

## Die Metadateneditoren der VFU soeb 3

Jensen, Uwe; Schweers, Stefan; Carevic, Zeljko

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

**Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:**

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Jensen, U., Schweers, S., & Carevic, Z. (2014). *Die Metadateneditoren der VFU soeb 3*. (GESIS-Technical Reports, 2014/14). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-402984>

### Nutzungsbedingungen:

*Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.*

*Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.*

### Terms of use:

*This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.*

*By using this particular document, you accept the above-stated conditions of use.*

## Die Metadateneditoren der VFU soeb 3

*Uwe Jensen, Stefan Schweers, Zeljko Carevic*



GESIS–Technical Reports 2014|14

## Die Metadateneditoren der VFU soeb 3

*Uwe Jensen, Stefan Schweers, Zeljko Carevic*

## **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](mailto: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

---

<b>1</b>	<b>Einführung</b> .....	<b>5</b>
<b>2</b>	<b>Die Metadateneditoren der VFU soeb 3 - eine Einführung</b> .....	<b>8</b>
2.1	Der Zugang zu den Metadateneditoren im Portal der VFU .....	8
2.2	Metadateneditor Studie .....	10
2.3	Metadateneditor Datennutzung.....	11
2.4	Metadateneditor Syntax .....	12
<b>3</b>	<b>Metadaten zur Studie erfassen - Editor Studie</b> .....	<b>13</b>
3.1	Eine neue Studienbeschreibung erstellen .....	13
<b>4</b>	<b>Metadaten zur Datennutzung erfassen - Editor Datennutzung</b> .....	<b>25</b>
4.1	Eine neue Datennutzung beschreiben.....	25
<b>5</b>	<b>Metadaten und Syntax bearbeiten - Editor Syntax</b> .....	<b>30</b>
5.1	Anwendungsfall 1: Neue Syntaxdatei anlegen.....	30
5.2	Anwendungsfall 2: Syntaxisinhalte bearbeiten.....	32
5.3	Anwendungsfall 3: Metadaten zur Syntax anzeigen und bearbeiten .....	33
5.4	Anwendungsfall 4: Metadaten neu erzeugter Variablen erfassen.....	38
<b>6</b>	<b>Gemeinsame Funktionen aller Editoren im Überblick</b> .....	<b>42</b>
6.1	Tooltips – Nutzerhilfe bei der Erfassung von Metadaten.....	42
6.2	Übersichtsbox – Liste der zuletzt verwendeten Objekte.....	42
6.3	Infobox – Anzeige von Metadaten zu einem erstellten Objekt.....	43
6.4	Versionierung – Metadaten vorhandener Objekte aktualisieren.....	43
6.5	Dateien zu Studie, Datennutzung und Syntax hochladen .....	45
6.6	Berechtigungen – Objekte und Dateien mit VFU Mitgliedern teilen .....	48
6.7	Duplizieren – Metadaten eines Objektes wiederverwenden.....	50
6.8	Löschen eines Objektes oder einer zum Objekt gehörenden Datei.....	51
<b>7</b>	<b>Die Einbettung der Metadateneditoren in die VFU</b> .....	<b>53</b>
<b>8</b>	<b>Gesucht, Gefunden: Metadaten im VFU Archiv praktisch nutzen</b> .....	<b>55</b>
	<b>Glossar: Studie, Studienbeschreibung, Datensatz, Datenfile/Datendatei</b> .....	<b>61</b>
	<b>Referenzen</b> .....	<b>63</b>



## 1 Einführung

---

Ziel des Projekts „Virtuelle Forschungsumgebung“ (VFU) <sup>1</sup> 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) <sup>2</sup> (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 <sup>3</sup> 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 <sup>4</sup> 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).

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

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).

---

<sup>1</sup> <http://www.soeb.de/vfu-soeb-3/>

<sup>2</sup> <http://www.soeb.de/ueber-soeb-3/>

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

<sup>4</sup> <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.<sup>5</sup>

Das für den Produktivbetrieb eingerichtete Portal „Virtuelle Forschungsumgebung für die sozioökonomische Berichterstattung“ ist für die registrierten Mitglieder des Forschungsverbundes Portal zugänglich<sup>6</sup>. 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-oekonomischen Panels, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

GESIS definierte die für die VFU benötigten Metadaten und spezifizierte diese in zusammenhängenden Metadatenschemata. Die Schemata sind im Report „Das Metadatenschema der VFU soeb 3“ (Jensen; Schweers. 2014). Darauf aufbauend wurden im Zuge weiterer fachwissenschaftlicher Spezifikationen mit den Projektpartnern weitere kollaborative IT-Werkzeuge, insbesondere drei Metadateneditoren von GESIS entwickelt und implementiert.

Ziel dieses Reports ist, die in der VFU implementierten Metadateneditoren und ihren technischen Kontext vorzustellen. Dazu werden als Ergebnisse der Entwicklung die Funktionen und Oberflächen der drei Editoren dargestellt. Die Felder zur Erfassung der Metadaten werden inhaltlich erläutert.

Kapitel 2 beschreibt einführend die wesentlichen Aufgaben der drei Metadaten Editoren und stellt die damit verbunden bzw. dokumentierten Objekte Studie, Datennutzung und Syntax vor. Gleichzeitig wird auf konzeptuelle Aspekte im Kontext des jeweils zugrunde liegenden Metadatenschemas hingewiesen.

Die Kapitel 3 bis 5 stellen die objektspezifischen Besonderheiten und Funktionen sowie die speziellen Felder zur Metadatenerfassung bzw. deren Verwaltung vor. Der Handlungskontext wird anhand von Abbildungen veranschaulicht und der Pfad beschrieben, um zu den einzelnen Funktionen eines Editors zu gelangen.

---

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

<sup>6</sup> <https://vfu.sofi.gwdg.de/>

---

In Kapitel 6 werden beispielhaft die Funktionen und Informationen dargestellt, die in allen Editoren implementiert sind. Dazu zählen etwa Infoboxen, das Hochladen von Dateien, die einem der drei Hauptobjekte zugeordnet werden, sowie Freigabe von Objekten und Dateien, etwa zum Syntax-Sharing.

Kapitel 7 stellt technische Aspekte der Editoren und ihre Einbettung in die VFU dar.

Schließlich werden in Kapitel 8 die Funktionen der Archivsuche zur Erschließung der dort gespeicherten Metadaten anhand von Beispielen vorgestellt.

Das Glossar: Studie, Studienbeschreibung, Datensatz, Datenfile / Datendatei erläutert anhand praktischer Beispiele, wie die Begriffe als Objekte bei der Metadatenerfassung insbesondere mit dem Editor Studie verwendet werden.

Hinweis: Die gezeigten Inhalte in den Abbildungen zu den Editoren dienen nur der Veranschaulichung durch Spielmaterial. Sie wurden auf einem Entwicklungsserver erstellt, der funktional dem Produktionssystem von soeb 3 entspricht, aber keine soeb 3 Projekteinhalte enthält.

## 2 Die Metadateneditoren der VFU soeb 3 – eine Einführung

Die einführende Darstellung der VFU Editoren zur Bearbeitung von VFU soeb 3 Metadaten vermittelt einen ersten Eindruck der wesentlichen Eigenschaften und Aufgaben dieser kollaborativen Metadatenwerkzeuge. Ergänzend wird auf die zugrundeliegenden Metadatenschemata der VFU soeb 3 (vgl. Jensen, Schweers. 2014) hingewiesen und Besonderheiten eines Schemas hervorgehoben.

### 2.1 Der Zugang zu den Metadateneditoren im Portal der VFU

Nach dem Einloggen des VFU Mitglieds öffnet sich der Startbildschirm der VFU. In der schwarz hinterlegten oberen Zeile befinden sich verschiedene Portlets (Komponenten der Benutzeroberfläche) mit den Funktionen der VFU.

Hier sind die kollaborativen Tools wie zum Beispiel Kalender, Absprachen, Forum und die Archivsuche zugänglich. Hinter dem Portlet „Daten & Analysen“ befindet sich der Zugang zu den Metadateneditoren. Die folgende Abbildung zeigt die Startseite der Editoren nach einem Klick auf „Daten & Analysen“ mit den folgenden Funktionen:

- Syntaxeditor starten
- Datennutzung beschreiben
- Datendatei hochladen / downloaden (selbstgenerierte Datensätze einer Datennutzung)
- Studie beschreiben (Beschreibung der Ausgangsdaten)



Abbildung 1

Die dargestellte Reihenfolge der Editoren spiegelt die Schwerpunkte der Metadatenerfassung durch die Forschenden in der VFU wieder. D. h. die Metadatenbearbeitung berührt aus dieser Nutzerperspektive hauptsächlich die Entwicklung (und das Teilen) von Syntaxen sowie konzeptuelle Angaben zur Nutzung der Ausgangsdaten. Die einmalige Dokumentation der Ausgangsdaten und partielle Aktualisierungen, die vorrangig im Projektmanagement erfolgen, tritt demgegenüber im Layout zurück.

Die Darstellung der Metadateneditoren folgt in diesem Report dem ablaufforientierten Schema der Nutzung von Forschungsdaten in einer VFU (vgl. folgende Abbildung aus Dieckmann u.a. 2010: 9).

Danach beginnt der Workflow mit den Ausgangsdaten, die in verschiedenen Schritten für die (iterative) Entwicklung von Syntaxen für die Datenanalysen bearbeitet werden.

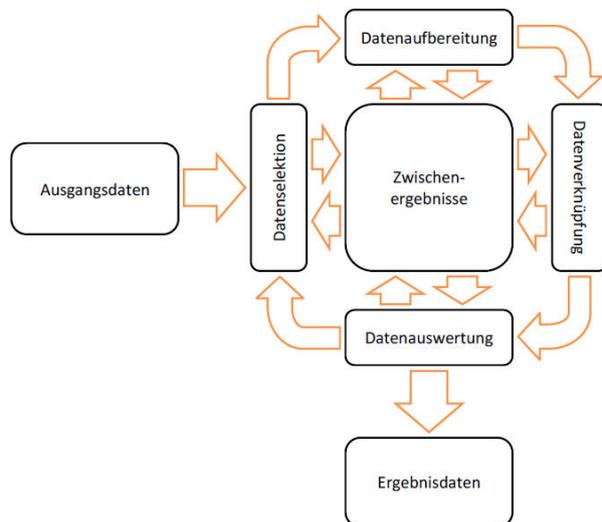


Abbildung 2: Dieckmann u.a. 2010: 9

Aus Sicht der Erfassung und Nutzung der Metadaten ergibt sich auch die Notwendigkeit, dass zumindest minimale Metadaten zur den Ausgangsdaten (Titel der Studie) mit dem Editor Studie erfasst sein müssen. Erst dann ist es möglich, diese Metadaten in den Editoren zur Syntax und zur Datennutzung wieder zu verwenden, um die drei Hauptobjekte miteinander verknüpfen zu können (vgl. Abbildung 3).

Die Metadatenwerkzeuge heißen analog zu den VFU soeb 3 Metadaten schemata (Jensen, Schweers, 2014):

- **Metadateneditor Studie**  
Beschreibt Ausgangsdaten und die Dokumentationen der Datensätze (einer Studie), die für Datenanalysen im soeb 3 Verbundprojekt genutzt werden. Es werden circa 50 Datensätze (ohne neue Versionen) von über 35 Datengebern in soeb 3 verwendet. Die Felder basieren auf dem Metadaten schema Studie (STB).
- **Metadateneditor Datennutzung**  
Beschreibt Konzepte und Parameter der Datennutzung sowie selbstgenerierter Datensätze im VFU soeb 3 Projekt. Die Felder basieren auf dem Metadaten schema Datennutzung (DN).
- **Metadateneditor Syntax**  
Beschreibt Syntaxdateien in verschiedenen Entwicklungsstadien sowie dazugehörige Outputdateien. Die Felder basieren auf dem Metadaten schema Syntax (SX).

Je nach Kontext werden zur Vereinfachung die Abkürzungen STB für Studienbeschreibung, DN für Datennutzung und SX für Syntax genutzt. Die folgende Abbildung veranschaulicht die Zusammenhänge von Editoren und Metadaten schemata.

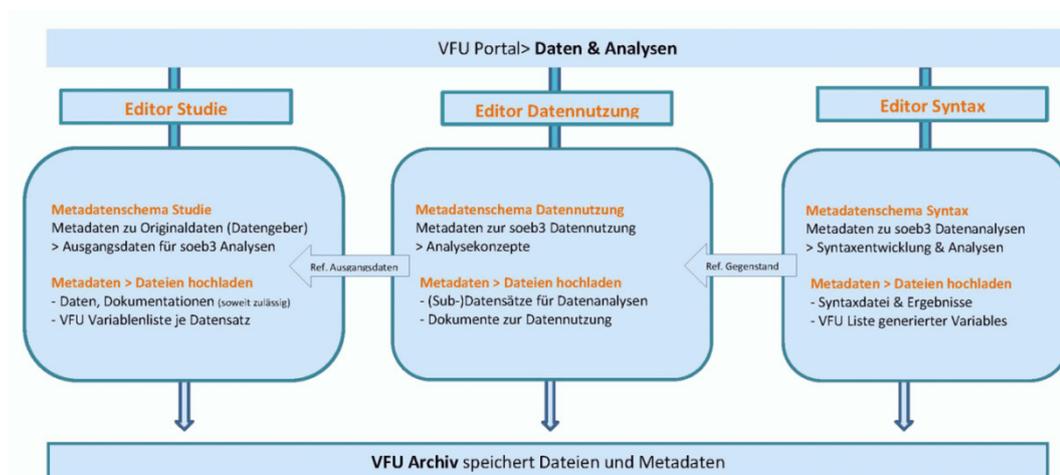


Abbildung 3: Editoren und Metadaten schemata im Überblick (Jensen. 2014)

Damit bilden die so erfassten Metadaten die Grundlage für eine VFU weite, integrierte Erschließung etwa von wichtigen Untersuchungsgegenständen oder selbst erzeugten Datensätzen, die auch für Analysen verschiedener Arbeitspakete von Relevanz sein könnten.

Gleichermaßen lassen sich auf diese Weise Syntaxen mit spezifischen Aufgaben finden, die für den eigenen Analysekontext von Interesse sein könnten bis hin zum Auffinden von konstruierten Variablen anderer Arbeitspakete, die auch im eigenem Analysekontext wiederverwendet werden könnten.

Die Erschließung der im VFU Archiv gespeicherten Metadaten und die entsprechenden Suchfunktionen werden in Kapitel 8 anhand von Beispielen beschrieben.

## 2.2 Metadateneditor Studie

Dieser Editor erfasst und verwaltet die Metadaten zu den sog. Ausgangsdaten, die extern von Datengebern bereitgestellt und für die Datenanalysen im Projekt VFU soeb 3 genutzt werden. Eine robuste Dokumentation der Ausgangsdaten ist von Relevanz, weil eine große Anzahl von Datensätzen von unterschiedlichen Datengebern genutzt werden (ca. 50 Datensätze von über 35 Datengebern; aktualisierte Daten oder neue Datensätze sind zusätzlich zu berücksichtigen).

Mit einer sog. Studienbeschreibung (STB) werden Informationen zum Entstehungskontext von Forschungsdaten, deren Zugänglichkeit und insbesondere deren fachlicher Inhalt sowie methodologische Eigenarten der Daten standardisiert auf Studienebene beschrieben. Metadaten schemata zur Dokumentation von Studien und Datensätzen werden seit langem z. B. in Datenkatalogen oder im Zusammenhang mit der Vergabe von PIDs (Persistent Identifier) zur Datenzitation eingesetzt. Diese standardisierten Schemata sind u. a. ein Bestandteil des seit langem etablierten sozialwissenschaftlichen DDI Dokumentationsstandard (DDI Alliance. 2014).

Zur Dokumentation der Ausgangsdaten, die für soeb 3 Datenanalysen genutzt werden, wurde als Grundlage das Metadaten schema STB als generischer Ansatz vorgeschlagen (Jensen. Schweers. 2014). Bei der konkreten Entwicklung des VFU soeb 3 Schemas wurde das Metadaten schema des GESIS Datenbestandskatalogs (Zenk-Möltgen. Habel. 2012) als Referenzschema genutzt und an den Bedarf der VFU soeb 3 angepasst und im Umfang entsprechend reduziert.

Die Dokumentation der für soeb 3 Datenanalysen genutzten Ausgangsdaten dient nicht nur der Nachweissicherung im Sinne Guter Wissenschaftlicher Praxis sondern sie erleichtert auch die Reproduzierbarkeit der Ergebnisse. Mit Blick auf die Archivfunktionalität der VFU sollen die entsprechenden Kontextinformationen zu den Ausgangsdaten dann auch gemeinsam mit Angaben zur Datennutzung und zur Syntax für die Nachnutzung z. B. durch Re-Analysen nach Ende des Projektes VFU soeb 3 zur Verfügung stehen.

Die Funktionen des Metadateneditors Studie beinhalten folgende Komponenten:

- Erfassen und Verwaltung von Metadaten des Datengebers zu den Ausgangsdaten, z.B.
  - Datenanbieter und Datenzugangsregeln
  - Bibliographie und Datenpublikation
  - Methodologie und Datensätze
- Upload von Daten und Material des Datengebers (soweit notwendig und rechtlich zulässig)
  - Dokumente des Datengebers
  - Datenfiles des Datengebers
- Das Erteilen von Berechtigungen zum Zugriff auf Daten und Materialien soweit notwendig

Die Nutzeroberflächen und die Felder zur Erfassung dieser Metadaten werden in Kapitel 3 detailliert beschrieben.

### 2.3 Metadateneditor Datennutzung

Mit dem Editor Datennutzung werden Metadaten zur Nutzung der Ausgangsdaten, insbesondere hinsichtlich des Untersuchungsgegenstandes und methodischer Parameter von geplanten Datenanalysen, im Projekt VFU soeb 3 erfasst. Darüber hinaus kann die detaillierte Beschreibung der verwendeten Analysedesigns sowie der angewendeten Methoden und statistischen Verfahren dauerhaft hinterlegt werden. Gleichzeitig können die verwendeten Ausgangsdaten, die in der Studienbeschreibung beschrieben werden, mit der Beschreibung der Datennutzung verknüpft werden.

Damit soll den VFU Nutzern und Nutzerinnen über entsprechende Recherchen mit Hilfe der Archivsuche z. B. eine Einschätzung ermöglicht werden, welche Fragestellungen mit welchen Daten untersucht werden bzw. mit welchen Syntaxen diese Fragen untersucht werden. Systematisch können damit etwa folgende Fragen kollaborativen Arbeitens einzelner Forscherinnen und Forschern bzw. entsprechende Arbeitsgruppen bzw. Arbeitspakete in einer VFU durch die Recherchefunktionen (vgl. Kap. 8) 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?
- Wann wurden Angaben zur Nutzung letztmals aktualisiert?

Weiterhin können Dateien, wie z. B. selbst generierte Datensätze und / oder Memos zu Datennutzung mit dem Editor verwaltet werden. Die Datendateien können aus Datenbanken wie z. B. Genesis des Statistischen Bundesamtes zusammengestellt worden sein oder es kann sich um gematchte Dateien aus Ausgangsdaten handeln, die für die eigentlichen Analyseläufe mit Syntaxdateien vorbereitet wurden.

Diese selbst generierten Dateien sind der Datennutzung des Nutzers/ der Nutzerin zugehörig und sind von den Ausgangsdaten abzugrenzen, welche in der Studienbeschreibung beschrieben werden. Das Metadatenkonzept zur Datennutzung wird außerdem von den Metadaten zu Syntaxen zu besserer Differenzierung abgegrenzt. Mit der Datennutzung werden Informationen zur Planung von Analysen erfasst, während die Metadaten zur Syntax konkret ausgeführte Datenanalysen dokumentieren.

Die Funktion des Metadateneditors Datennutzung beinhalten folgende Komponenten:

- Die Erfassung und Verwalten von Metadaten zu Konzepten der Datennutzung
- Das Hochladen und Verwalten von Dateien, z. B. selbstgenerierte Datensätze
- Das Freigeben und Teilen solcher Dateien mit anderen VFU Nutzerinnen und Nutzern

Kapitel 4 beschreibt detailliert die Nutzeroberflächen und die Felder zur Erfassung dieser Metadaten.

## 2.4 Metadateneditor Syntax

Mit dem Syntaxeditor werden sowohl Syntaxinhalte in entsprechenden Syntaxdateien sowie die Metadaten zu diesen Syntaxdateien erfasst.

Die Funktionen des Metadateneditors Syntax beinhalten folgende Komponenten:

- Die Erfassung und Verwalten von Metadaten zu Syntaxdatei
- Das Erstellen von neuen Syntaxdateien
- Das Bearbeiten von Syntaxinhalten
- Das Uploaden und Verwalten von Ergebnisdateien zur Syntaxdatei
  - Logfile
  - Arbeitsdatensatz / Fertiger Analysefile
  - Listing
  - Graphik
  - Tabelle/ Tabellendokument
- VFU- Variablenliste selbst generierter Variablen
- Das Teilen von Syntaxdateien mit anderen Nutzern und Nutzerinnen

Während für die Entwicklung der Metadaten schemata Studie und Datennutzung auf die Erfahrung des Metadatenstandards DDI zurückgegriffen werden konnte (DDI. 2014), existiert für Syntaxen ein vergleichbarer Standard nicht. Aus diesem Grund wurde für die Bedürfnisse der VFU ein Metadaten schema Syntax neu entwickelt. Dem Schema liegen verschiedene Anwendungsfälle für die Arbeit mit Syntaxen in der VFU zugrunde.

Die Nutzeroberflächen und Felder zur Erfassung der Metadaten werden in Kapitel 5 unter Berücksichtigung verschiedener Anwendungsfälle dargestellt.

### 3 Metadaten zur Studie erfassen – Editor Studie

Die Oberfläche zur Nutzung dieses Editors aus der Funktionsgruppe „Daten und Analysen“ startet mit der in der folgenden Abbildung dargestellten Funktion „Neue Studienbeschreibung erstellen“.

Wenn verfügbar, können von dieser Stelle aus auch die „Zuletzt verwendete Studienbeschreibungen“ geöffnet und bearbeitet werden. Diese Box ist in der aktuellen Ansicht leer. Vorhandene Studienbeschreibungen können auch über die Archivsuche (vgl. Kapitel 8) erschlossen und geöffnet werden.

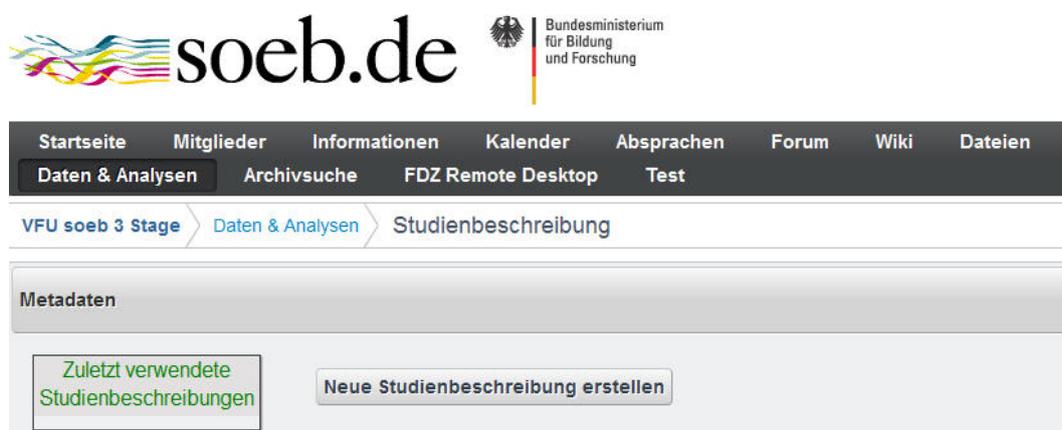


Abbildung 4

#### 3.1 Eine neue Studienbeschreibung erstellen

Die Aktion „Neue Studienbeschreibung erstellen“ öffnet die Dialoge zur Erfassung der Metadaten.



Abbildung 5

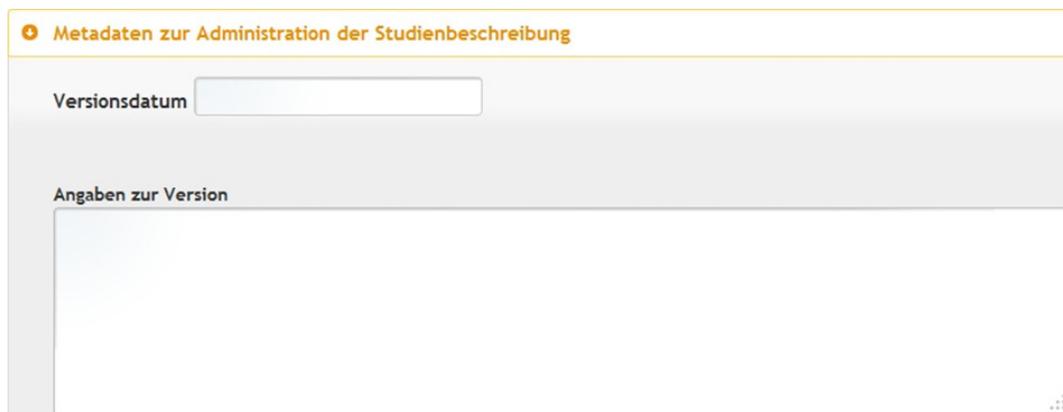
Die Metadatenfelder zur Studienbeschreibung sind in die vier Gruppen unterteilt:

- Administration der Studienbeschreibung
- Studie: Bibliographie und Datenpublikation
- Studie: Datengeber und Datenzugang
- Studie: Methodologie und Datensatz

Die Nutzeroberfläche und die Metadatenfelder in den einzelnen Gruppen werden im nächsten Abschnitt dargestellt.

### 3.1 Metadatenfelder der Studienbeschreibung

#### Gruppe 1 - Administration der Studienbeschreibung



The screenshot shows a web form titled "Metadaten zur Administration der Studienbeschreibung". It contains two main sections: "Versionsdatum" with a text input field, and "Angaben zur Version" with a large, empty text area for detailed notes.

Abbildung 6

#### 3.1.1 Versionsdatum

Das Versionsdatum wird manuell vom VFU Mitglied gesetzt. Es gibt Nutzer/in gesteuert an, wann die Studienbeschreibung zuletzt inhaltlich geändert wurde.

Im Gegensatz zum automatisch erfassten Datum „zuletzt geändert am“, welches in der linken Infobox (siehe vorherige Abbildung 6) erfasst wird, kann durch das Versionsdatum vom VFU Mitglied festgelegt werden, wann die Änderungen an einer Studienbeschreibung ausreichend sind, um ein neues Versionsdatum zu setzen.

#### 3.1.2 Angaben zur Version

Unter Angabe zur Version kann eine Beschreibung zum Versionsdatum angegeben werden. Dies können z. B. sein:

- Der Stand der Angaben des Datengebers zum Zeitpunkt der Erfassung der Metadaten in der VFU STB oder
- Informationen zur neuen Version des Datensatzes des Datengebers, der in der VFU benutzt wird.

## Gruppe 2 – Bibliographie und Datenpublikation der Studie (laut Datengeber)

### Abschnitt Bibliographie

The screenshot shows a web form titled 'Studie: Bibliographie und Datenpublikation'. Under the heading 'Bibliographische Metadaten', there are several input fields: 'Studientitel' (a large text box), 'Studie ID' (a text box), 'Art der Studien ID' (a dropdown menu with a '-' symbol), 'Inhalt' (a large text area), 'URL zum Inhalt der Studie' (a text box), 'Primärforscher/in Name' (a text box), and 'Institution' (a text box).

Abbildung 7

#### 3.1.3 Studientitel – erforderliche Angabe

Der Titel der Studie wird vom Datengeber vergeben, beispielsweise: „Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984–2012“.

Für die spätere Reproduzierbarkeit der Analysen ist es notwendig, dass genau der Titel der Studie angegeben wird, die auch in den Analysen verwendet werden.

#### 3.1.4 Studie ID

Bei der Studien-ID handelt es sich um eine eindeutige Kennung der Studie, die vom Datengeber vergeben wird. Eindeutig ist eine Studien-ID, weil sie von einem Datengeber nur einmal vergeben wird.

#### 3.1.5 Art der Studien ID

Bei der Studien-ID kann es sich z. B. um eine Studiennummer, wie sie von Archiven verwendet wird, oder eine Datensatznummer handeln, die z. B. von Forschungsdatenzentren genutzt werden.

#### 3.1.6 Studieninhalt

In der Regel können die Informationen zum Studieninhalt vom Datengeber übernommen werden, in denen Angaben zu den Themengebieten der Studie oder kodierten Variablen gemacht werden.

#### 3.1.7 Studieninhalt URL

Link zur Internetseite des Datengebers mit weiteren Informationen zum Inhalt der Studie.

### 3.1.8 Primärforscher Name

Unter Primärforscher sind diejenigen Personen zu verstehen, die an der wissenschaftlichen Entwicklung und Konzeption der Studie maßgeblich beteiligt sind. Zu unterscheiden ist der Primärforscher von den durchführenden Personen der Datenerhebung (bei Befragungen z. B. der Interviewer von TNS Infratest).

### 3.1.9 Primärforscher Institution

Hier wird ausgewiesen, welcher Institution ein Primärforscher angehört (Universität, außer-universitäre Forschungseinrichtung). Soweit keine Datenproduzenten als ‚Primärforscher‘ ausgewiesen sind kann hier die jeweils Institution dargestellt werden, die die Daten erzeugt oder bereitstellt, z. B. Institut für vergleichende Sozialforschung; Deutsche Bundesbank, Genesis Datenbank etc.

## Gruppe 2 – Bibliographie und Datenpublikation der Studie

### Abschnitt Datenpublikation

**Datenpublikation**

Publikationsdatum des Datensatzes

Versionsnummer des Datensatzes

Versionsgrund

Persisten Identifier (PID)  Art des PID:  PID URL

Datenzitation URL

Abbildung 8

### 3.1.10 Publikationsdatum des Datensatzes

Datum, an dem der Datengeber den Datensatz veröffentlicht hat. Diese Angabe findet sich auf der Informationsseite des Datengebers. Über eine Kalenderfunktion kann das Datum eingegeben werden.

### 3.1.11 Versionsnummer des Datensatzes

Neben dem Publikationsdatum gibt der Datengeber oftmals eine Versionsnummer des Datensatzes an, z. B. V1.1.

### 3.1.12 Versionsgrund

Bei mehreren Versionen eines Datensatzes können auch Versionsgründe des Datensatzes angegeben werden. Dies kann zum Beispiel die Korrektur eines Fehlers sein.

### 3.1.13 Persistent Identifier

Dauerhafter Identifikator des Datensatzes, z. B. „10.5684/soep.v29“

Eine PID wird analog zur ISBN eines Buches weltweit nur einmal vergeben wird, z. B. in Form einer DOI. Dadurch ist der Datensatz beim Datengeber dauerhaft erreichbar und durch weitere Metadaten zitierfähig vorgehalten. In Deutschland registriert die Agentur da|ra Daten der Sozial- und Wirtschaftswissenschaften und vergibt DOIs (DARA. 2014).

### 3.1.14 Art des PID – Kontrolliertes Vokabular

Die verschiedenen Arten des PID können per Auswahlliste aus einem kontrollierten Vokabular ausgewählt werden. Die häufigsten PIDs für Datensätze sind DOI und Handle. Aus der Liste können folgenden PIDs ausgewählt werden:

- ARK
- DOI
- EAN13
- EISSN
- Handle
- ISBN
- ISSN
- ISTC
- LISSN
- LSID
- PURL
- UPC
- URL
- URN

### 3.1.15 PID URL

Die URL einer PID verweist auf die Internetseite des Datengebers (Landing Page), auf der die Informationen zum Datensatz angeboten werden.

Bsp.: <http://dx.doi.org/10.5684/SOEP.V29.1> für soep.v29

### 3.1.16 Datenzitation URL

Macht ein Datengeber Vorgaben zur Datenzitation der Studie, kann eine URL zu den Zitationsvorgaben gesetzt werden.

## Gruppe 3 – Datenanbieter und Datenzugang zur Studie

**Studie: Datenanbieter und Datenzugang**

**Metadaten zum Datenanbieter**

Name

URL

Sprachversion der Metadaten:

**Metadaten zum Datenzugang**

Nutzungsbedingung:

Nutzungsbedingungen URL

Datenzugang ( Nach Art der Anonymisierung ):

Datenzugang URL

Abbildung 9

### 3.1.17 Name

Der Name des Datengebers, z. B. „DIW Berlin“ für das Sozio-oekonomisches Panel.

### 3.1.18 URL

Link zur Webseite des Datengebers.

### 3.1.19 Sprachversion der Metadaten – Kontrolliertes Vokabular

In der Regel liegen die Metadaten hauptsächlich in Deutsch oder in englischer Sprache vor. Es gibt aber auch Fälle, in denen die Metadaten teils in Deutsch und teils in Englisch vorliegen. Per Auswahlliste können drei verschiedene Sprachkombinationen ausgewählt werden, in der die Metadaten hauptsächlich vorliegen:

- deutsch
- englisch
- deutsch + englisch

### 3.1.20 Nutzungsbedingungen – Kontrolliertes Vokabular

Die Nutzung der Daten ist an rechtliche und formale Bedingungen geknüpft. Es wurden die folgenden Nutzungsbedingungen identifiziert, welche über eine Auswahlliste mit kontrolliertem Vokabular ausgewählt werden können. Angaben zur Zitation der Daten werden in Feld 3.16 erfasst.

- Frei, d.h. ohne vertragliche Nutzungsbedingungen und Zugang z. B. über Download
- Kontrolliert, wenn z. B. die Registrierung oder Zustimmung des Datengebers notwendig ist
- Vertrag, hier ist eine explizite vertragliche Regelung zur Nutzung der Daten notwendig
- Sonstiges

### 3.1.21 Nutzungsbedingungen URL

Link zu Internetseite des Datengebers, auf der die Nutzungsbedingungen beschrieben sind.

### 3.1.22 Datenzugang (nach Grad der Anonymisierung) – Kontrolliertes Vokabular

Zugang zu den Daten nach Grad der Anonymisierung und entsprechendem Zugangstyp. Hier stehen in einem kontrollierten Vokabular folgende Auswahlmöglichkeiten zur Verfügung:

- Absolut Anonymisiert (PUF)
- Faktisch Anonymisiert (SUF: Off-Site)
- Faktisch Anonymisiert (Gastaufenthalt)
- Projektbezogen (Gastaufenthalt On Site)
- Formal (kontrolliertes Fernrechnen On Site)

### 3.1.23 Datenzugang URL

Link zur Seite des Datengebers, auf der Informationen zum Datenzugang zu finden sind.

## Gruppe 4 – Methodologie und Datensätze zur Studie ( und Sonstige Nutzungsrechte)

### Abschnitt Methodologie

**Methodologie und Datensätze der Studie**

**Metadaten zur Methodologie**

Referenzzeitraum von:  bis:

Zeitliche Dimension der Datenerhebung:  Erhebungsverfahren:

Grundgesamtheit

Auswahlverfahren

Geographischer Raum

Abbildung 10

### 3.1.24 Referenzzeitraum von bis

Zeitraum, den die Daten inhaltlich abdecken, z. B. von 1984 bis 2012. Dieser Zeitraum kann über ein Kalenderfeld eingegeben werden.

### 3.1.25 Zeitliche Dimension der Datenerhebung- Kontrolliertes Vokabular

Angaben zur zeitlichen Ausdehnung, in die Daten erhoben werden. Die Angaben erfolgen in typisierter Form über ein kontrolliertes Vokabular:

- Längsschnitt, Trend, Kohorte, Zeitreihe
- Längsschnitt: Panel
- Querschnitt
- Andere

### 3.1.26 Erhebungsverfahren - Kontrolliertes Vokabular

In diesem Feld wird die Methode beschrieben, mit der die Daten erhoben wurden. Es werden Typen von Erhebungsverfahren über ein kontrolliertes Vokabular erfasst.

- Quantitatives Interview (Fragebogen)
- Qualitatives Interview (Leitfaden)
- Prozessgenerierte Daten
- Beobachtung
- Experiment
- Inhaltsanalyse

### 3.1.27 Grundgesamtheit

Beschreibung der (statistischen) Einheiten, die der Auswahl zur Grunde liegen.

Beispiel: „Privathaushalte mit Personen, die zum Befragungszeitpunkt das 18. Lebensjahr vollendet hatten“.

### 3.1.28 Auswahlverfahren

Beschreibung des Auswahlverfahrens zur Erstellung der Stichprobe (Stichprobenverfahren).

Beispiel: „1980 bis 1992 und 1998 mehrstufig geschichtete Zufallsauswahl aus allen Privathaushalten mit Personen, die zum Befragungszeitpunkt das 18. Lebensjahr vollendet hatten (ADM-Stichprobendesign)“.

### 3.1.29 Geographischer Raum

Gebietseinheit(en), die der Datenerhebung zugrunde liegen, z. B. Staaten und /oder staatliche Untereinheiten

## Gruppe 4 – Methodologie und Datensätze zur Studie(sowie Sonstige Nutzungsrechte)

### Abschnitt Datensätze

The screenshot displays a web-based metadata editor interface. It is divided into two main sections: 'Datensatz logisch' (Logical Dataset) and 'Datensatz technisch' (Technical Dataset). The 'Datensatz logisch' section includes a text area for 'Beschreibung Datensatz Struktur', a text input field for 'Datensatz Struktur URL', and two dropdown menus for 'Analyseeinheit Typ' and 'Daten Typ'. The 'Datensatz technisch' section includes a text area for 'Beschreibung Datenfile Struktur' and a text input field for 'Datensatz technisch URL'.

Abbildung 11

#### 3.1.30 Datensatz Struktur

Bei komplexen Daten, wie zum Beispiel dem SOEP, können hier Angaben zur Struktur der Daten gemacht werden. Bei weniger komplexen Daten sind Angaben zum Anonymisierungsgrad von Interesse.

#### 3.1.31 Datensatz Struktur URL

Link zur Internetseite des Datengebers mit Angaben zur Struktur der Daten.

#### 3.1.32 Analyseeinheit Typ- Kontrolliertes Vokabular

Daten können sich auf verschiedene Typen von Einheiten beziehen zu denen der Datensatz Aussagen trifft. Hierbei können auch Mehrfachnennungen vorkommen. Zum Beispiel gibt es im SOEP einen Fragebogen zu Personen und einen zu Haushalten.

- Individuum,
- Organisation
- Familie/im selben Haushalt
- Haushalt/ Wohneinheit
- Ereignis/ Prozess
- Geographische Einheit
- Zeiteinheit

- Texteinheit
- Gruppe
- Objekt
- Sonstiges

#### **3.1.33 Daten Typ- Kontrolliertes Vokabular**

Der formale Typ der Daten richtet sich danach, wie die Daten erhoben wurden. Folgende Kategorien können über das kontrollierte Vokabular ausgewählt werden:

- Befragungsdaten
- Prozessdaten
- Statistiken
- Indikatoren
- Andere

#### **3.1.34 Beschreibung Datenfile Struktur**

An dieser Stelle können Angaben zu einzelnen Datenfiles gemacht werden, die zur Studie gehören. Angaben sind z. B.: Dateinamen, Dateiformat, Größenangabe, Sprachversion je Datenfile.

#### **3.1.35 Datensatz technisch URL**

Link auf eine Übersichtseite des Datengebers, z. B. zur Struktur der Datenfiles.

**Gruppe 4 – Methodologie und Datensätze zur Studie(sowie Sonstige Nutzungsrechte)****Abschnitt Sonstige Nutzungsrechte / Dokumente / Literatur**

The image shows a screenshot of a metadata editor interface. It is divided into three main sections, each with a green header:

- Sonstige Nutzungsrechte**: Contains a dropdown menu labeled "Nutzungsrechte Webseiteninhalte:" with a "-" selected, and a text input field labeled "URL".
- Dokumente**: Contains a large text area labeled "Beschreibung Dokumente:" and a text input field labeled "Dokumente URL".
- Literatur**: Contains a large text area labeled "Beschreibung Literatur:" and a text input field labeled "Dokumente URL".

Abbildung 12

**3.1.36 Nutzungsrechte Webseiteninhalte**

Angaben zu Nutzungsrechten von Material und Webinhalten (ohne Daten) sowie Zitierregeln für diese Materialien. Aus einer Auswahlliste können folgende Antwortmöglichkeiten ausgewählt werden:

- Deutsch (Copyright, Rechtshinweise, Urheberrechte, Impressum)
- English (Disclaimer, Terms of Condition, Condition of use)

**3.1.37 URL**

Link zu den Nutzungsrechten von Material und Webinhalten beim Datengeber.

**3.1.38 Beschreibung Dokumente**

Beschreibung von Materialien und Dokumentationen zur Studie und der Daten.

**3.1.39 URL**

Link zur Seite mit Dokumentationen

Beispiel: [http://www.diw.de/de/diw\\_02.c.222857.de/dokumente.html](http://www.diw.de/de/diw_02.c.222857.de/dokumente.html)

**3.1.40 Beschreibung Literatur**

Literaturangaben oder Bibliographie des Datengebers zu wissenschaftlichen Publikationen, die sich primär auf diese Studie bzw. deren Daten beziehen.

**3.1.41 URL**

Link zur Seite mit Angaben zur Literatur oder Bibliographie.

Die erfassten oder geänderten Metadaten werden abschließend durch „Speichern“ im VFU soeb 3 Archiv gesichert und mit einer Meldung bestätigt.

## 4 Metadaten zur Datennutzung erfassen – Editor Datennutzung

Die Oberfläche zur Nutzung dieses Editors aus der Funktionsgruppe „Daten und Analysen“ umfasst die in der folgenden Abbildung dargestellte Funktion „Neue Datennutzung erstellen“.

Wenn verfügbar, können auch „Zuletzt verwendete Datennutzungen“ geöffnet und bearbeitet werden. Diese Box ist der aktuellen Ansicht leer.



Abbildung 13

### 4.1 Eine neue Datennutzung beschreiben

Die Aktion „Neue Datennutzung erstellen“ öffnet die Dialoge zur Erfassung der Metadaten, wie die nächste Abbildung zeigt.



Abbildung 14

Die Metadatenfelder für die Datennutzung sind durch zwei Gruppen gegliedert:

- Datennutzung
- Methodische Angaben

Die Nutzeroberfläche und die Metadatenfelder in den zwei Gruppen werden im nächsten Abschnitt dargestellt.

#### 4.1. Metadatenfelder zur Datennutzung

##### Gruppe 1 - Datennutzung

##### Abschnitt Administrative Angaben

🔍 Datennutzung

**Administrative Informationen**

Titel der Datennutzung

Inhalt

Versionsdatum

Datum Beschreibung

Abbildung 15

##### 4.1.1. Titel der Datennutzung - erforderliche Angabe

Titel zum Thema des Untersuchungsgegenstands/-ziels, auf das sich die Datennutzung bezieht.

##### 4.1.2. Inhalt

Beschreibung des Untersuchungsgegenstands/-ziels, für die Daten benutzt werden. Hier können Angaben zur genauen Fragestellung, der methodischen Vorgehensweise, generierten Variablen usw. gemacht werden.

##### 4.1.3. Versionsdatum

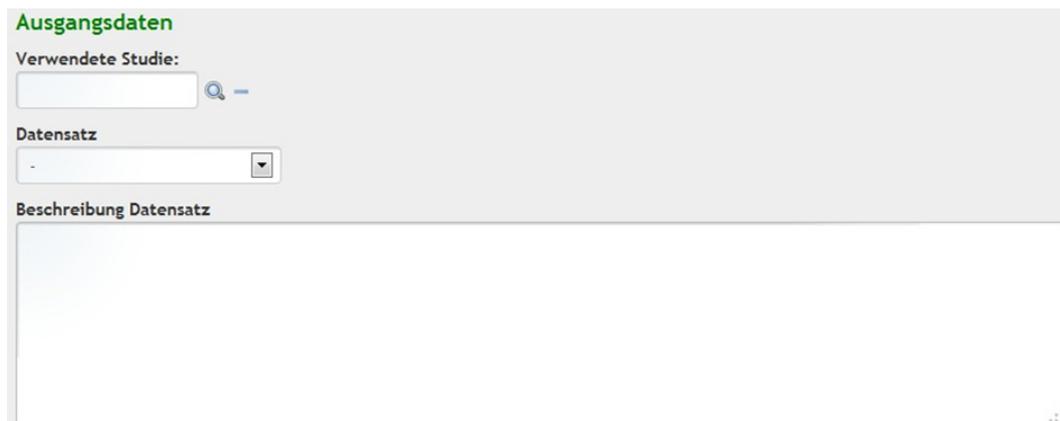
Die Versionierung erfolgt bei der Datennutzung über ein explizit gesetztes neues Datum und eine Beschreibung der Änderungen zu diesem Datum im nächsten Feld.

#### 4.1.4. Datum Beschreibung

Zu dem Versionsdatum kann eine Beschreibung eingegeben werden, in der die wichtigsten Änderungen zum neuen Datum beschrieben werden.

### Gruppe 1 - Datennutzung

#### Abschnitt Ausgangsdaten



The screenshot shows a web interface for editing metadata. At the top, the section is titled 'Ausgangsdaten' in green. Below this, there are three main components: 1. 'Verwendete Studie:': A search input field with a magnifying glass icon and a minus sign. 2. 'Datensatz': A dropdown menu with a downward arrow. 3. 'Beschreibung Datensatz': A large, empty text area for entering a description.

Abbildung 16

#### 4.1.5. Verwendete Studien

Im Feld Verwendete Studien wird der Untersuchungsgegenstand, der im Objekt Datennutzung beschrieben wird, mit einer bestimmten Studie (und deren Daten) durch Auswahl der entsprechenden Studienbeschreibung verknüpft.

Ziel dieser Verknüpfung ist es, in den übergeordneten Ergebnissen der Archivsuche anzuzeigen, welche Person und oder Arbeitspaket für welchen Untersuchungsgegenstand welche Studie, sprich welche Datenätze dieser Studie, als Ausgangsdaten nutzt.

Für die Verknüpfung zwischen Untersuchungsgegenstand und Studie wird auf die Lupe geklickt und in das aufklappende Suchfeld der Titel der Studie eingegeben (vgl. folgende Abbildung).



The screenshot shows a web interface for editing metadata. At the top, the section is titled 'Datennutzung'. Below this, there are three main components: 1. A button labeled 'Zurück zu Metadaten' with a left-pointing arrow. 2. A search input field labeled 'Suchbegriff:' with a magnifying glass icon. 3. A table with two columns: 'Titel' and 'Beschreibung'.

Abbildung 17

Hinweis: Der Titel der Studie muss exakt in das Suchfeld eingegeben werden, damit die passende Studie gefunden werden kann. Es empfiehlt sich, den Namen der Studie mit Strg + C und Strg + V in das Suchfeld hineinzukopieren.

Zur Erleichterung der Erfassung kann der Titel einer Datennutzung auch in der Archivsuche (vgl. Kap. 8) gefunden und dann kopiert werden. Dazu kann die Archivsuche in einem weiteren Browserfenster geöffnet werden.

#### 4.1.6. Datensatz – Kontrolliertes Vokabular

Typ des für Analysen benutzten Datensatzes der Studie. Durch Syntax erzeugte temporäre Arbeitsdatensätze werden im Syntaxeditor dokumentiert.

- Subdatensatz (aus Einzeldatenfile erzeugt)
- Einzeldatensatz (original Datenfile)
- Selbst generierter Datensatz (Daten aus Datenbank(en), etc.)
- Aus Originaldaten gematchter Datenfile (vorab erzeugt)

#### 4.1.7. Beschreibung Datensatz

Angaben zur Eigenart des gewählten Datensatzes und die geplante Nutzung in Datenanalysen. Beispiele sind:

- Subset
- selbstgenerierter Datensatz
- Long-; Wide Format
- Nutzung von Spelldaten

### Gruppe 2 – Methodische Angaben

The screenshot shows a web-based form titled "Methodische Angaben". It is organized into five distinct sections, each with a label and an input area:

- Zeitraum:** A large, empty text input field.
- Geographischer Raum:** A large, empty text input field.
- Kleinste Gebietseinheit:** A dropdown menu with a small arrow icon on the right.
- Analyseeinheit:** A large, empty text input field.
- Methoden:** A large, empty text input field.

Abbildung 18

#### 4.1.8. Zeitraum

Beschreibung für welchen Zeitraum (Referenzzeitraum) die Daten genutzt werden (sollen). Zum Beispiel werden die Daten des SOEP von 1990 bis 2010 genutzt. Hierbei ist unerheblich,

über welchen Zeitraum die Daten beim Datengeber zur Verfügung stehen sondern nur für welchen Zeitraum sie tatsächlich für die Analysen genutzt werden.

#### 4.1.9. Geographischer Raum

Beschreibung für welches Gebiet (Staat / Untereinheit) die Daten genutzt werden (sollen).

#### 4.1.10. Kleinste Gebietseinheit – – Kontrolliertes Vokabular

Beschreibung der kleinsten Gebietseinheit, zu der Daten genutzt werden. Aus einem kontrollierten Vokabular können folgende Gebietseinheiten ausgewählt werden.

##### Administrative Gebietseinheit

- Regierungsbezirk
- Region
- Kreis
- Gemeinde
- Andere Gebietseinheit: Erläuterung unter geographischer Raum

##### Statistische Gebietseinheit

- ROR
- BIK
- Andere Gebietseinheit: Erläuterung unter geographischer Raum

##### Weitere Gebietseinheit

- PLZ
- Finanzamtsbezirk
- Gerichtsbezirk
- Arbeitsagenturbezirk
- Bundestagswahlkreis
- Landtagswahlkreis
- Kammerbezirk HWK
- Kammerbezirk IHK
- Andere Gebietseinheit: Erläuterung unter geographischer Raum

#### 4.1.11. Analyseeinheit

Beschreibung für welche Analyseeinheit(en) die Daten genutzt werden (sollen).

#### 4.1.12. Methoden

Beschreibung welche statistischen Methoden die bei der Untersuchung / Datenanalyse angewendet werden (sollen).

Die erfassten oder geänderten Metadaten werden abschließend durch „Speichern“ im VFU soeb 3 Archiv gesichert und mit einer Meldung bestätigt.

## 5 Metadaten und Syntax bearbeiten – Editor Syntax

Im Vergleich zu den Metadateneditoren Studie und Datennutzung werden im Editor für Syntaxdateien nicht nur Metadaten erfasst, sondern es können auch die Inhalte einer Syntax editiert werden.

Aus diesem Grund weicht die Menüführung des Syntaxeditors etwas von den anderen Editoren ab. Deshalb werden in diesem Kapitel die folgenden vier Anwendungsfälle getrennt beschrieben, um sie übersichtlicher darstellen zu können.

- Anwendungsfall 1: Syntaxdatei anlegen,
- Anwendungsfall 2: Syntaxinhalte bearbeiten und
- Anwendungsfall 3: Metadaten zur Syntaxdatei erfassen, sowie
- Anwendungsfall 4: Metadaten zu neu erzeugten Variablen erfassen



Abbildung 19

### 5.1 Anwendungsfall 1: Neue Syntaxdatei anlegen

Der Syntaxeditor bietet zwei Möglichkeiten eine Syntaxdatei in der VFU soeb 3 neu anzulegen.

1. Im Editor kann der Inhalt einer Syntaxdatei manuell eingegeben werden. Dabei ist es auch möglich, den Inhalt einer Syntaxdatei in den Editor zu kopieren.
2. Komfortabler als den Inhalt einer Syntax in den Editor zu kopieren ist es, den Inhalt einer Syntax in den Editor hochzuladen. (Kapitel 5.1 Möglichkeit 2)

#### Möglichkeit 1: Syntaxdatei in der VFU erstellen

Eine Syntaxdatei wird in der VFU Umgebung wie folgt neu erstellt.

- Syntaxeditor starten auswählen und „Neue Syntax erstellen“ anklicken.

Es öffnet sich eine neue Oberfläche (siehe nächste Abbildung)

- Im Feld „Syntaxtitel“ wird ein aussagekräftiger Titel vergeben (Pflichtfeld)
- Syntaxinhalt kann in den Editor kopiert oder manuell eingegeben werden.

Abbildung 20

### Möglichkeit 2: Lokale Syntaxdatei hochladen

Um eine lokal erstellte Syntaxdatei in die VFU Umgebung hochzuladen, wird der Syntaxeditor gestartet und dann der Befehl „Syntax hochladen“ gewählt (vgl. Abbildung 21).

Es öffnet sich die Dialogbox „Syntax hochladen. Mit „Durchsuchen“ wird die lokale Datei ausgewählt und mit dem „OK & Hochladen“ bestätigt.

Abbildung 21

Anschließend wird der Inhalt der hochgeladenen Syntaxdatei im Syntaxeditor geöffnet, wie die nächste Abbildung zeigt.

- Nun wird ein „Syntaxtitel“ eingegeben(erforderliche Angabe).  
Neben dem Feld werden außerdem weitere Metadaten zu Version (als Datum), Bearbeiter, Arbeitspaket etc. angezeigt.
- Die Syntax kann bearbeitet werden.  
Weiterhin werden Syntaxbefehle farbig hervorgehoben. Die Syntax kann durchsucht und Fragmente können ersetzt werden, wie in Kapitel 5.2 erläutert wird.
- Das Fenster wird durch Speichern oder Abbrechen (in Abb. nicht sichtbar) geschlossen.



Abbildung 22

## 5.2 Anwendungsfall 2: Syntaxinhalte bearbeiten

Zur Unterstützung der manuellen Erstellung und Bearbeitung von Syntaxen sind im Editor die Funktionen „Suchen + Ersetzen“ und Syntax-Highlighting implementiert.

### Syntaxfragmente Suchen und Ersetzen

Der Inhalt der Syntaxdatei lässt sich per Volltext durch entsprechende Eingabe im Feld „Suchbegriff“ durchsuchen. Durch die Pfeile neben dem Feld Suchbegriff ist das Navigieren zwischen den Treffern in der Syntaxdatei möglich. Hinter dem Lupensymbol neben den Pfeilen befindet sich eine Suchen und Ersetzen Funktion. Nach dem Klicken auf dieses Symbol kann ein Suchbegriff eingegeben werden. Es werden alle Zeilen im Editor angezeigt, in denen dieser Suchbegriff vorkommt. Wird zusätzlich noch ein Begriff unter Ersetzen eingetragen, werden alle gesuchten Begriffe durch den neuen Begriff ersetzt.

### Syntax-Highlighting

Unter Syntax-Highlighting ist zu verstehen, dass in der Syntax zentrale Befehle von Statistikprogrammen farblich hervorgehoben werden. Dies erhöht die Übersichtlichkeit einer Syntaxdatei. Syntax-Highlighting wird für die Programme R, SPSS und Stata unterstützt. Unter Syntax-Highlighting für kann die gewünschte Programmiersprache ausgewählt werden, voreingestellt ist Stata.

### Syntax Layout

Das Layout der Syntaxdatei lässt sich durch Schriftgröße und Farbschema ändern. Neben dem Knopf Schriftgröße lässt sich die Darstellungsgröße des Syntaxinhalts ändern. Unter Farbschema kann die Farbe geändert werden.

### 5.3 Anwendungsfall 3: Metadaten zur Syntax anzeigen und bearbeiten

Eine vorhandene Syntaxdatei, zu der Metadaten eingegeben werden sollen, kann über die Übersichtsbbox „Zuletzt verwendete Syntaxen“ durch Anklicken im Editor geöffnet werden. Syntaxen können auch über die die Archivsuche erschlossen und soweit entsprechende Rechte vorliegen, von einem Mitglied in der VFU genutzt werden.



Abbildung 23

Nach dem Öffnen der zuletzt verwendete Syntax „test\_01“ öffnet sich die folgende Oberfläche.

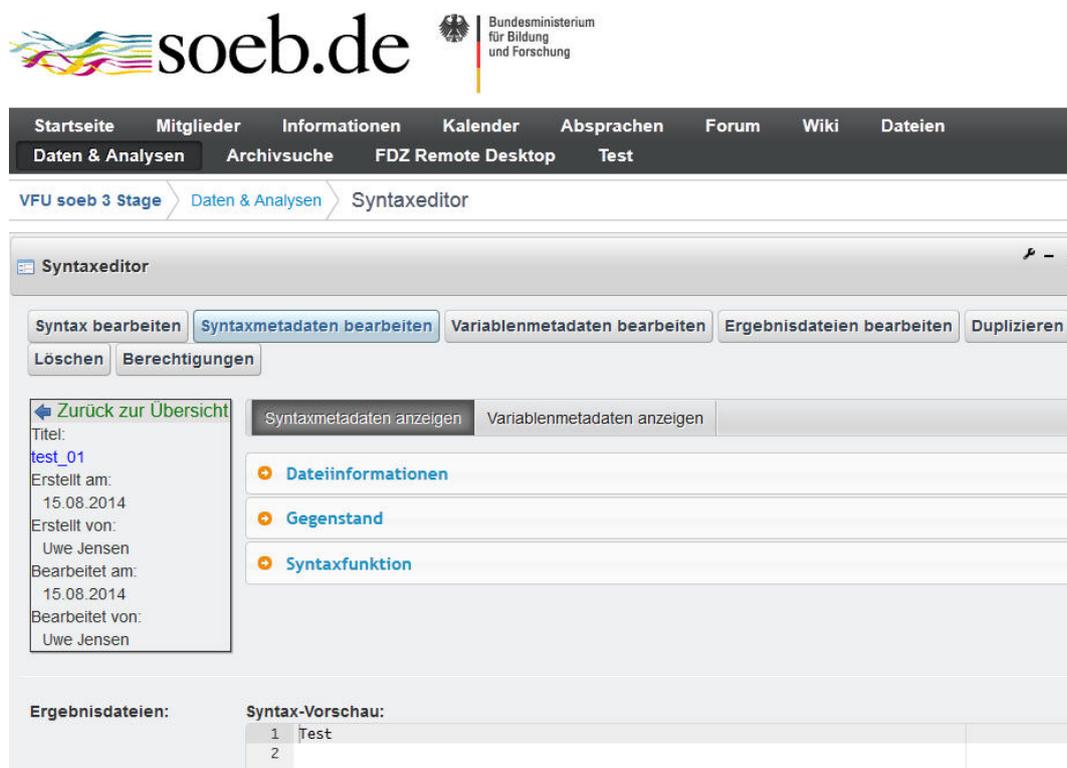


Abbildung 24

In dieser Ansicht des Syntaxeditors (Abb. 24) werden zunächst die Metadaten einer zuvor geöffneten Syntax und die Syntaxinhalte nur angezeigt, d.h. sie sind in dieser Oberfläche nicht editierbar.

1. Syntaxmetadaten werden innerhalb der folgenden Gruppen angezeigt:
  - Dateiinformationen / Gegenstand / Syntaxfunktion
2. Metadaten zu Variablen werden angezeigt, wenn diese zusätzlich aus Ausgangsvariablen in dieser Syntax erzeugt wurden.
3. Im unteren Ausschnitt zeigt das Syntax-Vorschau Fenster den Inhalt der Syntax.
4. Wurden Ergebnisdateien zu dieser Syntaxdatei in die VFU hochgeladen, werden diese unter dem Eintrag „Ergebnisdateien:“ angezeigt.

### Funktionen zur Bearbeitung von Syntax im Überblick

Im oberen Bereich der Oberfläche zur Syntaxbearbeitung (Abb. 27) sind sieben Funktionsbutton platziert. Mit diesen Befehlen können Metadaten, Inhalte der Syntax oder Dateiobjekte der geöffneten Syntax bearbeitet werden.

Die Funktion Syntax bearbeiten wurde in Kap. 5.2 bereits vorgestellt. Die Funktion zur Bearbeitung der Syntaxmetadaten und die dazugehörigen Metadatenfelder werden in diesem Kapitel beschreiben. Die Darstellung der Funktionen „Variablenmetadaten bearbeiten“ erfolgt dann in Kapitel 5.4.

Die Funktion „Ergebnisdateien bearbeiten“, „Duplizieren“ (des aktuellen Objektes Syntax), „Löschen“ (des aktuellen Objektes Syntax) sowie „Berechtigungen (für das aktuelle Objektes Syntax erteilen)“ erfolgt übergreifend für alle Objekte (Studie, Datennutzung, Syntax) in Kapitel 6.

Über den Befehl „Syntaxmetadaten bearbeiten“ wird die folgende Oberfläche geöffnet. Hier können die Metadaten zur geladenen Syntaxdatei erfasst oder aktualisiert werden. Der Inhalt der Syntax wird als Kontext (nicht editierbar) im Oberflächenbereich „Syntax Vorschau“ angezeigt.

The screenshot displays the 'Syntaxeditor' interface on the soeb.de website. At the top, there is a navigation menu with items like 'Startseite', 'Mitglieder', 'Informationen', 'Kalender', 'Absprachen', 'Forum', 'Wiki', and 'Dateien'. Below this is a secondary menu with 'Daten & Analysen', 'Archivsuche', 'FDZ Remote Desktop', and 'Test'. The breadcrumb trail indicates the current location: 'VFU soeb 3 Stage > Daten & Analysen > Syntaxeditor'. The main content area is titled 'Syntaxeditor' and contains a 'Speichern' button, a 'Zurück zur Syntax-Vorschau' link, and three expandable sections: 'Dateiinformationen', 'Gegenstand', and 'Syntaxfunktion'. On the left side, there is a metadata box for the file 'test\_01\_hochladen', showing details such as 'Erstellt am: 15.08.2014' and 'Erstellt von: Uwe Jensen'.

Abbildung 25

Die Metadatenfelder werden durch die folgenden Gruppen strukturiert:

- Dateiinformationen,
- Gegenstand und
- Syntaxfunktion

Die Nutzeroberfläche und Metadatenfelder dieser Gruppen werden im nächsten Abschnitt dargestellt.

### 5.3. Metadatenfelder zur Syntaxdatei

#### Gruppe 1 - Dateiinformationen

The screenshot shows a web-based metadata editor for a syntax file. It features a header with a yellow bar and the text 'Dateiinformationen'. Below this, there are several input fields: 'Syntaxtitel' with the value 'Testsyntax', 'Kurzbeschreibung' (empty), 'Versions-Beschreibung' (empty), 'Versionsdatum' with the value '28.05.2014', and 'Statistiksoftware' with a dropdown menu set to 'STATA'.

Abbildung 26

#### 5.3.1. Syntaxtitel – erforderliche Angabe

Titel der Syntaxdatei.

#### 5.3.2. Kurzbeschreibung

Allgemeine Beschreibung der Aufgabe und des Ziels der Syntax (Abstract). Eine detaillierte Beschreibung kann im Abschnitt „Syntax: Funktion“ erfasst werden.

#### 5.3.3. Versions-Beschreibung

Angaben zu den inhaltlichen Änderungen seit dem letzten Versionsdatum (Versionsbeschreibung).

#### 5.3.4. Versions-Datum

Datum, an dem die Angaben zur Syntax zuletzt inhaltlich aktualisiert wurden (Versionsdatum). Voreingestellt ist das aktuelle Datum.

#### 5.3.5. Statistiksoftware – Kontrolliertes Vokabular

Namen der Statistiksoftware, mit der die Syntax läuft.

- SPSS
- Stata
- R

## Gruppe 2 – Gegenstand

Abbildung 27

### 5.3.6. Gegenstand/ referenzierte Datennutzung

Im Feld Gegenstand/ referenzierte Datennutzung wird die in der VFU gespeicherte Syntaxdatei mit einer Beschreibung zur Datennutzung eines Nutzers/ einer Nutzerin verknüpft.

Ziel dieser Verknüpfung ist es, in den Ergebnissen der Archivsuche anzeigen lassen zu können wer für welchen Untersuchungsgegenstand welche Syntaxdateien hochgeladen hat. Für die Verknüpfung zwischen Untersuchungsgegenstand und Syntaxdatei wird auf die Lupe geklickt und in das aufklappende Suchfeld der Titel der Datennutzung eingegeben. (vgl. Abb. 28).

Abbildung 28

Hinweis: Der Titel der Datennutzung muss exakt eingegeben werden, damit die Datennutzung gefunden werden kann. Es empfiehlt sich, den Namen der Datennutzung mit Strg + C und Strg + V in das Suchfeld zu kopieren.

Zur Erleichterung der Erfassung kann der Syntaxtitel auch in der Archivsuche (vgl. Kap. 8) gefunden und dann kopiert werden. Dazu kann die Archivsuche in einem weiteren Browserfenster geöffnet werden.

### 5.3.7. Arbeitspaket

Name des soeb 3 Arbeitspaketes, in dem der Untersuchungsgegenstand erfasst wurde. Hier sind bisher alle AP von soeb 3 mit Nummern erfasst (AP 01 bis AP 18).

## Gruppe 3 – Syntaxfunktion

Abbildung 29

### 5.3.8. Funktion – Kontrolliertes Vokabular

Beschreibung der Aufgabe bzw. Funktion der Syntax im Workflow der Datenanalyse mit Hilfe von Kategorien. Es können folgende Kategorien ausgewählt werden, wobei eine Syntaxdatei immer nur eine Funktion haben kann:

- Datenselektion                      Subset aus Ausgangsdaten / Variablen erstellen
- Datenaufbereitung                Variablen recodieren etc.
- Datenverknüpfung                Analysefile erstellen
- Datenauswertung                 Analyse mit Syntax ausführen

### 5.3.9. Beschreibung Datenselektion

Fachliche Beschreibung welcher Datensatz wie eingelesen wird (Inputdaten).

### 5.3.10. Beschreibung Datenaufbereitung

Fachliche Beschreibung welche Variablen wie angepasst werden (Recodes etc.).

### 5.3.11. Beschreibung Datenverknüpfung

Fachliche Beschreibung welcher Analysefile erstellt wird.

### 5.3.12. Beschreibung Datenauswertung

Fachliche Beschreibung welche Analysemethode wie angewendet wird.

Die erfassten oder geänderten Metadaten werden abschließend durch „Speichern“ im VFU soeb 3 Archiv gesichert und mit einer Meldung bestätigt.

#### 5.4 Anwendungsfall 4: Metadaten neu erzeugter Variablen erfassen

Der Syntaxeditor bietet die zusätzliche Möglichkeit, Metadaten gesondert zu einzelnen Variablen einer Syntax zu erfassen, die aus Ausgangsvariablen neu erzeugt bzw. modifiziert wurden.

Die Syntaxdatei, zu der Variablenmetadaten eingegeben werden sollen, wird über die Übersichtsbox „Zuletzt verwendete Syntaxen“ oder die Archivsuche ausgewählt.



Abbildung 30

Nach dem Öffnen der zuletzt verwendete Syntax „test\_01“ öffnet sich die folgende Oberfläche.

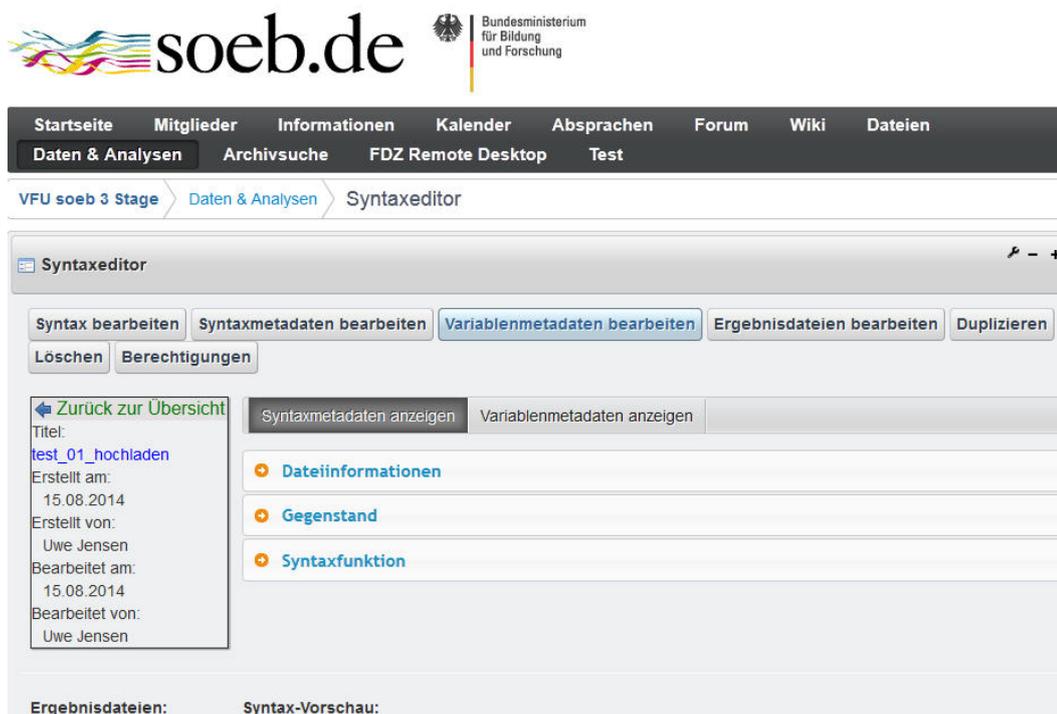


Abbildung 31

Nach Klicken auf den Knopf „Variablenmetadaten bearbeiten“ wird die folgende Oberfläche geöffnet.



Abbildung 32

- Durch „Neue Variable beschreiben“ wird eine neue Variable angelegt, d. h. es werden erstmalig Metadaten zu dieser Variable erfasst.

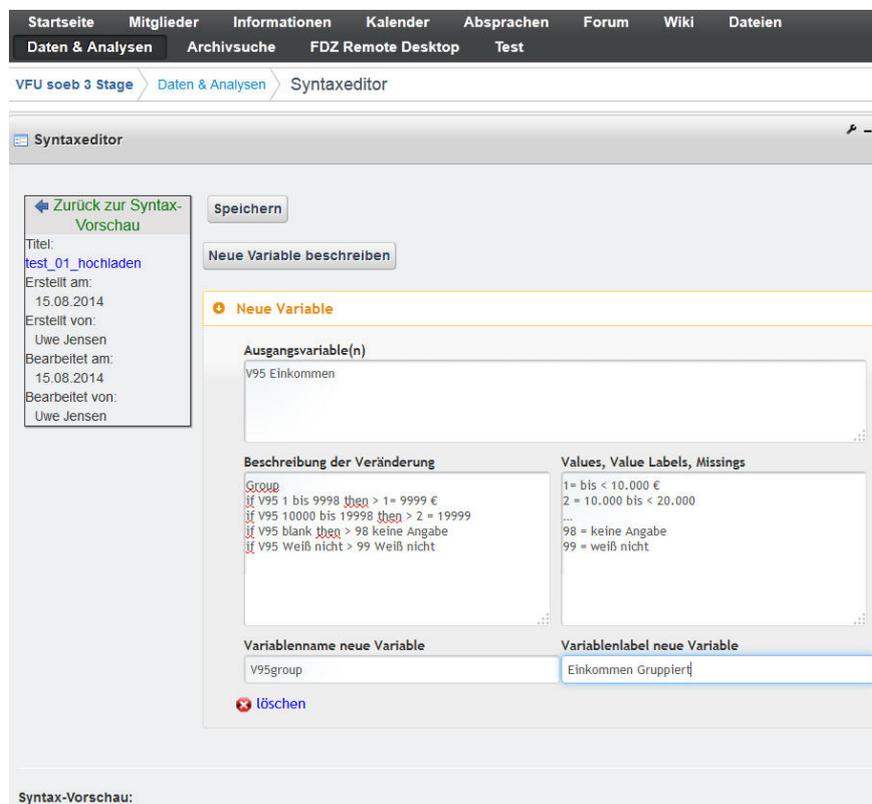


Abbildung 33

Die Metafelder „Neue Variable“ werden im Folgenden beschreiben.

#### 5.4. Metadatenfelder Neue Variable einer geöffneten Syntaxdatei

##### 5.4.1. Name Ausgangsvariable

Name der Variable(n), die zur Bildung einer neuen Variable verwendet wurden.

Beispiel: Aus den Variablen V1 und V2 wird die Variable V3 gebildet. An dieser Stelle sind als Ausgangsvariablen V1 und V2 einzutragen.

##### 5.4.2. Beschreibung der Veränderung

Fachliche Beschreibung welche Variablen wie angepasst werden (Recodes etc.).

##### 5.4.3. Values, Value Labels und Missings

Die Werte und zugehörige Labels der Variable werden in einem Freitextfeld beschrieben.

Beispiel Einkommensklassen:

- 1 = bis < 10.000 €
- 2 = 10.000 bis < 20.000
- ...
- 98 = keine Angabe
- 99 = weiß nicht

##### 5.4.4. Variablen Name neue Variable

Name der neuen Variable.

##### 5.4.5. Variablen Label neue Variable

Label der neuen Variable.

Es können mehrere Variable nacheinander angelegt werden. Die erfassten oder geänderten Metadaten werden abschließend durch „Speichern“ im VFU soeb 3 Archiv gesichert.

Das Speichern wird mit einer Meldung bestätigt, die in der nächsten Oberfläche angezeigt wird.

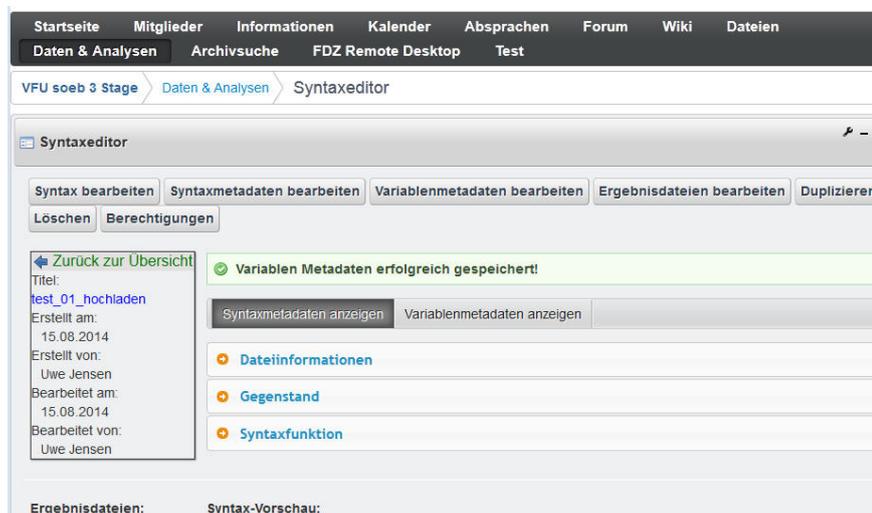


Abbildung 34

---

Gleichzeitig stehen in dieser Ansicht des Syntaxeditors wieder alle sieben Funktionen zur Bearbeitung der Syntax zur Verfügung.

- Mit der Funktion „Ergebnisdateien bearbeiten“ können automatisch erzeugte Dateien an die zugrundeliegende Syntax angehängt werden, die während des Syntaxlaufes entstehen (Protokolldateien, Ergebnisdateien, etc.).

Die in allen Editoren verfügbare Funktion, eine Datei an das Objekt Studie, Datennutzung oder Syntax anzuhängen, wird in Kap. 6.5 beschreiben.

- Die Funktion „Berechtigungen“, die das Teilen von Objekten oder Dateien mit anderen Personen oder Arbeitspaketen ermöglicht, wird zusammenhängend in Kap 6.6 beschrieben.
- Die Funktion „Duplizieren“ erlaubt das Erstellen von 1:1 Abbildern der Originalmetadaten, eines Objektes, die in einem neuen Objekt zum Teil wiederverwendet und zum Teil verändert werden sollen. Dies wird zusammenfassend für alle Objekte in Kapitel 6.7 dargestellt.
- Die Funktion „Löschen“ von Objekten Objekt Studie, Datennutzung oder Syntax beschreibt beispielhaft Kapitel 6.8.

## 6 Gemeinsame Funktionen aller Editoren im Überblick

### 6.1 Tooltips – Nutzerhilfe bei der Erfassung von Metadaten

Alle Felder zur Erfassung von Metadaten in den drei Metadateneditoren werden mit sog. Tooltips inhaltlich erläutert. Die Erklärung wird angezeigt, wenn mit der Maus ein Feld zur Metadatenerfassung berührt wird.

Das Beispiel zeigt die Erklärung zu den Feldern „Versionsbeschreibung“ und „Versionsdatum“ (vgl. auch Kap. 6.4 Versionierung).

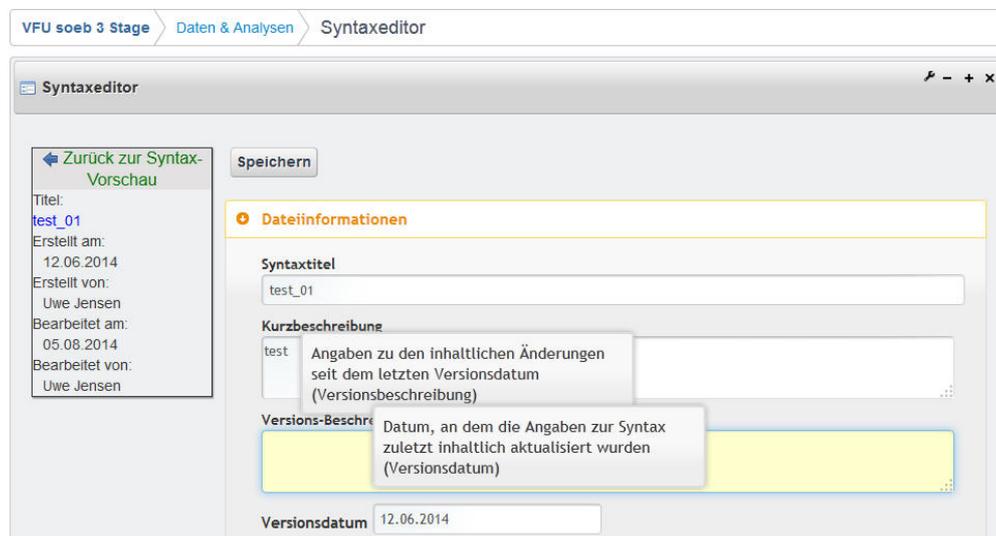


Abbildung 35

### 6.2 Übersichtsbox – Liste der zuletzt verwendeten Objekte

In jedem Editor werden bis zu 10 zuletzt verwendete Objekte in der jeweiligen Startoberfläche in einer Übersichtsbox dargestellt. Darüber hinaus können alle mit den Metadateneditoren erstellten Objekte über die VFU Archivsuche erschlossen werden (vgl. Kapitel 8)

Im Beispiel der unteren Grafik werden alle zuletzt verwendenden Syntaxen im Syntaxeditor angezeigt.



Abbildung 36

### 6.3 Infobox – Anzeige von Metadaten zu einem erstellten Objekt

In allen drei Metadateneditoren gibt es eine Infobox zu einem erstellten Objekt, wie das Beispiel einer Studienbeschreibung des SOEP 1984 bis 2011 zeigt Abbildung 37.

Die Übersichtsbox erscheint, sobald ein neues Objekt mit dem Editor erstellt oder ein bestehendes Objekt (über die Suche oder die Übersichtsbox) aufgerufen wurde.

Neben dem Titel des Objekts, sofern dieser schon eingegeben wurde, werden in der Box Metadaten angezeigt, die beim Erstellen, Speichern und Bearbeiten von Metadaten automatisch aus entsprechenden Systeminformationen erzeugt und angezeigt werden. Diese Metadaten sind

- Erstellt am
- Erstellt von
- Bearbeitet am
- Bearbeitet von

Abbildung 37

Mit der Funktion „Zurück zur Übersicht“ gelangt man zur Startoberfläche des jeweiligen Editors und seinen Grundfunktionen zum Auswählen eines vorhandenen Objektes oder zum Anlegen eines neuen Objektes.

### 6.4 Versionierung – Metadaten vorhandener Objekte aktualisieren

Zunächst wird das zu aktualisierende Objekt (STB, DN; SX) entweder über die Archivsuche oder über die Übersichtsbox zuletzt verwendeter Objekte im entsprechenden Editors geöffnet (vgl. Kap. 6.2 Übersichtsbox).

Die folgende Abbildung zeigt ein Beispiel, bei dem die Syntax „Test\_01“ im Syntaxeditor aufgerufen wurde. Nachdem das zu editierende Objekt ausgewählt wurde, wird über den Button „Syntax Metadaten bearbeiten“ die entsprechende Oberfläche aktiviert.



Abbildung 38

### Versionsangaben erfassen

Das ausgewählte Objekt öffnet sich im Editier-Modus und die Metadaten können bearbeitet werden. Das editierte Objekt kann explizit als neue Version (vgl. nächste Abbildung) beschrieben werden, indem bei der Bearbeitung ein neues:

- „Versionsdatum“ gesetzt wird und eine
- „Versionsbeschreibung“ die Änderungen seit der letzten Version beschreibt.
- Mit Speichern wird das Objekt unter dem gleichem Titel gesichert.

Hinweis: Alle Metadatenfelder werden mit sog. Tooltips inhaltlich erläutert (vgl. Kap. 6.1).

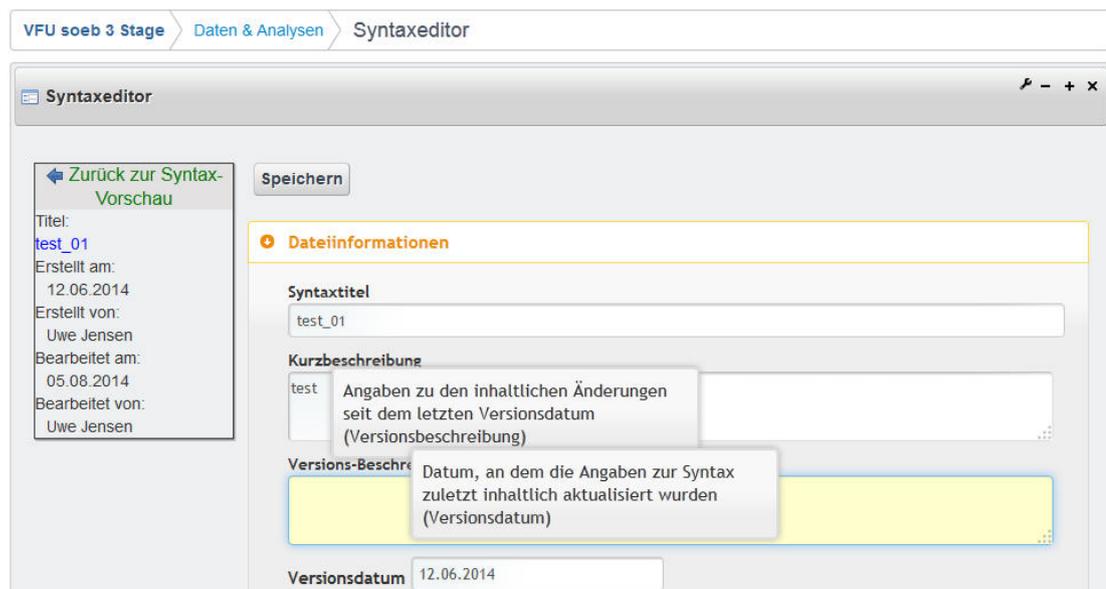


Abbildung 39

## 6.5 Dateien zu Studie, Datennutzung und Syntax hochladen

Zu den Objekten Studie und Syntax können lokale Dateien innerhalb der Funktionalität des jeweiligen Editors in die VFU hochgeladen werden. Selbsterzeugte Datendateien, die logisch zur Datennutzung gehören, werden aufgrund von Nutzerwünschen gesondert über die zusätzliche Funktionalität „Datendatei hochladen/ downloaden“ bearbeitet.

Zunächst wird der Editor des jeweiligen Hauptobjektes ausgewählt,

- zu der eine Datei hochgeladen werden soll (STB, SX) bzw.
- die Funktion „Datendatei hochladen/downloaden“ im Kontext des Objektes DN.

Das folgende Beispiel zeigt den Upload für Studiendokumente im Kontext des Objektes Studienbeschreibung. Die Funktionsweise in den anderen Editoren ist entsprechend angelegt.



Abbildung 40

### Anwendungsfall: Datei zu einer Studienbeschreibung hochladen

Mit Hilfe des Funktionsbutton „Dateien bearbeiten“ öffnet sich eine neue Oberfläche, in der ausgewählt wird ob

- eine Datei mit Material zur Studie (z. B. ein Codebuch) oder
- ein Datenfile der Studie hochgeladen werden soll (soweit rechtlich zulässig).

Im folgenden Beispiel soll ein Dokument zur Studie in die VFU hochgeladen werden.

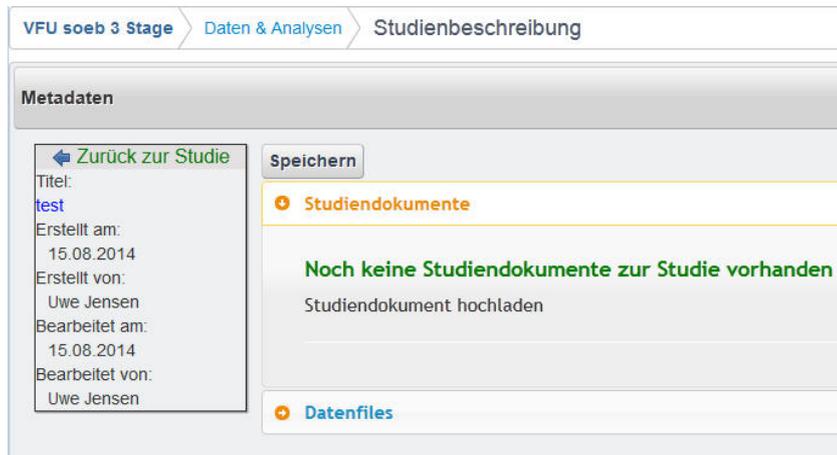


Abbildung 41

Es öffnet sich die Dialogbox „Durchsuchen“, mit der eine lokale Datei durch folgende Schritte ausgewählt und hochgeladen wird:

- Lokales Verzeichnis wählen
- Dokument auswählen und mit „OK & Hochladen“ bestätigen



Abbildung 42

Danach müssen zur hochgeladenen Datei die entsprechenden Metadaten eingegeben werden. Dazu zählen Dateiname, Inhalt, Typ, Nutzungsbedingungen (vgl. nächste Abbildung 43)

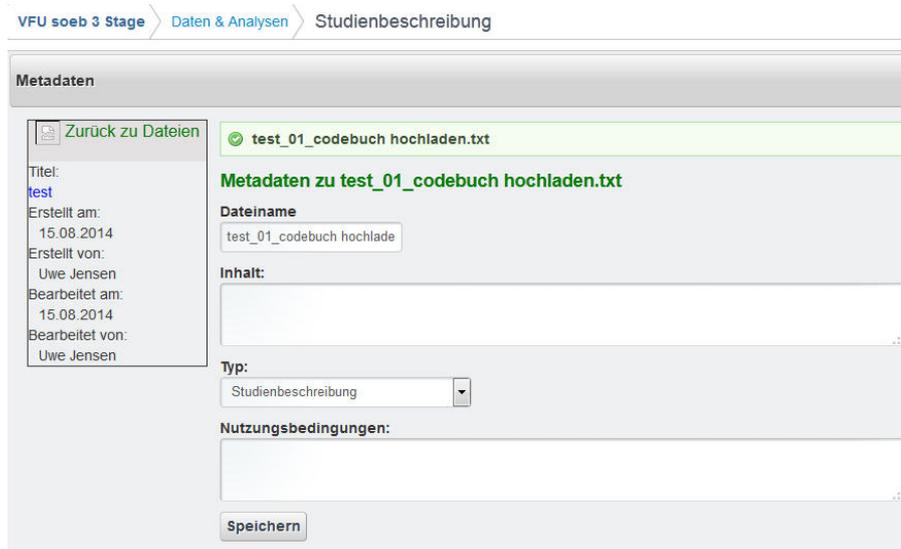


Abbildung 43

## 6.5 Metadatenfelder zu hochgeladenen Dateien

### 6.5.1 Dateiname - erforderliche Angabe

Name der hochzuladenden Datei. Der lokale Name kann übernommen werden.

### 6.5.2 Inhalt

Kurzbeschreibung des Inhalts der Datei.

### 6.5.3 Typ - Kontrolliertes Vokabular

Einordnung des Dokumententyps anhand von Kategorien.

Die vorgegebenen Kategorien unterscheiden sich je nach Editor Studie, Datennutzung (und Datendateien hochladen) und Syntax wie im Folgenden dargestellt.

**Objekt Studie – Dateien mit Material zur Studie- Typ:**

- Studienbeschreibung
- Kurzbeschreibung
- Fragebogen
- Methodenbericht
- Methodenreport
- Methodenbeschreibung
- Erhebungsdesign und –Instrumente
- Codebuch
- Codeplan
- Datenhandbuch
- Schlüsselverzeichnis
- Datensatzbeschreibung
- Datenreport
- Variablenliste

**Objekt Studie – Dateifile einer Studie- Typ:**

- Datenfile ist Teil eines Datensatzes
- Einzel-Datenfile

**Objekt Datennutzung: Datendatei hochladen/ downloaden – Typ:**

- Subdatensatz
- Selbst generierter Datensatz
- Gematchter Datenfile

**Objekt Syntax: zusätzliche Dateien, die während des Syntaxlaufes erzeugt wurden– Typ:**

- Ergebnisdatei
- Logfile

- Datensatz
- Anderer Typ

#### 6.5.4 Nutzungsbedingungen

Nutzungsbedingungen des Dokuments in der VFU laut Urheber des Dokuments.

Die neu erfassten oder geänderten Metadaten werden abschließend durch „Speichern“ im Archiv gesichert.

### 6.6 Berechtigungen – Objekte und Dateien mit VFU Mitgliedern teilen

Mit den Metadateneditoren erstellte oder hochgeladene Objekte können im Kontext der Funktionalität des jeweiligen Editors nur vom jeweiligen ‚Besitzer‘ des Objektes mit den standardmäßig erteilten erweiterten Rechten bearbeitet und verwaltet werden:

- Lesen, schreiben, Objekte Duplizieren, Inhalte kopieren
- dem Objekt eine Datei hinzufügen oder eine zugehörige Datei löschen
- Hauptobjekt löschen
- Download einer Datei
- Freigeben
- Verwalten (selber Rechte vergeben)

Die Rechte zum Lesen, Schreiben und Verwalten (Erteilen der Rechte Lesen, Schreiben etc.) von Objekten wurden durch umfassende Regeln zum Rechtemanagement in der VFU soeb 3 definiert. Dabei wurde auch festgelegt, dass alle Metadaten – im Gegensatz zu Inhalten in einer Datei – für alle registrierten Personen in der VFU sichtbar sind. Grundlegender Gedanke ist, dass die jeweiligen Inhalte von Dateien besonderes Wissen enthalten und damit geistiges Eigentum darstellen. Dies betrifft im speziellen die Dateien zu den Objekten Datennutzung und Syntax, die von den Forschenden in der VFU als Ergebnis ihrer Forschungstätigkeit entwickelt werden.

#### Anwendungsfall: Syntaxen mit VFU Mitgliedern und Arbeitspakete teilen

Wie erwähnt, können alle VFU Mitglieder gemäß den Standardeinstellungen die Metadaten beispielsweise zu einer Syntax, aber nicht den Inhalt der Syntax einsehen. Das Downloaden der Datei oder das Kopieren des Syntaxinhalts ist selbstverständlich auch nicht möglich.

- Mit der Funktion „Berechtigungen“ kann diese Voreinstellung geändert werden.

Dazu wird über den zuständigen Editor das entsprechende Objekt ausgewählt. z. B. eine Syntaxdatei, für die neue Berechtigungen gesetzt werden sollen.



Abbildung 44

Über den Funktionsbutton „Berechtigungen“ (Abbildung oben) öffnet sich eine neue Oberfläche mit allen Nutzern, wie die nächste Abbildung zeigt.

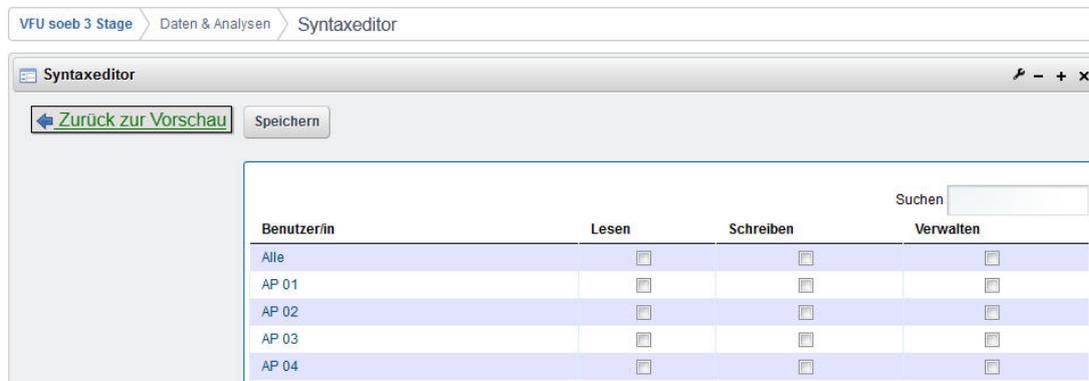


Abbildung 45

### Personenliste und Rechtevergabe

Die Oberfläche enthält eine Liste mit allen in der VFU registrierten Einzelpersonen (in der Abbildung nicht sichtbar) und allen definierten soeb 3 Arbeitspaketen.

Zu jedem Eintrag „Benutzer/in“ gibt es eine Box zum Aktivieren der Rechte „Lesen“, „Schreiben“ und „Verwalten“.

Einzelpersonen und Arbeitspaketen können entsprechende Rechte, z. B. zum Teilen von Syntaxen, auf verschiedene Weise erteilt werden.

#### Fall 1: Rechte selektiv vergeben

- Dazu werden Lesen-, Schreiben- und Verwalten-Rechte hinter den jeweiligen Namen durch ein Häkchen in der entsprechenden Rechtebox ausgewählt und mit Speichern bestätigt.

#### Fall 2: Rechte an Alle vergeben

Zur Arbeitserleichterung bei der Rechtevergabe wurde in der ersten Zeile eine Nutzergruppe „Alle“ implementiert.

- Mit dieser Funktion werden allen Personen die jeweils gewünschten Rechte für Lesen, Schreiben und / oder Verwalten durch Aktivierung der entsprechenden Box zugeteilt.

#### Fall 3: Person Suchen und Rechte vergeben

Das Feld „Suche“ dient der Erleichterung, wenn einzelnen Personen Rechte zugewiesen werden sollen (Abbildung 46).

- Werden hier die ersten Buchstaben des Namens eingegeben (z. B. „car“), reduziert sich die Anzeige auf den Namen des Nutzers „Carevic“ auf den die Auswahl im Beispiel unten zutrifft.



Abbildung 46

## 6.7 Duplizieren – Metadaten eines Objektes wiederverwenden

Zur Reduzierung von Arbeitsaufwand können die Objekte STB, DN und Syntax dupliziert werden, wenn sie neu angelegt werden müssen, aber einzelne Metadaten weiter benutzt werden sollen. D. h. ein Teil der erfassten Metadaten wird erhalten, während andere Metadaten verändert werden müssen.

### Anwendungsfall: Metadaten zu einer Syntaxdatei duplizieren

Eine Person hat in der VFU eine Syntaxdatei (Basissyntax) erstellt. Diese Syntaxdatei möchte sie mit anderen Personen teilen. Gleichzeitig muss sie die Basissyntax für die modifizierte Analysezwecke anpassen. Deshalb erstellt sie ein Duplikat der Basissyntax.

- Dazu wird die Syntaxdatei mit der Basissyntax geöffnet. Mit dem Befehl „Duplizieren“ wird ein Duplikat der Ursprungsdatei erstellt, wie die nächste Abbildung gezeigt.
- Die Person wird aufgefordert, dem Objekt einen neuen Namen unter „Neuer Titel“ zu geben. Dies ist eine ‚Muss Angabe‘, weil das Objekt sonst nicht dupliziert werden kann. Anschließend wird das Duplikat mit neuem Titel gespeichert,

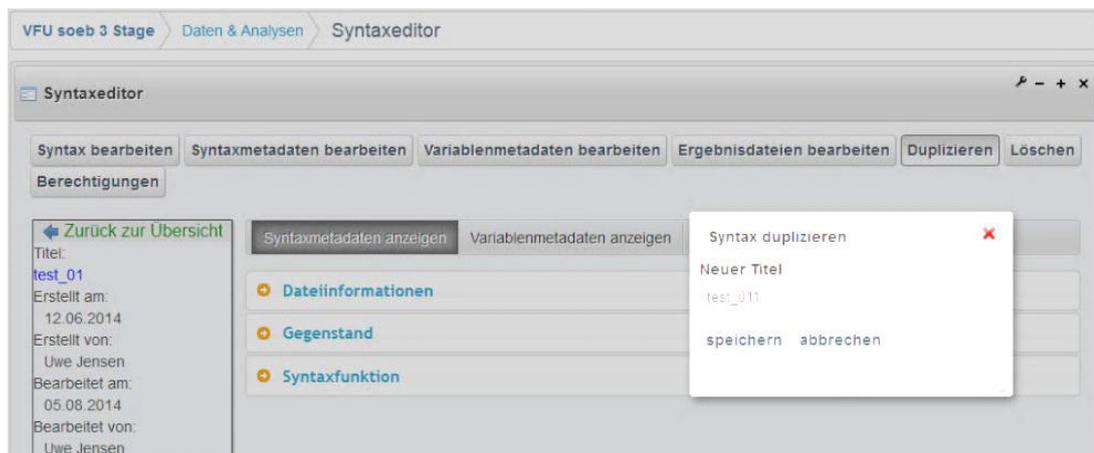


Abbildung 47

Jetzt können die Metadaten und die Inhalte der Syntax mit dem Syntaxeditor(vgl. Kapitel 5) angepasst werden, die für den besonderen Analysezweck notwendig sind.

### History Funktion

Nach dem Duplizieren kann die Ursprungssyntax mit den zugehörigen Metadaten über einen Link in der Infobox „Basiert auf“ aufgerufen werden; z. B. in der Grafik links unten „test\_01“

Damit ist es möglich, Metadaten der Ursprungssyntax aufrufen sowie die Syntax auch direkt zu öffnen (vgl. heller Ausschnitt in der folgenden Abbildung).

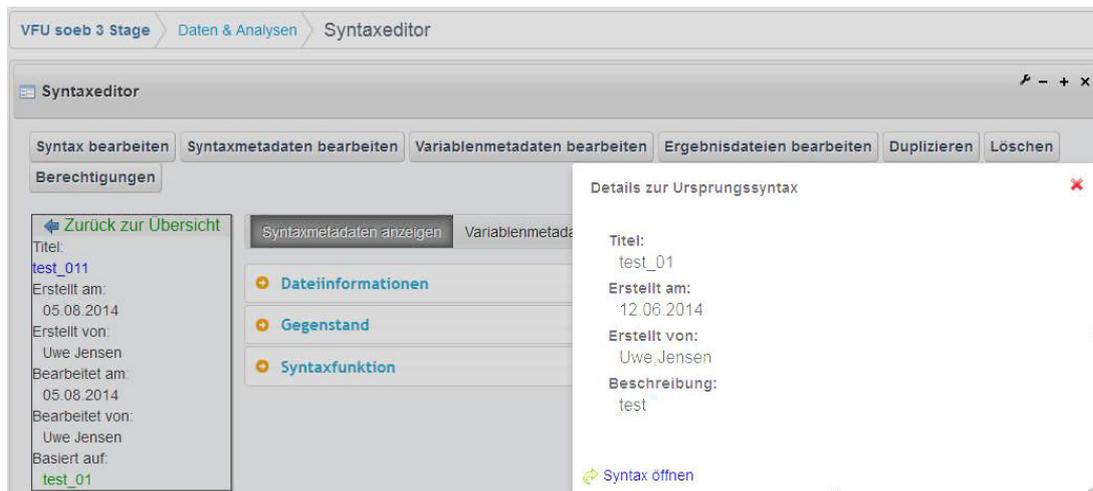


Abbildung 48

## 6.8 Löschen eines Objektes oder einer zum Objekt gehörenden Datei

Es können sowohl ganze Hauptobjekte (STB, DN, SX) komplett sowie einzelne Dateien, die einem der Hauptobjekte zugeordnet sind, selektiv gelöscht werden.

Beim Löschen eines Hauptobjekts werden auch alle mit dem Objekt verknüpften Dateien gelöscht.

### Anwendungsfall: Datei zu einer Studienbeschreibung löschen

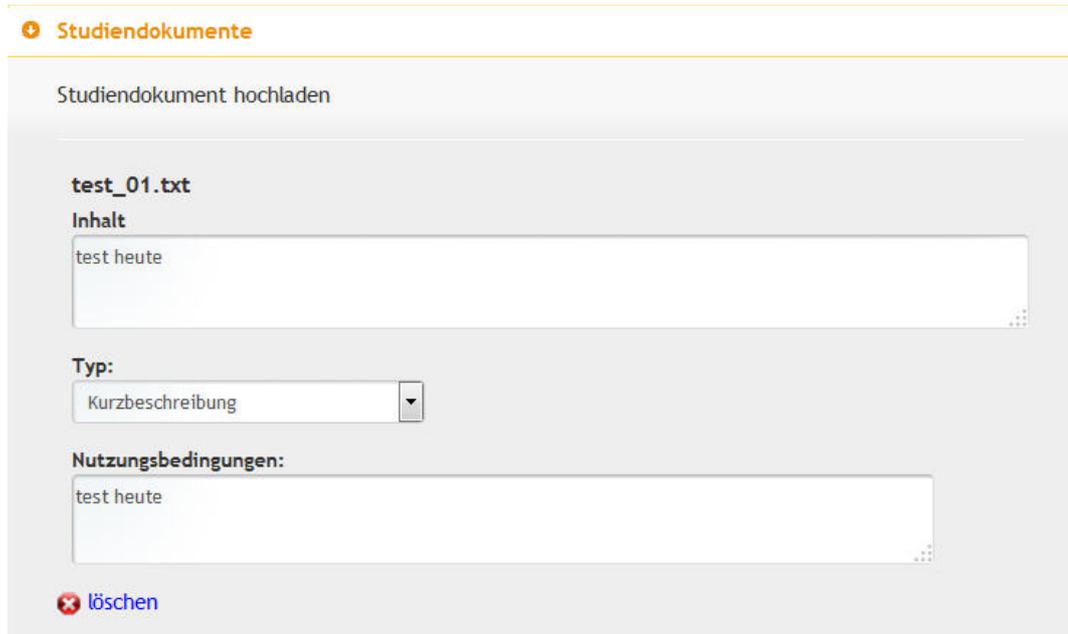
Zunächst wird das Hauptobjekt über den üblichen Funktionspfad ausgewählt, z. B. eine Studienbeschreibung:

- Daten & Analysen > Studienbeschreibung > Zuletzt verwendetes Objekt auswählen
- „Dateien bearbeiten“ wählen (Hinweis: „Löschen“ betrifft die gesamte STB)



Abbildung 49

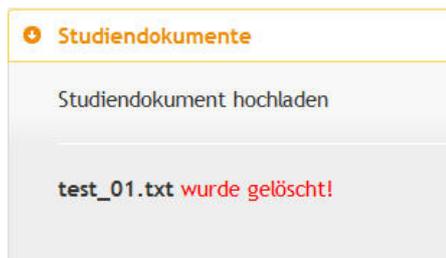
- Nach dem Ausführen des Befehls „Dateien bearbeiten“, kann z. B. eine Datei des Typs „Studiendokumente“ ausgewählt werden. Die zugehörigen Metadaten werden angezeigt. Hinweis: Alternativ könnte bei einer Studienbeschreibung auch eine Datei vom Typ „Datenfile“ ausgewählt werden.



The screenshot shows a web interface for managing study documents. At the top, there is a header with a plus icon and the text 'Studiendokumente'. Below this is a section titled 'Studiendokument hochladen'. The main content area displays a document titled 'test\_01.txt'. Underneath the title, there is a section labeled 'Inhalt' with a text area containing 'test heute'. Below the content, there is a 'Typ:' label followed by a dropdown menu currently set to 'Kurzbeschreibung'. At the bottom of the document entry, there is a section labeled 'Nutzungsbedingungen:' with another text area containing 'test heute'. At the very bottom left of the document entry, there is a red 'X' icon followed by the text 'löschen'.

Abbildung 50

- Das Objekt wird über die Funktion „Löschen“ unwiderruflich gelöscht. Die erfolgreiche Aktion wird mit einem Informationstext bestätigt.



The screenshot shows the same 'Studiendokumente' interface as in the previous image. The document entry for 'test\_01.txt' is no longer visible. Instead, a red text message 'test\_01.txt wurde gelöscht!' is displayed in the main content area, indicating that the document has been successfully deleted.

Abbildung 51

## 7 Die Einbettung der Metadateneditoren in die VFU

Das folgende Diagramm beschreibt den Aufbau und Integration der Metadateneditoren in die technische Umgebung der VFU.

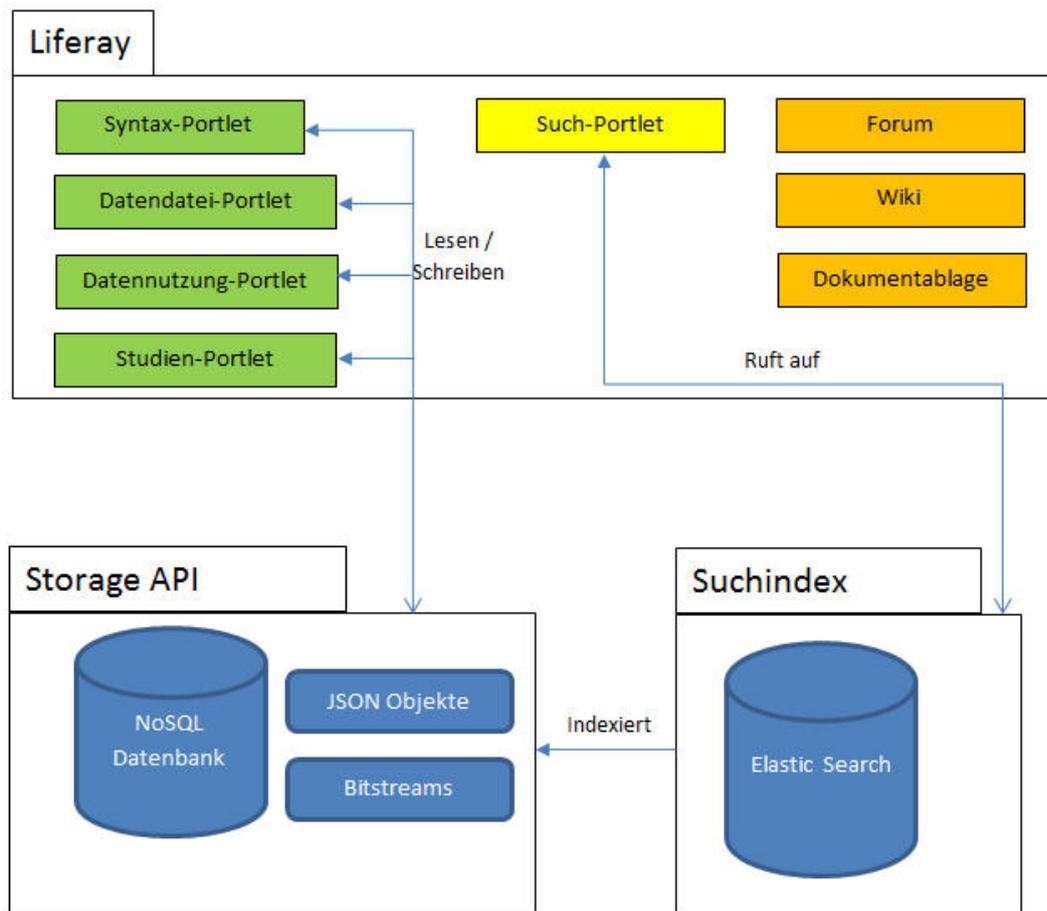


Abbildung 52

Die Editoren wurden als *Spring MVC Portlets* (Version 3.2.0) für die Kooperationsplattform *Liferay* (Version 6.2) entwickelt. Ein Portlet ist ein Teilmodul innerhalb von Liferay und greift regelmäßig auf verschiedene Liferay Funktionalitäten und andere maßgeschneiderte VFU-Module zu. Insbesondere gilt dies für die Datenhaltung und die Suche, deren Zusammenspiel im Folgenden erläutert werden soll. Die Module selbst wurden von der GWDG entwickelt, die für die Middleware-Entwicklung sowie die Basisfunktionalität des eingesetzten Portals zuständig war (vgl. Pohl. 2014).

Im Editor erstellte oder veränderte Objekte werden vom Editor in das JSON-Format gewandelt und an die Storage API geschickt. Nach Bedarf werden weitere Objekte aus dem Storage ebenso im JSON-Format gelesen, angezeigt und können dann im Editor bearbeitet werden. Das folgende Listing ist ein verkürztes Beispiel für ein Syntax-Objekt.

```
{
  "id": "IDdib7jeh",
  "editorFirstName": "Zeljko",
  "editorLastName": "Carevic",
  "created": "12.06.2014",
  "title": "Testobjekt",
  "version": "12.06.2014",
  "resultFileCollection": "IDouh57du",
  "typ": "syntax",
  ....
}
```

Abbildung 53: Listing JSON Darstellung einer Beispielsyntax

Sämtliche Operationen zum Lesen, Schreiben, Löschen und Verwalten von Inhalten laufen über die Storage-API, die Metadaten, Zugriffsberechtigungen und Bitstreams in einer NoSQL Datenbank speichert. Komplexere Datenstrukturen, wie Listen, Mengen und Unterobjekte werden durch Referenzierungen gelöst.

Um im Editor eine Referenzierung zwischen den unterschiedlichen Objekten zu ermöglichen, werden von der Storage-API jedem Objekt eindeutige IDs zugewiesen. Diese ID wird dann direkt in die Metadaten eingetragen, um die Referenzierung zu ermöglichen. So verweist die Beispielsyntax aus Listing (Abb. 53) auf eine Collection, die eine Liste von l's zu Ergebnisdateien (resultFileCollection) umfasst.

Die Referenzierung zwischen zusammengehörigen Objekten wie beispielsweise einer Syntax und seinen Ergebnisdateien erfolgt automatisch, um den Forschenden Arbeit zu ersparen. Referenzen die automatisch erstellt werden sind:

- Ergebnisdateien und Variablen für den Syntax-Editor
- Studiendokumente und Studiendateien für den Studien-Editor
- Selbst generierte Datendateien für den Datendateien-Editor

Einige Referenzierungen zwischen Objekten, beispielsweise zwischen einer Syntax und einer Datennutzung, müssen explizit durch den Nutzer / die Nutzerin vorgenommen werden. Zu diesem Zweck sucht der Anwender über einen Suchschlitz nach dem gewünschten Objekte und wählt dieses aus. Bei der Suche werden nur Objekte dargestellt, die eine Referenzierung zu dem Ursprungsobjekt erlauben (bspw. Syntax -> Datennutzung).

Durch die Verwendung von JSON als Ablageformat für Metadaten ist perspektivisch auch ein Exportieren der Metadaten möglich. Dies sollte dann aber in der Storage API realisiert werden.

Ein weiteres besonderes Merkmal von JSON ist die Flexibilität in der Datenstruktur, die von der Storage API auch voll unterstützt wird. Das Metadatenschema kann sich mit der Zeit ändern. Im Gegensatz zu herkömmlichen Datenbanken schlagen sich diese Änderungen bei No-SQL-Datenbanken jedoch nicht in der Datenbank selbst nieder, sondern können komplett unabhängig vom Datenhalter durchgeführt werden.

Zu diesem Zweck ist ein Migrationsskript entwickelt und getestet worden, dass bestehende Metadaten-Informationen auf ein aktualisiertes Metadatenschema migriert. Die Editoren müssen dann selbstverständlich auch angepasst werden, ebenso die Suche.

Die Suche wird in der VFU mit der *Elastic Search* Technologie realisiert. Diese indexiert sämtliche gespeicherten Metadaten über alle Objekttypen. Im Falle von Änderungen müssen die Angaben zu den relevanten Metadaten natürlich angepasst werden. Davon abgesehen funktioniert die Suche weitestgehend getrennt von den Editoren.

## 8 Gesucht, Gefunden: Metadaten im VFU Archiv praktisch nutzen

---

Die mit den Metadateneditoren erfassten Informationen über Syntaxen, methodische Strategien der Datennutzung und den genutzten Ausgangsdaten im soeb 3 Forschungsverbund dienen verschiedenen Zielen.

Sie unterstützen einmal den täglichen Forschungsbetrieb mit seinen üblichen Arbeitsabläufen bei der Durchführung kollaborativer Datenanalysen. Die Tätigkeiten werden von achtzehn wissenschaftlichen Arbeitspaketen von über fünfunddreißig Mitgliedern aus siebzehn wissenschaftlichen Einrichtungen durchgeführt. Die durch die Forscherinnen und Forscher erfassten Metadaten und Dateien zur Entwicklung und Durchführung von Analysen zu bestimmten Forschungsfragen und Untersuchungsgegenständen werden im Archiv der VFU dauerhaft und langfristig gespeichert. Gleichzeitig werden die analysezentrierten Informationen zu Ausgangsdaten, Datennutzung und Syntax durch spezifische Archivfunktionen systematisch miteinander verknüpft, um den Forschungsprozess durch neue Möglichkeiten gemeinsamen Arbeitens an Datenanalysen zu erleichtern und zu erweitern.

Das Archivsystem und seine besonderen Funktionen wurden aus den genannten Gründen zusätzlich zu dem ebenfalls verfügbaren VFU Dateisystem entwickelt. Dieses projektbezogene Dateisystem speichert, verwaltet und erschließt analog der Dateiablage eines Betriebssystems alle Arten von nicht analysebezogenen Dateien und Metadaten, die im soeb 3 Forschungsverbund während der Projektlaufzeit entstehen. Diese Projektdokumente können auch mit den anderen VFU Mitgliedern geteilt und kollaborativ bearbeitet werden. Auf das Dateisystem und seine besonderen Funktionen wird hier nicht weiter eingegangen.

Der Mehrwert der Archivfunktionen entwickelt sich in der täglichen Forschungstätigkeit durch die (wachsende) Anzahl und inhaltliche Qualität der erfassten Metadaten auf verschiedenen Ebenen. So können für Datenanalysen genutzte Ausgangsdaten, Untersuchungsziele, Strategien der Datennutzung, methodische Analyseparameter, eingesetzte statistische Methoden und die konkret benötigten Syntaxen und deren Entwicklung systematisch erfasst und durch Archivfunktionen verknüpft werden.

Die Verknüpfung der drei Hauptobjekte, deren Metadaten und Dateien sowie ihrer Autorinnen und Autoren stellt ein umfassendes Netzwerk erschließbaren Wissens dar, dass gleichzeitig die daten- und analysebezogenen Arbeitsprozesse nachhaltig sichert und nachvollziehbar für die im Projekt tätigen Wissenschaftlerinnen und Wissenschaftler zur gemeinsamen Nutzung bereitstellt.

So können eigene Syntax-Dateien bei Bedarf innerhalb eines Arbeitspaketes gemeinsam bzw. abwechselnd bearbeitet und entwickelt werden. Gleichzeitig können sie mit anderen Forschenden unter Wahrung der eigenen Autorenschaft gezielt geteilt und weiterentwickelt werden.

Die zu einer Syntaxdatei gehörenden Ergebnis- und Protokolldateien gestalten bereits erzielte Ergebnisse auch für potentielle Nachnutzer der Syntax in transparenter Weise. Umgekehrt lassen sich die zugrunde liegenden Syntaxen zu den Ergebnisdateien über die Archivsuche erschließen. Weiterhin können auch Teilinformationen etwa zu Bildungsvorschriften und Konzepte selbst generierter Variablen einer Syntax gefunden werden. Schließlich kann auch für neu konstruierte Variablen, wie etwa bei Typologien, eine gemeinsame Codierungspraxis zur Erleichterung von Harmonisierungsbemühungen abgesprochen werden. Eine solche Standardvariable bzw. Syntaxmodul zu ihrer Erzeugung können dann zur gemeinsamen Nutzung im Archiv abgelegt und zur Nachnutzung gesucht und bereitgestellt werden.

Durch die Verknüpfung einer Syntax mit einer Datennutzung werden Informationen zum Analysedesign genauso transparent angeboten. Schließlich verweist die Beschreibung einer Datennutzung auch auf die Ausgangsdaten sowie die Besonderheiten ihrer Nutzung und insbesondere auf die Version des

tatsächlich genutzten Datensatzes. Oder es wird auf mögliche Besonderheiten eines selbst zusammengestellten Datensatzes hingewiesen, der mit dieser Datennutzung verbunden und im Archiv gespeichert ist.

Diese Dateien können, soweit rechtlich zulässig, in der VFU gespeichert werden und z. B. aus generierten Subsets auf Basis von Ausgangsdaten bestehen oder selbst erzeugte Datensätze darstellen. Solche Daten basieren z. B. auf Datenbanken, die vom Datengeber nicht nutzungsbeschränkt angeboten werden, wie z. B. Daten aus der Genesis-Online Datenbank des statistischen Bundesamtes. Solche von Mitgliedern der VFU selbst erzeugten Datensätze können nach entsprechender Dokumentation (Editor Datennutzung)) über die Archivsuche erschlossen und bei Bedarf anderen Personen oder Arbeitspaketen zur Nutzung bereitgestellt werden.

Auch bei diesem Beispiel erzeugen die archivbasierten Verknüpfungen einen nahtlosen Überblick über den datenbasierten Analysezusammenhang. Dazu werden erstens selbsterzeugte Datenfiles z. B. aus der Genesis Datenbank mit der speziellen Datennutzung verbunden. Zweitens wird von der Datennutzung aus auf die Angaben zu den Ausgangsdaten in der Studienbeschreibung verwiesen (Datensatz: Genesis-Online Datenbank; Datengeber: Statistisches Bundesamtes). Schließlich können Syntaxen, die diese Daten in Analysen benutzen, mit den methodischen Angaben in der Datennutzung verknüpft werden.

So tragen diese Metadaten insgesamt zur Dokumentation des Forschungsprozesses bei. Alle relevanten Quellen des Prozesses einer spezifischen Datenanalyse - Syntax, Datennutzung und Ausgangsdaten - können während der gesamten Laufzeit des Forschungsprojektes systematisch durch die entsprechenden Metadaten von allen Beteiligten erfasst und gesichert werden. Umgekehrt steht das sich schrittweise verdichtende Netzwerk von Metadaten für die kollaborative Entwicklung und Nutzung des Analyseinstrumentariums durch die Mehrwertfunktionen des Archivs bereit. Subsets aus diesen Metadaten können dann auch zur Nachnutzung nach Projektende zur Verfügung gestellt werden.

Die folgende Abbildung stellt die wesentlichen Zusammenhänge der durch die Editoren erfassten Metadaten und deren Verknüpfung dar, die im Archiv gespeichert und durchsucht werden können.

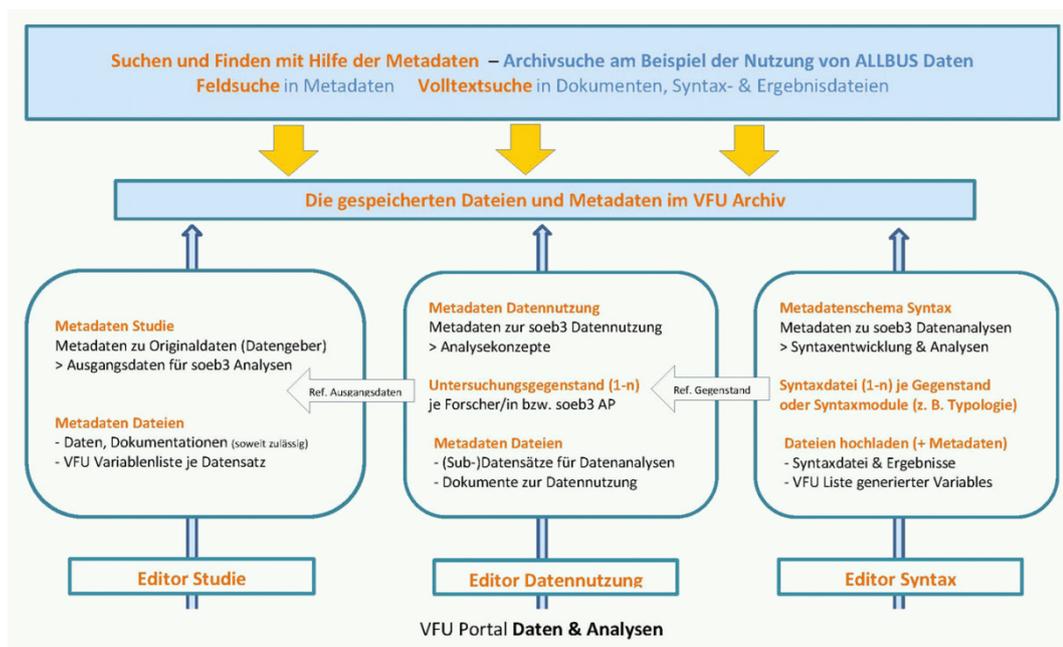


Abbildung 54

Die Suchoberfläche zur Suche in den Metadaten, die mit Hilfe der der Editoren erfasst und im Archiv gespeichert werden, bietet sowohl eine einfache Suche und eine erweiterte Suche an. Sie werden nun kurz beschrieben.

Anschließend werden mögliche Suchergebnisse und die Funktionalität zur Nutzung der Suchergebnisse anhand eines praktischen Beispiels vorgestellt.



Abbildung 55: Die Suchoberfläche „Einfach Suche“ im Archivbestand

In der einfachen Suche können Schlüsselwörter unter Nutzung der Platzhalter „\*“ (ein oder mehr beliebige Zeichen) und / oder „?“ (genau ein Zeichen) eingegeben werden.

Gleichzeitig kann zur Funktion der erweiterten Suche gewechselt werden, die die nächste Abbildung darstellt.



Abbildung 56: Die Suchoberfläche „Erweiterte Suche“ im Archivbestand

Die folgenden Optionen können einzeln oder kombiniert ausgewählt werden:

- Dateityp - Auswahlliste:  
Es kann gezielt nach den Objekten Studien, Syntaxen, Datennutzungen, Datendateien oder Variablen gesucht werden.
- Erstellt von - Freitext:  
Angabe zu einer Person, die das Objekt erstellt hat
- Titel - Freitext:  
Angabe zum Titel eines oder mehrerer Objekte
- Erstelldatum - Kalender:  
Datum, an dem das Objekt erstellt wurde. Das Datum kann über einen Kalender ausgewählt werden.
- Letzte Änderung in der VFU - Kalender:  
Datum, an dem das Objekt zuletzt in der VFU geändert wurde. Das Datum kann über einen Kalender ausgewählt werden.
- Letzte Änderung Datenanbieter - Kalender:  
Datum, an dem das Objekt Studie zuletzt vom Datengeber geändert wurde. Das Datum kann über einen Kalender ausgewählt werden.
- Inhalt - Freitext:  
Angaben zum Inhalt des Objektes
- Bearbeitet von - Freitext:  
Person, die das das Objekt zuletzt bearbeitet und gespeichert hat.
- Arbeitspaket - Auswahlliste:  
Angaben zum Arbeitspaket, dem eine Person angehört, die das Objekt bearbeitet hat.

Jeder Feldeintrag oder gewählte Listeneintrag kann durch folgende Operatoren spezifiziert werden:

- Muss enthalten
- Darf nicht enthalten
- Könnte enthalten

Die Ergebnisse einer Archivsuche werden auf zwei Arten präsentiert (vgl. nächste Abbildung).

Einmal wird das Gesamtergebnis der Suche im ersten Reiter „Alles“ mit Anzahl der Funde („6 Ergebnisse“) abgebildet. Außerdem werden die Funde je Objekt Studien, Syntaxen Datennutzung, Datendateien und Variablen zusammen mit der Anzahl der jeweiligen Funde dargestellt.

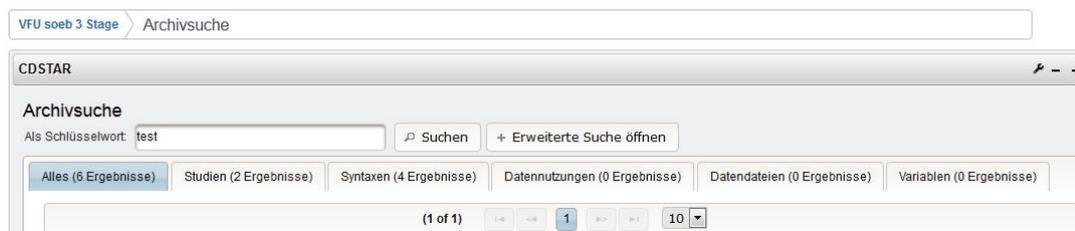


Abbildung 57: Ergebnisdarstellung einer Archivsuche

Die folgenden zwei Abbildungen zeigen die Fülle und Differenzierungsmöglichkeiten am Beispiel von unterschiedlichsten Objekten, die zum ALLBUS (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften) erfasst wurden.

Die folgende Ansicht zeigt fünf (von sechs) gefundenen Objekten; u.a. die Studienbeschreibung des ALLBUScompact – Kumulation 1980 bis 2010. Weiterhin wird das Codebuch zum ALLBUScompact gefunden, das der Studienbeschreibung sachlich zugeordnet und entsprechend mit dieser verknüpft ist.

Abbildung 58: Beispiel Suche „\*ALLBUS\*“ und Anzeige „Alle“ Ergebnisse.

In der vorherigen Abbildung werden weiterhin 3 Elemente angezeigt, die dem Objektbereich „Syntax“ zugeordnet sind. Diese Elemente können auch als eigenständiger Zusammenhang über den Reiter „Syntax“ direkt angezeigt werden.

Abbildung 59: Suche nach \*ALLBUS\* und Anzeige der Ergebnisse im Objektreiter „Syntaxen“

Das Beispiel in der vorherigen Abbildung zeigt eine zum ALLBUS gespeicherte Syntaxdatei sowie zwei zur Syntax „AP00\_Syntax\_Parteienpraferenz\_ALLBUS“ gehörende Ergebnisdateien.

Über die reine Ergebnisanzeige hinaus ist auch möglich, die entsprechenden Objekte (Studie, Datennutzung und Syntax) sowie diesen Objekten zugeordnete Einzeldateien im jeweils zuständigen Metadateneditor zu öffnen und zu bearbeiten.

Die folgende Abbildung zeigt dies am Beispiel der in der Suche gefundenen ALLBUS Syntaxdatei. Durch Anklicken des Objektstitels „AP00\_Syntax\_Partiepraferenz\_ALLBUS“ wird diese im Syntaxeditor angezeigt. Über die sieben Funktionen dieses Editors können die Syntax oder dazugehörige Dateien nun bearbeitet werden.

The screenshot shows the 'soeb.de' web application interface. At the top, there is a navigation menu with items like 'Startseite', 'Mitglieder', 'Informationen', 'Kalender', 'Absprachen', 'Forum', 'Wiki', 'Dateien', and 'Daten & Analysen'. Below this, a breadcrumb trail indicates the current location: 'VFU soeb 3 Stage > Daten & Analysen > Syntaxeditor'. The main window is titled 'Syntaxeditor' and contains several tabs: 'Syntax bearbeiten', 'Metadaten bearbeiten', 'Ergebnisdateien bearbeiten', 'Duplizieren', 'Löschen', and 'Berechtigungen'. The 'Ergebnisdateien bearbeiten' tab is selected, showing a list of linked files: 'AP00\_Graphik\_Parteienpraferenz\_ALLBUS' and 'AP00\_Logfile\_Parteienpraferenz\_ALLBUS'. To the left, a metadata box shows details for the selected file: 'Titel: AP00\_Syntax\_Partienpraferenz\_ALLBUS', 'Erstellt am: 20.01.2014', 'Erstellt von: Stefan Schweers', 'Bearbeitet am: 22.01.2014', and 'Bearbeitet von: Uwe Jensen'. The 'Syntax-Vorschau' section displays a SPSS command script:

```

1 GET
2 FILE='C:\Users\schweesn\Desktop\Dateien_Abschlussstagung_soeb3\ZA4575_v1-0-1.sav'.
3 DATASET NAME DataSet1 WINDOW=FRONT.
4
5
6 USE ALL.
7 COMPUTE filter_S=(v2 = 2008).
8 VARIABLE LABELS filter_S 'v2 = 2008 (FILTER)'.
9 VALUE LABELS filter_S 0 'Not Selected' 1 'Selected'.
10 FORMATS filter_S (f1.0).
11 FILTER BY filter_S.
12 EXECUTE.
13 GRAPH
14 /BAR(SIMPLE)=PCT BY v23
15 /TITLE='Parteienpraferenz= 2008'.

```

Abbildung 60: Anzeige der gefundenen ALLBUS Syntax im Syntaxeditor und verknüpfte Ergebnisdateien

---

## Glossar: Studie, Studienbeschreibung, Datensatz, Datenfile/Datendatei

---

### Studie – Wo kommen die Ausgangsdaten her?

Im Kontext von Metadaten und deren Bezug zum DDI Standard wird der Begriff Studie im weitesten Sinne benutzt, um den methodischen und organisatorischen Kontext der Datengenerierung der Ausgangsdaten, wie sie in der VFU soeb 3 benutzt werden, zu erfassen.

Zu Beginn der DDI Entwicklung (DDI Codebook) wurde vor allem Bezug auf Projekte und Studien genommen, in denen Daten mit Methoden der empirischen Sozialforschung gewonnen wurden. Mit der Entwicklung der aktuellen DDI LifeCycle Version hat sich der Umfang der dokumentierten Datentypen systematisch auf komplexe Datensätze mit Individualdaten und / oder Aggregatdaten ausgeweitet.

Daten werden durch ein von Regeln geleitetes, wissenschaftliches Vorgehen erzeugt. So werden Daten aus Umfragen über ein oder mehrere Jahre mit Hilfe eines bestimmten Studiendesign unter Einsatz bestimmter Methoden und Instrumenten erhoben. Entsprechend werden Aggregatdaten nach bestimmten Regeln aus Individualdaten aus unterschiedlichen Quellen erzeugt.

### Studienbeschreibung – Wie werden die Ausgangsdaten dokumentiert?

Die sog. Studienbeschreibung (STB) beschreibt den Kontext der Entstehung solcher (Umfrage-)Daten in formalisierter Form durch einzelne Metadatenelemente. Sie wird analog auch für komplexere Studien und Datensätze wie Paneldaten oder Aggregatdaten in der VFU genutzt, um zumindest in vereinfachter Form Kontextinformationen für diese komplexen Datentypen zu erfassen.

Das Schema wird seit den 1970er Jahren im Rahmen der Entwicklung des DDI Standards für elektronische Informationssysteme u. a. von Archiven und Datenanbietern benutzt (Bauske. 2000). Zur Dokumentation von Studien und Daten zählen in der wissenschaftlichen Praxis

- „die Dokumentation der Studie als Ganzes durch eine detaillierte Studienbeschreibung, die Dokumentation der Verfahren und Instrumente der Datenerhebung sowie der Charakteristika der Samples (Studienebene),
- die Dokumentation der Struktur und Eigenart der Datensätze (Datensatzebene)
- sowie die Beschreibung der Variablen einer Datendatei (Variablenebene)“. (Jensen et al. 2011, S. 83f.).

Darüber hinaus werden die Dokumentationen zu einer Studie und deren Daten (Methodenbericht, Variablendokumentation, Schlüsselverzeichnisse etc.) zusätzlich durch entsprechende Metadaten erfasst.

Eine neue Studienbeschreibung wird dann erstellt, wenn z. B. innerhalb eines Umfrageprogrammes durch eine Querschnittstudie neue Daten erhoben wurden. So stellen die Daten der ALLBUS aus den Erhebungen in den Jahren 2010 und 2012 verschiedene Datensätze dar, die durch gesonderte Studienbeschreibungen dokumentiert werden.

Werden demgegenüber nur einzelne Variablen eines Datensatzes korrigiert oder ergänzt, wird dies durch eine neue Version des Datensatz in einer vorhandenen Studienbeschreibung kenntlich gemacht. Werden zur Zitation von Daten zum Beispiel DOIs vergeben, führt eine Änderung an den Daten auch zu einer neuen DOI, die in der Beschreibung der Studie angegeben wird. Die Studienbeschreibung selbst wird üblicherweise nicht versioniert.

Auch komplexe Daten wie das SOEP werden als Versionen in Verbindung mit einer DOI angeboten, um die Daten oder deren Korrekturen innerhalb eines Jahrgangs zu kennzeichnen und einen benutzten Datensatz auch eindeutig zitieren zu können.

Würden z. B. Daten des SOEP 1984 bis 2011 (v28) und SOEP 1984 bis 2012 (v29) in der VFU genutzt, werden zwei Studienbeschreibungen mit entsprechenden (Minimal-)Angaben einschließlich der SOEP DOI mit entsprechender Verlinkung angelegt, da hier Datensätze mit einer unterschiedlichen Datenbasis vorliegen. Innerhalb der Grundversion eines SOEP Datensatz (Originaldatensatz, z. B. v29) sind eventuell weitere Versionen, z. B. bei Fehlerupdates (z. B. v29.1) zu berücksichtigen. Werden im Projektverlauf dann auch die korrigierten Daten der Version 29.1 benutzt, wäre die ursprünglich angelegte Studienbeschreibung zur Version 29 durch entsprechende (Kurz-) Angaben zu aktualisieren<sup>7</sup>.

#### Datensatz – Wie sind die Ausgangsdaten organisiert?

Oftmals wird der Begriff Datensatz und Datenfile (siehe unten) im Forschungsalltag gleichgesetzt. Die Felder zur Erfassung von Metadaten im Editor Studie unterscheiden die Begriffe auf Grundlage der DDI Konventionen des entsprechenden Metadatenschemas.

Ein Datensatz bezeichnet im DDI Kontext die **logische** Repräsentanz der Daten und deren Struktur.

Im Kontext der Studienbeschreibung sind dies die Daten, die zu einem bestimmten Erhebungszeitraum (z. B. Erhebungsjahr / Welle) anhand einer spezifischen Stichprobe (Sample) durch die Studie erhoben werden. Analoges gilt für die Generierung von aggregierten Daten (Index, Indizes, Monatsstatistik, ...), die anhand einer Datenbasis erzeugt werden (z. B. GENESIS Statistiken aus MZ Daten)

Es wird immer ein „Satz von Daten“ generiert, der in einem abgrenzbaren Erhebungszusammenhang steht, wie z. B. die Welle einer Panel-Studie.

- Ein Datensatz mit Individualdaten ist grob definiert durch Messzeitpunkt(e) x Sample(s) eines Typs (z. B. Personendaten je geographischer Raum).

Aggregatdaten werden durch entsprechende Vorschriften aus Individualdaten berechnet.

- Ein Datensatz besteht aus einem oder mehreren Datenfiles

Der Datensatz wird genauer durch die Metadatenelemente u. a. hinsichtlich der Anzahl von Befragten, Variablen und der Organisation in Datenfiles logisch beschrieben.

Beispiel:

- Der Datensatz SOEP.v28 besteht aus ca. 300 Datenfiles.
- Der Datensatz Mikrozensus hat je Erhebungsjahr einen Datenfile je Zugangsform, im Sinne eines Public-Use-File, eines Scientific-Use-File und On-Site-Zugang. Da sich die Datensätze des SUF und PUF hinsichtlich der Anzahl der Befragten und Variablen unterscheiden, wären unterschiedlichen Studienbeschreibungen anzulegen, wenn beide Datensätze als Ausgangsdaten in der VFU genutzt würden.

#### Datenfile / Datendatei – Wie werden die Ausgangsdaten gespeichert?

Ein Datenfile bezeichnet im DDI Kontext die **physische** Repräsentanz der Daten.

Dabei ist das Dateiformat des Datenfiles für die jeweilige Nutzung mit einem kompatiblen Statistikprogramm von Relevanz. Demgegenüber ist das technische Dateiformat für den Datensatz nicht relevant: der SUF des Mikrozensus aus dem Jahr 2011 im Stata oder SPSS Format speichert den gleichen Satz von Daten.

Ein Datenfile oder auch Datendatei ist im VFU Metadatenkontext die Datei, die formatabhängig durch ein spezifisches Softwareprogramm zur Datenanalyse benutzt wird.

<sup>7</sup> Vgl. [http://www.diw.de/de/diw\\_01.c.387315.de/versionen / doi.html](http://www.diw.de/de/diw_01.c.387315.de/versionen / doi.html)

---

## 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](http://www.soeb.de/fileadmin/redaktion/downloads/VirtAug/VirtAug_Abschlussbericht.pdf)
- Bauske, Franz. 2000  
Das Studienbeschreibungsschema des Zentralarchivs. In: ZA-Information, 2000, Nr. 47, S. 73-80.  
[http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za\\_information/ZA-Info-47.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za_information/ZA-Info-47.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](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/>
- 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/abschlusstagung/>  
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](http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU_Tagung_Jensen_final.pdf)
- Jensen, Uwe; Schweers, Stefan. 2014  
Das erweiterte Metadatenschema der VFU soeb 3  
GESIS-Technical Reports – 2014|13  
[http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2014/TechnicalReport\\_2014-13.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2014/TechnicalReport_2014-13.pdf)
- Jensen, Uwe. Katsanidou, Alexia. Zenk-Möltgen, Wolfgang. 2011.  
Metadaten und Standards. In: Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (Hrsg.). Handbuch Forschungsdatenmanagement. 2011, Bad Honnef, S. 83-100:  
<http://opus4.kobv.de/opus4-fhpotsdam/files/208/HandbuchForschungsdatenmanagement.pdf>
- Pohl, Christof. 2014  
Fachtagung "Das Portal, die Daten und wir - Eine Virtuelle Forschungsumgebung für die digitale Infrastruktur". Berlin 24.01.2014  
Präsentation: Die Architektur der Virtuellen Forschungsumgebung.  
[http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU\\_Tagung\\_Pohl\\_final.pdf](http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU_Tagung_Pohl_final.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](http://www.soeb.de/fileadmin/redaktion/downloads/VFU/Grundlegende_Informationen_zur_Nutzung_der_VFU_v01.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](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](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-01.pdf)