

# AUFFINDEN

FORSCHUNGSDATEN IN DEN

# ZITIEREN

SOZIAL- UND WIRTSCHAFTSWISSENSCHAFTEN

# DOKUMENTIEREN



**ZBW**

Leibniz-Informationszentrum  
Wirtschaft  
Leibniz Information Centre  
for Economics

**gesis**  
Leibniz-Institut  
für Sozialwissenschaften

RatSWD.

Rat für Sozial- und  
WirtschaftsDaten



# VORWORT

zur 2. Auflage



**Prof. York Sure-Vetter**



**Prof. Regina Riphahn**



**Prof. Klaus Tochtermann**

Der vorliegende Wegweiser »**Auffinden – Zitieren – Dokumentieren: Forschungsdaten in den Sozial- und Wirtschaftswissenschaften**« greift einige grundsätzliche Aspekte im Umgang mit quantitativen Forschungsdaten auf und orientiert sich dabei an den Regeln guter wissenschaftlicher Praxis (vgl. DFG, 2013; DFG, 1998). Der Wegweiser soll nicht nur Nachwuchswissenschaftler beim Umgang mit Forschungsdaten unterstützen, sondern ist nach unserer Überzeugung auch für etablierte Wissenschaftler hilfreich. Sicher kann dieser Wegweiser nicht alle Fragen beantworten, die in diesem Zusammenhang bestehen, aber er liefert nützliche Hinweise zu elementaren Fragestellungen.

Wie auch in anderen Wissenschaftsdisziplinen bilden Forschungsdaten in den Sozial- und Wirtschaftswissenschaften eine immer wichtiger werdende Grundlage für wissenschaftliche Erkenntnisse.

Eine wesentliche Ursache für die steigende Bedeutung liegt in der besseren und einfacheren Verfügbarkeit von Forschungsdaten. So haben sich durch das Internet die Möglichkeiten zur Nachnutzung bereits existierender Datensätze deutlich verbessert. Statistische Ämter, aber auch Forschungsinstitutionen und Infrastruktureinrichtungen in den Sozial- und Wirtschaftswissenschaften stellen heute eine Vielzahl von Daten-

sätzen zu unterschiedlichsten Themen für die wissenschaftliche Forschung bereit. Folglich ist es heute relativ einfach, Studien und Analysen auf Basis bereits bestehender Datensätze durchzuführen. Mit Hilfe geeigneter Statistik-Programme (wie SPSS, Stata oder R) lassen sich solche Daten analysieren. Diese Art der Analyse findet sich vor allem in Dissertationen, aber auch bereits in etlichen Bachelor- und Masterarbeiten.

Auch machen die stetig wachsenden technischen Möglichkeiten immer komplexere Analysen möglich. All dies trägt dazu bei, dass in den Sozial- und Wirtschaftswissenschaften verstärkt empirisch geforscht wird.

Auch die Forschungsförderer erkennen die verstärkte Bedeutung von Forschungsdaten: Die Deutsche Forschungsgemeinschaft (DFG) beispielsweise betont sowohl die gesteigerten Erkenntnismöglichkeiten der Wissenschaft durch Nachnutzung von Daten als auch monetäre Aspekte wie die Vermeidung von Kosten durch unnötige Doppeluntersuchungen (vgl. Winkler-Nees, 2012). Sie verlangt daher bereits seit einigen Jahren bei der Beantragung von Projekten, bei denen Daten erhoben werden, eine Erklärung zur (Nach-)Nutzung der gewonnenen Daten. Die National Science Foundation (NSF) schreibt in den USA seit 2011 gar das Vorlegen von Data-Management-Plänen für derartige Anträge zwingend vor.

**1.**

**Wo kann ich die für meine  
Fragestellung geeigneten Daten  
finden oder muss ich mangels  
existierender Daten selbst  
Daten erheben?**

**2.**

Wenn ich Daten von anderen  
nachnutze, wie muss ich diese  
Daten zitieren?

**3.**

Wie dokumentiere ich  
eigene Forschungsdaten,  
damit sie sinnvoll nach-  
genutzt werden können?

Auf europäischer Ebene wird im 8. Forschungsrahmenprogramm der EU, besser bekannt als »Horizon 2020«, nicht nur der freie Zugang zu wissenschaftlichen Publikationen festgeschrieben, sondern es werden auch die von einigen Beispielprojekten gewonnenen Forschungsdaten für andere Wissenschaftler im Internet zur Verfügung gestellt.

Auch verschiedene renommierte Fachzeitschriften verlangen – bei Einreichungen auf empirischer Basis – zunehmend häufiger eine Übermittlung der verwendeten Datensätze sowie der erstellten Syntax einer statistischen Auswertung.

In Hinblick auf die Entwicklung von Dokumentationsstandards sei angemerkt, dass die angemessene Dokumentation der »Produktion« empirischer Daten in den experimentellen Naturwissenschaften seit über 100 Jahren zur gängigen Praxis der Forschenden gehört: Labor-Bücher, die Experimente protokollieren, sind nichts anderes als Dokumentationen der Datenproduktion.

**Mit der Bedeutung, die der Datenanalyse und ihrer Nachprüfbarkeit inzwischen zukommt, geraten auch neue Fragestellungen ins Blickfeld (siehe Abbildung links).**

Konkret soll dieser Wegweiser zum Ersten dazu dienen, passende Forschungsdaten für eigene empirische Untersuchungen zu finden. Dazu liefert er einen Überblick über relevante Quellen solcher Daten. Zum Zweiten werden Hinweise gegeben, wie verwendete Forschungsdaten wissenschaftlich korrekt zitiert werden können und was bei der Zitation zu beachten ist. Zum Dritten werden abschließend Ratschläge für eine sinnvolle Dokumentation von selbst erhobenen Forschungsdaten gegeben.

Wir hoffen, dass dieser Wegweiser insbesondere Studierenden und Nachwuchswissenschaftlern eine Hilfestellung im Umgang mit Forschungsdaten in den Sozial- und Wirtschaftswissenschaften bietet und beim wissenschaftlich korrekten Umgang mit dieser wichtigen wissenschaftlichen Ressource unterstützt.

Abschließend möchten wir der Deutschen Forschungsgemeinschaft (DFG) für die Finanzierung des Projektes European Data Watch Extended (EDaWaX – [www.edawax.de](http://www.edawax.de)), welches zwischen 2011 und 2016 von der DFG gefördert wird, danken. Diese Handreichung entstand im Kontext des EDaWaX-Projektes. Es ist ein aus der Forschung stammender (erster) Wegweiser und soll kein Lehrbuch – das es noch zu schreiben gilt – ersetzen.

Die Idee für diesen Wegweiser wurde bei der Panel-Diskussion »Warum sollte ich meine Datensätze mit anderen teilen? Die empirische Wirtschaftsforschung und der Weg zu mehr Transparenz« auf der Jahrestagung des Vereins für Socialpolitik (VfS) in Düsseldorf 2013 geboren. Insofern gehört auch die Vereinigung der Volkswirte zu seinen Förderern.

Der vorliegende Text und die angegebenen Quellen wurden erstmals in dieser Form zusammengestellt. Die Autoren

freuen sich daher über Feedback, da es mit Sicherheit vielfältige Ergänzungs- und Überarbeitungsmöglichkeiten gibt. Daher unsere Bitte: Helfen Sie, diese Handreichung weiter zu entwickeln, indem Sie den Autoren Ihre Anregungen mailen oder Ihre Kommentare auf dem Blog [www.auffinden-zitieren-dokumentieren.de](http://www.auffinden-zitieren-dokumentieren.de) hinterlassen.

Kiel/Hamburg, Nürnberg/Berlin und Köln/Mannheim im Dezember 2014

**Klaus Tochtermann**

ZBW – Leibniz-Informationszentrum Wirtschaft  
[director@zbw.eu](mailto:director@zbw.eu)

**Regina Riphahn**

RatSWD – Rat für Sozial- und Wirtschaftsdaten  
[office@ratswd.de](mailto:office@ratswd.de)

**York Sure-Vetter**

GESIS – Leibniz-Institut für Sozialwissenschaften  
[praesident@gesis.org](mailto:praesident@gesis.org)

# **Wo finde ich relevante Forschungsdaten für meine wissenschaftliche Arbeit?**



Möchte man für eine wissenschaftliche (Qualifikations-)Arbeit eine empirische Auswertung durchführen, steht man vor der Frage, wo sich passende Daten für statistische Auswertungen finden lassen. Zwar liefern häufig der Betreuer einer solchen Arbeit, Kollegen und nicht zuletzt die einschlägige Literatur Hinweise auf mögliche Datenquellen, jedoch ist das Datenangebot in den Sozial- und Wirtschaftswissenschaften nicht nur auf den ersten Blick unübersichtlich.

**Die folgende Tabelle verdeutlicht beispielhaft die unterschiedliche Herkunft von quantitativen Forschungsdaten.** Qualitative Forschungsdaten werden in diesem Wegweiser weitgehend ausgeklammert. Beabsichtigt ist jedoch, in zukünftigen Versionen dieser Broschüre auch qualitative Forschungsdaten stärker zu thematisieren.

TABELLE 1

## Beispiele für Forschungsdaten in den Sozial- und Wirtschaftswissenschaften

### Statistische Daten

#### Statistische Ämter des Bundes und der Länder:

- Bildung
- Finanzen
- Gesundheit
- Landwirtschaft
- Umwelt/Energie
- VGR (BIP)
- Zahl der Beschäftigten
- Zahl der Einwohner/innen
- ...

#### OECD:

- Daten zu Mitgliedsstaaten
- Ökonomische Indikatoren
- Strukturdaten
- ...

#### Eurostat:

- Bevölkerung und soziale Bedingungen
- Industrie, Handel und Dienstleistungen
- Wirtschaft und Finanzen
- ...

### Survey-Daten

- ALLBUS
- Community Innovation Survey
- ESS
- EU-LFS
- EU-SILC
- GLES
- Gruppenbezogene Menschenfeindlichkeit
- ifo-Geschäftsindex
- Innovationsmonitor des IW
- Mannheimer Innovationspanel
- NEPS
- PAIRFAM
- PIAAC
- PISA
- SHARE
- SOEP
- Wahlstudien
- ZEW-Hightech-Gründungspanel
- ...

### Prozessproduzierte Mikrodaten

#### Öffentliche Verwaltung:

- Einkommens- und Vermögensstatistik
- Kraftfahrzeugdaten
- Melde daten
- Steuerdaten (Grundsteuer und Steuereinnahmen)
- ...

#### Wirtschaftsdaten:

- Aktienkurse
- Bilanzdaten
- Strukturmerkmale
- ...

#### Bundesagentur für Arbeit:

- Daten zu Arbeitsmarkt und Berufen
- Daten zu Arbeitsvermittlung
- Daten zu Entgeltersatzleistungen und Arbeitsförderungsmaßnahmen
- Sozialversicherungsdaten
- ...

Typisierung, nicht vollständig  
Quelle: Eigene Zusammenstellung/Legende

**ALLBUS** Allgemeine Bevölkerungs umfrage der Sozialwissenschaften

**BIP** Bruttoinlandsprodukt

**ESS** European Statistical System

**EU-LFS** European Union Labour Force Survey

**EU-SILC** European Union Statistics on Income and Living Conditions

**GLES** German Longitudinal Electi on Study / Nationale Wahlstudie

**Ifo** ifo Institut – Leibniz-Institut für Wirtschaftsforschung an der Universität München

**IW** Institut der deutschen Wirtschaft Köln e.V.

**NEPS** National Educational Panel Study / Nationales Bildungspanel

**OECD** Organisation for Economic Cooperation and Development / Organisation für wirtschaftliche Zusammenarbeit und Entwicklung

**PAIRFAM** Panel Analysis of Intimate Relationships and Family Dynamics / Das Beziehungs- und Familienpanel

**PIAAC** Programme for the International Assessment of Adult Competencies

**PISA** Programme for International Student Assessment

**SHARE** Survey of Health, Ageing and Retirement in Europe

**SOEP** Sozio-oekonomisches Panel

**VGR** Volkswirtschaftliche Gesamtrechnung

**ZEW** Zentrum für Europäische Wirtschaftsforschung, Mannheim

Die Zugriffsmöglichkeiten auf solche Daten unterscheiden sich erheblich: So stellen öffentliche Anbieter, insbesondere die Statistischen Ämter, aggregierte Forschungsdaten (**Makrodaten**, oft in Form von **Zeitreihen**) in der Regel kostenlos und direkt über das Internet bereit, während der Zugang zu datenschutzrechtlich sensiblen **Mikrodaten** (Survey-Daten und prozessproduzierte Daten) aus Datenschutzgründen reglementiert ist und oft auch (meist geringe) Kosten für die Nutzung anfallen (siehe Abschnitt »Zugangswege«). Im Falle von **kommerziellen Datenanbietern** besteht in manchen Fällen ein Zugang über wissenschaftliche Bibliotheken.

Für **Deutschland** sind insbesondere die wissenschaftlichen Mikrodaten von Relevanz, die durch die vom **Rat für Sozial- und Wirtschaftsdaten (RatSWD)** akkreditierten Forschungsdatenzentren für die Wissenschaft bereitgestellt werden. Die Forschungsdatenzentren bieten eine breite Palette an statistischen Mikrodaten aus verschiedenen Fachgebieten an. Auf dem Portal des RatSWD kann direkt nach Forschungsdaten gesucht werden.

Im Folgenden werden anhand der oben stehenden Typisierung einige relevante Anbieter von quantitativen Forschungsdaten vorgestellt. Bei einigen Anbietern kommt es dabei zu Überschneidungen zwischen den dargestellten Kategorien.

### a) Statistische Aggregatdaten

Aggregierte Daten (so genannte Makrodaten) fassen eine Vielzahl an Einzelbeobachtungen (Mikrodaten) in einem einzigen Wert zusammen. Beispiele hierfür sind das BIP, das die gesamte Wirtschaftsleistung eines Landes für einen bestimmten Zeitraum zusammenfasst, oder auch das durchschnittliche Haushaltseinkommen. Daten über Länder können wiederum zu Daten über Ländergemeinschaften, z.B. über die Mitgliedsstaaten der OECD, aggregiert werden.

Für bestimmte Fragestellungen ist es sinnvoll, so genannte disaggregierte Makrodaten zu nutzen. Diese sind dann, meist in räumlicher oder sektoraler Hinsicht, kleinteiliger definiert als Länderdaten, allerdings ohne die Granularität von Mikrodaten zu erreichen. Mikrodaten stellen somit gewissermaßen das Rohmaterial, d.h. direkt gemessene Kennwerte, für Makrodaten dar.

Die **Statistischen Ämter des Bundes und der Länder** sowie **Städte und Landkreise** sind ein guter Anlaufpunkt, wenn es um sozio-strukturelle oder wirtschaftliche Durchschnitts-Eigenschaften der in Deutschland lebenden Bevölkerung geht. Die Ämter stellen Informationen zu Themen wie Bevölkerung, Arbeitsmarkt, Wahlen, Bildung, Sozialleistungen, Gesundheit, Wohnen, Umwelt, aber auch zu Industriezweigen, Unternehmen, Außenhandel und öffentlichen Finanzen bereit. Auch die Bundesländer bieten statistische Daten auf ihren Websites an (> Übersicht der Websites der Statistischen Landesämter).

Zudem bietet die **Bundesagentur für Arbeit** umfangreiche Statistiken zu Arbeitslosen, Beschäftigten, zu Sozi-

alleistungsempfängern sowie zu Maßnahmen der Arbeitsförderung an.

Liegt der Fokus einer empirischen Arbeit auf Fragestellungen, für die internationale Daten benötigt werden, ist die Auswahl häufig noch schwieriger. Speziell komparative Studien stellen eine Herausforderung dar, insbesondere weil die Länder und die einzelnen Datenproduzenten zum Teil nach unterschiedlichen Definitionen und Methoden Daten erheben und zusammenstellen.

Für die **Europäische Union (EU)** lassen sich viele Daten über das Portal der Europäischen Statistikbehörde Eurostat abrufen. Die angebotenen Daten umfassen Themen wie Wirtschaft und Finanzen, Bevölkerung und soziale Bedingungen, Internationale Zusammenarbeit, Industrie, Handel und Dienstleistungen, Landwirtschaft sowie Forstwirtschaft und Fischerei (siehe auch hier).

Die **Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD)** veröffentlicht ebenfalls zahlreiche Daten. Thematisch werden Datensätze u.a. aus den Bereichen Landwirtschaft und Fischerei, Entwicklung, wirtschaftliche Prognosen, Ausbildung, Energie, Umwelt, Finanzen, Globalisierung, Gesundheit, volkswirtschaftliche Gesamtrechnung, Produktivität, Regionalstatistiken oder Sozialleistungen bereitgestellt. Neben den kostenfrei zur Verfügung gestellten Statistiken bietet die OECD iLibrary darüber hinaus kostenpflichtig weitergehende Informationen an. Ein kostenfreier Zugang zu Letzteren ist über einige wissenschaftliche Bibliotheken möglich.

Auch die **Weltbank** verfügt über zahlreiche frei zugängliche Datensätze zu Themen wie Landwirtschaft und ländliche Entwicklung, Armut, Klimawandel, Ausbildung, Umwelt, Finanzsektor, Gender, sozialen und städtischen Entwicklungen, Gesundheit, Infrastruktur, öffentlicher und privater Sektor u.v.m. (siehe auch: world bank data catalogue).

Einrichtungen der **Vereinten Nationen (UNO)** und deren Portal **UNdata** stellen ebenfalls Forschungsdaten bereit. Thematisch bieten diese und weitere Einrichtungen Daten zu ihren spezifischen Fachgebieten:

- die Food and Agriculture Organization (FAO)
- die International Labour Organization (ILO) und deren Portale Laborsta und Ilostat
- der International Monetary Fund (IMF) und deren e-Library
- die United Nations Educational, Scientific and Cultural Organization (UNESCO)
- die World Health Organization (WHO) und deren Datenportal
- die World Trade Organization (WTO)

Über die amtliche Statistik sind teilweise auch **Daten über Firmen**-gruppen verfügbar, da für einzelne Rechtsformen (z.B. AGs) umfassende Veröffentlichungs-

## Anbieter statistischer Aggregatdaten

### Anbieter statistischer Aggregatdaten mit Bezug auf Deutschland:

Datensätze des **Statistischen Bundesamts** sowie der **Statistischen Landesämter** bieten neben Sozial- und Wirtschaftsstatistiken auch Finanz- und Steuerstatistiken sowie Daten zu den Themen Agrar, Energie und Umwelt (> Übersicht zum Datenangebot der amtlichen Statistik).

Die in verschiedenen Bundesländern existenten **Open Data Portale** (z.B. für Hamburg und Berlin) stellen verschiedenste Daten (u.a. der öffentlichen Verwaltung) für die Allgemeinheit kostenfrei zur Verfügung.

- Das kommerzielle Angebot **Statista** bietet – zum Teil auch kostenlos – Zugang zu umfangreichen Datensätzen von Markt- und Meinungsforschungsinstitutionen sowie aus Wirtschaft und amtlicher Statistik an.

### Statistische Aggregatdaten mit internationalem Bezug bieten folgende Anbieter:

- Das Portal **Allthatstats** bietet Zugang zu Wirtschaftszeitreihen und Ländertabellen, die aus nationalen und internationalen sowie amtlichen und halbamtlchen Quellen stammen. Daneben sind Daten zu nationalen und internationalen Organisationen (z.B. OECD und Weltbank) sowie zu Staaten und Staatengemeinschaften wie der EU oder den USA abrufbar.
- Das Projekt **EUKlems** bietet Statistiken zu ökonomischem Wachstum, Produktivität, Arbeitsplatzentwicklung, Vermögensbildung und technologischem Wandel für alle EU-Staaten ab 1970.
- Die Website des **International Household Survey Networks** stellt Informationen zu Studien und Statistiken mit Schwerpunkt auf Schwellen- und Entwicklungsländern bereit (sowohl statistische Daten als auch Erhebungsdaten).
- Das **Inter-university Consortium for Political and Social Research (ICPSR)** hält zahlreiche Datensätze zu verschiedenen Ländern, teilweise zum direkten Download, vor.
- Das Portal **quandl.com** stellt nach eigenen Angaben gut neun Millionen Datensätze aus den Bereichen Finanz-, Wirtschafts- und Sozialwissenschaften kostenfrei bereit. Die Datensätze lassen sich u.a. auch in Excel, R, Stata und Matlab nutzen bzw. integrieren.
- Die **Statistischen Ämter des Bundes und der Länder** bieten eine Linkliste zu den statistischen Ämtern nahezu jeden Landes sowie zu Daten erhebenden supranationalen Organisationen.
- Die **Total Economy Database (TED)** von The Conference Board bietet eine Übersicht zu Wirtschaftsdaten von 123 Ländern ab 1950. Sie umfasst Angaben zu Bruttoinlandsprodukt, Bevölkerung, Arbeitsverhältnissen, Arbeitsqualität und -produktivität, Kapitaldienstleistungen und Produktivitätsfaktoren.
- Darüber hinaus bietet das EU-Projekt **Urban Audit** Statistiken und Informationen über die Lebensqualität in zahlreichen europäischen Städten.

## Adressen für Survey-Daten (Mikrodaten)

Einen Überblick über akkreditierte Forschungsdatenzentren und deren Angebot für die Wissenschaft bietet zunächst die Website des Rats für Sozial und Wirtschaftsdaten (**RatSWD**).

Darüber hinaus stellt insbesondere auch das Leibniz-Institut **GESIS** zahlreiche Daten von nationalen und international-vergleichenden Umfragen zu soziologischen und politikwissenschaftlichen Fragestellungen zur Verfügung, die beispielsweise über den Datenbestandskatalog recherchiert werden können.

Exemplarisch werden einige relevante Einzelanbieter von Mikrodaten vorgestellt, die zumeist einen **Bezug auf Deutschland** aufweisen.

- Die **Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)** wird seit über 30 Jahren von GESIS in Zusammenarbeit mit der sozialwissenschaftlichen Profession durchgeführt und misst die Einstellungen, das Verhalten und die Sozialstruktur der Bevölkerung in der Bundesrepublik Deutschland.
- Die **Deutsche Bundesbank** und deren **Forschungsdaten- und Servicezentrum (FDSZ)** bieten relevante finanzielle und außenwirtschaftliche Statistiken sowie umfangreiche Daten zu wirtschaftlichen Indikatoren und saisonbereinigten Konjunkturdaten an.
- Das **Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB)** bietet Forschenden Daten zu Personen, Haushalten und Betrieben sowie kombinierte Datensätze aus Betriebs- und Personendaten an.
- Das **Forschungsdatenzentrum des Deutschen Zentrums für Altersfragen (FDZ-DZA)** stellt Daten des Deutschen Alterssurveys (DEAS) und des Deutschen Freiwilligensurveys (FWS) für nicht-gewerbliche Zwecke kostenfrei zur Verfügung.
- Das **Forschungsdatenzentrum der Rentenversicherung (FDZ-RV)** bietet prozessproduzierte Mikrodaten zu den Leistungen der Rentenversicherung (z.B. Rehabilitation und Rente) sowie zu Anwartschaften aus sozialversicherungspflichtiger Beschäftigung, Kindererziehung und nicht erwerbsmäßiger Pflege.
- Das bei GESIS angesiedelte **Forschungsdatenzentrum Wahlen** bietet Zugang zu umfangreichen Daten zu Wahlen.
- Das **LMU-ifo Economics & Business Data Center (EBDC)** stellt Datensätze deutscher

Unternehmen bereit, die sowohl Umfragenedaten des ifo Instituts als auch externe Bilanzdaten der Firmendatenbanken Amadeus und Hoppenstedt enthalten.

- Der **Mikrozensus** (von den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder als Mikrodaten gebührenpflichtig bereitgestellt) ist eine amtliche Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt in der Bundesrepublik Deutschland, die vom Statistischen Bundesamt durchgeführt wird und an der jährlich 1% aller Haushalte in Deutschland beteiligt ist. Ein Auszug aus dem Mikrozensus ist als CAMPUS-File für Übungszwecke frei im Internet herunterladbar.
- Das **Sozio-ökonomische Panel (SOEP)** ist eine jährlich im Auftrag des Deutschen Instituts für Wirtschaftsforschung (DIW Berlin) durchgeführte Wiederholungsbefragung (»Panel«) von etwa 15.000 Haushalten, in denen fast 40.000 Menschen leben. Erfasst werden unter anderem Haushaltssammensetzung, Erwerbs- und Familienbiographien, Erwerbsbeteiligung und berufliche Mobilität, Einkommensverläufe, Gesundheit, Persönlichkeitsmerkmale und Lebenszufriedenheit.

### **Survey-Daten mit stärkerem internationalen Bezug bieten folgende Anbieter:**

- Das **Consortium of European Social Sciences Data Archives (CESSDA)** bietet einen integrierten Zugang zum Angebot sozialwissenschaftlicher Datenarchive in Europa mit dem Schwerpunkt auf Survey-Daten.
- Das **Forschungsdatenzentrum Internationale Umfrageprogramme** bei GESIS bietet Datenaufbereitung und Beratung für ausgewählte internationale Studienserien an.
- Das **Forschungsdatenzentrum SHARE** bietet Paneldaten zu über 50-Jährigen und älteren Personen in ganz Europa an. Schwerpunkte der SHARE-Erhebung sind Einkommen, Vermögen und Gesundheit.
- Die Website des **International Household Survey Networks** stellt Informationen zu Studien und Statistiken mit Schwerpunkt auf Schwellen- und Entwicklungsländern zur Verfügung (sowohl statistische Daten als auch Erhebungsdaten).
- Das **Inter-university Consortium for Political and Social Research (ICPSR)** hält zahlreiche Datensätze zu vielen Ländern, teilweise zum direkten Download, bereit.

## Adressen für prozessproduzierte Mikrodaten

Bei den **Statistischen Ämtern des Bundes und der Länder**, am **Institut für Arbeitsmarkt- und Berufsforschung (IAB)**, bei der **Bundesagentur für Arbeit (BA)**, beim **Forschungsdatenzentrum der Rentenversicherung (FDZ-RV)** sowie bei der **Deutschen Bundesbank** finden Sie eine Reihe von prozessproduzierten Unternehmens- bzw. Betriebsdaten, von denen viele über die institutionseigenen Forschungsdatenzentren (FDZ) für die Wissenschaft zugänglich sind. Sie gelten auch als Mikrodaten. Beispiele hierfür sind:

- das Projekt »**Kombinierte Firmendaten für Deutschland**« (KombiFiD)« und
- das Projekt »**Amtliche Firmendaten für Deutschland**« (AFID)«.

Auch das **Forschungsdatenzentrum der Rentenversicherung (FDZ-RV)** bietet prozessproduzierte Mikrodaten zu den Leistungen der Rentenversicherung (z.B. Rehabilitation und Rente) sowie zu Anwartschaften aus sozialversicherungspflichtiger Beschäftigung, Kindererziehung und nicht erwerbsmäßiger Pflege an. Zudem finden sich zahlreiche kommerzielle Anbieter in diesem Bereich:

- Die **Amadeus Datenbank** bietet Zugang zu Wirtschaftsdaten zu mehr als 20 Millionen öffentlichen und privaten Unternehmen aus **44 vornehmlich europäischen Ländern**.
- Die Firmendatenbank **Bisnode/Hoppenstedt** bietet Informationen über die 300.000 bedeutendsten Unternehmen und Institutionen in Deutschland. Abgebildet werden Unternehmen ab einer Million Euro Jahresumsatz und/oder mindestens zehn Beschäftigten.
- **Bloomberg** bietet Bestands- und Echtzeit-Datenlieferungen von Daten aus den verschiedenen Finanzbereichen (z.B. Bonds).
- Das **Bureau van Dijk** bietet u.a. Zugang zu nationalen und internationalen Firmen-, Banken-, Versichererdaten sowie zu Länderdaten mit volkswirtschaftlichen Indikatoren und Prognosen sowie M&A-Transaktionsdaten.
- **Thomson Reuters Datastream** bietet Zugriff auf Aktienkurse, internationale Marktindizes, Bonds, Termingeschäfte, Wechselkurse, Zinssätze, Dividenden sowie makroökonomische Zeitreihen.
- Die **Wharton Research Data Services (WRDS)** bieten Zugang zu mehr als 200 Terabyte Daten aus den Bereichen Accounting, Banking, Volkswirtschaft, Finanzwirtschaft, Versicherungsgewerbe, Marketing und Statistik.

*Informationen zur Verwendbarkeit der Daten und deren Nachnutzung entnehmen Sie bitte den jeweiligen AGBs.*

pflichten festgelegt wurden. Das **Handelsregister** verfügt beispielsweise flächendeckend über Daten zu Unternehmen (diese werden auch in Firmendatenbanken veröffentlicht) und es werden Daten zu Importen, Exporten, Steuereinnahmen, Direktinvestitionen und Bilanzen der einzelnen Branchen bereitgestellt. Hier sind aber nationale bzw. regionale Unterschiede in der Erhebung vieler Daten zu erkennen und diese sind i. d. R. nicht länderübergreifend in einer Datenbank zugänglich, geschweige denn vergleichbar.

### b) Survey-Daten

Survey-Daten liefern sehr detaillierte Informationen über einzelne Beobachtungseinheiten (insbesondere zu Personen, Privathaushalten, Betrieben, Unternehmen) und werden deshalb auch als Mikrodaten bezeichnet. Statistische Mikrodaten sind für die Forschung ausgesprochen attraktive Daten, denn sie ermöglichen es, Zusammenhänge und Merkmale teilweise bis auf Ebene des Individuums bzw. einer Organisation zu untersuchen. Dabei wird durch geeignete Anonymisierung der Daten die Anonymität der Beobachtungseinheiten voll gewahrt.

Die ÜBERSICHT 2 weist auch einige Anbieter von Erhebungs- bzw. Mikrodaten mit internationalem Bezug aus.<sup>1</sup>

### c) Prozessproduzierte Mikrodaten

Prozessproduzierte Mikrodaten entstehen zumeist durch Verwaltungsvorgänge, entstammen also in der Regel einem nicht primär wissenschaftlichen Hintergrund.

Einige dieser prozessproduzierten Daten werden auch von Unternehmen für Unternehmen und die Wissenschaft aufbereitet und teilweise durch Survey-Daten (etwa den Ergebnissen von Telefonbefragungen bei Firmen) ergänzt. Mikrodaten zu Abläufen, Produktionsprozessen und -kosten in Unternehmen enthalten oft wettbewerbsrelevante Informationen (z. B. Entwicklung des Kundenstammes) und sind somit speziell geschützt und nur schwer zugänglich. Meist sind derart sensible Daten nur im direkten Kontakt zu Firmen auswertbar und unterliegen in jedem Fall restriktiven Vertrags- und Nutzungsbedingungen.

### d) Weitere Ressourcen

In Ergänzung zu den oben genannten Quellen weisen wir auf zwei weitere Überblicksseiten mit Hinweisen zu Forschungsdatenressourcen hin:

- Die **American Economic Association (AEA)** verfügt auf ihrer Website über eine umfangreiche Linkssammlung von Datenquellen für Wirtschaftsforschende. Der Schwerpunkt liegt auf ökonomischen Daten zu den USA, jedoch gibt es auch dutzende Verweise auf andere wirtschaftlich relevante Datenquellen.

Eine interessante Sammlung von **Finanzdaten-Anbietern** liefert überdies die **Erasmus Universität Rotterdam**. Dort finden sich auch Übersichten zu **Datenbanken für die Sozialwissenschaften**.

<sup>1</sup> Wenn Sie beabsichtigten sollten, Mikrodaten zu nutzen, werfen Sie bitte auch einen Blick auf das Kapitel »Zugang zu Forschungsdaten«, das insbesondere für das deutsche Rechtsgebiet Gültigkeit hat.

Die vorgestellten Datenbanken sind im Regelfall kostenpflichtig.

## Zugangswege

Aufgrund von Datenschutz-Notwendigkeiten und anderen gesetzlichen Bestimmungen unterscheiden sich die Zugriffsmöglichkeiten auf sozial- und wirtschaftswissenschaftliche Forschungsdaten je nach Datentyp zum Teil erheblich. Darüber hinaus ist es für den Zugriff relevant, ob es sich um öffentlich erhobene und kostenfrei bzw. gegen eine geringe Gebühr bereit gestellte Daten handelt oder um Daten, die durch spezialisierte Unternehmen erhoben und/oder aufbereitet wurden und als Produkt verkauft werden. Für umfangreiche Datenmengen können die Gebühren, die statistische Ämter für Mikrodaten verlangen müssen, in einen vier- oder gar fünfstelligen Euro-Bereich gehen.

Aufgrund dieser Differenzierungen lassen sich die Zugangswege grob in drei Typen einteilen:

### 1. Aggregierte Statistik-Daten (»Makrodaten«)

Diese Daten sind meist kostenfrei und (teilweise nach Registrierung) online direkt zu beziehen. Genaue Informationen entnehmen Sie bitte dem jeweiligen Datenportal. Im Falle von disaggregierten Daten können darüber hinaus besondere Anforderungen an den Zugang bestehen.

### 2. Statistik-Mikrodaten

Für statistische Mikrodaten gelten besondere Schutzbestimmungen durch die Bundes- und Landesdatenschutzgesetze sowie das Bundesstatistikgesetz. So muss die Erhebung von personenbeziehbaren Daten immer an ein bestimmtes Forschungsziel geknüpft sein, welches als so relevant erachtet wird, dass die Gewinnung personenbeziehbarer Daten zu rechtfertigen ist.

Der Schutz von personenbeziehbaren Daten ist zudem ein gesetzlicher Auftrag, der durch eine Anonymisierung der Daten gewährleistet wird, wobei der Grad der Anonymisierung von der Art und der Kontrolle der Nutzung abhängt. Aus diesem Grund können statistische Mikrodaten nicht ohne Weiteres von den Websites der Forschungsinstitutionen heruntergeladen werden. Dies gilt in ähnlicher Form auch für Firmendaten, wie beispielsweise die Mikrodaten der Projekte AFID oder KombiFiD.

Grundsätzlich finden sich bei Organisationen, die solche Daten anbieten, verschiedene Stufen der Anonymisierung, die beispielhaft anhand einer Grafik der statistischen Ämter dargestellt werden (siehe Abbildung 1).

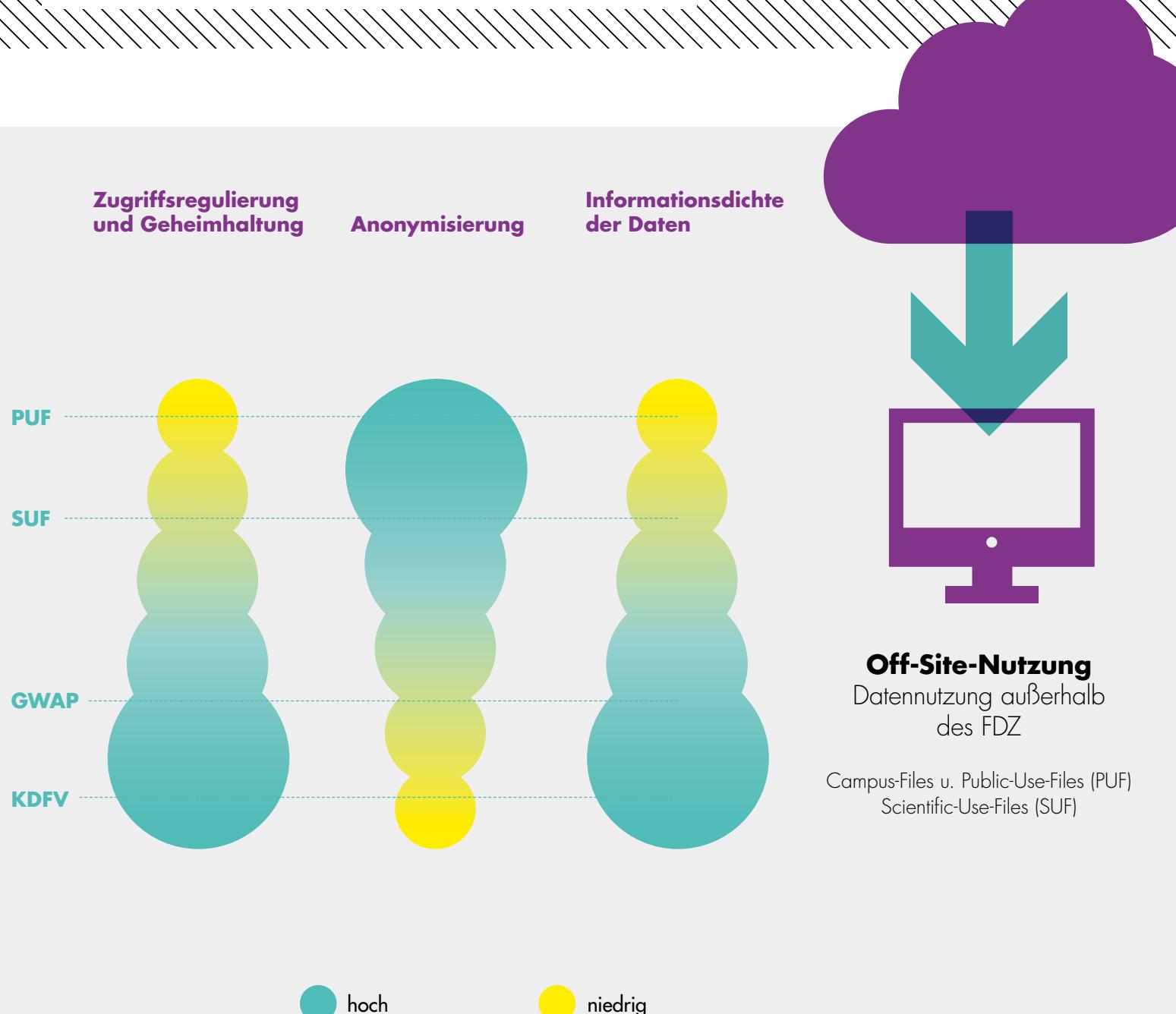
- Vollständig anonymisierte Mikrodatensätze werden auch für die Öffentlichkeit angeboten, es handelt sich hier um so genannte **Public-Use-Files** (PUF) oder um **CAMPUS-Files**. Diese Datensätze unterscheiden sich allerdings hinsichtlich ihres Informationsumfangs, der aus Datenschutzgründen stark reduziert ist, deutlich von den Scientific-Use-Files.



ABBILDUNG 1

## Nutzungswege für Mikrodaten der Statistischen Ämter

Quelle: Eigene Darstellung auf Basis von Köhler, F. (2012). 10 Jahre Forschungsdatenzentren der Statistischen Ämter. Statistische Monatshefte Niedersachsen, Ausgabe 6/2012, S. 335.



- Bei **Scientific-Use-Files** (SUF) handelt es sich um »faktisch anonymisierte« Mikrodatensätze, die z.B. von den Forschungsdatenzentren der Statistischen Ämter oder von spezialisierten Forschungsdatenzentren (siehe Abbildung 1) bereitgestellt werden. Nach Maßgabe des Bundesstatistikgesetzes dürfen solche faktisch anonymisierten Daten allerdings nur wissenschaftlichen Einrichtungen und nur zur Durchführung wissenschaftlicher Vorhaben zugänglich gemacht werden.
- Darüber hinaus bieten die meisten FDZs auf Antrag auch so genannte **Gastwissenschaftlerarbeitsplätze** (GWAP) an. Wird ein solcher Antrag genehmigt, können Forschende an besonders gesicherten PC-Arbeitsplätzen in den Forschungsdatenzentren mit schwach anonymisierten Daten arbeiten.
- Einen weiteren Weg stellt die **kontrollierte Datenfernverarbeitung** (KDFV) dar, die Zugang zu den formal anonymisierten Originaldaten bietet. Forschende können mit diesen Daten nicht selbst arbeiten, sondern erhalten stattdessen Dummy-Daten, die in Aufbau und Merkmalsausprägungen dem Originalmaterial ähneln. Mittels dieser Dummy-Dateien können Auswertungsskripte erstellt werden, mit denen Beschäftigte der Forschungsdatenzentren anschließend die Originaldaten auswerten. Die Datennutzer erhalten nach einer notwendigen Geheimhaltungsprüfung (einer Prüfung der zu übermittelnden Daten auf Re-Anonymisierungsrisiken) schließlich die Ergebnisse dieser Auswertung.

Ähnlich gestalten sich die Richtlinien für den Zugang zu den bei **Eurostat** verfügbaren anonymisierten Mikrodaten, die ebenfalls nur für Forschungszwecke gewährt wird.

### **Kosten**

Die Nutzung von wissenschaftsgetragenen Forschungsdaten (außerhalb der amtlichen Statistik) ist in der Regel für die Wissenschaft kostenfrei.

Für die von den Forschungsdatenzentren der Statistischen Ämter bereitgestellten Mikrodaten (über frei downloadbare CAMPUS-Files hinaus) wird derzeit pro Statistik, Erhebungsjahr und Nutzungsweg (Scientific-Use-File (SUF), Gastwissenschaftlerarbeitsplatz, kontrollierte Datenfernverarbeitung) ein Nutzungsentsgelt von ca. 250 Euro erhoben (vgl. <http://www.forschungsdatenzentrum.de/nutzungsentgelte.asp>).

Bislang konnte die Wissenschaft die Aufsichtsbehörden der statistischen Ämter und den Gesetzgeber noch nicht überzeugen, den »Daten-Service« für die Forschung kostenlos bzw. gegen eine marginale Gebühr bereitzustellen.

### **3. Daten kommerzieller Anbieter**

Die Nutzung kommerzieller Datenbanken erfordert im Regelfall den Abschluss eines Vertrages und ist nur gegen Gebühren möglich. Ausnahmen bilden kommerzielle Datenbanken, die Ihnen über Ihre wissenschaftliche Bibliothek bereitgestellt werden.

Fragen Sie zunächst bei Ihrer Bibliothek nach, ob entsprechende Zugangswege vorhanden sind. Darüber hinaus können Sie ggf. Spezialbibliotheken kontaktieren, die besondere Services für ihre Wissenschaftsbereiche anbieten (z. B. die ZBW – Leibniz-Informationszentrum Wirtschaft in Kiel und Hamburg, die Universitäts- und Staatsbibliothek Köln bzw. GESIS für den Bereich Sozialwissenschaften). Ohne entsprechende Unterstützung durch Bibliotheken bleibt zumeist nur ein (ggf. kostspieliger) Zugang über die Vertragsschließung mit einem kommerziellen Datenanbieter. Die Angaben zu den Kosten und Nutzungsbedingungen entnehmen Sie den entsprechenden Websites.

# Wie zitiere ich Forschungsdaten korrekt?



“

Eine korrekte Datenzitation ist nicht nur Ausdruck korrekten wissenschaftlichen Arbeitens. Durch die Datenzitation wird auch die Produktion von Forschungs- bzw. Statistik-Daten als eine primäre Wissenschaftsleistung anerkannt. Nicht nur Forschungspublikationen, sondern auch die im Forschungsprozess entstandenen Daten sollten im Sinne von »Good Scientific Practice« zitierbar sein und zitiert werden (vgl. Allianz der deutschen Wissenschaftsorganisationen, 2010).

Wie bei Textpublikationen ist die korrekte Zitation vor allem für den Datenproduzenten wichtig: Daten, die die Grundlage für wissenschaftliche Auswertungen und darauf basierende Publikationen bilden, schaffen einen Mehrwert für die Forschung. Daher sollte auch den Produzenten von Daten durch Zitationen eine entsprechende Anerkennung zuteil werden, dies umso mehr, da die Produktion und Dokumentation von Daten oftmals zu Lasten anderer wissenschaftlicher Leistungen (insbesondere Textpublikationen) gehen.

Die Entwicklung bibliometrischer Verfahren, die die Leistung der Daten-Produktion ähnlich den Impact-Faktoren bei Textpublikationen messbar machen, trägt zu einer erhöhten Relevanz von Datenzitationen bei. Die Grafik des Australian National Data Service (Abbildung 2) verdeutlicht diesen Prozess.

Wenn Sie entsprechende Daten gefunden und diese in Ihrer wissenschaftlichen Arbeit verwendet haben, stellt sich als Nächstes die Frage, wie Sie diese Quelle wissenschaftlich korrekt zitieren können.

Mittlerweile haben sich verschiedene Gruppen und wissenschaftliche Communities damit beschäftigt, Guidelines und Empfehlungen zur Zitation von Forschungsdaten zu erstellen.<sup>2</sup> Obwohl sich die Herangehensweisen prinzipiell ähneln, existiert (noch) kein einheitlicher Standard.

Eine Datenzitation benötigt mindestens die bei gedruckten Werken bekannten bibliografischen Angaben wie Autor, Titel, Verlag, Publikationsdatum. Zusätzlich ist in vielen Empfehlungen die persistente Identifikation (also die dauerhafte Speicherung eines Datensatzes unter einer sich nicht verändernden Webadresse) wesentlicher Bestandteil einer guten Datenzitation. Diese Empfehlung liegt in der Tatsache begründet, dass Informationen in der digitalen Welt dynamischer und fragiler sind als in gedruckten Veröffentlichungen. Digitale Daten sind leichter veränderbar und die dadurch entstandenen neuen Versionen sind nicht immer eindeutig identifizierbar. Zusätzlich kann sich der ursprüngliche Speicherort ändern, was die Auffindbarkeit der Daten ohne

<sup>2</sup> Beispiele finden sich beim ANDS, CODATA-ICSTI, da|ra, DataCite, Data-PASS, FORCE11, GESIS, IASSIST-SIGDC, ICPSR, IQSS, STD, Thomson Reuters Data Citation Index usw.

ABBILDUNG 2

## Wie der Prozess der Datenzitation funktioniert



Quelle: ands – Australian National Data Service, <http://ands.org.au>

## Das DOI-System

eine persistente Identifikation des Objektes fast unmöglich macht (vgl. CODATA-ICSTI, 2013).

Persistente Identifikatoren, die in etwa einer einmalig vergebenen ISBN- bzw. ISSN-Nummer bei gedruckten Werken entsprechen, sind ein Weg zur Lösung der geschilderten Problematik.

Für die Identifikation von elektronischen Textpublikationen existieren mittlerweile diverse Systeme von Persistenten Identifizieren, die technisch auch die Basis zur Identifizierung von Daten leisten können: Archival Research Key (ARK), Digital Object Identifier (DOI), Handle, Persistent URL (PURL), Uniform Resource Name (URN) und weitere. Auf einen gemeinsamen Standard haben sich die verschiedenen Nutzergemeinden jedoch noch nicht geeinigt, obwohl die Systeme im Prinzip gut ineinander überführbar wären.

In der ÜBERSICHT 4 wird die Systematik der Datenzitation am Beispiel des am häufigsten verwendeten persistenten Identifikators, des DOI, beispielhaft vorgestellt. Wir empfehlen das DOI-System, das auch bei der Zitierung von Forschungspublikationen eine wesentliche Rolle spielt.

### **Empfohlene Datenzitation**

Wenn Sie nun einen verwendeten Datensatz zitieren möchten, stehen Sie möglicherweise erst einmal vor dem Problem, dass Sie nicht genau wissen, wie Datensätze richtig zitiert werden können. Im Folgenden haben wir daher einige Beispiele für korrekte Zitationen von Forschungsdaten gesammelt, die Ihnen helfen, diese Quellen wissenschaftlich zu zitieren.

Da es noch keine verbindlichen Standards zur Datenzitation gibt, empfehlen wir eine Zitierung auf der Basis des da|ra-Metadatenschemas 3.1 (vgl. Helbig et al., 2014a; vgl. Helbig et al., 2014b). Dieses Vorgehen entspricht auch dem empfohlenen Standard von DataCite (vgl. DataCite Metadata Working Group, 2013) und wird mit den folgenden Elementen erstellt:

#### **Creator (Veröffentlichungsdatum):**

**Titel. Publikationsagent. Identifier.**

Es kann auch wünschenswert sein, zwei optionale Elemente – Version und Typ der Ressource – mit anzugeben. Dann ist die empfohlene Form wie folgt:

#### **Creator (Veröffentlichungsdatum):**

**Titel. Version. Publikationsagent. Typ der Ressource.**

**Identifier.**

Dabei gelten neben den Definitionen im Metadatenschema folgende zusätzliche Empfehlungen:

**Creator:** Maximal fünf Namen sind möglich, danach ist mit „et al.“ abzukürzen.

**Veröffentlichungsdatum:** Hier wird nur das Jahr der Veröffentlichung des Datensatzes angegeben.

**Titel:** Hier sind der Titel und eventuell weitere Titel der Studie/des Datensatzes anzugeben.

Das DOI-System ist ein verwaltetes System für die persistente Identifikation von Inhalten, die in digitalen Netzwerken angeboten werden. Es kann zur Identifizierung von physikalischen, digitalen oder anderen Objekten genutzt werden. Beispielsweise nutzen bereits viele wissenschaftliche Verlage DOI-Namen zur persistenten Addressierung ihrer Artikel. Die Identifikatoren (DOI-Namen) führen direkt zum Speicherort des bezeichneten Objektes.

### **Struktur und Resolving eines DOI-Namens**

Ein DOI-Name besteht immer aus einem Präfix und einem Suffix, wobei beide durch einen Schrägstrich getrennt werden und das Präfix stets mit »10.« beginnt (vgl. Abb. 3). Das Präfix ist dabei einem bestimmten Datenzentrum zugeordnet (in Abb. 3: »3478«), während das Suffix einer bestimmten Datenpublikation der Einrichtung zugewiesen ist. Die Kombination aus alphanumerischen Prä- und Suffixen ermöglicht die Bildung einer unbegrenzten Anzahl von eindeutigen DOI-Namen (vgl. International DOI Foundation, 2014).

Um einen DOI-Namen in die zugehörige URL aufzulösen (Englisch: to resolve), gibt es verschiedene Möglichkeiten. Zum Einen kann er über die für verschiedene Internetbrowser angebotenen Plug-ins direkt eingegeben und aufgelöst werden. Eine andere Möglichkeit ist die Verwendung des Proxy Servers des DOI-Systems. Letztlich führt auch die Eingabe des DOI-Namens zusammen mit der vorangestellten Zeichenkette <http://doi.org> (z.B. <http://doi.org/10.7806/zew.mip.2013.V1.suf>) in den Eingabeschlitz jedes beliebigen Browsers den Nutzer direkt zum Speicherort des Objektes bzw. zu einer Website, die dieses und die Zugangsbedingungen ausführlich beschreiben.

### **DataCite und da|ra**

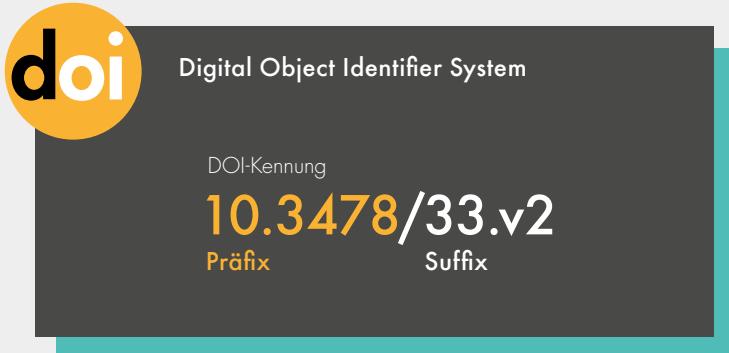
Der DOI-Registrierungsservice für Forschungsdaten wird von der bei der International DOI Foundation akkreditierten Registrierungsagentur DataCite zur Verfügung gestellt. DataCite ist ein 2009 in London gegründetes internationales Konsortium mit inzwischen 31 Mitgliedern aus 18 Ländern (Stand: 24. Juni 2014), die gemeinsam das Ziel verfolgen, die Akzeptanz von Forschungsdaten als eigenständige, zitierfähige wissenschaftliche Objekte zu fördern. Inzwischen wurden mehr als 4 Millionen DOIs vergeben.

Die Leibniz-Institute GESIS und ZBW sind seit 2010 Mitglieder von DataCite und bieten gemeinsam den DOI-Registrierungsservice für Sozial- und Wirtschaftsdaten da|ra an (vgl. Hausstein, 2012). Seit seiner Einführung hat sich da|ra innerhalb der nationalen Sozial- und Wirtschaftswissenschaften zur zentralen DOI-Vergabestelle entwickelt. Mehr als 30 Einrichtungen haben über da|ra mehr als 400.000 DOI-Registrierungen vorgenommen. Drei Viertel der beim RatSWD akkreditierten Forschungsdaten- und Datenservice-Zentren haben begonnen, ihre Datenbestände über da|ra zu registrieren.

## Zitation unter einer Tabelle (QUELLENANGABE)

ABBILDUNG 3

### Struktur eines DOI-Namens



**Version:** Die Version repräsentiert die Versionsnummer der Ressource.

**Publikationsagent:** Name des Datenzentrums/der Institution, das/die die Ressource veröffentlicht hat.

**Typ der Ressource:** Bezeichnung des generellen Typs einer Ressource. Hierzu kann unter den möglichen Ressourcentypen Sammlung, Datensatz, Text, Video, Bild, Audio und interaktive Ressource ausgewählt werden.

**Identifier:** Hier erscheint ein Persistent Identifier. Für Angaben mit DOI-Namen kann der Identifier optional im Originalformat und in einem Http-Format erscheinen. Um direkt auf die Quelle des Objektes verweisen zu können, sollte der DOI-Name entweder mit der URL des Resolvers abgedruckt (<http://doi.org/10.4232/1.10770>) oder – platzsparender – mit einem Hyperlink (<doi:10.4232/1.10770>) unterlegt werden.

Um diese Angaben für die zu zitierenden Forschungsdaten zu erhalten, empfiehlt es sich entweder das da|ra-Portal ([www.da-ra.de](http://www.da-ra.de)) oder den Metadata Store von DataCite (<https://mds.datacite.org>) zu konsultieren.

Zur Veranschaulichung folgen einige Beispiele für die Datenzitation. Sie beschreiben Daten, die über da|ra/ DataCite registriert wurden und zeigen, wie eine optimale Datenzitation aussieht.

**Fahrenberg, Jochen** (2010): Freiburger Beschwerdenliste FBL. Primärdaten der Normierungsstichprobe 1993. Version 1.0.0. ZPID- Leibniz-Zentrum für Psychologische Information und Dokumentation. Dataset.

<http://doi.org/10.5160/psychdata.fgjno5ano8>

**Rattinger, Hans; Roßteutscher, Sigrid; Schmitt-Beck, Rüdiger; Weßels, Bernhard** (2012): Wahlkampf-Panel (GLES 2009). Version: 3.0.0. GESIS Datenarchiv. Dataset.

Es ist üblich, unter den Tabellen und Grafiken einer Publikation die Quelle anzugeben. Dabei ist wie bei einem Verweis auf Literatur bzw. auf eine Datenquelle in einem Text zu verfahren, z.B.:

„Quelle: SOEPv29.1, Berechnungen des Autors“ oder „Source: SOEPv29.1, author's calculations.“

Wenn ein Datensatz mit einem DOI-Namen versehen ist, dann sollte dieser auch mit angegeben werden. Allerdings ist die Nennung des DOI-Namens bei den meisten Zeitschriften in der Quellenangabe von Tabellen noch nicht üblich. Möglich wäre z. B.:

„Quellen: SOEP (v29.1) (<doi:10.5684/soep.v29.1>), Berechnungen des Autors.“

In jedem Fall sollte die vollständige Zitationsangabe im Literaturverzeichnis aufgeführt werden.

**Schupp, Jürgen; Kroh, Martin; Goebel, Jan; Bartsch, Simone; Giesselmann, Marco et. al.** (2013): Sozio-ökonomisches Panel (SOEP), Daten der Jahre 1984-2012. Version: 29. SOEP- Sozio-ökonomisches Panel. Dataset. <doi:10.5684/soep.v29>.

Zusätzlich gilt es, zu beachten, dass manche Datenanbieter spezielle Zitierweisen wünschen bzw. zwingend als Teil der Datennutzungsbedingungen festgelegt haben (z. B. National Educational Panel Study (NEPS)). Um dem gerecht zu werden, sollte der DOI-Name der zu zitierenden Forschungsdaten zunächst aufgerufen werden (siehe «Struktur und Resolving eines DOI-Namens»). Unter der URL sind dann in der Regel die Nutzungsbedingungen (hier am Beispiel des NEPS) nachzulesen.

Darüber hinaus gibt es in Fachzeitschriften verschiedene Zitationsformate, die die oben beschriebene Zitationsweise abwandeln oder stark an die Zitation von Publikationen anpassen. Die Verwendung eines dieser Formate kann notwendig sein, wenn der Herausgeber einer Zeitschrift dieses Vorgehen verbindlich bei der Einreichung eines Beitrages vorschreibt. DataCite bietet in Kooperation mit crossref daher ein Werkzeug an, mit dessen Hilfe auf der Basis des DOI-Namens die Umwandlung in ein gewünschtes Format möglich wird. Der DOI Citation Formater unterstützt 500 in der Verlagswelt übliche Zitationsformate in 45 Sprachen.

## Veröffentlichung von Daten bei Zeitschriften

Einige Zeitschriften verlangen inzwischen bei Veröffentlichung von Artikeln, die auf der Auswertung von Daten beruhen, dass die verwendeten Daten sowie oft auch der Programmcode (do-File oder SPS-f ile) zusätzlich zum Artikelmanuskript miteingereicht und zu Replikationszwecken zur Verfügung gestellt werden. Bisher ist diese Art der Data Policy bei Zeitschriften im wirtschaftswissenschaftlichen Bereich nicht weit verbreitet: Vlaeminck und Siegert (2012) untersuchten 141 wirtschaftswissenschaftliche Fachzeitschriften, von denen 29 über Forschungsdatenrichtlinien verfügten, die zum Bereitstellen von Daten oder Programmcode verpflichteten. Für den Bereich der Politikwissenschaft untersuchten Gherghina und Katsanidou (2013) 120 Fachzeitschriften, von denen 18 über derartige Vorgaben verfügten.

### Data Policies einiger Zeitschriften in den Sozial- und Wirtschaftswissenschaften:

- American Economic Review ([AER](#))
- American Journal of Political Science ([AJPS](#))
- American Sociological Review ([ASR](#))
- Annual Review of Sociology ([ARS](#))
- Journal of the European Economic Association ([JEEA](#))
- Political Analysis ([PA](#))
- Royal Economic Society ([RES](#))

Dazu passend existiert mit dem [Replication Wiki](#) ein Projekt, in dem publizierte wirtschaftswissenschaftliche Forschung repliziert werden kann.

# Wie dokumentiere ich meine Daten und stelle sie zur Verfügung?



Einige Studierende oder Promovenden erheben oder sammeln im Zuge ihrer Qualifikationsarbeiten auch selbst Daten, erstellen neue Datensätze aus anderen, bereits existenten Datensätzen oder arbeiten in kleinen oder mittelgroßen Teams daran, entsprechende Daten zu generieren.

Das folgende Kapitel zeigt einige zentrale Aspekte auf, die es bei der Dokumentation Ihrer Daten zu beachten gilt.

Daran anschließend werden Optionen für die Archivierung von selbst erstellten Datensätzen aufgezeigt.

### a) Daten dokumentieren

Den eigenen Datensatz sinnvoll zu dokumentieren sollte dem Datenproduzenten in Hinblick auf die gute wissenschaftliche Praxis sowie aufgrund von Reproduzierbarkeit und Transparenz gegenüber Dritten eine Herzensangelegenheit sein. Leider sind Fragen der Dokumentation von Forschungsdaten in den Sozial- und Wirtschaftswissenschaften noch zu wenig in der akademischen Lehre verankert. Dieser Wegweiser soll nicht zuletzt dazu dienen, in die Curricula entsprechender Studiengänge einzugehen.

Da eine gute Dokumentation Zeit erfordert, wird sie leider oft kurzfristig zugunsten vermeintlich wichtigerer Aufgaben vernachlässigt. Datenproduzenten sollte freilich klar sein: Eine gute Dokumentation macht es externen Datennutzern einfacher, die Daten zu re-analysieren und die vom Datenproduzenten geleistete Arbeit mit einer Referenz, also einem Zitat, zu honorieren. Fehlt die Dokumentation, verschenkt der Datenproduzent eine mögliche Anerkennung seiner Arbeit (»credit«) durch Dritte.

Eine Dokumentation kann in Form eines simplen Dokuments, eines Working Papers oder auch in speziellen Fachzeitschriften (z.B. in der Rubrik »European Data Watch« der Fachzeitschrift Schmollers Jahrbuch) zitierbar veröffentlicht werden (vgl. Wagner, 2010). Die echte Veröffentlichung in einer Fachzeitschrift ist bislang allerdings nur »großen« Datensätzen vorbehalten, die eine breite(re) Nachnutzung erwarten lassen. Zum Zwecke einer möglichst umfangreichen Nachnutzung ist zudem ein guter Zugang zur Dokumentation über eine Projektwebsite hilfreich (vgl. z. B. die Dokumentationsseite des SHARE-Projekts).

Hauptziel einer Dokumentation ist es, die Entstehung des Datensatzes nachvollziehbar zu machen und ihn so zu beschreiben, dass Dritte damit arbeiten können. Der Aufwand, der dafür nötig ist, hängt zum Einen vom Umfang des Datensatzes selber ab. Andererseits bestimmt auch der Umfang der eigenen Arbeiten, wie ausführlich eine Dokumentation sein muss. So kann z. B. schon der Berechnungscode (z. B. do-file, sps-file) ausreichen, um einem Nachnutzer zu zeigen, wie ein Datensatz auf Basis frei verfügbarer und gut dokumentierter Primärdaten erstellt wurde. Zudem gibt es einige übergeordnete Informationen zu Datensätzen, die pauschal zur Verfügung gestellt werden sollten. Diese Informationen helfen den möglichen Nachnutzern bei der Entscheidung, ob die Daten relevant sein können. Folgende Punkte lassen sich darunter fassen:

### 1. Inhalt

Potentielle Nachnutzer eines Datensatzes werden im Allgemeinen versuchen, Angaben und Informationen über den Inhalt eines Datensatzes zu finden. Hilfreich dafür sind schlagwortartige Beschreibungen (z.B. »Arbeitsmarkt«, »Partnerschaften«, »Wahlen«, »Xenophobie«, »Investitionsgüter«) ebenso wie die Angabe von standardisierten inhaltsbezogenen Codes, z. B. JEL-Codes (ein Klassifikationsschema für Forschungsinhalte der US-Ökonomenvereinigung American Economic Association), die eine Einordnung in bestimmte Forschungsfelder erlauben. Der Nachteil dieser spezifischen Codes ist allerdings, dass ein Datenproduzent manchmal nicht abschätzen kann, in welchen ihm unbekannten bzw. wenig vertrauten Forschungsfeldern seine Daten für andere nutzbar sein könnten – beispielsweise wurde eine Studie der ALLBUS (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften) mit Wetterdaten verknüpft (Mutz & Kämpfer, 2011). Daher empfiehlt es sich, ein Abstract zu schreiben, das den Dateninhalt genauer spezifiziert als es ein einzelnes Schlagwort kann. Hier findet sich ein gutes Beispiel für das Abstract eines Datensatzes.

### 2. Beobachtungseinheit

Die Beobachtungseinheit ist die kleinste Ebene, die im Datensatz vorhanden ist. Sie muss in der Dokumentation klar benannt und beschrieben werden. Im sozial- und wirtschaftswissenschaftlichen Kontext können dies Länder, Personen oder Güter sein. Wenn man sich einen Datensatz in einem beliebigen Tabellenkalkulationsprogramm vorstellt, mit Variablen bzw. Untersuchungsgegenständen in Spalten und Beobachtungseinheiten in Zeilen (auch als »rechteckiges Format« bezeichnet), dann befindet sich die Information der kleinsten Ebene in der Regel in einer Zeile. Mehrere Beobachtungseinheiten können zusammengefasst werden, z.B. mehrere Personen zu einem Haushalt, mehrere Haushalte zu einem Bundesland etc. Zusätzlich wird häufig eine zeitliche Dimension berücksichtigt, beispielsweise können Datensätze von Personen, die über mehrere Jahre beobachtet werden, in der Einheit »Personen-Jahre« abgelegt werden.

### 3. Datengrundlage

Als Nächstes muss der potentielle Nutzer informiert werden, ob es sich bei den Daten um eine Vollerhebung oder um eine Stichprobe aus einer Grundgesamtheit handelt. Hierdurch erhält er im Idealfall direkt die Information darüber, welche Aussagen aufgrund der Daten überhaupt möglich sind. Beispielsweise kann es sich um eine Vollerhebung der Kinder einer Schule, deren Essgewohnheiten in der Mittagspause erfragt werden, handeln. Diese Daten können dann nicht herangezogen werden, um allgemein Aussagen über Essgewohnheiten von Kindern außerhalb dieser einen Schule zu treffen.

# Repositorien, die Daten aus eigenen Studien annehmen

Bei Stichproben ist eine Definition der Grundgesamtheit sowie die Frage, wie versucht wurde, die Stichprobe aus der Grundgesamtheit abzuleiten, essentiell. Man könnte z.B. Privathaushalte in Deutschland als Grundgesamtheit definieren. Dann aber ist es schwer, die Stichprobe zu realisieren, weil es in Deutschland (wie in den meisten Ländern) keine offizielle Liste aller Haushalte gibt, aus der man eine Stichprobe bilden könnte. Bei einer Stichprobe stellt sich deswegen immer die Frage, wie sie erhoben wurde. Handelt es sich um eine Zufallsstichprobe, um eine Quotenstichprobe oder um eine Ziehung ohne Vorgaben (vgl. auch Schnell, 2012)?

Die Art der Stichprobe hat wiederum Einfluss auf die Aussagekraft der Daten – und somit auch auf die Breite der Fragestellungen, für die eine Nachnutzung der Daten sinnvoll ist. Zur Einschätzung der Validität der Daten sind Angaben zum Prozess der Erhebung essentiell. So sollte z.B. dokumentiert werden, wie viele Einheiten (etwa Personen oder Betriebe) ursprünglich befragt werden sollten (»Bruttosample«) und wie viele letztendlich teilgenommen haben (»Nettosample«).

## 4. Erhebungsmethode

Daten können ganz unterschiedlich gewonnen werden und in verschiedenen Formen vorliegen. Dies genau darzulegen ist wichtig, um die Daten richtig interpretieren sowie deren Reliabilität (Messgenauigkeit) und Validität (Aussagekraft) einschätzen zu können. Beispielsweise lassen sich Zeitungsauschnitte zu einem Thema als Daten erfassen, Interviews mit Personen (die quantitativ oder qualitativ sein können) oder Suchanfragen auf Internetseiten können dabei eine Datengrundlage bilden. Insbesondere durch die fortschreitende Digitalisierung unseres Alltags lassen sich immer mehr Wege finden, an Daten zu kommen und diese zu wissenschaftlichen Zwecken zu nutzen. Umso wichtiger wird in diesem Zusammenhang die Dokumentation der Erhebungsmethode (für Standarderhebungsmethoden in persönlichen Interviews, siehe z. B. Schnell, 2012), so dass zusätzliche Informationen auch aus Fragebögen, Skalenhandbüchern, Testbeschreibungen, Kodierungsvorschriften, Übersetzungshilfen oder Anschreiben gezogen werden können – kurzum alles, was den Prozess der Datenerstellung für den Nutzer konkretisiert.

## 5. Umfang

Der Umfang der Daten ist wesentlich, wenn über den weiteren Gebrauch entschieden wird. Dabei geht es zum einen um die Anzahl an Beobachtungen: Wie viele Schüler wurden in einer Befragung erfasst? Wie viele Länder sind enthalten? Usw.

Wesentlich wichtiger ist aber, wie der in Punkt 1 angegebene Inhalt erfasst wird, also wie viele Variablen im Datensatz enthalten sind und was sie konkret messen. Hier kann eine veröffentlichte Aufsatz-Dokumentation, die den Lesern einen ersten Überblick geben soll,

### Öffentliche Forschungsdatenrepositorien in Deutschland

- [datorium](#) (von GESIS bereitgestellt)
- Datenservicezentrum für Betriebs- und Organisationsdaten ([DSZ-BO](#))
- Verbund Forschungsdaten [Bildung](#) (betrieben vom FDZ Bildung des DIPF)

### Einige englischsprachige öffentliche Repositorien im Ausland

- Australian Data Archive ([ADA](#)); Australien
- Danish Data Archive ([DDA](#)); Dänemark
- Data Archiving and Networked Services ([DANS](#)); Niederlande
- [DataHub](#) (Open Knowledge Foundation); Großbritannien
- Inter-university Consortium for Political and Social Research ([ICPSR](#)); USA
- UK Data Archive ([UKDA](#)); Großbritannien

### Universitäten mit Datenrepositorien in Deutschland

(Datenannahme beschränkt auf Universitätsangehörige)

- [LMU München](#)
- [TU Berlin](#)
- [Universität Bielefeld](#) (> [Data Policy](#))
- [Universität Heidelberg](#) (> [Data Policy](#))
- [Universität Mannheim](#) (> verfügt zudem über ein Forschungsdaten Service Center)

in der Regel nicht weit ins Detail gehen (die Vorstellungen der Datensätze in Schmollers Jahrbuch haben beispielsweise nur einen Umfang von bis zu 15 Seiten). Weiterführende Dokumentationen sind dann für die tatsächlichen Nutzer gedacht, die Genaueres über die Erhebung erfahren möchten, z. B. wie das Bruttosozialprodukt eines Landes erhoben wurde oder in welchem Format Jahresdaten abgelegt wurden. Hierfür ist die Erstellung eines so genannten Codebuches bzw. Datenhandbuchs sinnvoll. Ein Beispiel für ein sehr ausführliches Codebuch findet sich beim SOEP: »[Codebook: Household level questionnaires](#)«.

## 6. Zugang

Zu guter Letzt ist es wichtig, anzugeben, ob und wie ein Nachnutzer an die betreffenden Daten gelangen kann. Zunächst muss dabei ein Ansprechpartner oder eine Institution genannt werden, der oder die verantwortlich für den Zugang und den Vertrieb der Daten ist (falls dies vorgesehen ist). Die meisten Datensätze können nicht einfach öffentlich zur Verfügung gestellt werden, denn auch bei selbst erstellten Daten müssen datenschutzrechtliche Bestimmungen eingehalten werden. Beispielsweise müssen personenbezogene Daten für eine Weitergabe anonymisiert werden, d.h. sämtliche Möglichkeiten zur Identifikation von Individuen müssen entfernt werden (vgl. Statistisches

Bundesamt – Forschungsdatenzentrum, 2013). Dies bedeutet unter Umständen auch, bestimmte seltene Merkmale bzw. Kombinationen von Merkmalen zu entfernen. Erste Ansprechpartner für Fragen in diesem Zusammenhang sind die Datenschützer der jeweiligen Institution, die im Zweifel immer vor einer Studie mit selbst erhobenen Daten kontaktiert werden sollten.

Immer häufiger ist es möglich, Daten per Download bereitzustellen und dafür besondere Zertifikate auszugeben (in der Regel auf Basis eines Nutzungsvertrags). Dabei ist der Unterschied zwischen kommerziellen und wissenschaftlichen Nutzern, für die meist unterschiedliche Bedingungen gesetzt werden, zu beachten. Auch Kosten der Nachnutzung, die selbst bei grundsätzlich kostenfreien Daten allein durch den Versand entstehen können, sind zu benennen. Besonders im universitären Umfeld ist es wichtig, ob es eine Version der Daten für die Lehre gibt, die datenschutzrechtlich weniger sensibel ist (CAMPUS-File) und die ggf. für Studierende verbilligt oder vollständig kostenfrei abgegeben wird (z.B. per Downloadmöglichkeit).

Die Punkte 1 bis 6 geben einen Überblick, der den potentiellen Nachnutzern relativ schnell eine Entscheidung darüber erlaubt, ob die Daten für sie relevant sind. Je nach Art des Datensatzes gibt es dabei noch weitere Qualitätskriterien, die angegeben werden können. So kann (und sollte) zum Beispiel bei Umfragen der Anteil an fehlenden Antworten (»Item-Nonresponse«) oder der Anteil der Nicht-Teilnehmer (»Unit-Nonresponse«) dokumentiert werden. Je ausführlicher diese Dokumentation ist, desto wahrscheinlicher wird es, dass die Dokumentation und der Datensatz zitiert werden – nicht zuletzt dadurch, dass man den Nachnutzern in der Dokumentation auf die (bereits oben erwähnte) persistente Identifizierung (DOI) hinweist und einen Zitationsvorschlag für den Datensatz macht.

### **b) Selbst erstellte Datensätze anderen zur Verfügung stellen**

Es gibt in den Sozial- und Wirtschaftswissenschaften bereits einige Einrichtungen, die eine Infrastruktur bereitstellen, um Daten in so genannten Repositorien<sup>3</sup> einzupflegen und sie damit der wissenschaftlichen Öffentlichkeit zur Verfügung zu stellen. Beispielsweise werden im Folgenden das Repository »datorium« von GESIS, das Datenservicezentrum für Betriebs- und Organisationsdaten ([DSZ-BO](#)) der Universität Bielefeld sowie das Portal [Forschungsdaten Bildung](#) vorgestellt, die entsprechende Services bereitstellen. **datorium** ist ein von GESIS seit 2014 bereitgestelltes Repository für Forscher aus den Sozial- und Wirtschaftswissenschaften. Hier können Forschungsdaten von Forschern selbst eingestellt werden. Bestimmte Angaben sind Pflicht (z.B. Titel und Primärforscher), andere können zur weiteren Dokumentation gemacht werden (z.B. Software-Formate oder Fragebogeninhalte). datorium

<sup>3</sup> Repositorien sind an Universitäten oder Forschungseinrichtungen betriebene Dokumentenserver, auf denen wissenschaftliche Materialien archiviert und weltweit entgeltfrei zugänglich gemacht werden. Mehr Informationen zum Thema Repositorien gibt es [hier](#).

übernimmt für die Forscher die Funktion des Archivs – die eingestellten Daten werden mit einem DOI versehen und danach dauerhaft in der einmal eingestellten Form verfügbar gehalten. datorium macht die Daten öffentlich verfügbar, zitierfähig und stellt über eine standardisierte Maske zudem vergleichbare Metadaten zu verschiedenen Datensätzen zur Verfügung (vgl. Zenk-Möltgen & Linne, 2013).

Die Universität Bielefeld unterhält ein Datenservicezentrum für Betriebs- und Organisationsdaten ([DSZ-BO](#)), das Organisationsdaten Nachnutzern zur Verfügung stellt, nachdem der eigentliche Zweck (beispielsweise eine interne Mitarbeiterbefragung) erreicht wurde. Das DSZ-BO archiviert Daten verschiedenster Ursprünge, z.B. prozessproduzierte Daten oder Daten aus Fallstudien, die auch außerhalb von Deutschland erhoben worden sein können. Ein potentieller Datengeber muss seine Studie zunächst in ein Studienportal aufnehmen lassen, so dass andere Wissenschaftler auf diese Studie Bezug nehmen können. Im zweiten Schritt werden die Daten archiviert, wobei auch Metadaten (standardisierte Beschreibungen zum Datensatz) generiert werden. Die Daten können schließlich auch weitergegeben werden, wenn dies z.B. für Sekundäranalysen gewünscht wird. Für einen solchen Schritt werden durch das DSZ-BO zusätzlich Anonymisierungsmaßnahmen ergriffen (vgl. DSZ-BO, 2014).

Vornehmlich für qualitative Daten existiert das Portal **Forschungsdaten Bildung**. Beim Verbund Forschungsdaten Bildung handelt es sich um ein Gemeinschaftsprojekt von GESIS, IQB – Institut zur Qualitätsentwicklung im Bildungswesen – und DIPP - Deutsches Institut für Internationale Pädagogische Forschung. Im Auftrag des Bundesministeriums für Bildung und Forschung (BMBF) sollen sie die Forschungsdaten sichern, welche im BMBF-Rahmenprogramm zur Förderung der empirischen Bildungsforschung generiert wurden. Soweit möglich sollen diese Daten außerdem für die Nachnutzung verfügbar gemacht werden.

In der Forschungslandschaft gibt es mittlerweile immer mehr Universitäten, die Speichermöglichkeiten für Forschungsdaten anbieten, z.B. die LMU München (bisher noch keine Daten im sozial- oder wirtschaftswissenschaftlichen Bereich) oder die TU Berlin (im Sommer 2014 noch im Aufbau begriffen). Die Angebote ähneln den oben beschriebenen Repositorien, d.h. Wissenschaftler können ihre Forschungsdaten zur Verfügung stellen, allerdings ist meist die Zugehörigkeit zur Universität Voraussetzung. Die Daten sind dann – abhängig von der jeweiligen Rechtevergabe – auch für Wissenschaftler anderer Institutionen frei zugänglich.

Einen besonderen Service bietet die Website von [re3data.org](#): Die Betreiber dieser Seite haben sich zum Ziel gesetzt, sämtliche Repositorien für Forschungsdaten in einer Datenbank zu erfassen und Besuchern der Seite an einer Stelle verfügbar zu machen. Allerdings sei hier darauf verwiesen, dass zu diesen Repositorien auch die Einrichtungen gehören (wie z.B. Forschungsdatenzentren), die Primärdaten zur Verfügung stellen und somit keine »fremden« Daten in ihren Bestand aufnehmen (siehe auch Vlaeminck & Wagner, 2014).

# IMPRESSUM

## **Herausgeber**

ZBW – Leibniz-Informationszentrum Wirtschaft  
GESIS – Leibniz-Institut für Sozialwissenschaften  
RatSWD – Rat für Sozial- und Wirtschaftsdaten

## **Autorinnen und Autoren**

Die Verantwortung für die einzelnen Kapitel liegt bei den Autorinnen und Autoren:

### **ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg**

- Sven Vlaeminck ([s.vlaeminck@zbw.eu](mailto:s.vlaeminck@zbw.eu))
- Ralf Toepfer ([r.toepfer@zbw.eu](mailto:r.toepfer@zbw.eu))
- Prof. Dr. Klaus Tochtermann ([director@zbw.eu](mailto:director@zbw.eu))

### **GESIS – Leibniz-Institut für Sozialwissenschaften, Köln, Mannheim und Berlin**

- Brigitte Hausstein ([brigitte.hausstein@gesis.org](mailto:brigitte.hausstein@gesis.org))
- Prof. Dr. York Sure-Vetter ([praesident@gesis.org](mailto:praesident@gesis.org))

### **RatSWD – Rat für Sozial- und Wirtschaftsdaten, Berlin**

- Prof. Regina T. Riphahn, Ph.D. ([office@ratswd.de](mailto:office@ratswd.de))

### **DIW Berlin – Deutsches Institut für Wirtschaftsforschung, Berlin**

- Mathis Fräßdorf, Ph.D. ([mfraessdorf@diw.de](mailto:mfraessdorf@diw.de))
- Prof. Dr. Gert G. Wagner ([gagner@diw.de](mailto:gagner@diw.de))

## **Zitationsvorschlag (Printausgabe)**

ZBW, GESIS & RatSWD (2015). Auffinden, Zitieren, Dokumentieren: Forschungsdaten in den Sozial- und Wirtschaftswissenschaften, Hamburg.

## **Zitationsvorschlag (Digitalausgabe)**

ZBW, GESIS & RatSWD (2015). Auffinden, Zitieren, Dokumentieren: Forschungsdaten in den Sozial- und Wirtschaftswissenschaften, doi:10.4232/10.fisuzida2015.2

## **ISBN (Printausgabe)**

978-3-00-046925-1

## **DOI (Digitalausgabe)**

10.4232/10.fisuzida2015.2

## **Website**

<http://www.auffinden-zitieren-dokumentieren.de>

## **Lizenz**

Die Inhalte dieses Wegweisers sind unter der Creative-Commons-Lizenz [CC-BY-SA 3.0 DE](https://creativecommons.org/licenses/by-sa/3.0/de/) lizenziert.

## **Danksagungen**

Die Autorinnen und Autoren des Wegweisers sprechen folgenden Personen ihren Dank für hilfreiche inhaltliche Anregungen und Ergänzungen aus:

- Dr. Brigitte Preissl  
*ZBW – Leibniz-Informationszentrum Wirtschaft*
- Dr. Daniel Scheuregger  
*CEDEFOP, Thessaloniki, Griechenland*
- Heinz-Jürgen Bove  
*Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, Berlin*

Zudem danken wir den Mitgliedern des **Ständigen Ausschusses Forschungsdaten-Infrastruktur (FDI)** beim Rat für Sozial- und Wirtschaftsdaten (RatSWD) für viele hilfreiche Ergänzungen und Kommentare.

## **Darüber hinaus danken die Autorinnen und Autoren folgenden Mitarbeiterinnen und Mitarbeitern der ZBW – Leibniz-Informationszentrum Wirtschaft**

- Lisa-Kristin Herrmann und
- Claudia Sittner für das Redigieren des Wegweisers,
- Stefan Werner für die Gestaltung sowie
- Clemens Anders für die Umsetzung des Webauftritts.

Aufgrund der besseren Lesbarkeit wird in den Texten nur die männliche Form verwendet. Die weibliche Form ist selbstverständlich immer mit eingeschlossen.

## **Bildquelle Umschlag**

»The 3D visualization of overwhelming data«

Elif Ayiter/Alpha Auer / flickr.com

[https://secure.flickr.com/photos/alpha\\_auer/5068482201/](https://secure.flickr.com/photos/alpha_auer/5068482201/)

Nutzungslizenz:

<https://creativecommons.org/licenses/by-nc-nd/2.0/>

# LITERATUR-VERZEICHNIS

Allianz der deutschen Wissenschaftsorganisationen (2010). Grundsätze zum Umgang mit Forschungsdaten.

<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaetze/>

CODATA-ICSTI (2013). Task Group on Data Citation Standards and Practices: Out of Cite, Out of Mind. The Current State of Practice, Policy, and Technology for the Citation of Data. In: Data Science Journal (12), CIDCR1-CIDCR75.

<http://doi.org/10.2481/dsj.OSOM13-043>

DataCite Metadata Working Group (2013). DataCite Metadata Schema for the Publication and Citation of Research Data. DataCite.

<http://doi.org/10.5438/0008>

Datenservicezentrum Betriebs- und Organisationsdaten an der Universität Bielefeld: Informationen für DatengeberInnen. Servicestufen am DSZ-BO – Der Ablauf in einzelnen Schritten. DSZ-BO Info Nr. 1, 2014.

[http://www.uni-bielefeld.de/dsz-bo/pdf/DSZ-BO\\_Datengeber.pdf](http://www.uni-bielefeld.de/dsz-bo/pdf/DSZ-BO_Datengeber.pdf)

Deutsche Forschungsgemeinschaft (1998). Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“; Denkschrift der Deutschen Forschungsgemeinschaft.

[http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_0198.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf)

Deutsche Forschungsgemeinschaft (2013). Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Denkschrift; Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“ = Proposals for safeguarding good scientific practice: memorandum; recommendations of the Commission on Professional Self Regulation in Science. Wiley-VCH, Weinheim.

Gherghina, S. & Katsanidou, A. (2013). Data availability in political science journals. European Political Science (12), S. 333–349.

<http://dx.doi.org/10.1057/eps.2013.8>

Hausstein, B. (2012). Die Vergabe von DOI-Namen für Sozial- und Wirtschaftsdaten. Serviceleistungen der Registrierungsagentur da|ra. RatSWD Working Paper Series 193.

<http://doi.org/10.2139/ssrn.2008192>

Helbig, K. & Hausstein, B. (2014a). Best practice guide for the registration of resources with da|ra. GESIS - Technical Reports 2014/18.

<http://dx.doi.org/10.4232/10.bpg.1.o>

Helbig, K.; Hausstein, B.; Koch, U.; Meichsner, J. & Kempf, A. (2014b). da|ra Metadata Schema. Version 3.1. GESIS Technical Report 2014|7. GESIS – Leibniz-Institut für Sozialwissenschaften.

<http://doi.org/10.4232/10.mdsdoc.3.1>

International DOI Foundation (2014). DOI® Handbook.

<http://doi.org/10.1000/182>

Köhler, F. (2012). 10 Jahre Forschungsdatenzentren der Statistischen Ämter. Statistische Monatshefte Niedersachsen, 66(6).

<http://www.statistik.niedersachsen.de/download/68944>

Mutz, M. & Kämpfer, S. (2011). ...und nun zum Wetter: Beeinflusst die Wetterlage die Einschätzung von politischen und wirtschaftlichen Sachverhalten? Zeitschrift für Soziologie, 40(4), 208–226.

Schnell, R. (2012). Survey Interviews. Methoden standardisierter Befragungen. VS Verlag, Wiesbaden.

Statistisches Bundesamt – Forschungsdatenzentrum (2013). Konzept zur Anonymisierung der Mikrozensus 1973, 1982 und 1987 BRD zur Verwendung als Public-Use-Files (PUF).

[http://www.forschungsdatenzentrum.de/bestand/mikrozensus/puf/1987/fdz\\_mikrozensus\\_puf\\_1973\\_1982\\_1987\\_anonymkonzept.pdf](http://www.forschungsdatenzentrum.de/bestand/mikrozensus/puf/1987/fdz_mikrozensus_puf_1973_1982_1987_anonymkonzept.pdf)

Vlaeminck, S. & Siegert, O. (2012). Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften? Eine Analyse mit Fokus auf die Wirtschaftswissenschaften. RatSWD Working Paper Series 210.

<http://www.econstor.eu/dspace/handle/10419/75330>

Vlaeminck, S. & Wagner, G. (2014). On the role of research data centres in the management of publication-related research data. LIBER Quarterly, 23(4), 336-357.  
<http://liber.library.uu.nl/index.php/lq/article/view/9356>

Wagner, J. (2010). Zehn Jahre European Data Watch: Dokumentation von Datensätzen für die empirische Wirtschafts- und Sozialforschung und Zugangswege zu den Daten. AStA Wirtschafts- und Sozialstatistisches Archiv, 4(2), 141–149.  
<http://doi.org/10.1007/s11943-010-0079-6>

Winkler-Nees, S. (2012). Informationsinfrastrukturen für Forschungsdaten Ansätze und Strategien der DFG. Vortragsfolien von den Open Access Tagen 2012 in Wien.  
[http://open-access.net/fileadmin/downloads/OAT12/OAT12\\_Winkler-Nees.pdf](http://open-access.net/fileadmin/downloads/OAT12/OAT12_Winkler-Nees.pdf)

Zenk-Möltgen, W. & Linne, M. (2013). datorium - ein neuer Service für Archivierung und Zugang zu sozialwissenschaftlichen Forschungsdaten. In: Beiträge des Workshops „Digitale Langzeitarchivierung“ auf der Informatik 2013, 14-22.  
<http://files.d-nb.de/nestor/edition/Sonderheft1-Informatik2013.pdf>

## Internet

Allgemeine Bevölkerungsumfragen Sozialwissenschaften (ALLBUS):  
<http://www.gesis.org/allbus>

Australian National Data Service (ANDS):  
<http://ands.org.au>

California Digital Library (CDL) – ARK:  
<http://www.cdlib.org/inside/digilib/ark/>

CODATA-ICSTI:  
<http://www.codata.org/task-groups/data-citation-standards-and-practices>

Community Innovation Survey:  
[http://ec.europa.eu/eurostat/web/microdata/community\\_innovation\\_survey](http://ec.europa.eu/eurostat/web/microdata/community_innovation_survey)

Corporation for National Research Initiatives (CNRI):  
<http://www.cnri.reston.va.us>

da|ra:  
<http://www.da-ra.de>

DataCite:  
<http://www.datacite.org>

Data-PASS: Citation:  
<http://www.data-pass.org/citations.html>

datorium, GESIS Datenarchiv für Sozialwissenschaften:  
<https://datorium.gesis.org/>

Digital Object Identifier (DOI) System:  
<http://www.doi.org/>

European Data Watch Extended (EDaWaX):  
<http://www.edawax.de>

FORCE11:  
<https://www.force11.org>

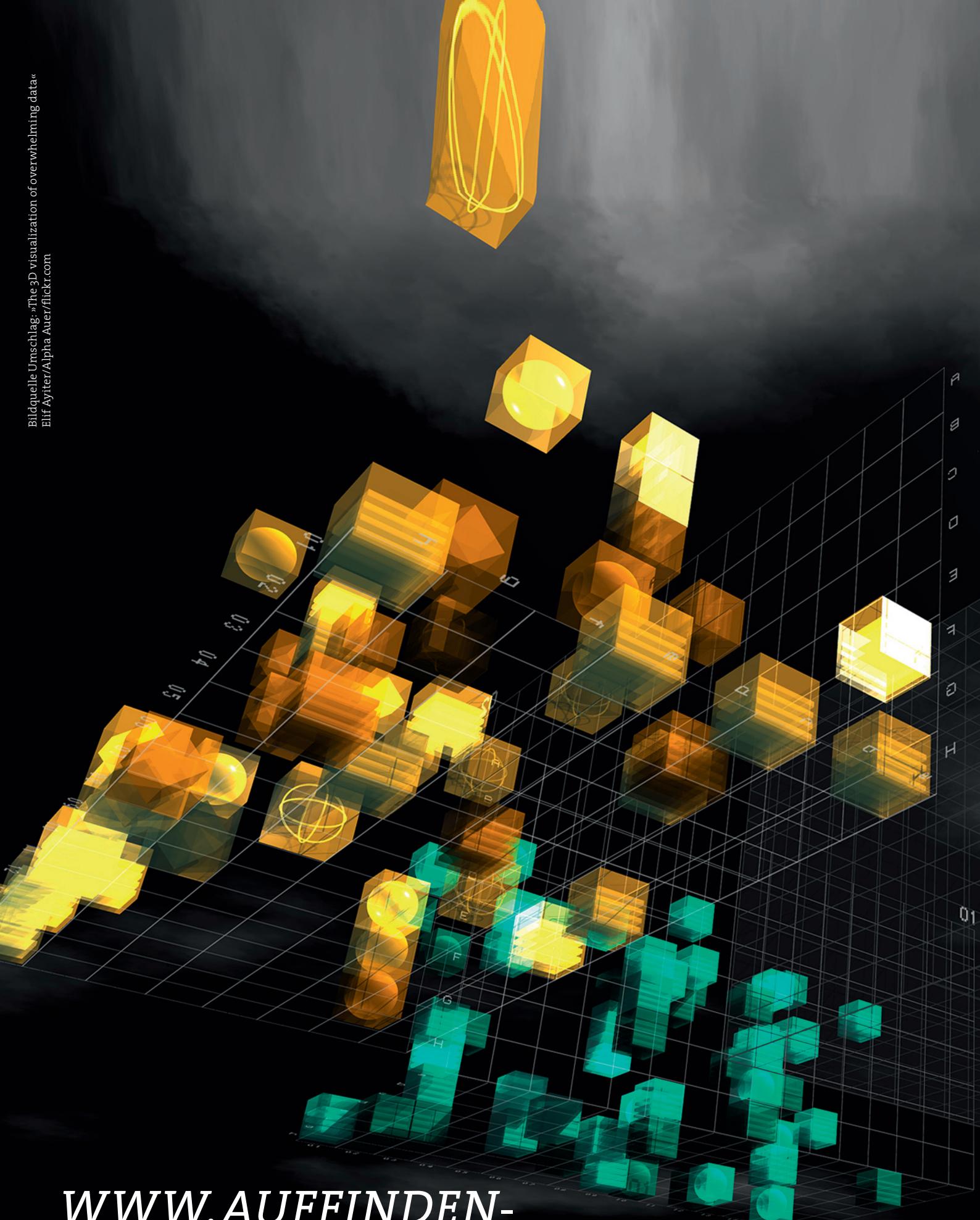
German Longitudinal Election Study / Nationale Wahlstudie ( GLES):  
<http://www.gles.eu/>

GESIS - Leibniz Institut für Sozialwissenschaften:  
<http://www.gesis.org>

Gruppenbezogene Menschenfeindlichkeit (GMF):  
<http://www.uni-bielefeld.de/ikg/projekte/GMF/>

ICPSR Data Citation:  
<https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/citations.html>

- Ifo-Geschäfts klimaindex:  
<http://www.cesifo-group.de/de/ifoHome/facts/Survey-Results/Business-Climate.html>
- Institute for Quantitative Social Science (IQSS) / The Data-verse network: Data Citation:  
<http://best-practices.dataverse.org/data-citation/>
- International Association of Scientific, Technical & Medical Publishers (STM):  
<http://www.stm-assoc.org/>
- Library of Congress: Structure of the LC Control Number:  
[http://www.loc.gov/marc/lccn\\_structure.html](http://www.loc.gov/marc/lccn_structure.html)
- Life Sciences Identifiers (LSID):  
<http://sourceforge.net/projects/lsids/>
- Mannheimer-Innovationspanel:  
<http://www.zew.de/de/publikationen/innovationserhebungen/innovationserhebungen.php3>
- National Educational Panel Study / Nationales Bildungspanel (NEPS):  
<https://www.neps-data.de/>
- Panel Analysis of Intimate Relationships and Family Dynamics / Das Beziehungs- und Familienpanel (PAIRFAM):  
<http://www.pairfam.de/>
- Persistent Uniform Resource Locator (PURL):  
<http://purl.oclc.org/docs/index.html>
- Programme for International Student Assessment (PISA):  
<http://www.oecd.org/pisa/>
- Programme for the International Assessment of Adult Competencies (PIAAC):  
<http://www.gesis.org/piaac>
- Rat für Sozial- und Wirtschaftsdaten (RatSWD):  
<http://www.ratswd.de>
- re3data.org:  
<http://www.re3data.org/>
- Sozio-oekonomisches Panel (SOEP):  
<http://www.soep.de>
- Special Interest Group on Data Citation (IASSIST-SIGDC):  
<http://iassistdata.org/community/sigdc>
- Survey of Health, Ageing and Retirement in Europe (SHARE):  
<http://www.share-project.org/>
- The Handle System:  
<http://www.handle.net>
- Thomson Reuters Data Citation Index:  
[http://wokinfo.com/products\\_tools/multidisciplinary/dci/](http://wokinfo.com/products_tools/multidisciplinary/dci/)
- Uniform Resource Name (URN) Syntax:  
<http://tools.ietf.org/html/rfc2141>
- ZBW - Leibniz-Informationszentrum Wirtschaft:  
<http://www.zbw.eu>
- ZEW-Hightech-Gründungspanel:  
<http://www.zew.de/de/projekte/631>



**WWW.AUFFINDEN-  
ZITIEREN-DOKUMENTIEREN.DE**