	Beschreibung zur Syntax, mit der ein oder mehrere Datenfiles (Inputdaten) eingelesen werden	
10.1	Input Daten Beschreibung	Angabe zum benutzten Datensatz	FT
10.2	Input Daten Dateinamen	Dateinamen des eingelesenen Datenfile (der Datenfiles)	FT
11	Syntax Datenaufbereitung	Beschreibung zur Syntax, mit der die eingelesenen Daten weiter aufbereitet werden, z. B. durch Recodierung	
11.1	Beschreibung	Allgemeine Beschreibung zur Datenaufbereitung Eine spezifische Beschreibung der Erzeugung einzelner neuer Variablen im Zuge der Datenaufbereitung ist mit den Elementen der folgenden Gruppe möglich	FT
12	Syntax Datenaufbereitung Neue Variablen	Beschreibung zur Syntax, mit der aus Ausgangsvariablen des eingelesenen Datenfiles neue Variable erzeugt werden (Recode, Generate, etc.) <ul style="list-style-type: none"> • Es können 0-n neue Variablen je Syntax dokumentiert werden 	
12.1	Konzept	Fachliche Beschreibung, wie und zu welchem Zweck die neue Variable gebildet wird (n=1 je neue Variable) <ul style="list-style-type: none"> • Z. B. Konzeptbeschreibung, z. B. zu einer Typologie • Z. B. Anweisung zur Recodierung der Ausgangsvariable(n), um eine neue Variable zu bilden 	FT
12.2	Variable Name	Name jeder neuen Variable (n=1)	FT
12.3	Variable Label	Label jeder neuen Variable (n=1)	FT
12.4	Value	Wert in der neuen Variable (1-n je neuer Variable)	FT
12.5	Value Label	Value Label (n=1 je neuem Value)	FT
12.6	Missing Value	Fehlender Wert in der neuen Variable (1-n je neuer Variable)	FT

Nr.	Elementname	Definition und Anmerkungen (MD-SX)	FA
12.7	Missing Value Label	Missing Value Label (n=1 je neuem Missing Value)	FT
13	Syntax Datenverknüpfung	Beschreibung zur Syntax, mit der der Datensatz mit eingelesenen und / oder modifizierten Variablen erstellt wird (Analysefile).	
13.1	Beschreibung	Beschreibung des neu erstellten Datensatzes / Analysefile	FT
13.2	Dateinamen	Name des Datenfile, der durch die Syntax neu erstellt wird	FT
14	Syntax Datenauswertung	Beschreibung zur Syntax, mit der die Datenanalyse ausgeführt wird	
14.1	Beschreibung	Beschreibung der angewendeten Rechen- / Analysemethoden	FT

5.3 Metadatenelemente zum Speichern von Dateien zu einer Syntax

Erforderlich sind Angaben zu 1.3 Arbeitspaket, 2.1 Referenz, 3.1 Titel und 3.4 Dateiname (M Spalte FA). Empfohlen werden Angaben zu allen weiteren editierbaren Elementen (FT / KV in Spalte FA).

Aus konzeptuellen Gründen wird in dieses Subschema auch der Upload einer Syntaxdatei mit aufgenommen. Es handelt dabei nur um die Elemente zur Beschreibung der Syntaxdatei selbst. Alle Elemente mit den Angaben zur Beschreibung der Syntax (Inhalt, Funktion usw.) und der mit ihr verbundenen Syntaxdatei sind im Schema Syntax (MD-SX) aufgeführt. Aus diesem Grund wird im folgenden Schema auf die Elementnummern im Schema Syntax verwiesen. Wie in den Fällen der Schemata Studie und Datennutzung kann auch hier eine gespeicherte Datei nur mit der Beschreibung einer Syntax verknüpft werden, wenn diese bereit angelegt ist.

Tabelle 6: Metadatenelemente zum Speichern einer Datei im VFU Archiv und Referenzierung mit der Beschreibung zur Syntax

Nr.	Elementname	Definition und Anmerkungen (SX-Upload)	FA
	Administration SX-Upload	Angaben zu VFU Mitglied und Arbeitspaket, in dessen Kontext der Upload von Dateien in das VFU Archiv erfolgt	
1	VFU Person	Mitglied der VFU, das die Datei im VFU Archiv speichert	
1.1.	Vornamen	Vorname der Person	AP
1.2.	Nachnamen	Nachname der Person	AP
1.3	Arbeitspaket	Arbeitspaket in dessen Kontext das Mitglied die Datei speichert <ul style="list-style-type: none"> Auswahlliste: Namen der Arbeitspakete 	M KV
	Datei	Angaben zur Datei, die gespeichert wird	
2	Datei Typ	Typ der Datei, die im Archiv gespeichert wird: Auswahlliste (KV) <ul style="list-style-type: none"> Syntaxdatei Outputdateien zu einer Syntaxdatei 	KV

Nr.	Elementname	Definition und Anmerkungen (SX-Upload)	FA
2.1	Referenz Titel Syntax	Titel der Syntax, mit der die gespeicherte Outputdatei verknüpft wird. Nur gültig für den entsprechenden Typ unter Nr. 2. <ul style="list-style-type: none"> Auswahlliste: Titel der Syntaxbeschreibung (MD-SX Nr. 7.1) 	M KV
2.2	Ref Parent-ID	Verknüpfung der Datei mit der System-ID der gewählten Syntax	
3	Eigenschaft	Weitere Angaben zur gespeicherten Datei	
3.1 (7.1)	Titel	Angabe eines Titel der Datei Falls Syntaxdatei: MD-SX Nr. 7.1	M FT
3.2	Typ	Spezifizierung des Objekttyps gemäß der Angabe zu Nr. 2 Angaben erfolgen entweder zu 3.2.1 (Syntaxdatei) oder zu 3.2.2 (Outputdatei)	
3.2.1 (5.1)	Typ Syntaxdatei	Typ des Syntaxfiles Auswahlliste (KV) > MD-SX Nr. 5.1 <ul style="list-style-type: none"> Einzelsyntax: Datei ist eine einzelne Syntax Syntaxset: Datei ist Teil eines Sets von Syntaxdateien Syntax-Modul: z. B. eine Haushaltstypologie 	KV
3.2.2	Typ Outputdatei	Typ der Outputdatei: <ul style="list-style-type: none"> Dokumentation des Laufes der Syntax: Logfile Datensatz: Arbeitsdatensatz, Fertiger Analysefile, Ergebnisdatei: Listings, Tabelle/ Tabellendokument, Andere Variablenliste: VFU-generierte Variablen 	KV
3.3 (7.2)	Inhalt	Beschreibung zur gespeicherten Datei > Falls Syntaxdatei: MD-SX Nr. 7.2	FT
3.4 (7.3)	Dateiname	Vollständiger Dateinamen (Name.Dateiformat) > Falls Syntaxdatei: MD-SX Nr. 7.3	M FT
3.4.1	Dateigröße	Angaben in GB / Mbyte/Kbyte	AA
3.4.2 (7.4)	Speicherdatum	Datum an dem die Datei im Archiv gespeichert wurde > Falls Syntaxdatei: MD-SX Nr. 7.4	AA
3.4.3	Fingerprint	Fingerprint der gespeicherten Datei (wenn implementiert)	
3.5 (8.1)	Freigabe Nutzungsbedingung	Bedingung zur Nutzung der Datei nach Freigabe für weitere(s) VFU Mitglied(er) und / oder VFU AP(s) > Falls Syntaxdatei: MD-SX Nr. 8.1	FT
3.6 (8.2)	Freigabe	Zuweisung definierter VFU-Rechte je Person / je Arbeitspaket <ul style="list-style-type: none"> Laut VFU Konzept zum Rollen- und Rechtemanagement > Falls Syntaxdatei: MD-SX Nr. 8.2	KV
3.6.1 (8.3)	Freigabedatum	Datum, an dem die Freigabe erteilt wurde. Angabe erfolgt automatisch nach Speichern von 3.6 > Falls Syntaxdatei: MD-SX Nr. 8.3	AA

Referenzen

- Bartelheimer, Peter. Schmidt, Tanja. 2011
Modellprojekt "Kollaborative Datenauswertung und Virtuelle Arbeitsumgebung" - VirtAug - Abschlussbericht. soeb-Arbeitspapier 2011-1.
http://www.soeb.de/fileadmin/redaktion/downloads/VirtAug/VirtAug_Abschlussbericht.pdf
- Dickmann, Frank. Enke, Harry. Harms, Patrick. 2010
Technische Evaluation der Grid-Technologie für das Modellprojekt Kollaborative Datenauswertung und virtuelle Arbeitsumgebung – VirtAug. SOEB Arbeitspapier 2010-1.
http://www.soeb.de/fileadmin/redaktion/downloads/VirtAug/Expertise_VirtAug.pdf
- DDI Alliance. 2014.
Data Documentation Initiative <http://www.ddialliance.org/>
- DDI 2 - Codebook version
<http://www.ddialliance.org/Specification/DDI-Codebook/>
- DDI Empfehlungen zu kontrollierten Vokabularen:
<http://www.ddialliance.org/Specification/DDI-CV/>
- Friedhoff, Stefan. Meier zu Verl, Christian. Pietsch, Christian. Meyer, Christian.. Vompras, Johanna. Liebig, Stefan. 2013
SFB 882 Working Paper Series, 16. Bielefeld: DFG Research Center (SFB) 882 From Heterogeneities to Inequalities.
<urn:nbn:de:0070-pub-25600355>
- Hausstein, Brigitte. Quitzsch, Nicole. Jeude, Kirsten. Schleinstein, Natalija. Zenk-Möltgen, Wolfgang. 2012. da|ra Metadata Schema Version 2.2.1 (Released August 8th, 2012)
GESIS-Technical Reports 2013|03
<http://dx.doi.org/10.4232/10.mdsdoc.2.2.1>
- Hausstein, Brigitte. Schleinstein, Natalija. Koch, Ute. Meichsner, Jana. Becker, Kerstin. Stahn, Lena-Luise. 2014. da|ra Metadata Schema. Version 3.0. (Released March 11th, 2014)
GESIS-Technical Technical Reports 2014|07
<DOI: 10.4232/10.mdsxsd.3.0>
- Jensen, Uwe. 2014
Fachtagung "Das Portal, die Daten und wir - Eine Virtuelle Forschungsumgebung für die digitale Infrastruktur". Berlin 24.01.2014. <http://www.soeb.de/vfu-soeb-3/abschlussstagung/>
Panel 4: Gesucht, gefunden – Metadaten zur Nutzung von Forschungsdaten
Präsentation: Einführung in das Metadatenschema der VFU soeb 3
http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU_Tagung_Jensen_final.pdf
- Jensen, Uwe. EDDI 2014
6th Annual European DDI User Conference. London. December 2. -3.2014.
Vortrag (eingereicht) „Metadata Requirements to document Data Analyses and Syntax Files in a Virtual Research Environment (VRE) – The use case soeb 3“
<http://www.eddi-conferences.eu/ocs/index.php/eddi/eddi14>

- Jensen, Uwe; Schweers, Stefan; Carevic, Zeljko. 2014
Die Metadateneditoren der VFU soeb 3
GESIS-Technical Reports – 2014|14
http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2014/TechnicalReport_2014-14.pdf
- Schmidt, Tanja. 2014a
Grundlegende Informationen zur Benutzung der Virtuellen Forschungsumgebung (VFU) für soeb 3 (Version 0.1). Januar 2014.
SOFI - Soziologisches Forschungsinstitut Göttingen an der Georg-August-Universität
http://www.soeb.de/fileadmin/redaktion/downloads/VFU/Grundlegende_Informationen_zur_Nutzung_der_VFU_v01.pdf
- Schmidt, Tanja. 2014b
Fachtagung "Das Portal, die Daten und wir - Eine Virtuelle Forschungsumgebung für die digitale Infrastruktur". Berlin 24.01.2014: <http://www.soeb.de/vfu-soeb-3/abschlussstagung/>
Panel 2: Erprobt – Nutzungserfahrungen und Forschungspraxis
Präsentation: „Ergebnisse aus zwei Nutzungsstudien“
http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU_Tagung_Schmidt_final.pdf
- SOFI. 2013.
Soziologisches Forschungsinstitut Göttingen an der Georg-August-Universität
Projektbeschreibung: Verbundprojekt „Virtuelle Forschungsumgebung für die sozioökonomische Berichterstattung“ (VFU soeb 3)
http://www.soeb.de/fileadmin/redaktion/downloads/VFU/Projektbeschreibung_VFU_SC_081013.pdf
- VFU soeb 3. Projekt „Virtuelle Forschungsumgebung“ (VFU)
VFU Projektseite:
<http://www.soeb.de/vfu-soeb-3/>
Über soeb 3 - Dritter Bericht zur sozioökonomischen Entwicklung in Deutschland:
<http://www.soeb.de/ueber-soeb-3/>
- Zenk-Möltgen, Wolfgang. Habel, Norma. 2012
Der GESIS Datenbestandskatalog und sein Metadatenschema - Version 1.8.
GESIS-Technical Reports 2012|01.
http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-01.pdf