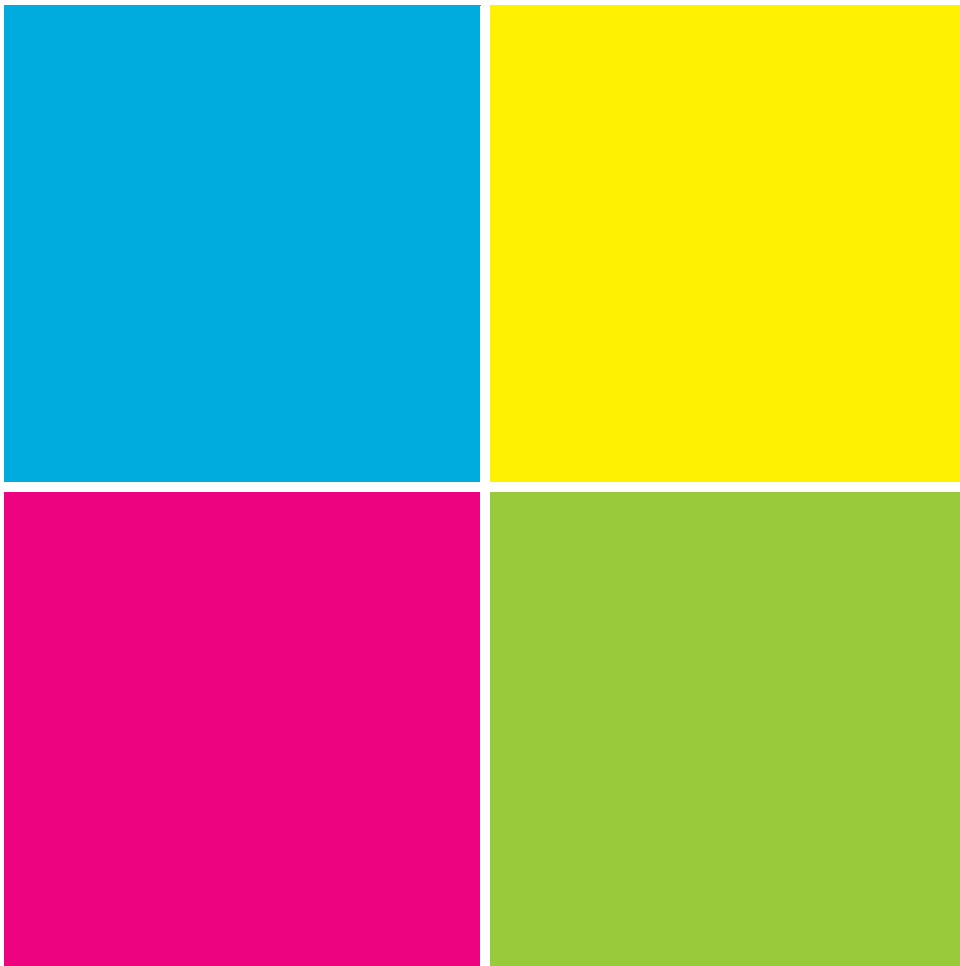


Tätigkeitsbericht

der akkreditierten
Forschungsdatenzentren des
Rates für Sozial- und Wirtschaftsdaten
(RatSWD) für das Jahr 2015



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Inhaltsverzeichnis

	Vorwort der Vorsitzenden	4
1	Überblick über die Forschungsdateninfrastruktur (FDI) des Rates für Sozial- und Wirtschaftsdaten (RatSWD)	6
1.1	Entstehung, Motivation und Datenangebot	6
1.2	Struktur der Forschungsdatenzentren (FDZ)	11
1.3	Datenschutz, Archivierung und Qualitätssicherung	11
1.4	Bereitgestelltes Datenangebot	14
1.5	Serviceangebot	16
2	Aktuelle Entwicklungen der Forschungsdateninfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften	18
2.1	Persistente Identifikatoren (PID): ein Meilenstein zur Professionalisierung des Forschungsdatenmanagements	18
2.2	Niederschwellige Datennutzung durch innovative Lösungen des Remote-Access.....	19

Vorwort der Vorsitzenden

■ Dieser Tätigkeitsbericht stellt erstmals umfassend die Bandbreite und das Service- und Datenangebot der beim Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditierten Forschungsdatenzentren (FDZ) dar.

Seit seiner Gründung im Jahr 2004 setzt sich der RatSWD für eine Verbesserung des Datenzugangs für die sozial-, verhaltens- und wirtschaftswissenschaftliche Forschung in Deutschland ein. Hierfür wurde u. a. das Modell des Forschungsdatenzentrums geschaffen, das der Wissenschaft einen flexiblen und umfangreichen Datenzugang ermöglicht.

Durch den Aufbau und die Akkreditierung von mittlerweile 30 Forschungsdatenzentren wurde eine Forschungsdateninfrastruktur geschaffen, die nicht nur im Vergleich zu anderen Forschungsgebieten in Deutschland sondern auch im internationalen Kontext einmalig ist. Das Spektrum des Datenangebots deckt nahezu alle inhaltlichen Forschungsfelder der beteiligten Fachdisziplinen ab.

Die Forschungsdatenzentren werden durch den RatSWD nach einheitlichen und transparenten Kriterien akkreditiert. Ziel ist es, gleichzeitig eine hohe Qualität des Datenzugangs sowie die strikte Einhaltung von geltenden Datenschutzbestimmungen zu gewährleisten. Damit stehen den Forschenden nicht nur qualitätsgeprüfte Forschungsdaten zur Verfügung. Die Forschungsdatenzentren stellen zudem die Einhaltung der für diese Daten geltenden Datenschutzbestimmungen sicher.

Seit 2010 treffen sich die Vertreterinnen und Vertreter der durch den RatSWD akkreditierten Forschungsdatenzentren regelmäßig im Ständigen Ausschuss Forschungsdateninfrastruktur (FDI Ausschuss). Dieses operativ orientierte Gremium ergänzt den forschungsstrategisch ausgerichteten RatSWD. Es vertritt alle Belange der akkreditierten Forschungsdatenzentren, die über den RatSWD an Zuwendungsgebende, Politik, Wissenschaft, Datenschutz und Ethikkommissionen auf nationaler wie internationaler Ebene kommuniziert werden. Gemeinsam mit dem FDI Ausschuss erarbeitet der RatSWD Empfehlungen zur Weiterentwicklung und Qualitätssicherung der Forschungsdatenzentren, um sie an neue Herausforderungen, wie z. B. die Langzeitarchivierung, anzupassen und (Harmonisierungs-)Potenziale vollständig zu erschließen.¹

Um diese Potenziale zu erfassen und um die Qualitätssicherung der Daten zu überprüfen, beteiligen sich alle akkreditierten Forschungsdatenzentren an einem jährlichen Berichtswesen. Sie geben Auskunft über ihren Datenbestand, den Umfang der Datennutzung und des Dienstleistungsangebots, ihre Forschungstätigkeit, ihre qualitätssichernden Maßnahmen und ihre Personalausstattung. In den Jahren 2015 und 2016 wurden die Erfahrungen des bisherigen Berichtswesens evaluiert und sowohl der Akkreditierungsprozess als auch die regelmäßige Qualitätssicherung überarbeitet. Das Ergebnis sind aktualisierte bzw. konkretisierte Akkreditierungskriterien sowie ein transparenteres Berichtswesen.²

1 Vgl. RatSWD (2015). Stellungnahme des RatSWD zur Archivierung und Sekundärnutzung von Daten der qualitativen Sozialforschung. http://www.ratswd.de/dl/RatSWD_Stellungnahme_QualiDaten.pdf (11.01.2017)

RatSWD (2016). Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften. Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte. http://www.ratswd.de/dl/RatSWD_Output3_Forschungsdatenmanagement.pdf (11.01.2017)

2 Die Akkreditierungskriterien sowie die Prozesse zur Qualitätssicherung können online eingesehen werden, unter: <http://www.ratswd.de/forschungsdaten/qualitaetssicherung>.

Im Rahmen dieses Überarbeitungsprozesses hat der RatSWD beschlossen, eine Monitoringkommission – bestehend aus vier Mitgliedern des FDI Ausschusses – einzurichten, die zeitgleich zur 6. Berufenungsperiode des RatSWD ab 2017 eingesetzt wird und u. a. die jährlichen Berichte kontrolliert sowie einen jährlichen Tätigkeitsbericht aller Forschungsdatenzentren verfassen wird.

Der hier vorliegende erste Tätigkeitsbericht wurde von einer Kommission (der Pilot-Monitoringkommission) erarbeitet, an der sich alle Mitglieder des FDI Ausschusses beteiligen konnten.³ Ziel war es, damit eine Vorlage zu erarbeiten, die das breite Leistungsspektrum der Forschungsdatenzentren erfasst und als Orientierung für künftige Berichte dienen kann.

Kapitel 1 veranschaulicht die Leistungsfähigkeit der Forschungsdatenzentren und basiert auf einer Fragebogenerhebung unter allen akkreditierten Forschungsdatenzentren. Die Angaben beziehen sich hier – sofern nicht anders angegeben – auf das Jahr 2015. Zudem sind in diesem Kapitel auch Aspekte dargestellt, die einmalig bei der Erstakkreditierung eines Forschungsdatenzentrums erfasst werden, wie beispielsweise die Motivation zur Gründung eines Forschungsdatenzentrums.

Kapitel 2 widmet sich aktuellen Entwicklungen in der Forschungsdatenlandschaft. Ein zentrales Entwicklungsfeld der letzten Jahre ist die Entwicklung und Nutzung persistenter Identifikatoren für Datensätze (wie beispielsweise DOI); diese erleichtern die Zitation der durch die Forschungsdatenzentren angebotenen Datensätze für die Wissenschaft und erlauben, die Arbeiten in den Forschungsdatenzentren angemessen zu würdigen. Ein weiteres Entwicklungsfeld ist die Entwicklung von Remote-Access Lösungen in den Forschungsdatenzentren, die die Datennutzung unabhängig vom physischen Standort der jeweiligen Forschungsdatenzentren ermöglichen.

Der RatSWD und der FDI Ausschuss bedanken sich sehr herzlich bei den Mitgliedern der Pilot-Monitoringkommission für ihr Engagement und das Verfassen dieses Berichtes.

Prof. Regina T. Riphahn, Ph.D.
(Vorsitzende des RatSWD)

Stefan Bender
(Vorsitzender des RatSWD)

Heike Habla
(Vorsitzende des FDI Ausschusses)

Prof. Dr. Stefan Liebig
(Vorsitzender des FDI Ausschusses)

³ Vgl. die Liste der Mitglieder dieser Kommission auf Seite 22.

1 Überblick über die Forschungsdateninfrastruktur (FDI) des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

1.1 Entstehung, Motivation und Datenangebot

Wie alles begann:

Auf Initiative der Wissenschaft⁴ wurde 1999 die „Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik“ (KVI) gegründet. Dieses Gremium legte im März 2001 sein 300 Seiten umfassendes Gutachten vor. Eine der zentralen institutionellen Empfehlungen der Kommission war die Einrichtung von Forschungsdatenzentren (FDZ) bei den öffentlichen Datenproduzenten sowie die Gründung zentraler Datenservicezentren (DSZ). Im selben Jahr wurde der Gründungsausschuss (GA) des Rates für Sozial- und Wirtschaftsdaten (RatSWD) konstituiert.

Weiterentwicklung und Sicherung:

Zwischen 2001 und 2008 entwickelte sich die Forschungsdateninfrastruktur durch die Gründung verschiedener Forschungsdatenzentren stetig weiter.

Im Jahr 2004 wurde der RatSWD gegründet.

Um den produktiven Dialog zwischen den FDZ und DSZ zu fördern, wurde im Jahre 2009 der „Ständige Ausschuss Forschungsdateninfrastruktur (FDI Ausschuss)“ eingerichtet. Hauptaufgabe dieses Gremiums ist die kontinuierliche Sicherung und Verbesserung der Forschungsdateninfrastruktur, also die Qualität und Quantität des Datenangebotes und des Datenzugangs voranzutreiben.

Zwecks Qualitätssicherung der Forschungsdateninfrastruktur hat der RatSWD im Jahr 2010 Mindeststandards und Kriterien für die Akkreditierung erarbeitet.

Aktueller Stand:

Heute steht der Forschung eine hochentwickelte und nachhaltige Forschungsinfrastruktur zur Verfügung, die „bottom up“ aus der Wissenschaft selbst entstanden ist. Ein dezentrales Netzwerk von 30 vom RatSWD akkreditierten Forschungsdatenzentren ermöglicht einen kostengünstigen und einfachen Zugang zu einer Vielzahl von Forschungsdaten.

Nicht nur die Zahl der Forschungsdatenzentren, sondern auch die Bandbreite der Disziplinen hat sich stetig erhöht. Neben den Sozial- und Wirtschaftsdaten wird mit einem erweiterten Datenangebot, das auch Gesundheits- und Bildungsdaten sowie Daten für die psychologische Forschung umfasst, auf neue Anforderungen innovativer Forschung reagiert. Inzwischen stehen nicht nur quantitative, sondern auch qualitative Daten zur Verfügung. Insgesamt werden über die akkreditierten Forschungsdatenzentren der Forschungsdateninfrastruktur des RatSWD rund 3000 Datensätze für die wissenschaftliche Forschung angeboten, die in einem Zeitraum von über 50 Jahren erhoben wurden.

⁴ Richard Hauser, Gert G. Wagner und Klaus F. Zimmermann: „Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung“ 1998. / Wolfgang Zapf et al: „Memorandum zur Verbesserung der Zugangsmöglichkeiten zu Mikrodaten der amtlichen Statistik“. In: ZUMA-Nachrichten, Nr. 39, S. 172–176.

Abb. 1: Entwicklung der Forschungsdateninfrastruktur (FDI) des RatSWD

- 1999** Empfehlung der „Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik“ (KVI) zur **Einrichtung von Forschungsdatenzentren (FDZ)**
- 2001** Konstituierung des Gründungsausschusses des RatSWD und **Gründung⁵** des **FDZ des Statistischen Bundesamtes**
- 2002** **Gründung⁵** des **FDZ der Statistischen Landesämter**
- 2003** **Gründung⁵** des **FDZ GML** und des **FDZ IZA (IDSC)**
- 2004** **Gründung⁵** des **FDZ BA im IAB**, des **FDZ-RV** und des **RatSWD**
- 2008** **Akkreditierung** des **BIBB-FDZ** und des **FDZ IQB**
- 2009** **Einrichtung** des Ständigen Ausschuss Forschungsdateninfrastruktur (**FDI Ausschuss**) des RatSWD
Akkreditierung des **FDZ SOEP**, des **FDZ SHARE**, des **FDZ Internationale Umfrageprogramme**, des **FDZ Wahlen** und des **FDZ ALLBUS**
- 2010** **Erarbeitung von Kriterien für die Akkreditierung** durch den RatSWD
Akkreditierung des **FDZ-BO**, des **FDZ-DZA**, des **FDZ PsychData** des **ZPID**, des **FDZ-pairfam** und des **FDZ Ruhr am RWI**
- 2011** **Akkreditierung** des **EBDC**, des **FDZ Gesundheitsmonitoring am RKI** und des **FDZ BZgA**
- 2012** **Akkreditierung** des **FDZ Wissenschaftsstatistik**, des **FDZ Bildung**, des **FDZ-LifBi** und des **ZEW-FDZ**
- 2013** **Akkreditierung** des **FDZ-DJI** und des **FDZ SFB 882**
- 2014** **Akkreditierung** des **FDZ-AGD** und des **FDZ PIAAC**
- 2015** **Akkreditierung** des **FDSZ Bundesbank** und des **FDZ-IWH**

Die Motive zur Gründung eines Forschungsdatenzentrums in Institutionen waren sehr heterogen: Die einen folgten Empfehlungen von Kommissionen (z. B. der KVI), des Wissenschaftsrates (WR) oder von wissenschaftlichen Beraterkreisen, andere erhielten aus der eigenen Einrichtung den Auftrag zur Gründung eines Forschungsdatenzentrums zum Zweck der Forschungsförderung. Allen gemein war jedoch das Bestreben, die Forschungsdateninfrastruktur in Deutschland durch die Bereitstellung von Mikrodaten für die Wissenschaft zu erweitern und zu stärken.

Die Akkreditierung durch den RatSWD bedeutet für alle oben aufgeführten Forschungsdatenzentren ein Qualitätssiegel. Zudem erhoffen sie sich durch die Vernetzung innerhalb der Forschungsdateninfrastruktur „best practice“-Lösungen für die Weiterentwicklungen der eigenen Infrastruktur sowie einen Erfahrungs- und Wissensaustausch.

⁵ Die Forschungsdatenzentren Statistisches Bundesamt, Statistische Ämter der Länder, GML, IZA, BA im IAB und RV sind bereits vor der Gründung des RatSWD eingerichtet worden und gehörten somit ab 2004 zur Forschungsdateninfrastruktur des RatSWD. Bei diesen FDZ wurde das Jahr der Gründung angegeben. Alle anderen Forschungsdatenzentren wurden nach 2004 beim RatSWD akkreditiert. Bei diesen FDZ ist jeweils das Jahr der Akkreditierung angegeben.

Abb. 2: Datenangebot der akkreditierten Forschungsdatenzentren⁶

Forschungsdatenzentrum des Statistischen Bundesamtes (FDZ-Bund) und Forschungsdatenzentrum der Statistischen Ämter der Länder (FDZ-Länder)	Flächendeckender Zugang zu Daten aus den Bereichen: Soziales, Gesundheit, Wirtschaft, Agrarwirtschaft, Umwelt, Rechtspflege und Steuern.
Forschungsdatenzentrum German Microdata Lab bei GESIS (FDZ GML)	Erschließung von Mikrodaten der amtlichen Statistik, darauf bezogener forschungsbasierter Service sowie Entwicklung von Instrumenten zur Umsetzung sozialwissenschaftlicher Konzepte.
Internationales Forschungsdatenzentrum des Forschungsinstituts zur Zukunft der Arbeit (IZA) (FDZ IZA, IDSC)	Nationale und internationale Datensätze zum Arbeitsmarkt mit standardisierten Informationen.
Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB) (FDZ BA im IAB)	Daten zu Personen, Haushalten und Betrieben sowie kombinierte Datensätze aus Befragungsdaten und administrativen Daten im Bereich der Sozialversicherung und Arbeitsmarkt- und Berufsforschung.
Forschungsdatenzentrum der Rentenversicherung (FDZ-RV)	Daten zu Versicherungskonten der in der gesetzlichen Rentenversicherung Versicherten, in denen der Versicherungsverlauf sowie Leistungen an die Versicherten aus Rente und Rehabilitation gespeichert sind.
Forschungsdatenzentrum im Bundesinstitut für Berufsbildung (BIBB-FDZ)	Personen- und Betriebsdatensätze der Berufsbildungsforschung, die sich inhaltlich hauptsächlich mit dem Erwerb und der Verwertung von beruflichen Kenntnissen und Kompetenzen beschäftigen.
Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen (FDZ IQB)	Deutsche Datensätze aus den großen nationalen und internationalen Schulleistungsstudien sowie aus nationalen Studien mit Kompetenzmessungen im Bildungsbereich.
Forschungsdatenzentrum des Sozio-oekonomischen Panels (FDZ SOEP)	Daten aus repräsentativ erhobenen Wiederholungsbefragungen von Privathaushalten, die SOEP-CORE Stichprobe umfasst Themen wie Einkommen, Erwerbstätigkeit, Bildung und Gesundheit. Zudem gibt es die längsschnittliche Innovationsstichprobe (SOEP-IS), die es externen Forschenden ermöglicht, eigene Forschungsvorhaben einzubringen.
Forschungsdatenzentrum des Survey of Health, Ageing and Retirement in Europe (SHARE) (FDZ SHARE)	Datensätze der Längsschnittstudie „SHARE“ zum Gesundheitszustand, der sozioökonomischen Lage sowie den sozialen und familiären Netzwerken von Personen ab 50 Jahren in 20 europäischen Ländern und in Israel.
Forschungsdatenzentrum Internationale Umfrageprogramme bei GESIS (FDZ Internationale Umfrageprogramme)	International vergleichende Umfragedaten aus über 70 Ländern der Welt zu fast allen Themen der Sozialwissenschaften: CSES, EVS, Eurobarometer, Europa-Wahlen, ISSP.

⁶ Die Reihenfolge orientiert sich am Zeitpunkt der Akkreditierung beim RatSWD (siehe Abb. 1 auf Seite 7). Es handelt sich im Folgenden um eine Kurzdarstellung des Datenangebots. Das gesamte Datenangebot kann online eingesehen werden unter: <http://www.ratswd.de/forschungsdaten/fdz>.

Forschungsdatenzentrum Wahlen bei GESIS (FDZ Wahlen)	Zugang zu nationalen Wahlstudien (Bundes- und Landtagswahlen), Politbarometer, Forsa-Bus, ARD Deutschlandtrend. Das derzeit größte Projekt im FDZ Wahlen ist die German Longitudinal Election Study (GLES).
Forschungsdatenzentrum ALLBUS bei GESIS (FDZ ALLBUS)	Daten aus der allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) zu Einstellungen, Verhaltensweisen und Sozialstruktur der Bevölkerung in Deutschland.
Forschungsdatenzentrum Betriebs- und Organisations-Daten (FDZ-BO)	Quantitative und qualitative Betriebs- und Organisationsdaten, verknüpfte Personen- und Betriebsdaten sowie Daten aus Mitarbeiter- und Mitgliederbefragungen.
Forschungsdatenzentrum des Deutschen Zentrums für Altersfragen (FDZ-DZA)	Daten der Langzeitstudien „Deutscher Alterssurvey (DEAS)“, zum Wandel der Lebenssituationen und Altersverläufe der Bevölkerung in der zweiten Lebenshälfte sowie des Deutschen Freiwilligen surveys (FWS) - einer repräsentativen Bevölkerungsumfrage zum freiwilligen Engagement in Deutschland .
Forschungsdatenzentrum PsychData des Leibniz-Zentrums für Psychologische Information und Dokumentation (FDZ PsychData des ZPID)	Bündelung quantitativer Datensätze sowohl aus der Grundlagenforschung als auch aus der angewandten Psychologie, Ausrichtung des Sammelauftrages insbesondere auf Längsschnittstudien, groß angelegte Umfragestudien und Testentwicklungen.
Forschungsdatenzentrum des Beziehungs- und Familienpanels (FDZ-pairfam)	Datensätze aus der repräsentativen, interdisziplinären Längsschnittstudie zu partnerschaftlichen und familialen Lebensformen in Deutschland: „Panel Analysis of Intimate Relationships and Family Dynamics“ (pairfam).
Forschungsdatenzentrum Ruhr am RWI-Leibniz-Institut für Wirtschaftsforschung (FDZ Ruhr am RWI)	Spezialisierung auf regionale Daten: Angebot von Daten auf Ein-Quadratkilometer-Rastern. Neben der Geo-Referenzierung von Daten auf wissenschaftlicher Basis werden unterschiedliche Individual- und Unternehmensdaten aus Forschungsprojekten des RWI zur Verfügung gestellt.
LMU-ifo Economics & Business Data Center (EBDC)	Datensätze deutscher Unternehmen, dazu zählen die vom ifo Institut erhobenen Wirtschaftsdaten zur Konjunkturentwicklung, zur Innovationsfreudigkeit und zum Investitionsverhalten sowie externe Daten zur Unternehmensfinanzierung und zur Governance-Struktur von Unternehmen.
Forschungsdatenzentrum „Gesundheitsmonitoring“ am Robert Koch-Institut (RKI) (FDZ Gesundheitsmonitoring am RKI)	Daten zum Gesundheitszustand und zum Gesundheitsverhalten der Bevölkerung, die auf der Basis von bundesweit repräsentativen Studien gewonnen wurden.
Forschungsdatenzentrum der Bundeszentrale für gesundheitliche Aufklärung (FDZ BZgA)	Daten aus regelmäßig wiederholten bundesweiten Repräsentativbefragungen zur Erreichbarkeit der Bevölkerung mit Aufklärungs- und Präventionsmaßnahmen sowie des Wissens, der Einstellungen und des Gesundheitsverhaltens der Bevölkerung bezüglich der Gesundheitsthemen der BZgA.

Forschungsdatenzentrum Wissenschaftsstatistik des Stifterverbandes für die Deutsche Wissenschaft (FDZ Wissenschaftsstatistik)	Daten über die Forschungs- und Entwicklungsaktivitäten der deutschen Wirtschaft, zum finanziellen Volumen, der Struktur und der regionalen Verteilung von Forschung und Entwicklung (FuE) sowie der in diesem Bereich tätigen Personen.
Forschungsdatenzentrum Bildung am Deutschen Institut für Internationale Pädagogische Forschung (DIPF) (FDZ Bildung)	Datensätze zur Bildungsforschung, insbesondere im Feld der qualitativen Forschungsdaten.
Forschungsdatenzentrum des Leibniz-Instituts für Bildungsverläufe e.V. an der Otto-Friedrich-Universität Bamberg (FDZ-LifBi)	Längsschnittdaten des Nationalen Bildungspanels (NEPS - National Educational Panel Study) zu Kompetenzentwicklungen, Bildungsprozessen, Bildungsentscheidungen und Bildungsrenditen in formalen, nicht-formalen und informellen Kontexten.
Forschungsdatenzentrum des Zentrums für europäische Wirtschaftsforschung (ZEW-FDZ)	Individualdaten von ZEW-Befragungen.
Forschungsdatenzentrum des Deutschen Jugendinstituts (FDZ-DJI)	Daten aus den seit 1988 regelmäßig durchgeführten Surveys zum Aufwachsen von Kindern und Jugendlichen sowie zu Lebenslagen von Erwachsenen und Familien.
FDZ SFB 882 (wurde 2016 aufgelöst ⁷)	Qualitative und quantitative Datensätze aus dem Bereich der Ungleichheitsforschung.
Forschungsdatenzentrum Archiv für Gesprochenes Deutsch am Institut für Deutsche Sprache (FDZ AGD)	Daten zu gesprochenem Deutsch in Interaktionen (Gesprächskorpora) sowie inländische und ausländische Varietäten des Deutschen (Variationskorpora).
Forschungsdatenzentrum PIAAC bei GESIS (FDZ PIAAC)	Nationale und internationale Datensätze des PIAAC (Programme for the Assessment of Adult Competencies). Für Deutschland stehen des Weiteren Regionaldaten sowie Längsschnittdaten zur Verfügung.
Forschungsdaten- und Servicezentrum der Bundesbank (FDSZ Bundesbank)	Anonymisierte Datensätze zu Banken, Wertpapieren, Investmentfonds, Unternehmen und privaten Haushalten.
Forschungsdatenzentrum des Leibniz-Instituts für Wirtschaftsforschung Halle (FDZ-IWH)	Unternehmensdaten aus Panel- und Längsschnittstudien zu Entwicklungstendenzen in der Industrie und im Baugewerbe Ostdeutschlands und zur Standortwahl multinationaler Unternehmen in den neuen Bundesländern und in Mittel- und Osteuropa.

⁷ Je nach Art und Grundlage wurden die Daten des FDZ SFB 882 an verschiedene Organisationen weitergegeben: IAB Daten gingen an das FDZ IAB; qualitative Daten mit Organisationsbezug gingen an das FDZ BO; die restlichen Daten gingen an das fakultätsinterne Archiv SOBI an der Universität Bielefeld (im Aufbau befindlich). (Stand 09.01.2017)

1.2 Struktur der Forschungsdatenzentren (FDZ)

1.2.1 Personal

Die 30 Forschungsdatenzentren beschäftigen insgesamt 222,67 Personen (in Vollzeitäquivalenten, VZÄ) auf nicht mit Drittmitteln geförderten Stellen. Die Anzahl der Beschäftigten variiert dabei sehr stark zwischen den Forschungsdatenzentren, wie die Spannweite der Angaben verdeutlicht. Zwei Drittel der Forschungsdatenzentren beschäftigen darüber hinaus studentische Hilfskräfte.

Abb. 3: Personalbestand in Vollzeitäquivalenten im Jahr 2015

Angaben von 30 Forschungsdatenzentren (n=30).

	Anzahl in VZÄ	Spannweite
Mitarbeitende gesamt	222,67	0,25 – 35
davon wissenschaftliche Mitarbeitende	133,67	0,25 – 23
davon nicht-wissenschaftliche Mitarbeitende	57,30	0 – 12
Studentische Hilfskräfte	31,70	0 – 7

1.2.2 Selbstständige Forschung von wissenschaftlichem Personal

Von 30 Forschungsdatenzentren geben 24 an, wissenschaftlich ausgebildetes Personal zu haben, welches mit den angebotenen Daten selbstständig forscht. Grundsätzlich ist es wünschenswert, dass die Forschungsdatenzentren eigenständige Forschungsaktivitäten durchführen, damit die Mitarbeitenden mit den relevanten Verfahren und Problemlagen vertraut werden und in der Lage sind, Daten-, Service- und Beratungstätigkeiten auf angemessenem wissenschaftlichem Niveau durchführen zu können. Die Forschungstätigkeit in den Datenzentren kann sich auf inhaltliche Fragestellungen beziehen, sie kann methodologische Themen verfolgen oder auf Weiterentwicklungen im technischen Bereich gerichtet sein.

1.2.3 Forschungsk Kooperationen

27 der 30 Forschungsdatenzentren waren 2015 in Forschungsk Kooperationen eingebunden. Von den drei Forschungsdatenzentren, die nicht in Forschungsk Kooperationen eingebunden waren, gaben zwei an, dies zu planen.

1.3 Datenschutz, Archivierung und Qualitätssicherung

1.3.1 Konzepte für die langfristige Verfügbarmachung der Daten

Die langfristige Verfügbarkeit der für die Forschung verwendeten Daten ist Teil der guten wissenschaftlichen Praxis. Immer mehr wissenschaftliche Zeitschriften erwarten, dass publizierte empirische Ergebnisse repliziert werden können. Somit wird es zu einer Aufgabe von Forschungsdatenzentren, Datenbestände langfristig zu archivieren. Dies erfordert, Daten so zu erhalten, dass deren Lesbarkeit und Interpretierbarkeit auch über längere Zeiträume garantiert werden kann.

In den einzelnen Forschungsdatenzentren werden verschiedene Verfahren angewendet, um die Daten in Datenbanken, auf Servern oder in spezialisierten Datenarchiven mit regelmäßigen Sicherungskopien vorzuhalten. Eine besondere Herausforderung an die Archivierung ist es sicherzustellen, dass die Daten-Dateien auch nach vielen Jahren noch gelesen werden können. Dies ist durch neue Versionen gängiger Softwareprodukte nicht immer gewährleistet.

Spezialisierte Datenarchive wenden eine ortsverteilte und redundante Speicherung auf unterschiedlichen Medien an, d. h. die archivierten Daten werden dupliziert an mehreren physischen Orten gleichzeitig und unter dem Einsatz verschiedener Speichermedien (z. B. Festplattenlaufwerke, CDs, etc.) vorgehalten.

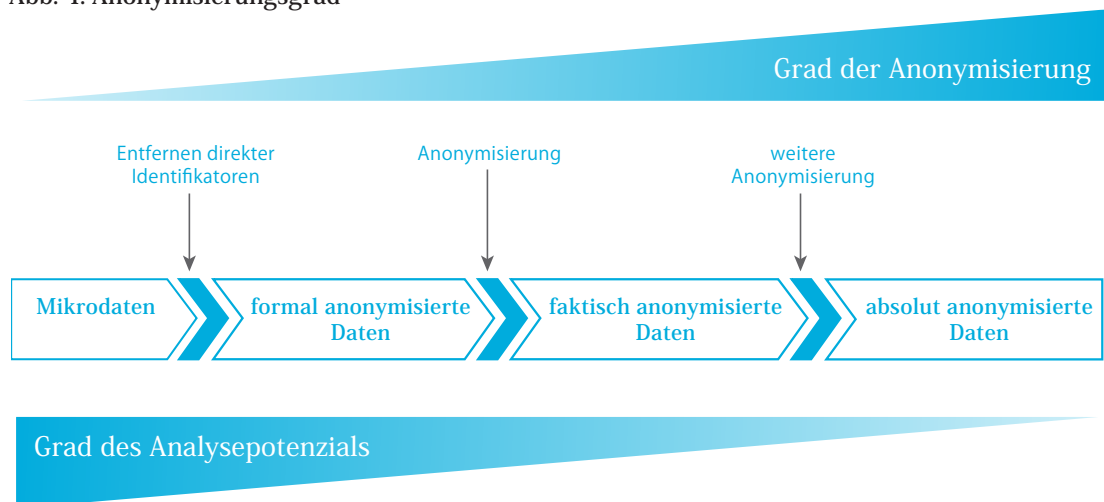
1.3.2 Wahrung des Datenschutzes und der Forschungsinteressen

Um einerseits den Datenschutz gemäß den geltenden gesetzlichen Regelungen sicherzustellen und andererseits eine möglichst umfangreiche Nutzung der Daten durch die Forschung zu ermöglichen, wurden in den Forschungsdatenzentren verschiedene aufeinander abgestimmte Konzepte zur Anonymisierung und zu den Datenzugangswegen etabliert. Dabei gilt als Grundregel: Je restriktiver der Datenzugangsweg gestaltet ist, desto schwächer können die an den Daten selbst vorgenommenen Anonymisierungsmaßnahmen sein.

In vielen Forschungsdatenzentren liegen sensible Einzeldaten zu natürlichen Personen, Haushalten oder Unternehmen bzw. zu juristischen Personen vor. Alle Einzelangaben, insbesondere personenbeziehbare Daten, bedürfen eines besonderen Schutzes. Dementsprechend werden von den Forschungsdatenzentren bei der Bereitstellung von Daten für Forschungsvorhaben Schutzmaßnahmen gemäß dem Bundesdatenschutzgesetz, den Landesdatenschutzgesetzen oder anderen rechtlichen Grundlagen ergriffen. Hierzu zählen insbesondere Anonymisierungsmaßnahmen.

Bezüglich der Anonymisierungsmaßnahmen ist zwischen formaler Anonymisierung (Entfernung von Namen, Adressen und direkten Identifikatoren), faktischer Anonymisierung (Einzelangaben können nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden) und absoluter Anonymisierung (keine Reidentifikation möglich) zu unterscheiden.

Abb. 4: Anonymisierungsgrad



Quelle: FDZ der Statistischen Ämter des Bundes und der Länder.

In der Regel stellen organisatorische Maßnahmen in den Forschungsdatenzentren sicher, dass Namen und Adressen stets getrennt von anderen Daten gespeichert bzw. sogar gelöscht werden. Bei Forschungsdatenzentren, deren Heimateinrichtungen eigene Personen- und Haushaltsstudien durchführt, ist es die übliche Vorgehensweise, dass Namen und Adressen der Befragten bei dem Umfrageinstitut verbleiben, das die Erhebung durchgeführt hat und nicht an das Forschungsdatenzentrum übermittelt werden.

Die Anonymisierungsmaßnahmen, die entsprechend rechtlicher Vorgaben zur faktischen oder absoluten Anonymisierung der Daten durchgeführt werden, reduzieren die Möglichkeit der Reidentifikation einzelner Merkmalsträger wie Personen, Haushalte oder Unternehmen. Häufig eingesetzte Verfahren sind dabei das Löschen oder Vergrößern von Variablen mit hohem Reidentifikationspotenzial wie beispielsweise Regionalvariablen, das Alter von Personen, Branchen von Unternehmen, das Ziehen von Substichproben sowie die Aggregation von Merkmalen. Personenbeziehbare Daten können je nach Gesetzesgrundlage auch unanonymisiert für die Forschung zur Verfügung gestellt werden, wenn ein explizites Einverständnis der betroffenen – in der Regel der befragten – Personen vorliegt.

1.3.3 Datenzugangswege

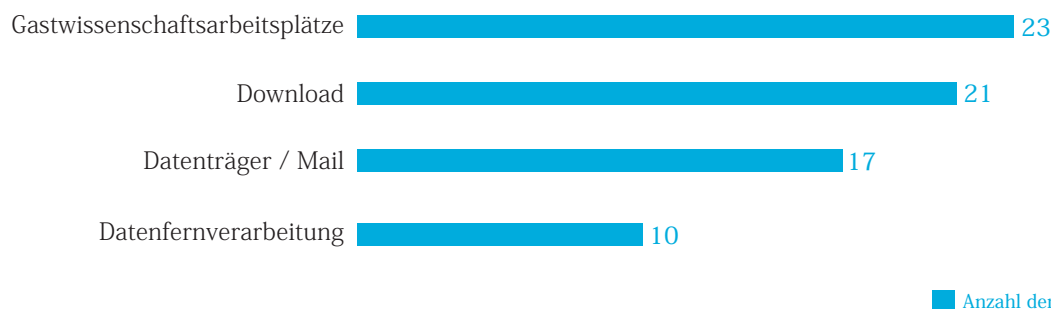
Der Zugang zu Daten ist Forschenden nur über Datennutzungsverträge möglich, in denen sowohl die Möglichkeiten der Datennutzung als auch die einzuhaltenden Bestimmungen beim Umgang mit den Daten festgehalten werden. Dabei wird seitens der Forschungsdatenzentren kontrolliert, ob die Nutzung der Daten im Rahmen der wissenschaftlichen Forschung erfolgt, da in der Regel nur in diesem Fall ein Zugang möglich ist.

Bei den Datenzugangsweisen ist grundsätzlich die Datennutzung On-Site (innerhalb der Räumlichkeiten des Forschungsdatenzentrums) und Off-Site (außerhalb des Forschungsdatenzentrums) zu unterscheiden.

Für die On-Site-Datennutzung stehen in insgesamt 23 Forschungsdatenzentren speziell gesicherte Arbeitsplätze zur Verfügung, an denen Forschende auf vertrauliche Daten zugreifen können. Zehn Forschungsdatenzentren bieten die Datennutzung über eine Datenfernverarbeitung⁸ an. Dies ist ein Datenzugangsweg, bei dem die Forschenden die schwach anonymisierten Daten im Forschungsdatenzentrum auswerten können, ohne selbst vor Ort zu sein. Der Zugriff erfolgt über eine Remoteverbindung oder durch die Übergabe von Analyse-Syntaxen. Im zweiten Fall schreiben die Forschenden an ihren eigenen Arbeitsplätzen den Programmcode für die Datenauswertung (eventuell mit Hilfe eines Strukturdatensatzes, an dem sie die Programmsyntax testen können) und übermitteln ihn an das Forschungsdatenzentrum, in welchem die Originaldaten liegen. Je nach Datensensibilität und/oder Rechtsgrundlage werden die Ergebnisse nach Ausführung der Programmcodes im Forschungsdatenzentrum geprüft (Outputkontrolle), bevor sie an den Forschenden zurückgeschickt werden. Man spricht dann von kontrollierter Datenfernverarbeitung. Off-Site-Datennutzungen bestehen im Download von Datensätzen oder der Zusendung via Mail bzw. eines Datenträgers. Insgesamt 24 Forschungsdatenzentren ermöglichen dies über Scientific Use Files. Es handelt sich hierbei um Forschungsdatensätze, die gegenüber dem Datenangebot in den gesicherten Räumen des Forschungsdatenzentrums stärker anonymisiert sind (meist faktische Anonymisierung), aber dennoch ein hohes Analysepotenzial bieten. In vier Forschungsdatenzentren werden zusätzlich Campus Files angeboten, die für die universitäre Lehre bestimmt sind und in der Regel im Vergleich zu einem Scientific Use Files noch stärker anonymisiert sind. Zwölf Forschungsdatenzentren bieten Public Use Files an. Diese sind so stark anonymisiert, dass keine Nutzungseinschränkungen bestehen und eine Weitergabe auch außerhalb der wissenschaftlichen Forschung möglich ist.

Abb. 5: Angebotene Datenzugangswege im Jahr 2015

Angaben von 30 Forschungsdatenzentren (n=30).



⁸ Je nach Definition der Forschungsdatenzentren wird die Datenfernverarbeitung unter On-Site Zugang oder Off-Site Zugang subsumiert.

1.3.4 Maßnahmen zur Qualitätssicherung der Daten

Bei den Forschungsdatenzentren, deren Heimatinstitution die Daten selbst erhebt bzw. erheben lässt, beginnt die Qualitätssicherung bereits beim Testen der Erhebungsinstrumente und den einzelnen Schritten der Datenerhebung. Die erhobenen Daten durchlaufen in der Regel verschiedene Aufbereitungsschritte, in denen die Vollständigkeit, die Konsistenz und die Plausibilität geprüft und gefundene Datenfehler und Datenlücken nach Möglichkeit behoben werden, bevor diese Daten an das Forschungsdatenzentrum übergeben werden. Zu den durchgeführten Maßnahmen gehört zudem die Dokumentierung, die Verschlagwortung und die Vergabe von Versionsnummern. Auch das Feedback der Nutzenden wird von vielen Forschungsdatenzentren herangezogen, um die Datenqualität zu verbessern. Verschiedene Forschungsdatenzentren dokumentieren ihre Qualitätssicherungsmaßnahmen zudem regelmäßig in Qualitätsberichten oder in Datenreporten zu einzelnen Datenprodukten.

1.4 Bereitgestelltes Datenangebot

1.4.1 Umfang des Datenangebots

Zum Stichtag (31.12.2015) haben die Forschungsdatenzentren 2.862 Datensätze angeboten. Dabei gibt es in den Forschungsdatenzentren eine große Spannweite, beginnend von einem bis zu 1.500 Datensätzen. Ein Datensatz kann dabei aus mehreren Einzelstudien bestehen, so dass die Anzahl bereitgestellter Studien und insbesondere von Dateien wesentlich höher liegt.

1.4.2 Sperrfristen für Datensätze

Sieben Forschungsdatenzentren nennen Sperrfristen, bei den übrigen 23 existieren keine, d.h. die Daten werden unmittelbar nach der Aufbereitung im Forschungsdatenzentrum zur Verfügung gestellt. Sperrfristen können sich auch nur auf Teile der Datengrundlage beziehen. In einigen Fällen kann die Sperre dauerhaft sein.

1.4.3 Gebühren der Datenbereitstellung

Von 30 Forschungsdatenzentren geben 24 an, dass sie keine Gebühren erheben. Die genannten Gebühren der sechs Forschungsdatenzentren sind häufig niedrig (im zweistelligen Euro-Bereich) und orientieren sich an den Kosten für die Medien und die Vertragslegung.

1.4.4 Zugangsberechtigter Personenkreis

Die Forschungsdatenzentren bedienen einen sehr breiten Kreis von potenziellen Nutzerinnen und Nutzern aus dem Bereich Wissenschaft und Forschung. Hauptamtlich an anerkannten öffentlichen Forschungseinrichtungen oder Universitäten arbeitende Personen sind bei allen Forschungsdatenzentren antragsberechtigt.

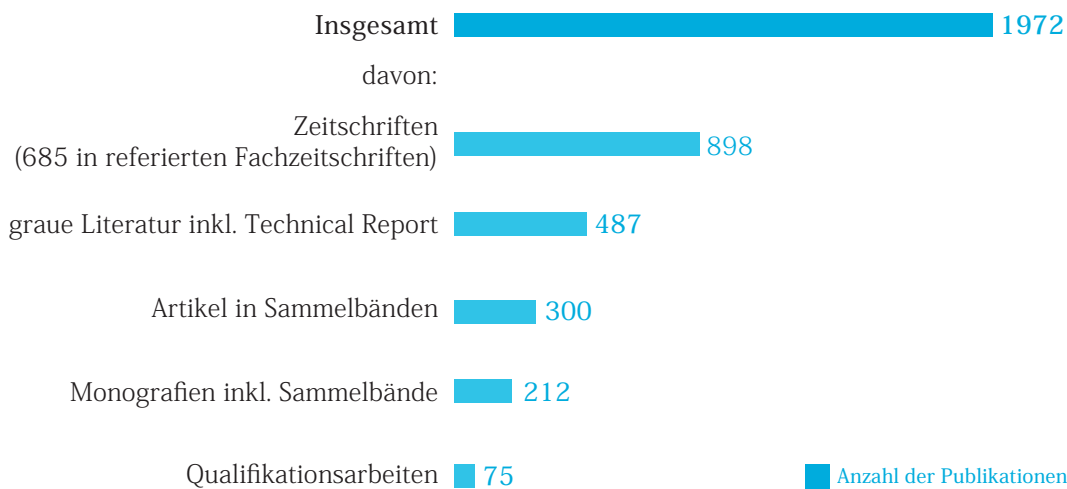
1.4.5 Forschungoutput

Insgesamt wurden 1.972 Publikationen für das Berichtsjahr 2015 dokumentiert, welche die von den Forschungsdatenzentren bereitgestellten Daten verwenden. Sieben Forschungsdatenzentren erfassen solche Publikationen nicht.

Da Forschende die Abgabe von Belegexemplaren sehr häufig versäumen und nicht jedes Forschungsdatenzentrum die Ressourcen für eine umfangreiche Recherche besitzt, ist von einer **erheblichen Untererfassung** auszugehen.

Abb. 6: Publikationen im Jahr 2015 auf Basis der bereitgestellten Forschungsdaten

Angaben von 30 Forschungsdatenzentren (n=30).



1.4.6 Nutzungsumfang

Zu den zentralen Aufgaben der Forschungsdatenzentren gehört es, für die Forschung einen umfangreichen und flexiblen Datenzugang zu schaffen und diesen ständig auszubauen und zu verbessern. Für Forschende werden – unter anderem aufgrund unterschiedlicher Datenschutzerfordernungen und rechtlicher Regelungen – verschiedene Zugangsmöglichkeiten angeboten. Daneben variiert zwischen den Einrichtungen die Erfassung von Verträgen, Projekten und Datennutzenden. Doppelzählungen bei Verträgen und Projekten können daher nicht ausgeschlossen werden.

Downloads ohne vorherige Registrierung stehen nur für einige Public Use Files und einige qualitative Daten zur Verfügung. Die Nutzung dieser Angebote wird nicht von allen Forschungsdatenzentren erfasst. In der Regel sind Downloads von Daten erst nach Registrierung bzw. nach Vertragsabschluss möglich. Insgesamt wurden 67.146 Downloads von Daten im Jahr 2015 erfasst.

Aus Datenschutzgründen enthalten Datenbereitstellungs- und Nutzungsverträge einen expliziten Bezug zu einem Forschungsprojekt (Zweckbindung), d. h., für jedes Forschungsprojekt wird ein Vertrag ausgestellt. Die Definitionen von „Projekt“ umfassen wissenschaftliche Projekte im engeren Sinne – d. h. dass zu einem Forschungsthema einer Einrichtung ein oder mehrere Datennutzende Zugang erhalten, wobei häufig eine Mehrfachnutzung der Daten in den Forschungseinrichtungen möglich ist.

Insgesamt überwiegen Verträge mit einzelnen Forschenden. Diese Verträge können einen oder mehrere Datensätze umfassen. Da nicht für alle Forschungsdatenzentren zu den genannten Downloads auch Angaben zu Projekten vorliegen, dürfte die Anzahl der mit den Daten der Forschungsdatenzentren bearbeiteten Projekte ebenfalls untererfasst sein. Nicht zuletzt dürfte die tatsächliche Datennutzung durch einzelne Forschende deutlich über der Zahl von Verträgen und Projekten liegen.

Abb. 7: Nutzungsumfang des Datenangebots

Angaben von 30 Forschungsdatenzentren (n=30).

	Anzahl
Downloads	67.146
Verträge	5.613
Wissenschaftliche Projekte	1.254

1.4.7 Neuerungen beim Datenzugang

Die aktuellen Entwicklungen in den Forschungsdatenzentren zeigen, dass sowohl das Datenangebot ständig erweitert als auch der Zugang zu den Daten verbessert wird. Neuerungen im Datenzugang zeichnen sich häufig durch die Verknüpfung von Daten eines Forschungsdatenzentrums mit den Daten anderer Forschungsdatenzentren oder Daten aus anderen Quellen aus. Damit wird sichtbar, dass die Vernetzung der Forschungsdatenzentren die Analysepotentiale bestehender Datenquellen erhöhen können. Das wird mittel- oder langfristig auch zu einer stärkeren Nutzung der Daten führen. Beispiele für Verknüpfungen mit Registerdaten sind die IAB-SOEP Migrationsstichprobe oder der SHARE-RV Datensatz und für georeferenzierte Datensätze die Erhebungswellen des Sozio-oekonomischen Panels (SOEP) seit 2000 sowie der ALLBUS 2014.

Zudem wurde über zahlreiche Verbesserungen in der Datendokumentation und Recherche berichtet. Das betrifft bspw. die Überarbeitung von Suchoberflächen (Forschungsdatenzentrum Betriebs- und Organisations-Daten), die Entwicklung erweiterter Metadaten schemata (Forschungsdatenzentrum Bildung am Deutschen Institut für Internationale Pädagogische Forschung) und die Entwicklung von Werkzeugen zur Datendokumentation (Forschungsdatenzentrum PsychData des Leibniz-Zentrums für Psychologische Information und Dokumentation).

1.5 Serviceangebot

1.5.1 Bereitstellung von Dokumentationen und Arbeitshilfen

Nahezu alle akkreditierten Forschungsdatenzentren stellen Dokumentationsmaterialien zur Verfügung. Lediglich bei einem ist die Veröffentlichung der Materialien noch in Planung.

In der Regel werden die Dokumentationen als PDF-Dateien über die Webseite zum kostenlosen Download angeboten („Open Access“). Der Umfang der Materialien unterscheidet sich je nach Datentypus. So werden für Befragungsstudien in der Regel mindestens Fragebögen und Codebücher bereitgestellt und teilweise um Methodenberichte, Datensatzbeschreibungen, Qualitätsberichte, Klassifikationen, Technical Paper und Variablenlisten ergänzt. Einige Forschungsdatenzentren bieten zudem englische Versionen der Dokumentationen an. Häufig werden zudem Arbeitshilfen, wie etwa Syntaxen, oder theorieorientierte Dokumente bereitgestellt, die über den inhaltlichen Hintergrund einzelner Fragen oder Variablen informieren. In der Regel werden diese Materialien online über die Webseiten angeboten und können ohne eine vorherige Registrierung als Open Access Material bezogen werden. Werden die Daten in einem Paket ausgeliefert, so umfasst dies meistens auch die verfügbaren Arbeitshilfen.

1.5.2 Bekanntmachung des Datenangebotes

Die Forschungsdatenzentren nutzen verschiedene Möglichkeiten, um ihre Forschungsdaten in der wissenschaftlichen Community vorzustellen. An erster Stelle erfolgt dies über die **eigene Webseite**. Ein Großteil der Forschungsdatenzentren gibt zudem **eigene Newsletter** heraus, andere Forschungsdatenzentren nutzen den **Newsletter der Einrichtung**, zu der sie gehören. Newsletter spielen eine wichtige Rolle, wenn es darum geht, über neue Daten- und Serviceangebote der Forschungsdatenzentren zu informieren. Für neu entstandene Forschungsdatenzentren sind sie eine gute Gelegenheit, ihr Angebot zu kommunizieren. Bereits länger bestehende Forschungsdatenzentren informieren auch über **Mailinglisten** bzw. **Rundmails** zu neuen oder aktualisierten Daten. **Vorträge** und **Konferenzen** bilden eine weitere Möglichkeit, über das Datenangebot zu informieren und den Nutzerkreis zu erweitern. Forschungsdatenzentren präsentieren sich auf **Fachmessen** oder **Workshops** in Form von Vorträgen, mit **Flyern** und **Postern**. Auch **Lehrveranstaltungen an Universitäten** werden genutzt, um Studierende anzusprechen. Aus einigen Forschungsdatenzentren kommt der Hinweis, dass sie **Plattformen wie den RatSWD** bzw. seine Newsletter oder dessen Internetpräsenz einbeziehen. Verlinkungen zu adressatenbezogenen Plattformen oder auch zu **Fachportalen** oder **Metadatenportalen** werden ebenfalls von einigen Forschungsdatenzentren aufgeführt. Zum Datenbestand bzw. zu seiner Erweiterung werden mitunter auch **Pressemitteilungen** veröffentlicht. Einige Forschungsdatenzentren geben eigene **Berichte** heraus. **Wissenschaftliche Publikationen** aus den Einrichtungen, zu denen das Forschungsdatenzentrum gehört oder/und Publikationen, in welchen ausdrücklich auf die Datenquellen bzw. den Daten-

bereitsteller hingewiesen wird, tragen ebenfalls zur Bekanntheit der Datenangebote bei. Einige wenige Forschungsdatenzentren nutzen die Möglichkeit, das Datenangebot über **Kataloge** zu verbreiten. Eher selten wird berichtet, dass **Social Media** wie Twitter oder Facebook eingesetzt werden.

1.5.3 Servicekonzepte

Alle Forschungsdatenzentren haben für die Belange, die von externen Forscherinnen und Forschern an sie herangetragen werden, Ansprechpartnerinnen und Ansprechpartner, die telefonisch oder per E-Mail kontaktiert werden können. Häufig fungieren diese Kontaktpersonen als zentrale Anlaufstelle und leiten das jeweilige Anliegen an Spezialisten und Spezialistinnen weiter. Die große Mehrheit der Forschungsdatenzentren bietet neben Auskunft auch Beratung an (z.B. zur Auswahl der Datensätze, zum Analysepotenzial oder zum Datenmanagement, inklusive notwendiger Maßnahmen in Bezug auf Datenschutz und Datensicherheit). Aus zahlreichen Forschungsdatenzentren wird gemeldet, dass regelmäßig Workshops und/oder Nutzertagungen stattfinden. Schulungen durch eigenes Personal sind ein weiteres Angebot, das von wenigen Forschungsdatenzentren bereitgestellt wird. Eine andere Möglichkeit, Datennutzende zu unterstützen, bietet die Angabe von Antworten auf FAQs, die ebenso wie die Bereitstellung schriftlichen Informationsmaterials von einigen Forschungsdatenzentren auf deren Homepage bereitgestellt wird.

1.5.4 Evaluation und Qualitätssicherung

Die Evaluierungen sind bei den Forschungsdatenzentren unterschiedlich stark institutionalisiert, auch ihre zeitlichen Abstände variieren.

Neben der Evaluierung spielt die Nutzerbefragung, z. B. durch Feedback-Fragebögen, eine größere Rolle. Die Ergebnisse fließen genauso in die stetige Verbesserung der Serviceleistungen ein wie das informelle Feedback, das von den Nutzenden gegeben wird. Standardisierte, in regelmäßigen Abständen durchgeführte Nutzerbefragungen sind für knapp die Hälfte der Forschungsdatenzentren ein Instrument zur Qualitätssicherung. Unter den Forschungsdatenzentren, die bisher keine Nutzerbefragungen durchführen, bekunden indes einige, dass sie diese in Zukunft etablieren wollen.

2 Aktuelle Entwicklungen der Forschungsdateninfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften

■ Seit der Gründung des Rates für Sozial- und Wirtschaftsdaten im Jahr 2004 wurde durch den Aufbau der mittlerweile 30 akkreditierten Forschungsdatenzentren eine Forschungsdateninfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften geschaffen, die nicht nur im Vergleich zu anderen Fachdisziplinen in Deutschland, sondern auch im internationalen Kontext einmalig ist. Das Spektrum des Datenangebots deckt nahezu alle inhaltlichen Forschungsfelder der beteiligten Fachdisziplinen ab und erstreckt sich auf alle Datenarten, derer man sich in der Sozial-, Verhaltens- und Wirtschaftsforschung bedient. Damit stehen den Forschenden nicht nur qualitativ hochwertige Forschungsdaten zur Verfügung, sondern die Forschungsdatenzentren gewährleisten auch die Einhaltung der für diese Daten geltenden Datenschutzbestimmungen. Kern dieser Entwicklung ist eine Professionalisierung des Forschungsdatenmanagements, die sich auf alle Phasen des „data life cycles“ bezieht und die zur Etablierung von Standards bei der Aufbereitung und Bereitstellung von Forschungsdaten geführt hat. In zwei Bereichen können dabei in den letzten Jahren besondere Entwicklungsfortschritte beobachtet werden. (1) Der Fortschritt in der Professionalisierung des Forschungsdatenmanagements wird in einer mittlerweile durch nahezu alle Forschungsdatenzentren vorgenommenen Vergabe Persistenter Identifikatoren (PID) deutlich. Damit etablieren sich gemeinsame Standards zur Dokumentation, Beschreibung und Referenzierung von Forschungsdaten. (2) Gleichsam am anderen Ende des data-life-cycles setzen neue Lösungen des Datenzugangs und der Datennutzung an. Neue institutionelle und technische Lösungen des Remote-Access eröffnen Datennutzungsmöglichkeiten unabhängig vom physischen Standort des jeweiligen Forschungsdatenzentrums. Für die Forschenden bedeutet dies eine entscheidende Verringerung des Aufwands bei der Nutzung von – auch datenschutzrechtlich sensiblen – Forschungsdaten. Beide Entwicklungsfelder werden im Folgenden kurz erläutert.

2.1 Persistente Identifikatoren (PID): ein Meilenstein zur Professionalisierung des Forschungsdatenmanagements

Persistente Identifikatoren werden verwendet, um digitale Objekte mit einem „elektronischen Standort“ zu versehen. Damit können sie immer wieder gefunden, eindeutig identifiziert und referenziert werden. Im Bereich des wissenschaftlichen Publizierens sind PID bereits alltäglich: Alle Artikel, die in den Fachzeitschriften der internationalen wissenschaftlichen Verlage erscheinen, können anhand eines PID eindeutig identifiziert und zitiert werden. Zumeist werden dafür sogenannte Digital Object Identifier (DOI) vergeben.

Auch für Forschungsdaten ist die Vergabe von Persistenten Identifikatoren aus mehreren Gründen sinnvoll:

- 1) PID sind dauerhaft angelegt. Während Links im Internet nach einer gewissen Zeit oft ins Leere laufen, bleiben PID mit dem „elektronischen Standort“ des Inhaltes verbunden.
- 2) PID erleichtern den Nutzenden die Kontrolle über die von ihnen verwendeten Daten. In vielen Fällen werden Datensätze aktualisiert, wenn Fehler in den Daten oder der Dokumentation gefunden wurden. Nicht in allen Fällen können die Nutzenden eines Datensatzes auf Änderungen in den Daten hingewiesen werden. Jede neue Version eines Datensatzes erhält einen neuen DOI. Wird der DOI im Datensatz als Variable gespeichert, kann der Nutzende eindeutig die Version seines Datensatzes bestimmen und sicherstellen, dass seine Auswertungen mit der aktuellsten Version des Datensatzes durchgeführt werden.

- 3) PID erleichtern das Zitieren von Daten. Gerade in langfristig angelegten Erhebungsprogrammen ist der Name des Programms oft nicht eindeutig zuzuordnen (z. B. ist der Verweis auf den Mikrozensus unter einer Tabelle nicht ausreichend). Über den PID können die Datensätze und ihre Versionen den Auswertungen einer Publikation eindeutig zugeordnet werden. Damit wird auch die Leistung der Datenerhebungen im wissenschaftlichen Prozess gewürdigt. Die Datenproduzenten können leichter nachvollziehen, für welche Publikationen ihre Daten verwendet wurden und damit den Nutzen der Daten dokumentieren.
- 4) PID erleichtern aus den genannten Gründen auch die Replikation von Ergebnissen aus vorliegenden Studien, weil der verwendete Datensatz eindeutig identifiziert wird.
- 5) PID erleichtern die Weitergabe von Metadaten und damit die Suche nach Forschungsdaten in gemeinsamen Suchportalen.
- 6) Die Verwendung von PID ermöglicht sowohl die Verlinkung von Daten und Publikationen als auch die Verbindung zu Autoren und Forschungsfördernden. Dadurch kann die Transparenz im Forschungsprozess weiter erhöht werden.

Auch der Rat für Informationsinfrastrukturen hebt in seinem Bericht⁹ die Nützlichkeit von PID für das Forschungsdatenmanagement hervor und unterstreicht, dass das Potential der PID bislang noch nicht ausgeschöpft wurde.

Der Rat für Sozial- und Wirtschaftsdaten unterstützt bereits seit einiger Zeit die Forschungsdatenzentren in ihren Bemühungen, PID für ihr Datenangebot anzubieten. Von den 30 Forschungsdatenzentren vergeben inzwischen 24 PID zu ihren gesamten Forschungsdaten oder zu Teilen dieser. Fünf weitere Forschungsdatenzentren planen die Vergabe von PID. Insbesondere für die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder sind die Planungen weit fortgeschritten, so dass wahrscheinlich im Laufe des Jahres 2017 weitere wichtige Datenbestände mit PID versehen werden. Am häufigsten wird die Datenregistrierungsagentur da|ra für die Registrierung von DOI-Namen verwendet (21 Forschungsdatenzentren). Aber auch andere Dienstleister wie z. B. TIB Hannover (FDZ SFB 882) oder handle.net (Forschungsdatenzentrum Archiv für Gesprochenes Deutsch am Institut für Deutsche Sprache) werden verwendet.

Die Vergabe von PID hat sich somit als Standard in den Forschungsdatenzentren etabliert. Die Vorteile der PID stehen damit sowohl den Infrastrukturen als auch den Nutzenden für die Suche und Zitation von Daten zur Verfügung.

2.2 Niederschwellige Datennutzung durch innovative Lösungen des Remote-Access

Der Großteil der in den Sozial-, Verhaltens- und Wirtschaftswissenschaften genutzten Forschungsdaten enthält – soweit es sich um Mikrodaten handelt – Informationen über natürliche oder juristische Personen. Die zentrale Herausforderung bei der Bereitstellung dieser Daten besteht darin, sie vor einer Re-Identifikation zu schützen, um damit den Datenschutzbestimmungen (bei natürlichen Personen) oder den Schutzforderungen aus anderen Rechtskreisen (juristischen Personen) zu entsprechen. Die Anforderungen dafür sind in den letzten Jahren weiter gestiegen, da zunehmend komplexe Forschungsdesigns realisiert werden, bei denen unter anderem auch durch die Verknüpfung unterschiedlicher Datenquellen eine Vielzahl an Informationen vorliegen. Bei der Bereitstellung dieser Forschungsdaten hat sich in den Forschungsdatenzentren ein dreistufiges Zugangsmodell etabliert: Bei sogenannten Public Use Files wird über datenverändernde Maßnahmen oder durch die Freigabe eines reduzierteren Umfangs an Informationen das Re-Identifikationsrisiko ausgeschaltet. Durch die Einschränkung des Informationsgehalts dieser Daten werden jedoch die Analysemöglichkeiten deutlich

⁹ RfII – Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen 2016, 160 S., hier Seite 31. <http://www.rfii.de/download/rfii-empfehlungen-2016/>

eingeschränkt. Geringere Eingriffe werden bei Scientific Use Files vorgenommen, sie fordern jedoch auf Seiten der Nutzenden umfassendere Schutzmaßnahmen. Beim dritten Zugangsweg kann schließlich eine Nutzung ausschließlich in entsprechend gesicherten Arbeitsräumen in den einzelnen Forschungsdatenzentren vor Ort erfolgen. Diese Daten sind in der Regel lediglich pseudonymisiert, d. h. für die Nutzenden sind die Klarnamen der betreffenden natürlichen oder juristischen Personen unkenntlich gemacht. Auf alle anderen Informationen können sie unmittelbar zugreifen. Den Forschenden ist es nicht möglich, selbst Ergebnisse nach außen zu übermitteln. Daher wird der von ihnen (z. B. für ihre Publikationen) benötigte Output zunächst von Mitarbeitenden des Forschungsdatenzentrums daraufhin überprüft, ob etwaige Rückschlüsse auf Einzelangaben möglich sind (sogenannte absolute Anonymität). Es handelt sich dabei um die sogenannte „Outputkontrolle“. Ist der Output unkritisch, wird er den Forschenden übermittelt, dies geschieht in der Regel elektronisch. Mittlerweile verfügen 23 der 30 Forschungsdatenzentren über derartige geschützte Arbeitsräume.

Gleichwohl stellt die Notwendigkeit der Nutzung von Forschungsdaten vor Ort ein Hindernis für – insbesondere auch im Ausland arbeitende – Forschende dar. Aus Sicht der Nutzenden wird der Zugriff auf Forschungsdaten vom eigenen Arbeitsplatz deutlich präferiert. Deshalb wurden bereits von Anfang an Angebote zur Datenfernverarbeitung entwickelt. Die Kontrolle der Auswertungsergebnisse im Rahmen der Datenfernverarbeitung ist jedoch aufwändig und damit auch kostenintensiv, weil diese über entsprechend geschultes Personal erfolgen muss.

Zwei Innovationen wurden in den letzten Jahren – u. a. durch entsprechende Kooperationsprojekte zwischen unterschiedlichen Forschungsdatenzentren – entwickelt, um einerseits den Aufwand der Kontrolle zu minimieren und andererseits einen ortsunabhängigen und gleichzeitig datenschutzkonformen Zugriff auf sensible Forschungsdaten zu ermöglichen: Analog zu dem beim Forschungsdatenzentrum des Sozio-oekonomischen Panels genutzten System LISSY hat das Internationale Forschungsdatenzentrum des Forschungsinstituts zur Zukunft der Arbeit (IZA) mit JoSuA (Job Submission Application) eine technische Lösung für die Verwaltung der Datenfernverarbeitung bei datenschutzrechtlich hoch sensiblen Forschungsdaten entwickelt, die sowohl für Nutzende als auch für das Forschungsdatenzentrum effizienter und komfortabler ist, als die zuvor in einigen FDZ praktizierte händische Lösung. Zugleich erfolgt über diese Software-Lösung eine Unterstützung des datenschutzrechtlichen Prüfprozesses der Analyseergebnisse der jeweiligen Nutzenden. Die Entwicklungen von JoSuA werden durch den Einsatz der Software im Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB) unterstützt und vorangetrieben. Das Forschungsdatenzentrum der Rentenversicherung hat eine Eigenentwicklung für das Fernrechnen im Einsatz. Diese wurde entwickelt, weil sie in die IT-Sicherheitsarchitektur der DRV Bund passt und daher sehr schnell einsatzfähig war. Die Nutzung ist teilautomatisiert.

Die Möglichkeit der Datenfernverarbeitung stellt zwar einen deutlich erleichterteren Zugang zu Forschungsdaten dar, ist jedoch aufgrund der zeitlichen Lücke zwischen Datenanalyse und Ausgabe bzw. Sichtung der Ergebnisse aus Nutzersicht suboptimal. Deshalb wurden bereits seit längerem Möglichkeiten des „Remote-Access“ entwickelt. Insbesondere durch die Kooperation einzelner Forschungsdatenzentren untereinander wurde ein Angebot geschaffen, bei dem Forschende an den gesicherten Arbeitsplätzen eines Forschungsdatenzentrums auch auf Daten anderer Forschungsdatenzentren zugreifen können („FDZ im FDZ“). Aktuell bestehen derartige Kooperationen beim Forschungsdatenzentrum des Sozio-oekonomischen Panels, dem Forschungsdatenzentrum der Statistischen Ämter der Länder, dem Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB) und dem Forschungsdatenzentrum Betriebs- und Organisationsdaten. Die technische Weiterentwicklung des Remote-Access macht es darüber hinaus möglich, dass Nutzende in entsprechenden Außenstellen direkten Zugriff auf Forschungsdaten erhalten. So verfügen die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder über Außenstellen an der Johann-Wolfgang-Goethe Universität, an der Technischen Universität Dresden, am DIW (Berlin) und am ifo (München). Das Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB) hat Außenstellen an 14 Standorten, neben Deutschland auch in den USA und Großbritannien.

Eine entscheidende Weiterentwicklung wurde vom Forschungsdatenzentrum des Leibniz-Instituts für Bildungsverläufe e.V. an der Otto-Friedrich-Universität Bamberg vorgenommen. Mit „RemoteNEPS“ wird eine sehr leistungsfähige und flexible Infrastruktur der kontrollierten Ferndatenverarbeitung zur Verfügung gestellt. Das innovative Element besteht hier darin, dass der Zugang zu den Forschungsdaten über ein persönliches Login und Passwort in Kombination mit einer zertifizierten biometrischen Authentifizierung via Tippverhalten („keystroke biometrics“) erfolgt. Damit kann – auch außerhalb gesicherter Arbeitsplätze in einem Forschungsdatenzentrum – sichergestellt werden, dass allein die Personen Zugriff auf die Daten haben, die entsprechend der abgeschlossenen Nutzungsverträge dazu berechtigt sind.

Mitwirkende bei der Erstellung

Mitglieder der Pilot-Monitoringkommission des RatSWD

Tatjana Mika (Vorsitz)

Forschungsdatenzentrum der Rentenversicherung (FDZ-RV)

Holger Quellenberg (Vorsitz)

Forschungsdatenzentrum des Deutschen Jugendinstituts (FDZ-DJI)

Dr. Nikos Askitas

Internationales Forschungsdatenzentrum des Forschungsinstituts zur Zukunft der Arbeit (FDZ IZA, IDSC)

Martin Eisele

Forschungsdaten- und Servicezentrum der Bundesbank (FDSZ Bundesbank)

Dr. Jan Goebel

Forschungsdatenzentrum des Sozio-oekonomischen Panels (FDZ SOEP)

Heike Habla

Forschungsdatenzentrum des Statistischen Bundesamtes (FDZ-Bund)

Dr. Cornelia Lang

Forschungsdatenzentrum des Leibniz-Instituts für Wirtschaftsforschung Halle (FDZ-IWH)

Prof. Dr. Stefan Liebig

Forschungsdatenzentrum Betriebs- und Organisations-Daten (FSZ-BO)

Dr. Bernhard Schimpl-Neimanns

Forschungsdatenzentrum German Microdata Lab (FDZ GML)

Dr. Pascal Siegers

Forschungsdatenzentrum ALLBUS bei GESIS (FDZ ALLBUS)

Ständige Gäste der Pilot-Monitoringkommission des RatSWD

Prof. Regina T. Riphahn, Ph.D.

Vorsitzende des RatSWD

Stefan Bender

Vorsitzender des RatSWD

Impressum

Herausgeber:

Rat für Sozial- und Wirtschaftsdaten (RatSWD)
Chausseestraße 111
10115 Berlin
office@ratswd.de
www.ratswd.de

Redaktion:

Sandra Hagedorn, Dr. Jörg Holthöfer

Gestaltung/Satz:

Markus Weiß | www.typogo.de

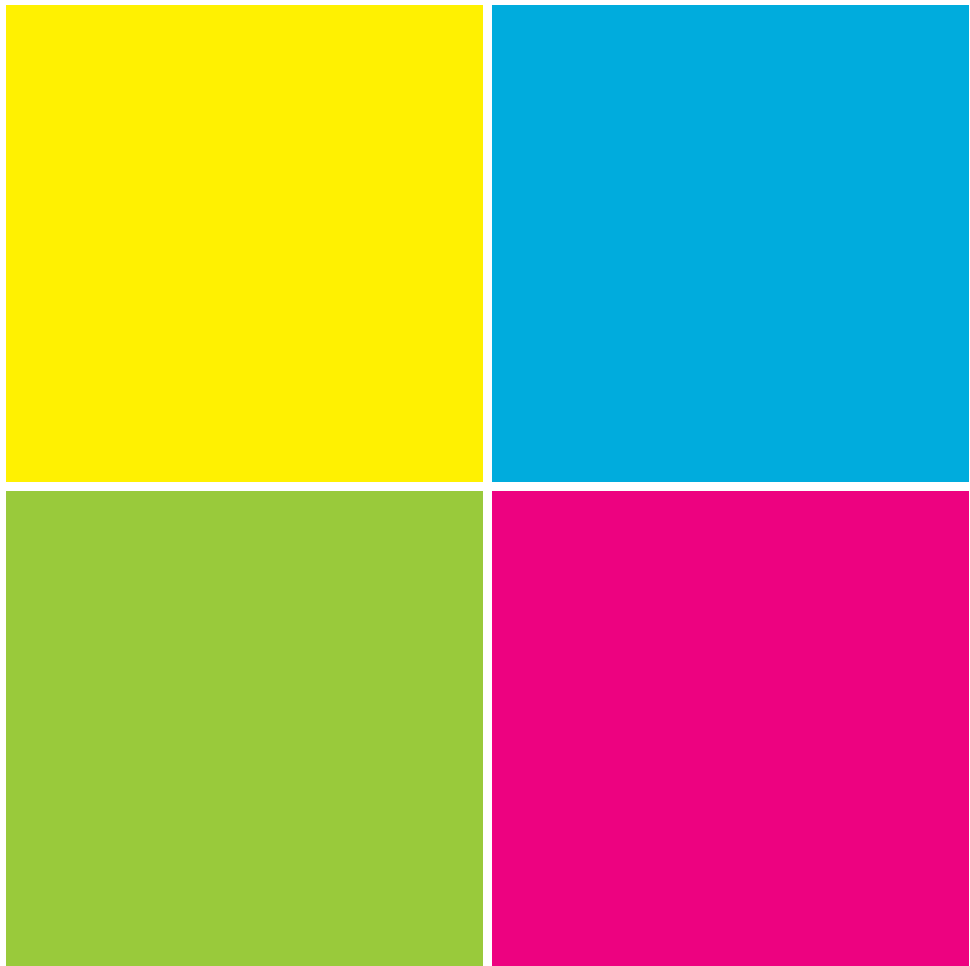
Berlin, Februar 2017

RatSWD Output:

Die RatSWD Output Series dokumentiert die Arbeit des RatSWD in seiner 5. Berufungsperiode (2014–2017). In ihr werden seine Stellungnahmen und Empfehlungen veröffentlicht und auf diesem Weg einer breiten Leserschaft zugänglich gemacht.

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01UW1402 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt, sofern nicht anders ausgewiesen, beim RatSWD.

doi: 10.17620/02671.3



www.ratswd.de