

Entwicklung und Betrieb eines Metadatenmanagementsystems für Forschungsdaten aus dem Bereich der Hochschul- und Wissenschaftsforschung – Lessons Learned

Karsten Stephan und René Reitmann

Jülich, den 04.06.2019

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung

- Kompetenzzentrum für Hochschul- und Wissenschaftsforschung mit Sitz in Hannover und Berlin
- Große Studienreihen
 - DZHW-Studienberechtigtenpanel, Sozialerhebung, DZHW-Absolventenpanel, DZHW-Wissenschaftsbefragung
 - Student Life Cycle, Die Studierendenbefragung in Deutschland, National Academics Panel Study
 - Eurostudent, Eurograduate
- Bestandsdaten seit den 1970er Jahren

Forschungsdatenzentrum für Hochschul- und Wissenschaftsforschung (FDZ-DZHW)

- Forschungsfeldbezogenes Forschungsdatenzentrum
- Aufbauprojekt von 2015 bis 2017 vom BMBF finanziert
- Eröffnung am 01.06.2017
- Akkreditierung durch den Rat für Sozial- und Wirtschaftsdaten (RatSWD)
- Aufnahme anonymisierter quantitativer und qualitativer Daten
 - aktuelle und zukünftige Studien des DZHW
 - Bestandsdaten des DZHW
 - Daten DZHW-externer Wissenschaftler*innen
- Datenherausgabe als Scientific Use File und Campus Use File über verschiedene Herausgabewege
- Beteiligung an gemeinschaftlicher Infrastrukturweiterentwicklung

Metadatenmanagementsystem (MDM)

- Zentraler Bestandteil der technischen Infrastruktur des Forschungsdatenzentrums zur Datenherausgabe und –aufnahme
- Wichtiges Werkzeug zur Unterstützung des Forschungsdatenmanagements im Haus
- Ausweisung von Forschungsdaten, strukturierte Erfassung und Editierung von Metadaten, Suche im Metadatenbestand
- Aufbau von 2015 bis 2017 (Datenspeicher und Suche) und 2017 bis 2019 (qualitätsgesicherte Datenaufnahme)

MDM – Vision

Verfügbarkeit eines Systems, welches ...

- ... eine möglichst weitreichende Transparenz zum Datenbestand des Forschungsfeldes und dem Entstehungsprozess der Daten herstellt
- ... die Nachvollziehbarkeit von Forschungsdaten und Forschungsprozessen gewährleistet
- ... Forschungsprozesse von Wissenschaftler*innen unterstützen kann

MDM – Anforderungen

konnten teilweise vor Projektstart spezifiziert werden, sind häufig erst im Verlauf des Projektes entstanden

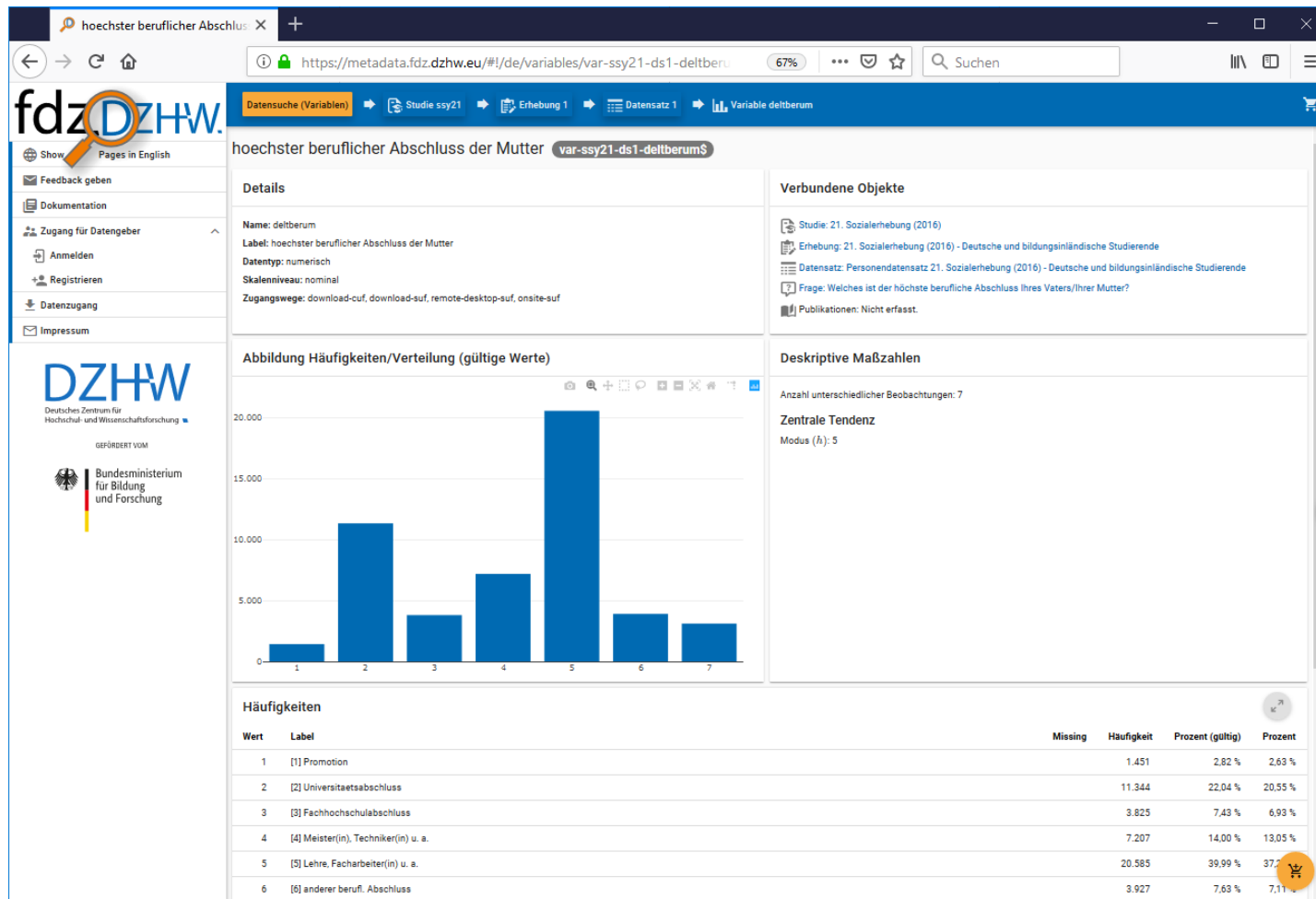
- Strukturierte Erfassung und Speicherung der Metadaten in einem zentralen Metadatenpeicher
- Strukturierte Suche in deutscher und englischer Sprache
- Unterstützung von Prozessen zur einfachen, qualitätsgesicherten, forschungsprozessbegleitenden Datenaufnahme
- Versionierung der Metadaten
- Möglichkeit zur technischen Anbindung an andere Systeme

MDM – Startseite mit Suchfunktionalität

The screenshot displays the MDM (Metadatamanagement) website interface. The browser address bar shows the URL <https://metadata.fdz.dzhw.eu/#/de/search?page=1&type=studies>. The page title is "Datensuche (Studien)". The left sidebar contains navigation links: "Show Pages in English", "Feedback geben", "Dokumentation", "Zugang für Datengeber", "Datenzugang", and "Impressum". The main content area lists search results for studies. The first result is "beeinträchtigt studieren - best2" (stu-best023) by Poskowsky, Jonas; Fromme, Christine; Heißenberg, Sonja; Zaussinger, Sarah; Brenner, Julia. The second result is "KomPaed 2014 - Tätigkeitsbezogene Kompetenzen in pädagogischen Handlungsfeldern (quantitative Teilstudie)" (stu-cmp20145) by Braun, Edith; Brachem, Julia-Carolin. The third result is "DZHW-Absolventenpanel 2005" (stu-gra20055) by Briedis, Kolja; Euler, Thorsten; Grotheer, Michael; Isleib, Sören; Minks, Karl-Heinz; Netz, Nicolai; Schaeper, Hildegard; Trennt, Fabian; Trommer, Maximilian. The fourth result is "DZHW-Absolventenpanel 2009" (stu-gra20095) by Brandt, Gesche; Briedis, Kolja; Fabian, Gregor; Klüver, Saskia; Rehn, Torsten; Trommer, Maximilian. The footer of the page shows the DZHW logo and the text "GEFÖRDET VOM Bundesministerium für Bildung und Forschung".

<https://metadata.fdz.dzhw.eu/>

MDM – Detailseite einer Variablen



[https://metadata.fdz.dzhw.eu/#!/de/variables/var-ssy21-ds1-deltberum\\$](https://metadata.fdz.dzhw.eu/#!/de/variables/var-ssy21-ds1-deltberum$)

MDM – Projekt-Cockpit

The screenshot displays the 'Projekt-Cockpit' web application for project 'ssy21t'. The browser address bar shows the URL <https://metadata.fdz.dzhw.eu/#i/de/projects/ssy21t>. The interface features a left sidebar with navigation links: 'Show Pages in English', 'karsten abmelden', 'Feedback geben', 'Aktuelles Datenaufbereitungsprojekt: ssy21t', 'Nicht freigegeben', 'Zugewiesen an Publisher', 'Hinweise für Datengeber', 'Projektübersicht', 'Dokumentation', 'Zugang für Datengeber', 'Konto bearbeiten', 'Passwort ändern', 'Administration', 'Datenzugang', and 'Impressum'. The main content area is titled 'Projekt-Cockpit ssy21t' and includes tabs for 'Status' and 'Einstellungen'. It displays several project components with their counts and status indicators:

- Studie (0)**: ☐ Datengeber Fertig, ☐ Publisher Fertig. Buttons: Löschen, Neu.
- Erhebungen (0)**: ☐ Datengeber Fertig, ☐ Publisher Fertig. Buttons: Löschen, Bearbeiten, Neu.
- Instrumente (0)**: ☐ Datengeber Fertig, ☐ Publisher Fertig. Buttons: Löschen, Bearbeiten, Neu.
- Fragen (268)**: ☐ Datengeber Fertig, ☒ Publisher Fertig. Buttons: Löschen, Hochladen.
- Datensätze (0)**: ☐ Datengeber Fertig, ☐ Publisher Fertig. Buttons: Löschen, Bearbeiten, Neu.
- Variablen (1730)**: ☐ Datengeber Fertig, ☒ Publisher Fertig. Buttons: Löschen, Hochladen.
- Publikationen (0)**: ☐ Datengeber Fertig, ☐ Publisher Fertig. Buttons: Löschen, Bearbeiten.

The footer of the interface includes the DZHW logo (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung) and the text 'GEFÖRDERT VOM Bundesministerium für Bildung und Forschung'.

MDM – Eingabemasken Studie

The screenshot shows a web browser window with the URL <https://metadata.fdz.dzhw.eu/#1/de/studies/new>. The page title is 'Studie anlegen: stu-ssy21t\$'. The form is divided into several sections:

- Header:** 'Studie anlegen: stu-ssy21t\$' and a search bar.
- Left Sidebar:** Contains navigation links like 'Datensuche (Alle)', 'Studie?', 'Hinweise für Datengeber', 'Projektübersicht', 'Dokumentation', 'Zugang für Datengeber', 'Konto bearbeiten', 'Passwort ändern', 'Administration', 'Datenzugang', and 'Impressum'. It also features the fdzDZHW logo and the DZHW logo.
- Main Form:**
 - Details:** A table with two columns for German and English input fields.

Titel (auf Deutsch) *	Titel (auf Englisch) *
Geben Sie den Titel der Studie auf Deutsch ein. 0 / 2048	Geben Sie den Titel der Studie auf Englisch ein. 0 / 2048
Studienreihe (auf Deutsch)	Studienreihe (auf Englisch)
Geben Sie, falls vorhanden, den Namen der Studienreihe auf Deutsch ein.	Geben Sie, falls vorhanden, den Namen der Studienreihe auf Englisch ein.
Erhebende Institution (auf Deutsch) *	Erhebende Institution (auf Englisch) *
Geben Sie den deutschen Namen der Institution ein, die die Erhebungen durchgeführt hat.	Geben Sie den englischen Namen der Institution ein, die die Erhebungen durchgeführt hat.
Gefördert von (auf Deutsch) *	Gefördert von (auf Englisch) *
Geben Sie den deutschen Namen des Geldgebers für diese Studie ein.	Geben Sie den englischen Namen des Geldgebers für diese Studie ein.
Erhebungsdesign *	
Wählen Sie das Erhebungsdesign dieser Studie aus.	
Anmerkungen (auf Deutsch)	Anmerkungen (auf Englisch)
Geben Sie zusätzliche Anmerkungen zur Studie auf Deutsch ein. 0 / 2048	Geben Sie zusätzliche Anmerkungen zur Studie auf Englisch ein. 0 / 2048
Datenverfügbarkeit *	
Wählen Sie den Status aus, der die aktuelle Verfügbarkeit der Daten am Besten beschreibt.	
 - Studienbeschreibung:** A section for describing the study.

MDM – Ein Datenprodukt in Warenkorb legen

The screenshot shows the 'Personendatensatz 21. Sozialerhebung (2016)' page. A modal window titled 'Optionen des zugehörigen Datenproduktes' is open, displaying the study name '21. Sozialerhebung (2016)' and the access path 'remote-desktop-suf'. It also shows the version '2.0.0 (aktuell)' and a dropdown menu for selecting a version. Below this, there is a text box asking for a CUF or SUF and a note about the data product. At the bottom of the modal, there are two buttons: 'Schließen' and 'In den Einkaufswagen'.

Personendatensatz 21. Sozialerhebung (2016) - Deutsche und bildungsinländische Studierende

dat-ssy21-ds1\$

Details

Typ: Personendatensatz

Format: breit

Verbundene Objekte

Studie: 21. Sozialerhebung (2016)

Optionen des zugehörigen Datenproduktes

Studie

21. Sozialerhebung (2016)

Zugangsweg zu den Datensätzen

remote-desktop-suf

Version der Datensätze

2.0.0 (aktuell)

Benötigen Sie einen CUF oder SUF und wie möchten Sie mit den Daten arbeiten?

Welche Version der Datensätze benötigen Sie?

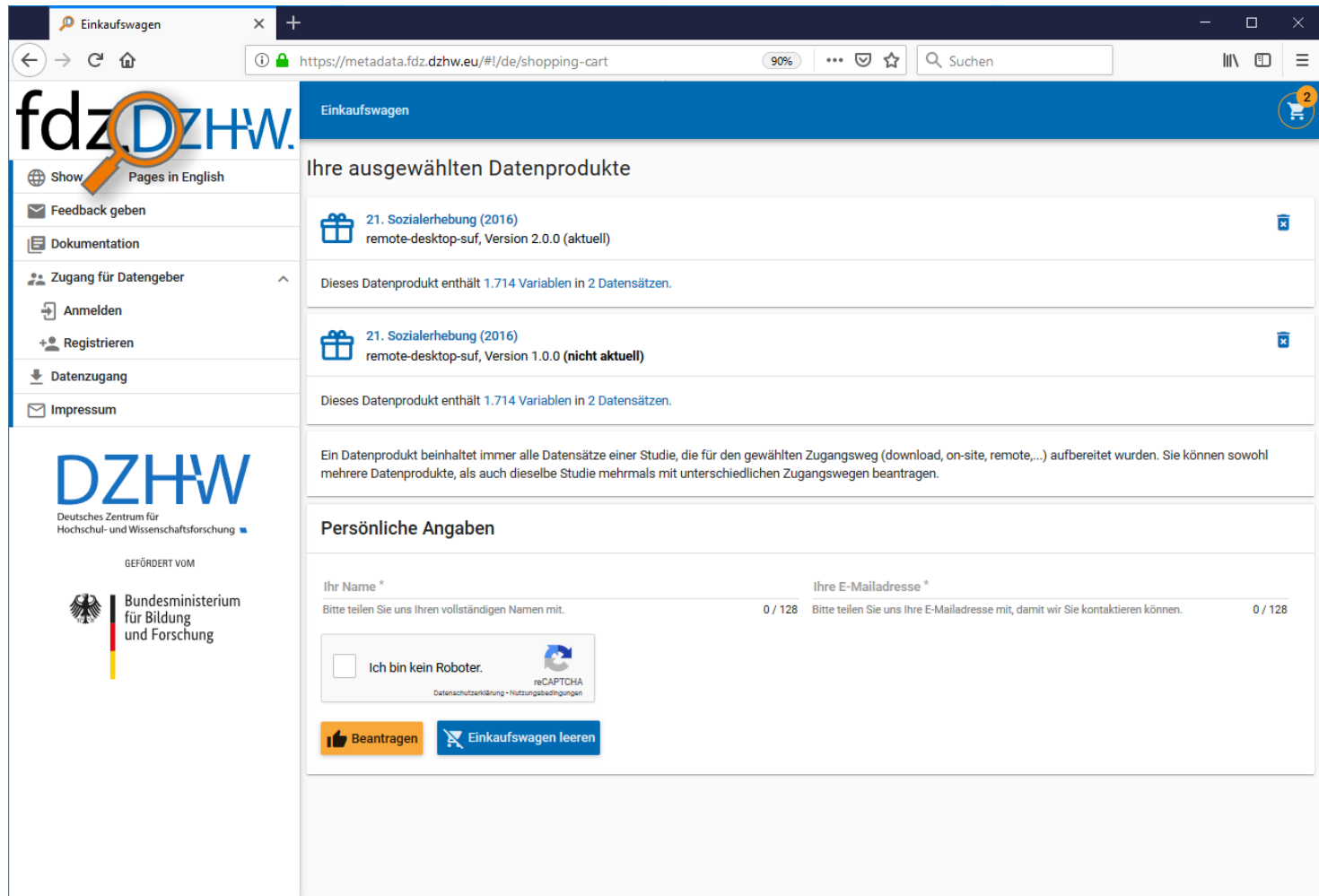
Ein Datenprodukt beinhaltet immer alle Datensätze einer Studie, die für den gewählten Zugangsweg (download, on-site, remote,...) aufbereitet wurden. Sie können sowohl mehrere Datenprodukte, als auch dieselbe Studie mehrmals mit unterschiedlichen Zugangswegen beantragen.

Schließen In den Einkaufswagen

Materialien zu dem Datensatz

Titel	Beschreibung	Dokumentensprache	Datei
Variablendokumentation: 21. Sozialerhebung (2016) Datensatz 1	Codebook des ersten Datensatzes von SSy21.	Deutsch	dsreport-ssy21-ds1-...df

MDM – Zwei Produktversionen im Warenkorb



The screenshot shows a web browser window with the URL `https://metadata.fdz.dzhw.eu/#/de/shopping-cart`. The page title is "Einkaufswagen" (Shopping Cart). The left sidebar contains navigation links: "Show Pages in English", "Feedback geben", "Dokumentation", "Zugang für Datengeber" (expanded), "Anmelden", "Registrieren", "Datenzugang", and "Impressum". The main content area is titled "Ihre ausgewählten Datenprodukte" (Your selected data products) and lists two items:

- 21. Sozialerhebung (2016) remote-desktop-suf, Version 2.0.0 (aktuell)
- 21. Sozialerhebung (2016) remote-desktop-suf, Version 1.0.0 (nicht aktuell)

Each item indicates it contains 1.714 variables in 2 data sentences. Below the list, a note states: "Ein Datenprodukt beinhaltet immer alle Datensätze einer Studie, die für den gewählten Zugangsweg (download, on-site, remote,...) aufbereitet wurden. Sie können sowohl mehrere Datenprodukte, als auch dieselbe Studie mehrmals mit unterschiedlichen Zugangswegen beantragen." The "Persönliche Angaben" (Personal Information) section includes fields for "Ihr Name *" and "Ihre E-Mailadresse *" (both with 0/128 characters), a reCAPTCHA checkbox, and buttons for "Beantragen" (Submit) and "Einkaufswagen leeren" (Empty cart). The footer features the DZHW logo and the text "GEFÖRDERT VOM Bundesministerium für Bildung und Forschung".

MDM – Rahmenbedingungen der Entwicklung

- 27 Monate Zeit bis zur Produktivsetzung der Metadatenuche
- Ein Team mit drei Softwareentwicklern (ein Senior- und zwei Juniorentwickler) und einem Sozialwissenschaftler
- Mittel für Beratung und Schulung
- Fachliche Unsicherheiten (DZHW-spezifische FDZ-Prozesse)
- Technologische Unsicherheit

Lessons Learned: Agiler Entwicklungsprozess

- Wahl eines agilen Entwicklungsmodells, welches Änderungen und Ergänzung von Anforderungen im Entwicklungsprozess berücksichtigt
- Vision als Ausgangspunkt von Spezifikation und Priorisierung
- Funktionalitäten zunächst basal umsetzen und evtl. später weiter ausdifferenzieren; sehr komplexe spät implementieren
- So schnell wie möglich ein lauffähiges System mit nützlicher Funktionalität implementieren, so spät wie möglich in Produktion gehen
- Soweit wie möglich auf existierende vielfach eingesetzte Komponenten und Frameworks zurückgreifen; Entwicklungs- und Betriebsinfrastruktur nicht selbst betreuen (... as a service)

Lessons Learned: Domänenexpertise

- Domänenexpertise für den Entwicklungsprozess
 - systematisch
 - kontinuierlich
 - in hohem zeitlichen Umfang
- Testen
 - durch Fachwissenschaftler*innen
 - im Idealfall zukünftige Nutzer*innen einbeziehen

=> Lösungsmöglichkeit: Einkaufen von Domänenexpertise

Lessons Learned: Dynamiken interdisziplinären Arbeitens

- Interdisziplinäre Teams bringen unterschiedliche Fachsprachen und Fachkulturen mit
- Von Beginn an
 - Wissen über „die andere Seite“ vermitteln
 - das gemeinsame Ziel definieren
 - eine gemeinsame Sprache entwickeln
 - interdisziplinäre Dynamiken berücksichtigen
 - hilfreich sind „Übersetzer*innen“

Fazit: Eigenentwicklung?

Ja, wenn ...

- ... eine klare Vision des zu entwickelnden Systems existiert
- ... softwaretechnische Expertise besteht bzw. eingekauft werden kann
- ... ausreichend Entwicklungskapazitäten vorhanden sind
- ... Domänenexpertise systematisch einbezogen werden kann

Vielen Dank für Ihre Aufmerksamkeit!

Dr. Karsten Stephan
stephan@dzhw.eu

René Reitmann
reitmann@dzhw.eu

Metadatenmanagementsystem:
<https://metadata.fdz.dzhw.eu>

Entwicklungsrepository:
<https://github.com/dzhw/metadatamanagement>