

Michaela Meyer

Sozialwissenschaftliche Fragestellungen zum Forschungsdatenmanagement am Beispiel von WissGrid

Projektarbeit im Rahmen des Moduls M4.2 / WS15 im
Studiengang M.A. Informationswissenschaften der FH Potsdam:
Konzeptionelle Entwicklung eines Werkzeugs
für die Planung des Forschungsdatenmanagements
Prof. Dr. Heike Neuroth

in Kooperation mit dem DFG-Projekt



<https://dmpwerkzeug.github.io/>

08.02.2016

Inhalt

Abkürzungsverzeichnis	I
Abbildungsverzeichnis.....	II
1 Einleitung.....	1
2 Forschungsdaten der Sozialwissenschaften.....	2
2.1 Definition	2
2.2 Qualitative und quantitative Forschung	2
2.3 Datenformate.....	3
2.4 Datenschutz	3
2.4.1 Anonymisierung.....	4
2.4.2 Pseudoanonymisierung	4
2.4.3 Transkription.....	4
2.5 Datenzugang.....	4
3 Metadaten, Standards und Richtlinien.....	5
3.1 DDI-Standard	5
3.1.1 DDI- Codebook (2.5).....	5
3.1.2 DDI- Lifecycle (3.2)	6
3.2 ELSST.....	6
4 Methoden und Instrumente	7
4.1 Befragung	7
4.1.1 Papierfragebogen	7
4.1.2 Computergestützte Befragungen.....	7
4.2 Beobachtung & Inhaltsanalyse.....	8
4.3 Erhebungsprogramme	8
4.3.1 ALLBUS.....	8
4.3.2 ISSP	9
5 Use Cases /Datensätze.....	9

5.1	Beispiel aus dem GESIS- Datenarchiv (ALLBUS 2014).....	9
5.2	Beispiel aus der Datenbank des Statistischen Bundesamtes (Zensus 2011)	11
5.3	GESIS vs. Statistisches Bundesamt.....	11
5.4	Fazit	12
6	WissGrid	12
6.1	Vorstellung des Fragenkatalogs.....	12
6.1.1	Teil I.....	13
6.1.2	Teil II.....	13
6.2	Anpassungen für die Sozialwissenschaften.....	14
6.2.1	Planung und Erstellung	14
6.2.2	Auswahl und Aufbewahrungsdauer	14
6.2.3	Ingest: Einspeisen und Verantwortungsübernahme	15
6.2.4	Zugriff und Nutzung.....	15
6.2.5	Management, Organisation und Policies.....	16
6.2.6	Kosten.....	16
7	Ergebnisse und Ausblick	16
8	Literaturverzeichnis.....	III
	Anhang	V

Abkürzungsverzeichnis

ALLBUS	Allgemeine Bevölkerungsumfrage der Sozialwissenschaften
BMBF	Bundesministerium für Bildung und Forschung
CAPI	Computer Assisted Personal Interview
CASI	Computer Assisted Self Interview
CATI	Computer Assisted Telephone Interview (CATI)
CESSDA	Council of European Social Science Data Archives
DDI	Data Documentation Initiative
DFG	Deutsche Forschungsgemeinschaft
ELSST	European Language Social Science Thesaurus
FDM	Forschungsdatenmanagement
ICPSR	Interuniversity Consortium for Political and Social Research
ISSP	International Social Survey Program
PAPI	Paper Assisted Personal Interview
RADIESCHEN	Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur
RatSWD	Rat für Sozial- und Wirtschaftsdaten
XML	Extensible Markup- Language

Abbildungsverzeichnis

Abbildung 1: Zugangskategorien des GESIS- Datenarchivs (Quelle: GESIS- Leibniz- Institut für Sozialwissenschaften)	10
Abbildung 2: Aufgaben im Lebenszyklus von Forschungsdaten.....	13
Abbildung 3: Übergreifende Aufgaben des Forschungsdatenmanagmenrts.....	13

1 Einleitung

Das Forschungsdatenmanagement (nachfolgend FDM) stellt eine große Herausforderung an alle beteiligten Akteure, insbesondere das disziplinspezifische, da Forschungsdaten nicht nur in den einzelnen (Teil-) Disziplinen in unterschiedlicher Form auftreten, sondern dadurch auch anders im Lebenszyklus¹ behandelt werden müssen. Daher wurden bereits Versuche unternommen, durch Richtlinien und Leitfäden eine gewisse Orientierung zu geben, welche Kriterien bei der Entstehung, Verarbeitung und Archivierung der Forschungsdaten zu beachten sind. Ein Beispiel dafür ist die Handreichung zu dem Projekt **WissGrid**², bei dem unter Zusammenarbeit verschiedener Forschungsdisziplinen ein Leitfaden und eine Checkliste zum FDM erstellt wurden. Weiterhin hat die **DFG** (Deutsche Forschungsgemeinschaft)³ Leitlinien zum Umgang mit Forschungsdaten⁴ herausgegeben und die Fachdisziplinen gleichzeitig dazu aufgerufen, über spezifische Regularien im Umgang mit Forschungsdaten zu diskutieren. In den „Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur“, kurz **RADIESCHEN**⁵, wird u. a. auch auf die Disziplin der Wirtschafts- und Sozialwissenschaften eingegangen. Große Unterstützung für die Sozialwissenschaften bietet das **GESIS** Leibniz-Institut für Sozialwissenschaften⁶, welches sich sowohl mit sozialwissenschaftlichen Inhalten als auch der Methodik auseinandersetzt. Des Weiteren unterstützt es Forscherinnen und Forscher bei ihrer Arbeit durch die Beratung im Forschungsprozess sowie der Erhebung, Aufbereitung und Archivierung von Forschungsdaten mit Schwerpunkt auf der empirischen Sozialforschung. Aus diesen Gründen bezieht sich die Projektarbeit auch des Öfteren auf die Erkenntnisse des GESIS, die in einem Leitfaden zum Management von Forschungsdaten⁷, in Bezug auf sozialwissenschaftliche Umfragedaten, festgehalten wurden.

Die vorliegende Arbeit wurde im Rahmen des DFG-Projekts „Konzeptionelle Entwicklung eines Werkzeugs für die Planung des Forschungsdatenmanagements“ verfasst. Sie soll zum einen das genannte Vorhaben unterstützen und zum anderen Forscherinnen

¹ Ludwig und Enke 2013, S.15

² Ludwig und Enke 2013

³ <http://www.dfg.de/>

⁴ DFG-Deutsche Forschungsgemeinschaft 2015

⁵ Klar und Enke 2013

⁶ <http://www.gesis.org/das-institut/>

⁷ Jensen 2012

und Forschern sowie Interessierten Empfehlungen zum Umgang mit sozialwissenschaftlichen Forschungsdaten geben. Dabei bezieht sich die Arbeit vornehmlich auf die WissGrid- Checkliste, die durch Vorschläge angepasst werden kann. Die Empfehlungen stützen sich dabei auf eine Analyse der Umsetzung von Richtlinien und Standards und der Ergebnisse aus dem Vergleich zweier Use Cases aus der empirischen Sozialforschung. Diese Vorschläge zur Formulierung von Fragen bzw. Ergänzungen für die Erstellung eines Forschungsdatenmanagement-Plans können nicht als vollständig betrachtet werden. Der Fokus liegt in dieser Arbeit auf der quantitativen Forschung sowie den wichtigsten Aufgaben im Umgang mit sozialwissenschaftlichen Forschungsdaten.

2 Forschungsdaten der Sozialwissenschaften

2.1 Definition

Sozialwissenschaftliche Forschungsdaten sind nicht eindeutig zu definieren und ihr Begriff in der Wissenschaft ist zahlreich diskutiert. Vorerst kann zwischen qualitativen und quantitativen Forschungsdaten unterschieden werden. Da diverse Wissenschaften in den Sozialwissenschaften vereint sind, werden verschiedene Methoden zur Gewinnung von Daten eingesetzt. Letztere können sich kontinuierlich durch Anpassungen verändern, wodurch eine Vereinheitlichung der Begrifflichkeiten sich als sehr schwierig gestaltet. Allgemein bilden sozialwissenschaftliche Forschungsdaten die Zusammenhänge des gesellschaftlichen Zusammenlebens ab und zeigen Ausschnitte der Wirklichkeit auf. Es bestehen jedoch Ansätze, signifikante Eigenschaften⁸ für sozialwissenschaftliche Forschungsdaten festzulegen⁹, die bereits deutlich machen, dass die Forschungsdaten nicht nur aus den Datensätzen bestehen, sondern zusätzlich die Dokumentation zum Erhebungsprozess nötig ist, um die Daten bewerten bzw. nachnutzen zu können.¹⁰

2.2 Qualitative und quantitative Forschung

Kennzeichnend für die qualitative Forschung sind Verfahren, die keine verbindlichen Instrumente verwenden und daher zu der offenen Methode zählen. Die Forschung kann so eher tiefere Zusammenhänge erschließen.

⁸ nestor-Arbeitsgruppe Digitale Bestandserhaltung 2012 (S.)

⁹ Schumann et al. 2015

¹⁰ Quandt und Mauer 2012

Im Vergleich zu der qualitativen Forschung ist bei den Methoden der Erhebung von quantitativen Daten die Standardisierung des Untersuchungskontextes besonders wichtig. So muss das zu verwendende Erhebungsinstrument eine Vereinheitlichung zulassen (z.B. muss für die gleiche Untersuchungsfrage auch der gleiche Fragebogen verwendet werden). Nicht nur das Instrument, sondern auch die Situation der Befragung muss dabei standardisiert sein.

Während die qualitative Forschung Daten verbalisiert oder grafisch verarbeitet, werden diese in der quantitativen Forschung in Zahlen gemessen und statistisch ausgewertet.¹¹ Sozialwissenschaftliche Forschungsdaten können auch Jahre oder sogar Jahrzehnte nach ihrer Erhebung von Nutzen sein (z.B. zur Untersuchung des sozialen Wandels).

2.3 Datenformate

Sozialwissenschaftliche Daten liegen in unterschiedlicher Art vor. Da wir uns im Bereich der empirischen Sozialforschung bewegen, werden vor allem Analyse-Systeme wie SPSS, Stata oder SAS mit dem zugehörigen Format verwendet.¹² Weiterhin nimmt die Nutzung von „R“¹³, einer Open Source Lösung, zu. In den meisten Fällen sind die Datenformate jedoch nicht oder nur teilweise zur Langzeitarchivierung geeignet. Um diese jedoch zu gewährleisten, müssen die Daten in Formate umgewandelt werden, die eine langfristige Speicherung gewährleisten können (z.B. Überführung der Daten in *.csv und deren Metadaten in textbasierte Steuerungsdateien, z.B. *.txt).¹⁴ Zur Abspeicherung ganzer Dokumentationen wird nur teilweise das Format XML verwendet, welches jedoch für die einheitliche Nachnutzung sinnvoll wäre.

2.4 Datenschutz

In den Sozialwissenschaften ist die Einhaltung des Datenschutzes einer der wichtigsten Aspekte, da immer wieder mit sensiblen bzw. personenbezogenen Daten gearbeitet wird. Welche Möglichkeiten sich bieten, um dieser Aufgabe gerecht zu werden, soll im Folgenden geklärt werden.

¹¹ Büttner et al. 2011

¹² Schumann und Watteler 2013

¹³ <https://www.r-project.org/>

¹⁴ Quandt und Mauer 2012

2.4.1 Anonymisierung

Um die eindeutige Zuordnung einer Person zu ihren Daten zu verhindern, werden nach der Beendigung der Forschungsarbeit die zugehörigen IDs (Identifier) in den Kontaktprotokollen bzw. gewisse gewählte Nummernfolgen, insofern es sich um eine telefonische Befragung gehandelt hat, gelöscht. Die Daten werden also anonymisiert und lassen sich nicht mehr eindeutig einer bestimmten Person zuordnen. Besondere Regelungen sind für Panel-Studien einzuhalten, bei denen die Anonymisierung eine Voraussetzung dafür bildet, dass die Daten von Dritten genutzt werden können. Des Weiteren kann von einer **faktischen Anonymisierung** gesprochen werden, wenn nur mit einem unverhältnismäßig hohen Aufwand eine bestimmte Person identifiziert werden kann.

2.4.2 Pseudoanonymisierung

Um eine Re-Identifikation von personenbezogenen Daten zu verhindern, wird der Datennutzer darauf hingewiesen, dass eine De-Anonymisierung verboten ist und gleichzeitig ist für ihn eine Verpflichtungserklärung zu unterschreiben. Des Weiteren werden nur Teile des Datensatzes zur Verfügung gestellt und eventuell werden die Daten sogar nur für bestimmte wissenschaftliche Zwecke freigegeben.

2.4.3 Transkription

Insbesondere bei qualitativen Interviews oder Gruppendiskussionen werden viele persönliche Details mit den Daten geliefert. Eine Anonymisierung kann daher sehr aufwendig sein und mit Verlusten in der Datenqualität einhergehen, wenn die Daten archiviert und für Re-Analysen genutzt werden sollen.¹⁵

2.5 Datenzugang

Bezüglich des Zugangs zu Forschungsdaten können verschiedene Stufen unterschieden werden. Sogenannte **Public Use Files** können beispielsweise uneingeschränkt von jedem eingesehen werden. **Scientific Use Files** sind wiederum nur für den wissenschaftlichen Gebrauch und nur eingeschränkt zugänglich. Des Weiteren können Daten auch durch hochrestriktive Nutzung nur in der Institution, welche die Daten innehat, in dafür vorgesehenen Räumlichkeiten, eingesehen werden. Ebenso können Dritte aber auch von der Nutzung der Daten komplett ausgeschlossen werden. Die Einschränkungen ergeben

¹⁵ Häder 2009

sich im Zusammenhang mit dem Urheberrecht oder gesetzlichen Bestimmungen wie dem Datenschutzrecht (z.B. Daten von Verwaltungen oder Mikrodaten der amtlichen Statistik).¹⁶

3 Metadaten, Standards und Richtlinien

Neben den Qualitätsstandards der Marktforschung (ISO 20252:2012)¹⁷ und ethischer Prinzipien (ESOMAR Kodex)¹⁸, die an dieser Stelle nur erwähnt sein sollen, aber keineswegs irrelevant für die Sozialwissenschaften sind, stehen im Folgenden insbesondere Standards und Richtlinien zur digitalen Langzeitarchivierung von sozialwissenschaftlichen Forschungsdaten im Fokus, die dann zu der Auswertung der Umsetzung führen sollen (Use Cases).

3.1 DDI-Standard

Die Data Documentation Initiative (DDI)¹⁹, ist 1995 aus einem Projekt des amerikanischen Datenarchivs ICPSR (Interuniversity Consortium for Political and Social Research)²⁰ entstanden, welches die standardisierte Dokumentation von sozialwissenschaftlichen Daten verbessern sollte. Im Jahr 2003 wurde dann die DDI-Alliance gegründet und eine systematische Entwicklung von Standards und Instrumenten vorangetrieben. Daraus entstand ein offener Standard zur Beschreibung von sozial- und wirtschaftswissenschaftlichen Daten, der in DDI- Codebook und DDI-Lifecycle unterschieden werden kann. Dabei soll der gesamte Lebenszyklus (data life cycle) der Forschungsdaten abgedeckt und durch XML (Extensible Markup- Language) mit Metadaten beschrieben werden. Der Standard wird u. a. von dem sozio-ökonomischen Panel²¹ sowie dem GESIS Leibniz- Institut für Sozialwissenschaften genutzt.

3.1.1 DDI- Codebook (2.5)

Die DDI-Codebook Spezifikation, kurz DDI-C, dient der Dokumentation einfacher Umfragen. Dabei können Mikrodaten, Aggregatdaten und geografische Angaben beschrieben werden.

¹⁶ Quandt und Mauer 2012

¹⁷ <https://www.iso.org/obp/ui/#iso:std:iso:20252:ed-2:v1:en>

¹⁸ ICC/ESOMAR 2008

¹⁹ <http://www.ddialliance.org/>

²⁰ <https://www.icpsr.umich.edu/>

²¹ <http://www.diw.de/soep>

3.1.2 DDI- Lifecycle (3.2)

Der DDI- Lifecycle (DDI-L) ist eine umfangreiche Erweiterung des DDI-C und führt den Ansatz über die Metadaten weiter. Er soll vor allem das Forschungsdatenmanagement in Projekten, Archiven sowie anderen beteiligten Institutionen unterstützen und dabei den gesamten Lebenszyklus der Forschungsdaten mit Hilfe von Metadaten abbilden. Sowohl einfache als auch komplexe Daten sollen dadurch möglichst vollständig und präzise durch reichhaltige Metadaten beschrieben werden und eine Überlieferung des Analysepotenzials der Daten für Nutzer gewährleisten. Außerdem ist die Wiederverwendbarkeit der Metadaten ein wichtiger Aspekt. Diese kann interessant werden, sobald z.B. Replikationen von Erhebungen in der Dokumentation der Daten vorgenommen werden. DDI-L ist so aufgebaut, dass Metadaten bereits bei der Entstehung erfasst und ausgetauscht werden können.

Die DDI- Spezifikationen erleichtern sowohl die Programmierung von Werkzeugen als auch die Verarbeitung und Erschließung von Informationen durch Verwendung von XML (Extensible Mark-up Language). Des Weiteren wird dadurch der flüssige Austausch der Metadaten zwischen den verschiedenen Organisationseinheiten im Lebenszyklus der Forschungsdaten gewährleistet.²²

Ein ähnlicher, bedeutender Standard neben DDI ist SDMX (Statistical Data and Metadata Exchange)²³, der besonders für statistische Aggregatdaten verwendet wird.

3.2 ELSST

Der European Language Social Science Thesaurus (ELSST)²⁴ ist aus dem britischen Thesaurus HASSET (Humanities and Social Science Electronic Thesaurus)²⁵ des UK Data Archive entstanden. Die Entwicklung zu einem multilingualen Thesaurus wurde durch Mitglieder des Europäischen Datenarchivverbunds CESSDA (Council of European Social Science Data Archives)²⁶ in mehreren Projekten umgesetzt. Mittlerweile ist die Suche in fünfzehn verschiedenen Sprachen mit einem Umfang von 3287 Deskriptoren und 2539 Synonymen möglich. Schwerpunktmäßig werden die Disziplinen Politik-

²² Gregory 2011

²³ <https://sdmx.org/>

²⁴ <https://elsst.ukdataservice.ac.uk/Thesaurus-search>

²⁵ <https://hasset.ukdataservice.ac.uk/>

²⁶ <http://cessda.net/>

wissenschaft, Soziologie, Wirtschaftswissenschaft, Bildungsforschung, Erziehungswissenschaft, Rechtswissenschaft, Bevölkerungsforschung, Gesundheitswesen sowie Arbeitsmarkt- und Berufsforschung abgedeckt und der Thesaurus enthält u. a. empirisch-sozialwissenschaftliche Methodenbegriffe. Er soll Disziplin- und Sprachgrenzen überwinden und dadurch einen besseren Datenzugang gewährleisten.²⁷ Er findet u. a. im CESSDA- Katalog und im GESIS- Datenarchiv Verwendung.

4 Methoden und Instrumente

4.1 Befragung

Die am häufigsten verwendete Erhebungsmethode ist die Befragung. Sie ermöglicht die Erfassung von subjektivem Sachverhalten (Einstellung, Werte, Erwartungen, Wissen). Es können die folgenden Arten der Befragung unterschieden werden.

4.1.1 Papierfragebogen

Das Paper und Pencil Interview (PAPI) kann auf verschiedene Arten durchgeführt werden. Zum einen kann es in einer mündlichen Befragung abgehalten werden, d. h., es wird ein persönliches Interview durchgeführt, bei dem der Befragte dem Interviewer direkt gegenüber sitzt. Weiterhin kann es in Form einer schriftlichen Gruppenbefragung oder in Form einer Mail Survey (Versand des Fragebogens über E-Mail) stattfinden. Möglich ist außerdem die Befragung am Telefon, indem der Interviewer den Fragebogen ausfüllt.

4.1.2 Computergestützte Befragungen

Wird die Befragung mit Hilfe von einem Computer durchgeführt, wird dabei zwischen dem Computer Assisted Personal Interview (CAPI), dem Computer Assisted Telephone Interview (CATI), sowie einem Computer Assisted Self Interview (CASI) und einem Online Survey (standardisiert) unterschieden.

Nach Strukturierungsgrad wird zwischen einem standardisierten Interview (quantitativ) sowie einem Leitfadengespräch und einem narrativen Interview unterschieden (beides qualitativ). Bei einer standardisierten Befragung werden dem Befragten immer die glei-

²⁷ Balkan et al. 2011

chen Fragen mit den zugehörigen Antwortalternativen vorgelegt, sodass die Situation immer konstant bleibt. So kann eine objektive und zuverlässige Datenerhebung gewährleistet werden.

Weiterhin kann in Bezug auf die Häufigkeit der Befragung bzw. die zeitliche Ausdehnung zwischen Querschnitt- und Längsschnittstudien unterschieden werden. Bei Querschnittstudien wird die Befragung einer Population nur einmalig durchgeführt, im Falle einer Längsschnittstudie geschieht dies mehrmals. Dabei treten wiederum verschiedene Formen der Untersuchung auf (Folge-Studie, Panel-Studie u. Trend- oder Zeitreihenanalysen).²⁸

4.2 Beobachtung & Inhaltsanalyse

Diese Formen der Datenerhebung sind der qualitativen Forschung zuzuordnen und sollen daher nur kurz angesprochen werden. Bei der Beobachtung finden eine indirekte Beobachtung von sprachlichen Äußerungen (z.B. während einer Befragung) und die Betrachtung von menschlichem Verhalten statt. Die Inhaltsanalyse nimmt eine systematische Analyse des offenbaren Inhalts von Kommunikation (Texte, Bilder u. Film) vor.

4.3 Erhebungsprogramme

4.3.1 ALLBUS

Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)²⁹ dient zur Erhebung von Daten über die Bevölkerung der Bundesrepublik Deutschland. Im Abstand von zwei Jahren (seit 1980) wird regelmäßig ein repräsentativer Querschnitt zu Einstellungen, Verhaltensweisen sowie sozialen Strukturen befragt. Das Fragenprogramm wird dabei teilweise variabel und teilweise konstant gehalten. Dabei werden jeweils pro Erhebung zwei Themenschwerpunkte bearbeitet. Nach der Aufbereitung und Dokumentation werden die Daten allen Interessenten aus Forschung und Lehre im GESIS- Datenarchiv³⁰ zur Verfügung gestellt.

²⁸ Wittenberg 2005

²⁹ Blohm et al. 2015

³⁰ <http://www.gesis.org/allbus/allbus-home/>

4.3.2 ISSP

Das International Social Survey Program (ISSP)³¹ wurde 1984 von Australien, Deutschland, Großbritannien und den USA als internationales Kooperationsprogramm gegründet. Inzwischen besteht es aus fünfzig Mitgliedsländern, die jährlich eine gemeinsame Umfrage zu sozialwissenschaftlich relevanten Themen durchführen. Dabei basiert das Programm auf freiwilliger Mitarbeit und Kooperation. Entscheidungen über Verpflichtungen einzelner Mitglieder oder der verwendeten Fragebögen werden gemeinsam in einer Abstimmung getroffen. Durch die andauernde Zusammenarbeit und die Anlegung der Umfragen auf Replikation wird eine international vergleichende Forschung möglich gemacht und diese in Forschungsprogramme der ISSP- Mitgliedsländer integriert.

5 Use Cases

Mit dem Hintergrund, wie sozialwissenschaftliche Forschungsdaten einzuordnen sind und welche Standards und Richtlinien für das Forschungsdatenmanagement von Bedeutung sind, sollen nun konkrete Beispiele aus der empirischen Sozialforschung betrachtet werden. Zum einen handelt es sich um eine Querschnittsstudie aus 2014 und zum anderen um den registergestützten Zensus aus 2011, der durch Vollerhebungen und Stichprobenerhebungen ergänzt wird.

5.1 Beispiel aus dem GESIS- Datenarchiv (ALLBUS 2014)³²

Über das GESIS- Datenarchiv erhält der Nutzer (z. T. eingeschränkter) Zugriff auf die Daten der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften. Beispielhaft soll die Studie aus dem Jahr 2014 (Screenshot s. Anhang) betrachtet werden.

Die Studie wird durch die **Dokumentation** der bibliografischen Angaben (**Zitation**, Studiennummer, Titel etc.), Beschreibungen des Inhalts (Schwerpunkte der Untersuchung), sowie der Methodologie (Untersuchungsgebiet, Auswahl, Erhebungsmethode, Erhebungszeitraum usw.) ausführlich beschrieben.

³¹ <http://www.issp.org/index.php>

³²

<https://dbk.gesis.org/dbksearch/sdesc2.asp?no=5241&search=Befragung&search2=&DB=d&tab=0¬abs=&nf=1&af=&ll=10>

Der **Zugang** zu den Daten ist über Zugangskategorien geregelt, die auch in der Benutzerordnung der Abteilung Datenarchiv für Sozialwissenschaften festgehalten sind (s. Abb. 1). Im vorliegenden Beispiel ist die Kategorie A zutreffend, einzige Beschränkung ist dann noch die Registrierung bzw. das Log-in im Datenarchiv, um die Daten auch herunterladen zu können.

Kategorie 0	Daten und Dokumente sind für jedermann freigegeben.
Kategorie A	Daten und Dokumente sind für die akademische Forschung und Lehre freigegeben.
Kategorie B	Daten und Dokumente sind für die akademische Forschung und Lehre freigegeben, wenn die Ergebnisse nicht veröffentlicht werden. Sollte eine Veröffentlichung oder eine weitergehende Verarbeitung der Ergebnisse geplant sein, ist eine Genehmigung über das Datenarchiv einzuholen.
Kategorie C	Daten und Dokumente sind für die akademische Forschung und Lehre nur nach schriftlicher Genehmigung des Datengebers zugänglich. Das Datenarchiv holt dazu schriftlich die Genehmigung unter Angabe des Benutzers und des Auswertungszweckes ein.

Abbildung 1: Zugangskategorien des GESIS- Datenarchivs (Quelle: GESIS Leibniz Institut für Sozialwissenschaften)

Die Datensätze werden in den **Formaten** von SPSS/SPSS portable und STATA geliefert und zudem in verschiedenen Sprachen bereit gestellt (bei internationalen Studien). Eine Liste der verwendeten Variablen wird als *.txt Textdokument und die Fragebögen in den **Sprachen** deutsch und englisch werden als PDF-Dokument bereitgestellt, ebenso die Codebücher, Methodenberichte und Studienbeschreibungen. Die Benennung der einzelnen Dokumente setzt sich dabei immer aus dem Studien- Identifikator und den Codes des Dokumenttyps zusammen (evtl. noch Länder-Kodes (Verwendung der ISO3166-2 oder ISO3166-3)³³ und Sprachversionen).

Weiterhin wird die Studienbeschreibung als DDI-C und DDI-L über den **Standard** der DDI zur Verfügung gestellt (in XML).

Außerdem besteht die Möglichkeit der **Suche** in Fragetexten (in Codebüchern und Handbüchern der Studie).

³³ http://www.iso.org/iso/country_codes

5.2 Beispiel aus der Datenbank des Statistischen Bundesamtes (Zensus 2011)³⁴

Der Einstieg in die Datenbank des Zensus 2011 erfolgt über verschiedene **Darstellungsoptionen**, welche die Daten jeweils passgerecht präsentieren sollen. „Ergebnisse einfach und schnell“ (Screenshot s. Anhang) bedeutet, dass Region und Thema gewählt werden können und dann vordefinierte Tabellen und Diagramme dargestellt werden.

In der Kategorie „Ergebnisse dynamisch und individuell“ (Screenshot s. Anhang) lassen sich variable Tabellen und Diagramme erstellen, sowie außerdem Gemeinde- und Regionalvergleiche anlegen. Dies geschieht über die Auswahl einer Region/eines Ortes, die statistische Einheit sowie die zugehörigen Merkmale.

Möchte der Nutzer die „Ergebnisse kartografisch und visuell“ (Screenshot s. Anhang), kann er verschiedene Indikatoren auf Landes-, Kreis- und Gemeindeebene darstellen lassen und hat die Möglichkeit, Klassenzahl und Klassengrenzen anzupassen.

Zusätzliche Informationen (z.B. zu den Merkmalen) können durch ein Mouse-Over erhalten werden.

Bei den ersten beiden Darstellungsoptionen werden dem Nutzer die Tabellen in den **Datenformaten** *.csv, *.pdf und *.xls angeboten. Das Diagramm kann außerdem als *.png heruntergeladen werden. Es handelt sich hierbei um bereits **ausgewertete Daten**. In der dritten Darstellungsoption wird die Möglichkeit zum Download der Karte als *.svg gegeben.

In den **Hinweisen** befinden sich Informationen zur jeweiligen Darstellung (z.B. Bedeutung der Bevölkerungsdichte) und **allgemeine Informationen** zum Zensus, der Erhebungsmethode und rechtlichen Grundlagen.

5.3 GESIS vs. Statistisches Bundesamt

Im Vergleich zwischen ALLBUS und dem Zensus fällt auf, dass die **Zielgruppe** offenbar sehr unterschiedlich ist. Bei ALLBUS wird auf eine **umfassende Dokumentation** der Studie Wert gelegt, die dann für **Forschung und Lehre** zugänglich sein soll. Bei dem Zensus liegt der Fokus auf einen möglichst einfachen und **visuell/grafisch aufbe-**

³⁴ <https://ergebnisse.zensus2011.de/#Home>:

reiteten Zugriff auf ausgewertete Daten, die für die **allgemeine Bevölkerung** zugänglich sind. Des Weiteren bemüht sich das GESIS- Datenarchiv darum, sich an **Richtlinien und Standards** der Sozialwissenschaften zu halten, während bei dem Statistischen Bundesamt z.B. kein Hinweis darauf existiert, ob die DDI Verwendung findet. In Bezug auf die **Nachnutzung** und die verwendeten **Datenformate** ist zu sagen, dass das Datenarchiv des GESIS übliche Formate verwendet und eine reibungslose Weitergabe der Daten gewährleistet ist (und der Schutz personenbezogener Daten bzw. sensibler Daten durch Zugriffsbeschränkungen geregelt werden). Währenddessen werden bei dem Statistischen Bundesamt nur ausgewertete Daten zur Verfügung gestellt (sind an **rechtliche Auflagen** gebunden). Zur Vergleichbarkeit der Daten und der Aufhebung von sprachlichen Grenzen verwendet das GESIS zweisprachige Dokumentationen- das Statistische Bundesamt bietet diese Möglichkeit überhaupt nicht. Insgesamt sind die Informationen der ALLBUS- Studie viel ausführlicher gehalten als die des Zensus (die Dokumentation aller Zensus-Erhebungen befindet sich an anderer Stelle auf der Seite des Statistischen Bundesamts, jedoch keine Beschreibung durch Metadaten oder Codebooks).

5.4 Fazit

Die Betrachtung der Use Cases hat gezeigt, dass vor allem der Zweck und damit auch die Zielgruppe sozialwissenschaftlicher Forschung völlig unterschiedlich ausfallen kann und die Nachnutzung der Forschungsdaten durch rechtliche Auflagen erheblich beeinträchtigt werden kann.

6 WissGrid

6.1 Vorstellung des Fragenkatalogs

Der vorliegende Leitfaden bzw. die Checkliste zum Forschungsdatenmanagement entstand im Rahmen des WissGrid- Projekts durch die Zusammenarbeit verschiedener Fachdisziplinen. Er unterteilt sich in zwei Aspekte: Zum einen behandelt er die Aufgaben im Lebenszyklus von Forschungsdaten und zum anderen beschäftigt er sich mit den übergreifenden Aufgaben des Forschungsdatenmanagements. Eine Kurzfassung der Checkliste zum FDM befindet sich im Anhang.³⁵

³⁵ Ludwig und Enke 2013 (S.98-100)

6.1.1 Teil I

Der erste Teil des Leitfadens beschäftigt sich mit der Untersuchung des Lebenszyklus von Forschungsdaten (s. Abb. 1) und den damit zu erfüllenden Aufgaben für Wissenschaftlerinnen und Wissenschaftler sowie auch Einrichtungen wie z.B. ein Rechenzentrum.

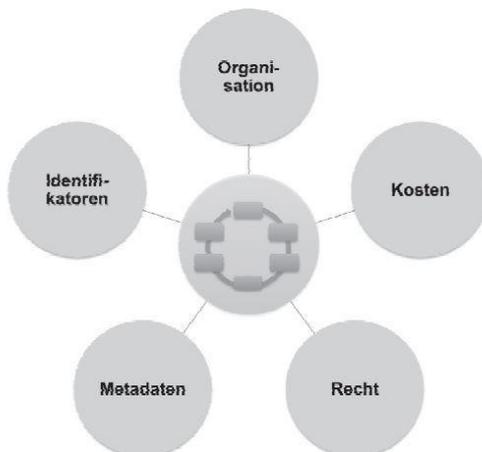


Dazu wird jeder einzelne Aspekt des Zyklus von der Planung/Erstellung der Daten bis hin zum letztendlichen Zugriff und der Nutzung behandelt und Orientierungsfragen gestellt, die bei der Erstellung eines FDM-Plans behilflich sein sollen.

Abbildung 2: Aufgaben im Lebenszyklus von Forschungsdaten (Quelle: Ludwig und Enke 2013)

6.1.2 Teil II

Im zweiten Teil werden übergreifende Themen behandelt, die für alle Aspekte des Lebenszyklus gelten (s. Abb. 2). Organisation (Verantwortliche Einrichtung sowie definierte und dokumentierte Prozesse),



Kosten (Finanzierung der Aufbewahrung von Forschungsdaten), Recht (z.B. Schutz personenbezogener Daten), Metadaten (für jede Phase des Zyklus unverzichtbar) sowie Identifikatoren (eindeutige Identifizierung der Forschungsdaten) bilden damit die Komponenten ab, die für ein erfolgreiches FDM wichtig sind.

Abbildung 3: Übergreifende Aufgaben des Forschungsdaten-Managements (Quelle: Ludwig und Enke 2013)

6.2 Anpassungen für die Sozialwissenschaften

Im Folgenden werden Anpassungen für den WissGrid- Fragenkatalog vorgeschlagen, ausgehend von den Aufgaben im Lebenszyklus von Forschungsdaten. Dazu wird zuerst die Originalfrage aufgeführt und nachfolgend die empfohlenen Ergänzungen oder Anpassungen der Frage oder der Aufgabe.

6.2.1 Planung und Erstellung

6.2.1.1 Arten von Daten

„a) Welche Datenarten werden verwendet bzw. erzeugt? (z. B. Beobachtungsdaten, Simulationsdaten, Video-Interviews, ...)“

Empfehlung:

Welche Erhebungsart wird verwendet (Umfrage, Interview, Beobachtung, Inhaltsanalyse)?

Ist die Erhebung standardisiert?

„c) Wer sind der/die Projektträger oder Finanzgeber?“

Empfehlung:

Wird das Projekt gefördert? Wenn ja, von wem (BMBF, DFG)?

Ist das Projekt Teil eines Förderprogramms (Horizont 2020³⁶)?

„f) In welchen Dateiformaten werden die Daten vorliegen? Mit welchen Datenformaten wird gearbeitet?“

Empfehlung:

Welche Software bzw. welches Datenformat wird zur Erfassung, Verarbeitung und Speicherung von Daten genutzt (SPSS, STATA, SAS, R, XML)?

6.2.2 Auswahl und Aufbewahrungsdauer

6.2.2.1 Aufbewahrung

„a) Wie lange sollen welche Daten aufbewahrt werden? (z. B. bis zum Ende des Projekts, zehn Jahre nach Ende des Projekts, bis zu einem bestimmten Ereignis, unbefristet, ...)“

³⁶ Bundesministerium für Bildung und Forschung (BMBF) 2014

Empfehlung:

Werden neben den Rohdaten auch Erhebungsinstrumente und Dokumentationen bei der Archivierung berücksichtigt (Methodenbericht, Codebooks, Fragebogen)?

6.2.3 Ingest: Einspeisen und Verantwortungsübernahme**6.2.3.1 Verfahren**

„e) Wie werden sensible Daten behandelt?“

Empfehlung:

Welche Verfahren werden verwendet, um personenbezogene Daten zu schützen (Anonymisierung, Pseudoanonymisierung, Transkription)?

6.2.4 Zugriff und Nutzung**6.2.4.1 Nachnutzung und Suchbarkeit**

„a) Können prinzipiell die Daten auch von Anderen innerhalb oder außerhalb des Projekts genutzt werden?“

Empfehlungen:

Sind die Daten in ihrer Rohform zugänglich (keine bereinigten oder veränderten Daten)?

Sind sie suchbar bzw. auffindbar (z.B. über ein Datenarchiv wie GESIS oder CESSDA; Forschungsdatenzentrum z.B. RatSWD; statistische Ämter z.B. statistisches Bundesamt; internationale Organisationen)?

Werden entsprechende Standards zur Wiederverwendung der Daten genutzt (DDI)?

„d) Welche Einrichtungen bzw. Gruppen werden voraussichtlich an den Daten interessiert sein?“

Empfehlungen:

Sind die Daten für Interessierte ansprechend aufbereitet bzw. welche Art der Aufbereitung sollte ausgewählt werden (verschiedene Darstellungsarten: Karte, Tabelle, Diagramm)?

Welche Zielgruppe soll erreicht werden (z.B. Forscherinnen und Forscher oder allgemeine Bevölkerung)?

6.2.6.2 Offener Zugang versus Zugriffsbeschränkungen

„d) Wie werden die Zugriffsbedingungen durchgesetzt und technisch implementiert? Wer ist für die Durchsetzung verantwortlich?“

Empfehlung:

Welche Maßnahmen werden zum Schutz sensibler Daten getroffen (Registrierung bzw. Log-in vor dem Zugriff auf die Daten, Einholen der Zustimmung des Datenarchivs, Nutzung nur mit Erlaubnis des Datenarchivs)?

Welche Stufe(n) des Datenzugangs gibt es (uneingeschränkt, eingeschränkt, hochrestriktive Nutzung, Ausschluss der Nutzung durch Dritte)?

6.2.5 Management, Organisation und Policies

6.2.6 Kosten

6.2.8.1 Kosten- und Aufwandsabschätzung

„a) Sind die Kosten und der Personalaufwand abgeschätzt worden? (z. B. anhand von Vergleichswerten, einer Lebenszyklus-Analyse, etc.)“

6.2.8.2 Anreize

„c) Sind die Kosten für das Datenmanagement ins Verhältnis zu den Kosten der Datenerzeugung gesetzt worden?“

Empfehlung:

Ist das Erhebungsinstrument im Verhältnis zu Personal- und Finanzkapazitäten angemessen ausgewählt?

7 Ergebnisse und Ausblick

Insgesamt hat sich bei der Betrachtung der praktischen Umsetzung des Forschungsdatenmanagement gezeigt, dass zumindest ein Teil dessen unterschiedlich aufgefasst bzw. durchgeführt wird. Angeführte Standards wie DDI und Datenformate zur besseren Nachnutzung wie XML werden teilweise nicht umgesetzt. Da die Sozialwissenschaften in dem gesamten Lebenszyklus der Forschungsdaten sehr komplex sind und eine Vereinheitlichung der gesamten Disziplin sich als äußerst schwierig gestaltet, kann nur auf die bereits vorgestellten Leitfäden, Richtlinien und Standards verwiesen werden und an

die Motivation einer jeden Forscherin und eines jeden Forschers appelliert werden. In ihrem eigenen Interesse sollten sie dafür Sorge tragen, den Forschungsprozess durch einen Forschungsdatenmanagementplan abzubilden, um dadurch insbesondere die Nachvollziehbarkeit und Nachnutzung der Forschungsarbeit zu gewährleisten. Ebenso in der Verantwortung befinden sich die Datenzentren und Datenarchive, die möglichst Richtlinien zur Annahme von Forschungsergebnissen herausgeben sollten, um eine Unvollständigkeit in der Dokumentation zu vermeiden.

8 Literaturverzeichnis

Balkan, Lorna; Jääskeläinen, Taina; Frentzou, Christina; Kappi, Chryssa (2011): European Language Social Science Thesaurus (ELSST): issues in designing a multilingual tool for social science researchers. Hg. v. University of Tartu. Estonia. Online verfügbar unter http://dspace.ut.ee/bitstream/handle/10062/17272/chat2011_submission_1.pdf?sequence=1&isAllowed=y, zuletzt geprüft am 26.01.2016.

Blohm, Michael; Hofstätter, Ute; Schmidt, Katharina; Junghänel, Jakob; Beck, Linda (2015): ALLBUS-Bibliographie. GESIS – Leibniz-Institut für Sozialwissenschaften. Online verfügbar unter <http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/allbus/Bibliographie/Biblio29.pdf>, zuletzt geprüft am 26.01.2016.

Bundesministerium für Bildung und Forschung (BMBF) (Hg.) (2014): Horizont 2020 im Blick. Online verfügbar unter https://www.bmbf.de/pub/horizont_202_im_Blick_2.Auflage.pdf, zuletzt geprüft am 26.01.2016.

Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars; Rümpel, Stephanie; Huschka, Denis; Oellers, Claudia et al. (2011): Handbuch Forschungsdatenmanagement. Hg. v. Stephan Büttner, Hans-Christoph Hobohm und Lars Müller. Bad Honnef. Online verfügbar unter [file:///C:/Users/Toshiba/Downloads/HandbuchForschungsdatenmanagement%20\(1\).pdf](file:///C:/Users/Toshiba/Downloads/HandbuchForschungsdatenmanagement%20(1).pdf), zuletzt geprüft am 26.01.2016.

DFG-Deutsche Forschungsgemeinschaft (2015): Leitlinien zum Umgang mit Forschungsdaten. Online verfügbar unter http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf, zuletzt geprüft am 26.01.2016.

Gregory, Arofan (2011): The Data Documentation Initiative(DDI): An Introduction for NationalStatistical Institutes. Open Data Foundation. Online verfügbar unter http://odaf.org/papers/DDI_Intro_forNSIs.pdf, zuletzt geprüft am 26.01.2016.

Häder, Michael (2009): Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. RatSWD-Rat für Sozial- und WirtschaftsDaten. Online verfügbar unter http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf, zuletzt geprüft am 26.01.2016.

ICC/ESOMAR (Hg.) (2008): ICC/ESOMAR International Code. On Market and Social Research. Online verfügbar unter https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ICCESOMAR_Code_German_.pdf, zuletzt geprüft am 26.01.2016.

Jensen, Uwe (2012): Leitlinien zum Management vonForschungsdaten. Sozialwissenschaftliche Umfragedaten. Unter Mitarbeit von Evelyn Brislinger, Kristina Hauser, Brigitte Hausstein, Alexia Katsanidou, Dafina Kurti, Laurence Horton et al. GESIS – Leibniz-Institut für Sozialwissenschaften. Köln. Online verfügbar unter http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_met_hodenberichte/2012/TechnicalReport_2012-07.pdf, zuletzt geprüft am 26.01.2016.

Klar, Jochen; **Enke**, Harry (2013): Projekt RADIESCHEN. Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur. Report „Organisation und Struktur“. Hg. v. Deutsches Geoforschungszentrum Helmholtz-Zentrum Potsdam. Online verfügbar unter http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:117051:3/component/escidoc:124050/ProjektRadieschen_Organisation_und_Struktur.pdf.

Ludwig, Jens; **Enke**, Harry (2013): Leitfaden zum Forschungsdatenmanagement. Handreichungen aus dem WissGrid-Projekt. Hg. v. Jens Ludwig und Harry Enke. Glückstadt. Online verfügbar unter http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf, zuletzt geprüft am 26.01.2016.

nestor-Arbeitsgruppe Digitale Bestandserhaltung (Hg.) (2012): Leitfaden zur digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung. Version 2.0. nestor. Online verfügbar unter http://files.dnb.de/nestor/materialien/nestor_mat_15_2.pdf, zuletzt geprüft am 26.01.2016.

Quandt, Markus; **Mauer**, Reiner (2012): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Kapitel 4 Sozialwissenschaften. Hg. v. Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump und Jens Ludwig. nestor. Boizenburg. Online verfügbar unter http://www.nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor_lza_forschungsdaten_bestandsaufnahme.pdf, zuletzt geprüft am 26.01.2016.

Schumann, Natascha; **Recker**, Astrid; Müller, Stefan; Mauer, Reiner (2015): Konzept für die Definition Signifikanter Eigenschaften von sozialwissenschaftlichen Forschungsdaten. GESIS – Leibniz-Institut für Sozialwissenschaften. Berlin. Online verfügbar unter <http://www.forschungsdaten.org/images/a/a7/03--schumann--signifikante-eigenschaften-sowi-forschungsdaten.pdf>, zuletzt geprüft am 26.01.2016.

Schumann, Natascha; **Watteler**, Oliver (2013): Forschungsdaten in den Sozialwissenschaften. In: *ZfBB-Zeitschrift für Bibliothekswesen und Bibliographie* 60, 12.12.2013 (6), S. 307–315. Online verfügbar unter http://zs.thulb.uni-jena.de/servlets/MCRFileNodeServlet/jportal_derivate_00235078/j13-h6-auf-2.pdf, zuletzt geprüft am 26.01.2016.

Wittenberg, Reinhard (2005): Einführung in die sozialwissenschaftlichen Methoden und ihre Anwendung in empirischen Untersuchungen I: Skript. Universität Erlangen-Nürnberg. Nürnberg. Online verfügbar unter http://www.ssoar.info/ssoar/bitstream/handle/document/31601/ssoar-2005-wittenberg-Einfuehrung_in_die_sozialwissenschaftlichen_Methoden.pdf?sequence=1, zuletzt geprüft am 26.01.2016.

Anhang

ALLBUS 2014

GESIS Datenbestandskatalog 2.1

Suche **Blättern** Übersicht News Über

Anmelden Warenkorb

Gruppen
Produkte
Studien
Geografische Codes

Mitglied der Leibniz-Gemeinschaft

da|ra Registrierungsagentur für Sozial- und Wirtschaftsdaten

FIND, ACCESS, AND REUSE DATA

DBKSearch 2.1 © GESIS

ZA5240: Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2014

Bibliographische Angaben Inhalt Methodologie **Daten & Dokumente** Errata & Versionen

Weitere Hinweise Veröffentlichungen Gruppen

Datensatz Anzahl der Einheiten: 3471
Anzahl der Variablen: 861
Analyse-System(e): SPSS, Stata

Zugangsklasse A - Daten und Dokumente sind für die akademische Forschung und Lehre freigegeben.

Download von Daten und Dokumenten **Alle Downloads aus diesem Katalog sind kostenlos. Datensätze der Zugangsklassen B und C müssen kostenpflichtig über den Warenkorb bestellt werden.**

Datensätze Fragebögen Codebücher DDI Dokumente

- ZA5240_missing_v2-1-0.sps (Datensatz) 34 KBytes
- ZA5240_v2-1-0.dta (Datensatz STATA) 5 MBytes
- ZA5240_v2-1-0.por (Datensatz SPSS Portable) 7 MBytes
- ZA5240_v2-1-0.sav (Datensatz SPSS) 5 MBytes
- ZA5240_Variablenliste_v2-1-0.txt (Datensatz) 50 KBytes

ZACAT Online Analyse, Datendownload von Subsets in verschiedenen Formaten und Suche in der Dokumentation auf Variablenebene.
[Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2014](#)

Sie können diese Studie kostenpflichtig über den Warenkorb bestellen. [zum Warenkorb hinzufügen](#)

Zensus 2011

Darstellung „Ergebnisse einfach und schnell“

ENGLISH | IMPRESSUM | KONTAKT | zensus2011.de

STARTSEITE **ERGEBNISSE EINFACH UND SCHNELL** ERGEBNISSE DYNAMISCH UND INDIVIDUELL ERGEBNISSE KARTOGRAFISCH UND VISUELL GLOSSAR LINKS

Startseite Ergebnisse einfach und schnell Deutschland

Regionen- und Themenauswahl [zurück zu den Ergebnisszusammenfassungen](#)

Geben Sie Ihren regionalen Bezug an

Wählen Sie Ihr Thema

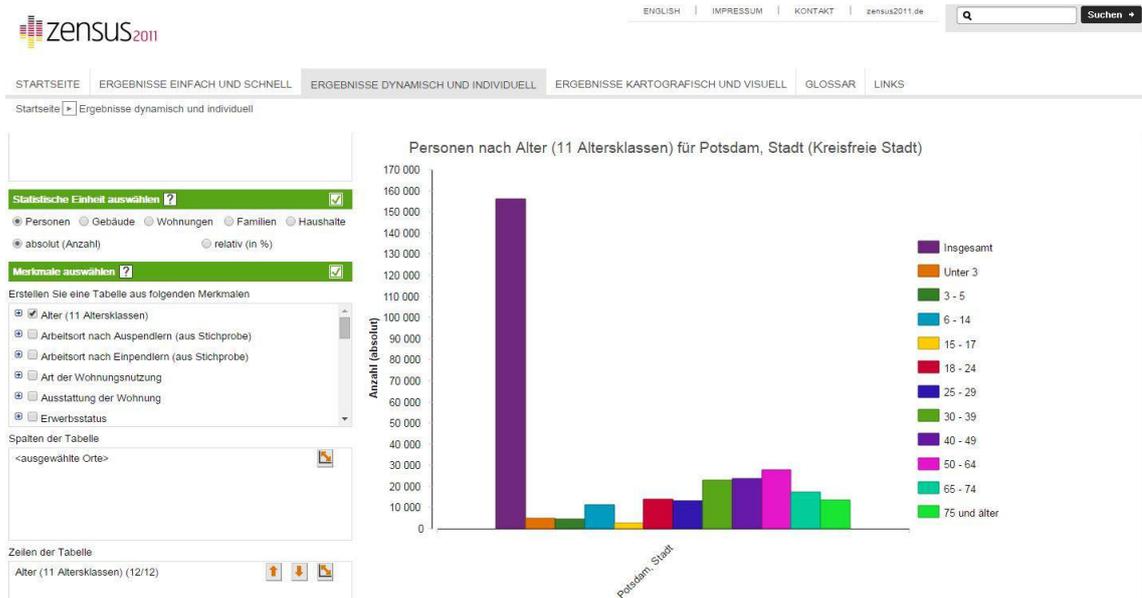
- Gebäude
- Wohnungen
- Haushalte
- Familien
- Demografie**
 - Einwohnerzahlen
 - Alter (5 Altersklassen)
 - Alter (11 Altersklassen)**
 - Alter (5er-Jahresgruppen)
 - Alter (10er-Jahresgruppen)
 - Alter (Infrastrukturrelevante Altersgruppen)
 - Geschlecht
 - Familienstand
 - Familienstand (ausführlich)
 - Staatsangehörigkeitsgruppen
 - Staatsangehörigkeit nach ausgewählten Ländern
 - Regionalergebnisse

Bevölkerung nach Alter (11 Altersklassen) und Typ des privaten Haushalts (nach Familien)
Auszählung aus dem bereinigten Registerbestand

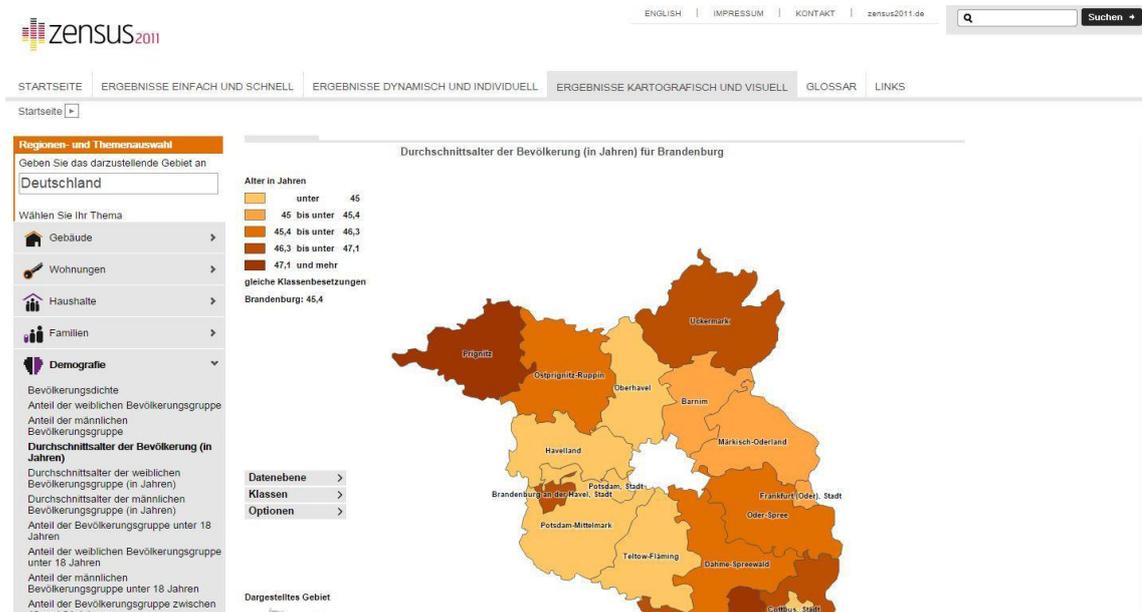
Typ des privaten Haushalts (nach Familien)	Insgesamt		Alter (11 Altersklassen)										
	Anzahl	%	Unter 3	3 - 5	6 - 14	15 - 17	18 - 24	25 - 29	30 - 39	40 - 49	50 - 64	65 - 74	75 und älter
Einpersonenhaushalte (Singlehaushalte)	13 435 845	17,1	-	-	-	-	1 106 963	1 231 842	1 758 654	2 033 705	2 676 735	1 929 255	2 652 720
Paare ohne Kind(er)	21 266 126	27,2	(14 694)	12 956	(40 660)	-	707 399	1 158 697	1 614 294	2 295 614	7 038 237	5 417 904	3 043 541
Paare mit Kind(ern)	34 890 902	44,3	1 681 731	1 681 001	5 454 221	1 759 668	3 319 612	1 717 654	4 962 876	7 418 649	5 236 095	1 072 166	587 029
Alleinerziehende Elternteile	7 039 163	8,9	261 412	306 918	1 197 398	462 768	897 905	392 994	745 539	1 186 289	886 633	315 693	386 614
Mehrpersonenhaushalte ohne Kernfamilie	1 940 946	2,5	(7 731)	8 224	36 110	24 828	378 514	300 446	287 672	262 032	291 225	154 724	189 540
Insgesamt	78 672 982	100,0	1 970 322	2 012 899	6 742 530	2 293 270	6 409 793	4 801 833	9 368 935	13 196 289	16 127 925	8 889 742	6 859 444

Ergebnis des Zensus 2011 zum Berichtszeitpunkt 9. Mai 2011.
© Statistische Ämter des Bundes und der Länder, 2014
Vervielfältigung und Verbreitung mit Quellenangabe gestattet.

Darstellung „Ergebnisse dynamisch und individuell“



Darstellung „Ergebnisse kartografisch und visuell“



Kurzfassung

1. *Planung und Erstellung*

- 1.1 Sind alle Rahmendaten des Projekts dokumentiert (Name, Ziele, Finanzgeber, Laufzeit, Partner, Leiter)?
- 1.2 Wie können bereits existierende Daten integriert/nachgenutzt werden?
- 1.3 Welche Bedeutung haben die Daten für das Vorhabensziel?
- 1.4 Wie lassen sich die verwendeten/erzeugten Daten charakterisieren (Datenarten, Formate, Reproduzierbarkeit)?
- 1.5 Wie werden die Daten erfasst/erstellt?
- 1.6 Wie groß ist die Datenmenge/Produktionsrate?
- 1.7 Wie erfolgt die Qualitätssicherung?

2. *Auswahl und Aufbewahrungsdauer*

- 2.1 Wieso müssen welche Daten aufbewahrt werden?
- 2.2 Wann, durch wen und womit erfolgt die Datenauswahl?
- 2.3 Wie lange müssen die Daten aufbewahrt werden?
- 2.4 Was geschieht bei Ablauf der Aufbewahrungsdauer?

3. *Ingest: Einspeisen und Verantwortungsübernahme*

- 3.1 Wann und wie werden die Daten übergeben/übertragen?
- 3.2 Wann und von wem werden welche Metadaten erfasst?
- 3.3 Wie werden Daten und Metadaten validiert?
- 3.4 Wie wird mit sensiblen Daten umgegangen?
- 3.5 Sind Rechte und Pflichten von Datenproduzent und -archiv geklärt (Protokollierung, Fehlerbehandlung)?

4. *Speicherung und Infrastruktur*

- 4.1 Wer ist während des Projekts und danach für die Speicherung der Daten verantwortlich?
- 4.2 Mit welchen Technologien und an welchen Orten werden die Daten gespeichert?
- 4.3 Werden regelmäßig Sicherheitskopien erstellt und überprüft?
- 4.4 Wie hoch ist die erwartete Datenmenge?
- 4.5 Gibt es besondere Infrastruktur-Anforderungen für Datentransfer, -zugriff und -nutzung? (Netzwerk-Bandbreite, Hardware etc.)

5. *Erhaltungsmaßnahmen*

- 5.1 Sind die eingesetzten Technologien sowie Abhängigkeiten, Nutzungszielgruppe und -anforderungen dokumentiert?
- 5.2 Wird regelmäßig überprüft, ob sich die Anforderungen, verfügbaren Technologien oder Abhängigkeiten verändert haben?
- 5.3 Wird die Aufbewahrungswürdigkeit regelmäßig überprüft?
- 5.4 Gibt es eine Nachfolgeregelung bei einem Wechsel der aufbewahrenden Institution?

6. *Zugriff und Nutzung*

- 6.1 Können die Daten auch von Anderen innerhalb oder außerhalb des Projekts genutzt werden?
- 6.2 Gibt es Verpflichtungen, Daten freizugeben oder nicht freizugeben?
- 6.3 Welche Einrichtungen/Gruppen werden an den Daten interessiert sein?
- 6.4 Wie werden Veröffentlichung, Suchbarkeit und Zugriff realisiert?
- 6.5 Zu welchem Zweck und mit welcher Software werden die Daten voraussichtlich genutzt?
- 6.6 Wie werden ggf. Erstnutzungsrecht, Nutzungsbeschränkungen oder Lizenzbedingungen durchgesetzt?
- 6.7 Spielt Interoperabilität eine Rolle?

7. *Management, Organisation und Policies*

- 7.1 In welcher Art von Repository werden die Daten aufbewahrt?
- 7.2 Ist der Bezugsrahmen national oder international?
- 7.3 Welche Institutionen sind für das Datenmanagement zuständig oder daran beteiligt?
- 7.4 Haben alle Beteiligten eingewilligt und sind deren Beiträge definiert?
- 7.5 Ist der Workflow des Datenmanagements beschrieben?
- 7.6 Sind alle Ressourcen beschrieben und abgeschätzt?
- 7.7 Welche Anforderungen/Auflagen/Policies müssen berücksichtigt/umgesetzt werden? Von wem?
- 7.8 Wie wird die Einhaltung der Planung überprüft?

8. *Kosten*

- 8.1 Wie hoch sind die Kosten und der Personalaufwand für das Datenmanagement während des Projektes / nach dem Projekt?
- 8.2 Wer übernimmt die Kosten?

- 8.3 Stehen alle Verantwortlichen und Beteiligten hinter den Plänen zum Datenmanagement?
9. *Rechtliche Aspekte von Forschungsdaten*
 - 9.1 Sind die Daten aufgrund des Datenschutzes oder aus anderen Gründen sensibel?
 - 9.2 Werden fremde Daten oder Software verwendet, welche dem Urheberrecht, dem Patentrecht o.Ä. unterliegen?
 - 9.3 Unterliegen eigene Daten oder Software dem Urheberrecht/Patentrecht und sind die Nutzungsbedingungen und Rechte geklärt?
 - 9.4 Werden alle notwendigen Maßnahmen zum Datenmanagement eingeräumt?
 - 9.5 Sind evtl. auslaufende Schutzfristen zu berücksichtigen?
 10. *Metadaten*
 - 10.1 Wozu sollen die Metadaten dienen oder benutzt werden?
 - 10.2 Was soll durch Metadaten beschrieben werden?
 - 10.3 Welche Semantik und Syntax wird verwendet?
 - 10.4 Inwieweit können die Metadaten automatisch erstellt werden?
 - 10.5 Welche Voraussetzungen bestehen für die Verarbeitung und das Verständnis der Metadaten?
 11. *Identifikatoren und Informationsobjekte*
 - 11.1 In welchem Verhältnis stehen die Informationsobjekte zueinander?
 - 11.2 Für welche Informationsobjekte werden dauerhafte Identifikatoren benötigt?
 - 11.3 Welche Identifikatoren werden benutzt?
 - 11.4 Wie werden die Identifikatoren nachgewiesen?
 - 11.5 Wer wird die Aktualisierung und Pflege von Identifikatoren vornehmen?

Formale Informationen zur Checkliste

Erstellungsdatum

An Erstellung beteiligte Personen