

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

Metadatamanagement (MDM) X +

https://metadata.fdz.dzhw.eu/#/de/search?page=1&type=studies 90% Suchen

fdzDZHW

Show Pages in English

Feedback geben

Dokumentation

Zugang für Datengeber

Datenzugang

Impressum

DZHW
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung

GEFÜRDERT VOM
Bundesministerium für Bildung und Forschung

Datensuche (Studien)

Studien (11) Erhebungen Instrumente Fragen Datensätze Variablen Publikationen

Suchen Sie nach Studien...

Filter für die Studiensuche...

beeinträchtigt studieren - best2 stu-best025
Poskowsky, Jonas; Fromme, Christine; Heißenberg, Sonja; Zaussinger, Sarah; Brenner, Julia

Mit best2 werden zum zweiten Mal bundesweit Daten zur Situation Studierender mit studienerschwerenden gesundheitlichen Beeinträchtigungen erhoben. Das Hauptaugenmerk der Befragung liegt auf den Schwierigkeiten und Barrieren, mit denen sich gesundheitlich B...

KomPaed 2014 - Tätigkeitsbezogene Kompetenzen in pädagogischen Handlungsfeldern (quantitative Teilstudie) stu-cmp20145
Braun, Edith; Brachem, Julia-Carolin

Das Projekt „KomPaed - Tätigkeitsbezogene Kompetenzen in pädagogischen Handlungsfeldern“ ist ein Verbundprojekt der BMBF-Förderinitiative „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor“ (KoKoHs) und umfasst eine quantitative und eine quali...

DZHW-Absolventenpanel 2005 stu-gra20055
Briedis, Kolja; Euler, Thorsten; Grotheer, Michael; Isleib, Sören; Minks, Karl-Heinz; Netz, Nicolai; Schaeper, Hildegard; Trennt, Fabian; Trommer, Maximilian

Das DZHW-Absolventenpanel 2005 ist Teil der DZHW-Absolventenstudienreihe, in der anhand von standardisierten Befragungen Informationen zu Studium, Berufseintritt, Berufsverlauf und Weiterqualifizierung von Hochschulabsolvent(inn)en erfasst werden. Das erst...

DZHW-Absolventenpanel 2009 stu-gra20095
Brandt, Gesche; Briedis, Kolja; Fabian, Gregor; Klüver, Saskia; Rehn, Torsten; Trommer, Maximilian

Das DZHW-Absolventenpanel 2009 ist Teil der DZHW-Absolventenstudienreihe, in der anhand von standardisierten Befragungen Informationen zu Studium, Berufseintritt, Berufsverlauf und Weiterqualifizierung von Hochschulabsolvent(inn)en erfasst werden. Das erst...

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

MDM – Detailseite einer Variablen

The screenshot shows the detail page for the variable 'hoehster beruflicher Abschluss der Mutter' (highest educational attainment of the mother) in the MDM system. The page includes a navigation breadcrumb, a search bar, and a sidebar with user options. The main content area is divided into several sections:

- Details:** Name: deltberum; Label: hoehster beruflicher Abschluss der Mutter; Datentyp: numerisch; Skalenniveau: nominal; Zugangswege: download-cuf, download-suf, remote-desktop-suf, onsite-suf.
- Verbundene Objekte:** Studie: 21. Sozialerhebung (2016); Erhebung: 21. Sozialerhebung (2016) - Deutsche und bildungsinländische Studierende; Datensatz: Personendatensatz 21. Sozialerhebung (2016) - Deutsche und bildungsinländische Studierende; Frage: Welches ist der höchste berufliche Abschluss Ihres Vaters/Ihrer Mutter?; Publikationen: Nicht erfasst.
- Abbildung Häufigkeiten/Verteilung (gültige Werte):** A bar chart showing the distribution of values 1 through 7. The y-axis represents frequency, ranging from 0 to 20,000. The highest frequency is for value 5, which exceeds 20,000.
- Deskriptive Maßzahlen:** Anzahl unterschiedlicher Beobachtungen: 7; Zentrale Tendenz: Modus (h): 5.
- Häufigkeiten:** A table showing the frequency and percentage of each value.

Wert	Label	Missing	Häufigkeit	Prozent (gültig)	Prozent
1	[1] Promotion		1.451	2,82 %	2,69 %
2	[2] Universitätsabschluss		11.344	22,04 %	20,55 %
3	[3] Fachhochschulabschluss		3.825	7,43 %	6,93 %
4	[4] Meister(in), Techniker(in) u. a.		7.207	14,00 %	13,05 %
5	[5] Lehre, Facharbeiter(in) u. a.		20.585	39,99 %	37,20 %
6	[6] anderer berufl. Abschluss		3.927	7,63 %	7,11 %

[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 interface for project 'ssy21t'. The browser address bar shows the URL: <https://metadata.fdz.dzhw.eu/#1/de/projects/ssy21t>. The interface features a blue header with the project name and navigation options like 'Status' and 'Einstellungen'. A left sidebar contains the 'fdzDZHW.' logo, a language selector, and a menu with items such as 'karsten abmelden', 'Feedback geben', 'Aktuelles Datenaufbereitungsprojekt: ssy21t', 'Hinweise für Datengeber', 'Projektübersicht', 'Dokumentation', 'Zugang für Datengeber', 'Konto bearbeiten', 'Passwort ändern', 'Administration', 'Datenzugang', and 'Impressum'. The main content area shows a dashboard with two status buttons at the top: 'Nicht freigegeben' (red) and 'Zugewiesen an Publisher' (blue). Below are several data cards: 'Studie (0)', 'Erhebungen (0)', 'Instrumente (0)', 'Fragen (268)', 'Datensätze (0)', 'Variablen (1730)', and 'Publikationen (0)'. Each card includes checkboxes for 'Datengeber Fertig' and 'Publisher Fertig', and action buttons like 'Löschen', 'Neu', 'Bearbeiten', and 'Hochladen'. The 'Fragen' and 'Variablen' cards show the 'Publisher Fertig' checkbox checked. At the bottom left, the DZHW logo and the text 'Deutsches Zentrum für Hochschul- und Wissenschaftsforschung' are visible, along with the logo of the 'Bundesministerium für Bildung und Forschung'.

MDM – Eingabemasken Studie

The screenshot shows a web browser window with the URL <https://metadata.fdz.dzhw.eu/#/de/studies/new>. The page title is "Studie anlegen: stu-ssy21t\$". The form is titled "Details" and contains the following fields:

Field (German)	Field (English)
Titel (auf Deutsch) * Geben Sie den Titel der Studie auf Deutsch ein. 0 / 2048	Titel (auf Englisch) * Geben Sie den Titel der Studie auf Englisch ein. 0 / 2048
Studienreihe (auf Deutsch) Geben Sie, falls vorhanden, den Namen der Studienreihe auf Deutsch ein.	Studienreihe (auf Englisch) Geben Sie, falls vorhanden, den Namen der Studienreihe auf Englisch ein.
Erhebende Institution (auf Deutsch) * Geben Sie den deutschen Namen der Institution ein, die die Erhebungen durchgeführt hat.	Erhebende Institution (auf Englisch) * Geben Sie den englischen Namen der Institution ein, die die Erhebungen durchgeführt hat.
Gefördert von (auf Deutsch) * Geben Sie den deutschen Namen des Geldgebers für diese Studie ein.	Gefördert von (auf Englisch) * Geben Sie den englischen Namen des Geldgebers für diese Studie ein.
Erhebungsdesign * Wählen Sie das Erhebungsdesign dieser Studie aus.	
Anmerkungen (auf Deutsch) Geben Sie zusätzliche Anmerkungen zur Studie auf Deutsch ein. 0 / 2048	Anmerkungen (auf Englisch) 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.	

Below the form is a section for "Studienbeschreibung".

The left sidebar contains navigation and user options:

- 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
- Impressum

The footer of the page includes the DZHW logo (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung) and the logo of the Bundesministerium für Bildung und Forschung.

MDM – Ein Datenprodukt in Warenkorb legen

The screenshot shows a web browser window displaying the metadata management system (MDM) interface. The main page is titled "Personendatensatz 21. Sozialerhebung (2016) - Deutsche und bildungsinländische Studierende" and shows details for the dataset "dat-ssy21-ds1\$". A modal dialog titled "Optionen des zugehörigen Datenproduktes" is open, allowing users to select the access method and version of the data product.

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-rdf

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 main content area is titled "Ihre ausgewählten Datenprodukte" (Your selected data products) and contains 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 is described as containing 1.714 variables in 2 data sentences. Below the items, 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." (A data product always contains all data records of a study, which have been prepared for the selected access method (download, on-site, remote,...). You can apply for both multiple data products, as well as the same study multiple times with different access methods.)

The "Persönliche Angaben" (Personal Information) section includes:

- Ihr Name * (0 / 128)
- Ihre E-Mailadresse * (0 / 128)
- reCAPTCHA security check with "Ich bin kein Roboter." checkbox and "Datenschutzerklärung - Nutzungsbedingungen" link.
- Buttons: "Beantragen" (Apply) and "Einkaufswagen leeren" (Empty cart).

The left sidebar contains navigation links: "Show Pages in English", "Feedback geben", "Dokumentation", "Zugang für Datengeber", "Anmelden", "Registrieren", "Datenzugang", and "Impressum". The footer of the sidebar 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>