

Daniel, Andreas et al.

**Working Paper**

## Die Erfassung von Publikationen aus der Datennutzung: Verfahren, Herausforderungen und Nutzen. Ein Erfahrungsbericht von Forschungsdatenzentren

RatSWD Working Paper, No. 281

**Provided in Cooperation with:**

German Data Forum (RatSWD)

*Suggested Citation:* Daniel, Andreas et al. (2023) : Die Erfassung von Publikationen aus der Datennutzung: Verfahren, Herausforderungen und Nutzen. Ein Erfahrungsbericht von Forschungsdatenzentren, RatSWD Working Paper, No. 281, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin, <https://doi.org/10.17620/02671.80>

This Version is available at:

<https://hdl.handle.net/10419/283927>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

# RatSWD Working Paper Series

RatSWD.  
Rat für Sozial- und  
WirtschaftsDaten

281

## Die Erfassung von Publikationen aus der Datennutzung: Verfahren, Herausforderungen und Nutzen.

Ein Erfahrungsbericht von  
Forschungsdatenzentren

Andreas Daniel, Stefan Jakowatz, Nadeshda Jung,  
Lydia Kleine, Aleksander Kocaj, Alexia Meyermann,  
Kati Mozygemba, Alexander Schuster

April 2023

[www.ratswd.de](http://www.ratswd.de)

# Working Paper Series

## des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

---

Die *RatSWD Working Papers*-Reihe startete Ende 2007. In dieser Online-Publikationsreihe werden konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

*RatSWD Working Papers* sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autorinnen bzw. Autoren dar und nicht die des RatSWD. Die Zuwendungsgeber des RatSWD haben die Publikationen nicht beeinflusst.

**Herausgeberin oder Herausgeber der RatSWD Working Papers-Reihe ist die/der Vorsitzende des RatSWD:**

seit 2020 Monika Jungbauer-Gans

2014–2020 Regina T. Riphahn

2009–2014 Gert G. Wagner

2007–2008 Heike Solga

# Die Erfassung von Publikationen aus der Datennutzung: Verfahren, Herausforderungen und Nutzen.

Ein Erfahrungsbericht von Forschungsdatenzentren

Andreas Daniel<sup>1\*</sup>, Stefan Jakowatz<sup>2</sup>, Nadeshda Jung<sup>3</sup>, Lydia Kleine<sup>4</sup>,  
Aleksander Kocaj<sup>5</sup>, Alexia Meyermann<sup>6</sup>, Kati Mozygamba<sup>7</sup>,  
Alexander Schuster<sup>6</sup> (alphabetisch)

**April 2023**

<https://doi.org/10.17620/02671.80>

<sup>1</sup> Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)

<sup>2</sup> Leibniz-Institut für Sozialwissenschaften (GESIS)

<sup>3</sup> Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF)

<sup>4</sup> Leibniz-Institut für Bildungsverläufe (LifBi)

<sup>5</sup> Institut zur Qualitätsentwicklung im Bildungswesen (IQB) & Zentrum für Internationale  
Bildungsvergleichsstudien (ZIB)

<sup>6</sup> Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF)

<sup>7</sup> Forschungsdatenzentrum Qualiservice, Universität Bremen

\* **Kontakt:** Dr. Andreas Daniel | [daniel@dzhw.eu](mailto:daniel@dzhw.eu)

Abstract .....	3
1. Einleitung .....	4
2. Wozu? - Funktionen der Erfassung von Publikationen .....	5
3. Wie? - Verfahren zur Erfassung von Publikationen .....	7
3.1. Publikationsmeldungen durch Datennutzer:innen .....	9
3.2. Maßnahmen zur Erhöhung der Meldungsneigung .....	11
3.3. Automatische Recherche von Publikationen.....	13
3.4. Manuelle Recherche von Publikationen.....	15
3.5. Maßnahmen zur Förderung korrekter Datenzitationen.....	16
4. Erkenntnisse, Diskussion und Ausblick .....	18
Literaturverzeichnis .....	23

## Abstract

Publikationen sind im Wissenschaftssystem das wesentliche Medium zur Kommunikation von Erkenntnissen und werden häufig als Indikator für wissenschaftliche Leistung herangezogen. Für Forschungsdatenzentren (FDZ) erfüllen Publikationen weitere Funktionen: Verknüpft mit den zugrundeliegenden Forschungsdaten dienen sie als wichtige Informationsquelle für potentielle Datennutzer:innen. Zudem wird die Anzahl der auf Basis von Nachnutzungen entstandenen Publikationen als ein Evaluationskriterium für die Datendienste der FDZ herangezogen und das, obwohl die Erfassung dieser Publikationen mit einer Reihe von Herausforderungen verbunden ist. Der vorliegende Artikel beschreibt die verschiedenen Vorgehensweisen von FDZ bei der Erfassung von Publikationen, die aus der Nutzung von Forschungsdaten entstehen, und diskutiert damit einhergehende Probleme der Fehl-, Über- und Untererfassung. Es zeigt sich, dass die verschiedenen Verfahren der Erfassung mit je unterschiedlichen Schwierigkeiten behaftet sind, die ihre Verwendung als Leistungsindikator relativieren. Einen Königsweg gibt es derzeit nicht; stattdessen wird eine Kombination aus unterschiedlichen Vorgehensweisen empfohlen, die auch in Abhängigkeit von der jeweiligen Größe des Datenbestands eines FDZ und den jeweiligen Nutzer:innen zu bewerten ist.

**Keywords:** Publikationserfassung, Datenangebot, Datennutzung, Leistungsindikator, Forschungsdatenzentren, Nachnutzung, Sekundäranalyse, Evaluation, Bibliographie, Forschungsdaten

# 1. Einleitung

Publikationen sind im Wissenschaftssystem das wesentliche Medium, über das Ergebnisse des wissenschaftlichen Arbeitens kommuniziert werden. Anzahl und Art der Publikationen sind auch als Indikator für wissenschaftliche Leistung relevant. Für Forschungsdatenzentren (FDZ) ist es sinnvoll, Publikationen zu erfassen, die auf Basis der durch sie bereitgestellten Forschungsdaten<sup>1</sup> entstanden sind, und sie mit diesen zu verknüpfen (sog. Publikations-Daten-Relationen).<sup>2</sup> Damit verbindet sich für Forschende der Mehrwert, besser beurteilen zu können, ob ein Datensatz zur eigenen Forschungsfrage passt (data fit) und mit welchen Erkenntnisgewinnen bestimmte Forschungsdaten analysiert wurden. Zudem erleichtert die Zuordnung von Publikationen zu Daten einerseits deren Auffindbarkeit, andererseits fördert die Verlinkung von Daten (z. B. durch permanente Identifikatoren) in Publikationen deren Zitationshäufigkeit (Colavizza et al., 2020) und ist damit im Sinne der FAIR-Prinzipien<sup>3</sup>. In einigen FDZ wird die Anzahl an Publikationen, die auf Basis der angebotenen Forschungsdaten erstellt wurden, auch als Leistungsindikator für Evaluationszwecke verwendet. Darüber hinaus lassen sich aus der Analyse entsprechender Publikationen in gewissem Ausmaß Anhaltspunkte zum Nutzungsverhalten ableiten, die in Bezug auf das Datenangebot bzw. die FDZ-Infrastruktur insgesamt Veränderungsbedarfe aufzeigen können und so für die interne Steuerung bedeutsam sind. So könnte z. B. das Angebot eines bestimmten Datentyps ausgebaut werden, wenn sich zeigt, dass er ein hohes Nachnutzungspotential bietet. Aus den vielfältigen Verwendungsmöglichkeiten resultiert das Interesse von FDZ, Publikations-Daten-Relationen möglichst vollständig zu erfassen.

Dabei sind die Erfassung von Publikationen und die Zuordnung zu den darin verwendeten Forschungsdaten mit einer Reihe von Herausforderungen verbunden: Um entsprechende Publikationslisten erstellen zu können, verpflichten manche FDZ ihre Nutzer:innen über die jeweiligen Nutzungsbedingungen dazu, in Publikationen die genutzten Daten gemäß Vorgabe zu zitieren und die Publikationen dem FDZ zu melden (siehe auch Pegelow et al., 2020). Dieser Vereinbarung kommen Nutzende jedoch in der Praxis häufig nicht oder nur unzureichend nach, was in zweifacher Hinsicht zu Problemen führt: Erstens sind die Publikationslisten zu den Daten durch ausbleibende Meldungen unvollständig und zweitens erschwert die fehlende oder unzureichende Zitation der Daten in den Publikationen die Recherche durch die FDZ selbst. FDZ haben daher weitere Wege implementiert und erprobt, um Publikationen zu erfassen.

Über diese Ansätze und die damit einhergehenden Erfahrungen sowie Vor- und Nachteile wird im Folgenden berichtet. Abschließend wird die Interpretierbarkeit der Anzahl von

---

<sup>1</sup> Unter bereitgestellten Forschungsdaten werden in diesem Artikel alle Datenbestände verstanden, die durch FDZ zur Nachnutzung zur Verfügung gestellt werden.

<sup>2</sup> Im vorliegenden Beitrag werden vor allem Erfahrungen aus Forschungsdatenzentren berichtet. Hierfür werden beispielhaft Prozesse einzelner beteiligter FDZ herausgegriffen. Die beschriebenen Erfahrungen und Empfehlungen sind jedoch grundsätzlich auch allgemein auf Einrichtungen der Forschungsdateninfrastruktur übertragbar.

<sup>3</sup> Das Akronym FAIR steht für Findable, Accessible, Interoperable, Reusable. Diese Prinzipien sollen eine offene Zugänglichkeit von Daten und Metadaten fördern. Siehe hierzu auch <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>

Publikationen als Indikator für die Bewertung der Datennachnutzung kritisch diskutiert und mögliche Alternativen aufgezeigt. Die an diesem Artikel beteiligten Autor:innen vertreten verschiedene FDZ, die im Rahmen des Verbunds Forschungsdaten Bildung (VerbundFDB)<sup>4</sup> kooperieren, um die Forschungsdateninfrastruktur für die Bildungsforschung gemeinsam weiter zu entwickeln (siehe Kasten).

### **VerbundFDB**

Der Verbund Forschungsdaten Bildung (VerbundFDB) ist ein Zusammenschluss von Forschungsdatenzentren, die gemeinsam wissenschaftliche Infrastrukturdienstleistungen für die empirische Bildungsforschung bereitstellen. Der VerbundFDB betreibt das Onlineportal [www.forschungsdaten-bildung.de](http://www.forschungsdaten-bildung.de), das Forschungsdaten standortübergreifend nachweist, sowie ein Angebot zur Meldung von Studien und Übermittlung von Daten beinhaltet. Forscher:innen haben im Onlineportal die Möglichkeit, ihre Daten abzugeben und über eine spezifische Suche, Daten unabhängig von ihrem jeweiligen Archivierungsort zu finden. Ein zweiter Schwerpunkt der Aktivitäten des VerbundFDB sind Schulungs- und Beratungsangebote sowie die Bereitstellung von Informationsmaterialien zum Forschungsdatenmanagement.

## **2. Wozu? - Funktionen der Erfassung von Publikationen**

Die Kernaufgabe von FDZ besteht in der Bereitstellung von Datenservices für die Forschung. Wesentlicher Teil dieser Dienste ist es, von Forscher:innen erstellte Datenbestände für die wissenschaftliche Nachnutzung bereitzustellen. Neben den Forschungsdaten selbst bieten FDZ mehr oder weniger differenzierte Kontextinformationen an, welche den Datenbestand in seinen Entstehungs- oder Verwendungskontext einbetten. Publikationen, die aus der Nutzung von Forschungsdaten entstanden sind, gehören zu diesen Kontextinformationen. Sie informieren nicht nur darüber, welche Forschungsfragen bereits mit den Daten bearbeitet wurden, sie zeigen darüber hinaus auch die Potentiale und Grenzen auf. Ebenso erleichtert eine direkte Verknüpfung eines Datenbestands und der zugehörigen Publikation die Reproduzierbarkeit und Replikation der dargestellten Forschungsergebnisse, die oft bei der Verwendung quantitativer Datensätze von Bedeutung ist. Es wird deutlich, dass die Verknüpfung von Publikationen mit den dazugehörigen Forschungsdaten für Forschende, die nach Daten suchen und diese verwenden, einen wichtigen Mehrwert bietet.

Publikationen kontextualisieren Forschungsdaten aber nicht nur, sondern sie erleichtern auch deren Auffindbarkeit, da die Suche nach forschungsrelevanten Daten eng mit der Recherche nach einschlägigen Publikationen verbunden ist. Die Bedeutung von Publikationen als Quelle für das Auffinden von nachnutzbaren Forschungsdaten zeigen auch Krämer et al. (2021). Insgesamt lässt sich eine wechselseitige Wirkung festhalten, da datenbezogene Publikationen von Wissenschaftler:innen bzw. in Journals mit hoher Reputation die Sichtbarkeit der genutzten Datenbestände erhöhen und umgekehrt die Erfassung und Listung von Publikationen zu

---

<sup>4</sup> [www.forschungsdaten-bildung.de](http://www.forschungsdaten-bildung.de)

(bekannten) Datenbeständen durch die FDZ zu einer erweiterten Sichtbarkeit der Forschungstätigkeit der entsprechenden Autor:innen beitragen.

Manche FDZ bieten auch den Service, Datengeber:innen über die Nachnutzung „ihrer“ Forschungsdaten und entsprechende Publikationen zu informieren und diese zur Verfügung zu stellen. Für diese Serviceleistung ist die Erfassung von datenbezogenen Publikationen eine notwendige Voraussetzung. Ein weiterer Aspekt besteht in dem Interesse der Datenproduzent:innen, informiert zu werden, wozu die eigenen, bei einem FDZ archivierten Datenbestände durch andere Forschende genutzt werden und welche weiteren Erkenntnisse mit diesen Daten gewonnen wurden.

Daneben liefern Angaben zur Nutzung von Datenbeständen Informationen, die von Relevanz für weitere (wissenschaftspolitische) Akteure und für die Evaluation der Arbeit eines FDZ sein können: *Externe Stakeholder* wie Beiräte oder Förderer sind interessiert an Informationen zum Output und zum Impact von infrastrukturellen Diensten. Hierfür können Publikationen ein wichtiger Indikator sein, der darüber informiert, ob die Datennutzung im Hinblick auf einen Erkenntniszuwachs in der wissenschaftlichen Community „produktiv“ war, was zu einem messbaren Output – einer wissenschaftlichen Publikation – geführt hat. Informationen zur Datennutzung besitzen – als Teil des Monitorings – auch Relevanz für die *interne Steuerung* und Qualitätssicherung (bspw. von großen Erhebungsreihen). Dabei ist zu beachten, dass Publikationsmeldungen hinsichtlich ihrer Eignung als Leistungskriterium Einschränkungen unterliegen, wie in den folgenden Kapiteln diskutiert wird.<sup>5 6</sup>

Schließlich steht die Erfassung von Publikationen auch im Dienst der guten wissenschaftlichen Praxis und stellt eine Datengrundlage für Metaforschung zur Belastbarkeit von Forschungsergebnissen dar. In den letzten Jahren wurden in der Psychologie und weiteren Disziplinen, deren Erkenntnisse auf quantitativen Auswertungsmethoden basieren, vermehrt Studien veröffentlicht, die die Belastbarkeit von Ergebnissen empirischer Studien in Frage stellen (z. B. Simmons et al., 2011; Ulrich et al., 2016). Als mögliche Gründe wurden u. a. die mangelnde Transparenz im Forschungsprozess und geringere Anreize im Wissenschaftssystem, alle Schritte im Forschungsprozess (bspw. Datengrundlage, Auswertungsverfahren) öffentlich zugänglich und reproduzierbar zu machen, diskutiert. Die transparente und nachvollziehbare Darstellung des Forschungsprozesses gewinnt auch in der empirischen Bildungsforschung immer mehr an Bedeutung (LeBeau et al., 2021). Durch die Erfassung von Publikationen, die aus Datennutzungsanträgen entstehen, werden Analysen zur Reproduzierbarkeit und Belastbarkeit von Forschungsergebnissen erleichtert, indem die Forschungsergebnisse mit den zugrundeliegenden Daten durch Reanalysen überprüft werden

---

<sup>5</sup> Grundsätzlich können verschiedene Indikatoren für die Messung der Datennutzung herangezogen werden. Neben den Publikationen mit den angebotenen Datensätzen, unterscheidet der RatSWD in seinem Tätigkeitsbericht die Anzahl (interner und externer) Datennutzenden und die Anzahl der Downloads von frei verfügbaren Datensätzen.

<sup>6</sup> Siehe allgemein zur Leistungsmessung in der außeruniversitären Forschung Pfeffer (2016).



können. Dadurch wird der Forschungsprozess im Sinne der guten wissenschaftlichen Praxis transparenter und evaluierbar gemacht.

Darüber hinaus ermöglicht die Erfassung von Publikationen Metaforschung zum sog. Publication Bias. Publication Bias bezeichnet die Tendenz, Ergebnisse in Abhängigkeit von ihrer Richtung (z. B. Ablehnung der Nullhypothese) und ihres Signifikanzniveaus (z. B.  $p < .05$ ) eher zu veröffentlichen (Dickersin, 1990). Demzufolge könnten sich Forscher:innen entscheiden, Ergebnisse ihrer Sekundäranalyse nicht zu publizieren, wenn sich ihre Ausgangshypothesen im Zuge der Datenanalyse nicht bestätigen. Auf der anderen Seite kann der Publication Bias auch bei Fachzeitschriften dazu führen, dass positive Ergebnisse bevorzugt veröffentlicht werden (Franco et al., 2014). Für eine direktere Abschätzung des Publication Bias können Datennutzungsanträge und daraus entstandene Publikationen in Beziehung gesetzt und um Fragen zum Nichtzustandekommen einer Publikation ergänzt werden.

#### **Übersicht verschiedener Funktionen der Publikationserfassung**

Das Ausmaß der Datennutzung, das über Publikationen repräsentiert wird, ist ein wichtiger Indikator für:

- (1) Forschungsaktivitäten; ein Service für die Forscher:innen, die hierdurch erfahren, welche Forschungsarbeiten mit bestimmten Datenbeständen durchgeführt und welche Forschungsergebnisse erzielt wurden (Informationsfunktion);
- (2) den Output oder Impact von Datennutzung; ein Service für externe Stakeholder wie Beiräte von FDZ (Evaluationsfunktion);
- (3) das Monitoring und die Qualitätssicherung von FDZ (v. a. für die interne Steuerung relevant) (Steuerungsfunktion).
- (4) Es dient auch als Datengrundlage für Metaforschung zur Belastbarkeit von Forschungsergebnissen (bspw. für Analysen zu Publication Bias und Reproduzierbarkeit von Forschungsergebnissen) (Begutachtungsfunktion).

### **3. Wie? - Verfahren zur Erfassung von Publikationen**

Bei den hier betrachteten Publikationen handelt es sich um Veröffentlichungen, in denen empirische Forschungsergebnisse aus Analysen bereits vorliegender Datenbestände der Fachöffentlichkeit berichtet werden. Diese Publikationen können sowohl von Sekundärnutzer:innen als auch von den Primärforscher:innen bzw. Datengeber:innen verfasst worden sein. In der Regel werden verschiedene Publikationstypen berücksichtigt: Zeitschriftenartikel (teilweise peer-reviewed), Monographien (meist inkl. Sammelbände), Sammelwerkbeiträge, Qualifikationsarbeiten (Promotions- und Habilitationsschriften, Master- und Bachelorarbeiten), graue Literatur (z. B. Working Paper, Technical Reports) etc. Welche Publikationstypen erfasst werden, unterscheidet sich zwischen den FDZ. Nicht alle FDZ berücksichtigen z. B. Bachelor- oder Masterarbeiten, Preprints oder Artikel in Konferenzbänden. Die Vorgehensweisen der FDZ, Publikationen zu erfassen, unterscheiden

sich zudem darin, ob Publikationen von den Datengebenden/Primärforschenden berücksichtigt, d. h. gezählt werden oder nicht.

Festzuhalten ist, dass die Unterscheidung zwischen Publikationen von Primärforscher:innen bzw. Datenproduzent:innen auf der einen und Sekundärnutzer:innen auf der anderen Seite in der Praxis schwierig bis unmöglich ist: Primärforscher:innen publizieren auch nach Ende der Projektlaufzeit sowie zu anderen als den im Primärforschungsprojekt vorgesehenen Zwecken und auch mit anderen Forschenden gemeinsam als Co-Autor:innen. Insofern sind Publikationen als ein Indikator für die Nutzung eines Datenbestands zu verstehen und nicht allein als ein Indikator für die Sekundär- oder Nachnutzung i. e. S. durch nicht an der Erhebung beteiligte Dritte.

Eine besondere Schwierigkeit ergibt sich, wenn Forscher:innen des Primärerhebungsprojektes bzw. die Datengeber:innen mit eigenen Datensatzversionen arbeiten, die von den über FDZ bereitgestellten Versionen abweichen. Die Publikationen beruhen dann auf Daten(satzversionen), die nur intern verfügbar sind, keine DOI haben und sich von den Scientific Use Files (SUF) unterscheiden, die über das FDZ bereitgestellt werden.

Die Erfassung von Publikationen kann in unterschiedlicher Granularität bzw. Informationstiefe erfolgen; je nachdem, welche Zielsetzungen das FDZ mit der Erfassung verfolgt. Die Bandbreite reicht von einer einfachen Zählung der Publikationen über vollständige bibliographische Angaben mit und ohne Datenzitation bis hin zu Literaturdatenbanken mit durchsuchbaren Metadaten (mit und ohne Datenverlinkungen) (siehe Tabelle 1). Je nach Informationsbedürfnis fällt die angestrebte Informationstiefe und Metadatenqualität unterschiedlich aus. Für das Berichtswesen und die *Evaluation* von FDZ kann es ggf. ausreichend sein, die Anzahl der Publikationen zu erfassen (so z. B. für den RatSWD-Tätigkeitsbericht<sup>7</sup>, der Leibniz-Gemeinschaft oder für Forschungsbeiräte von Umfrageprogrammen wie z. B. beim ALLBUS<sup>8</sup>). Sind die Publikationen Teil des *Informationsangebots* für Forschende, ist es wichtig, vollständige bibliographische Angaben zu den Publikationen bereitzustellen, damit diese eindeutig zu recherchieren sind – ggf. erweitert um eine Verlinkung der entsprechenden Nachweise zu spezifischen Literaturdatenbanken (ggf. inkl. Volltextzugriff). Eine solche wechselseitige Verlinkung von Daten- und Publikationsnachweisen gibt es bspw. zwischen forschungsdaten-bildung.de und der Literaturdatenbank auf fachportal-paedagogik.de. Die Verlinkung vom Datenbestand zum Publikationsnachweis im Fachportal Pädagogik erleichtert es, Zugriff auf zugehörige Publikationen zu erhalten, da - wo möglich - Volltexte (z. B. über das Open Access Repository pedocs) direkt eingebunden oder Informationen dazu geliefert werden, wo Publikationen verfügbar sind. Umgekehrt erleichtert die Verlinkung von Datenbeständen in der Literaturdatenbank nachzuvollziehen, auf welcher Datenbasis die Publikation beruht.

---

<sup>7</sup> [www.konsortswd.de/aktuelles/publikationen/taetigkeitsberichte/](http://www.konsortswd.de/aktuelles/publikationen/taetigkeitsberichte/)

<sup>8</sup> [www.gesis.org/allbus/allbus](http://www.gesis.org/allbus/allbus)

UMFANG DER PUBLIKATIONSERFASSUNG	ANWENDUNGSBEISPIELE	FUNKTION
Anzahl	Berichtswesen, z. B. RatSWD-Tätigkeitsbericht	Evaluation
Bibliographische Angaben	Publikationsliste, Bibliographie, z. B. FDZ-Websites oder forschungsdaten- bildung.de	Information
Durchsuchbare Metadaten	Literatur- oder Forschungsdatenbanken, z. B. fachportal-paedagogik.de oder GESIS.org	Information

Tabelle 1: Unterschiedliche Granularität bei der Darstellung von Publikationen

FDZ wenden üblicherweise verschiedene Wege zur Erfassung der Publikationen an, auf die wir im Folgenden näher eingehen möchten.

### 3.1. Publikationsmeldungen durch Datennutzer:innen

Ein bereits angesprochener Weg, Publikationen zu erfassen, ist es, Nutzer:innen im Rahmen von Nutzungsverträgen zu verpflichten, die auf Basis der verwendeten Daten erzeugten Publikationen an das FDZ zu melden (siehe auch Pegelow et al., 2020, S. 19). In der Benutzungsordnung von GESIS - Leibniz-Institut für Sozialwissenschaften heißt es beispielsweise: *„Der Benutzer verpflichtet sich, alle benutzten Unterlagen den wissenschaftlichen Gepflogenheiten entsprechend zu zitieren und dem Datenarchiv von seiner Publikation zwei Belegexemplare zu übereignen.“*<sup>9</sup>

In der Praxis kommen Forschende dieser Verpflichtung jedoch eher selten nach. In einer am Leibniz-Institut für Bildungsverläufe (LIfBi) durchgeführten Studie (Kleine, 2021) mit ca. 200 Datennutzer:innen der National Educational Panel Study NEPS (siehe Kasten) zeigte sich, dass nur etwa jede zweite Publikation selbständig an das FDZ gemeldet wurde. Für die Erfassung von Publikationen durch Eigenmeldung lässt sich damit konstatieren, dass dieser Weg durch den geringen Meldeanteil eine beträchtliche Untererfassung mit sich bringt. Auch andere, der an diesem Artikel beteiligten FDZ gehen von einer erheblichen Untererfassung aus.

<sup>9</sup> verfügbar unter:  
[www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/\\_bgordnung\\_bestellen/Benutzungsordnung.pdf](http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/Benutzungsordnung.pdf);  
 zugegriffen am 17.03.2022

**Studie "Publikationserfassung, Datenzitation und Nutzungsmonitoring" (PEN), 2020-2021:**

Der Schwerpunkt der Studie lag in der Ermittlung des Meldeanteils von Publikationen über Google Alerts sowie durch Eigenmeldungen der Autor:innen. Ausgehend von einer Zufallsstichprobe von 200 NEPS-Datennutzenden wurde eine Liste bereits bekannter Publikationen erstellt. In dieser Liste wurden 92 Publikationen identifiziert, die von Personen ohne Projektanbindung verfasst wurden und in die Kategorie Monografie, Einzelbeitrag in Sammelwerken und Zeitschriftenaufsatz fallen. Das Rechercheverfahren über Google Scholar erfolgte über einen Suchstring, welcher die DOI aller NEPS-Datenprodukte umfasst. Die Trefferliste wurde mit der bekannten Publikationsliste abgeglichen. Als zweite Informationsquelle fungierten die Eigenmeldungen durch die Autor:innen. Schließlich wurden im Rahmen einer Zitationsrecherche alle NEPS-Publikationen auf die Bestandteile einer korrekten Datenzitation überprüft. Die resultierenden Daten wurden deskriptiv ausgewertet (Kleine, 2021).

Warum melden Nutzer:innen ihre Publikationen nicht? Aufklärung verspricht ein Blick auf die Ursachen von (ausbleibenden) Publikationsmeldungen. Das FDZ am Institut zur Qualitätsentwicklung im Bildungswesen (IQB) und das FDZ am LIfBi haben im Jahr 2021 ihre Nutzer:innen dazu befragt (n=387, IQB; n=348, LIfBi). In beiden Umfragen gab die Mehrzahl der Datennutzenden an, die Meldung ihrer Publikation vergessen zu haben (37% bzw. 47%). Weitere Gründe für das Unterlassen einer Meldung, die bei der Befragung des FDZ-LIfBi genannt wurden, waren Unsicherheiten über die Form und den Adressaten der Meldung (23%), Unwissenheit über die Meldepflicht und die Einschätzung, dass der Beitrag nicht wichtig genug sei (je 17%). Einschränkend soll auf die Grenzen der Interpretierbarkeit hingewiesen werden: Entgegen der Ergebnisse der PEN-Studie von Kleine 2021, in der eine Selbstmeldequote von ca. 50% ermittelt wurde, gaben in der Nutzendenbefragung des FDZ-LIfBi 81% der Autor:innen an, ihre Publikationen aktiv an das FDZ gemeldet zu haben. Dies weist möglicherweise auf einen Non-Response-Bias hin: Autor:innen, welche mit den NEPS-Daten gearbeitet und die daraus entstandenen Publikationen gemeldet haben, beteiligten sich eher an der Umfrage, bzw. antworteten möglicherweise auch sozial erwünscht. Ferner lassen die Ergebnisse keine Schlüsse darüber zu, wie es zu dem Vergessen der Meldung kommt, und warum die Unsicherheiten im Meldeprozess nicht aktiv durch die Datennutzenden geklärt wurden. Es kann darüber hinaus angenommen werden, dass die ausbleibenden Meldungen aus der langen Zeitspanne (in der Regel ein oder mehrere Jahre) resultieren, die zwischen aktiver Datennutzung und Publikation vergehen. Mit der eigenverantwortlichen Meldung von Publikationen ist für die Forschenden stets ein gewisser Aufwand verbunden, da ein konkreter persönlicher Vorteil nicht direkt ersichtlich und das Nichtnachkommen der Verpflichtung nicht sanktionsbehaftet ist. Bei Autor:innenkollektiven kommt zudem das Problem der Verantwortungsdiffusion hinzu. Sicherlich sind auch unterschiedlich stark verankerte Kulturen des Data Sharing und damit einhergehend der Rolle und Bedeutung von FDZ für die eigene Arbeit relevant. Möglicherweise werden FDZ von den Forschenden selbst nicht als relevante Akteure im Forschungsdiskurs angesehen und damit als wenig bedeutend für die eigene Sichtbarkeit und Präsenz wahrgenommen. Den Datennutzenden ist möglicherweise nicht bekannt, dass eine Publikationsmeldung zu einer Aufnahme in die Publikationsliste des

zugehörigen Datensatzes und damit zu einer gesteigerten Sichtbarkeit der eigenen Forschung und zu mehr Zitationen führen kann (Colavizza et al., 2020).

### 3.2. Maßnahmen zur Erhöhung der Meldungsneigung

FDZ versuchen anhand verschiedener Maßnahmen, die Motivation zur Publikationsmeldung zu erhöhen. So erinnern manche FDZ ihre Nutzer:innen in regelmäßigen Abständen (per automatisch versandter E-Mail) an ihre Meldepflicht. Neben den *E-Mail-Benachrichtigungen* konkreter Nutzer:innen können auch allgemeine Aufforderungen erfolgen, die sich an die gesamte Gruppe der potentiellen Nutzer:innen richten, bspw. in Form von *Meldetools*, die über Websites angeboten werden. Das Rechercheportal des FDZ des Deutschen Zentrums für Hochschul- und Wissenschaftsforschung (FDZ-DZHW) enthält ein solches Meldetool (siehe Abbildung 1), das Nutzer:innen eine niedrigschwellige Möglichkeit bietet, Publikationen zu melden.<sup>10</sup>

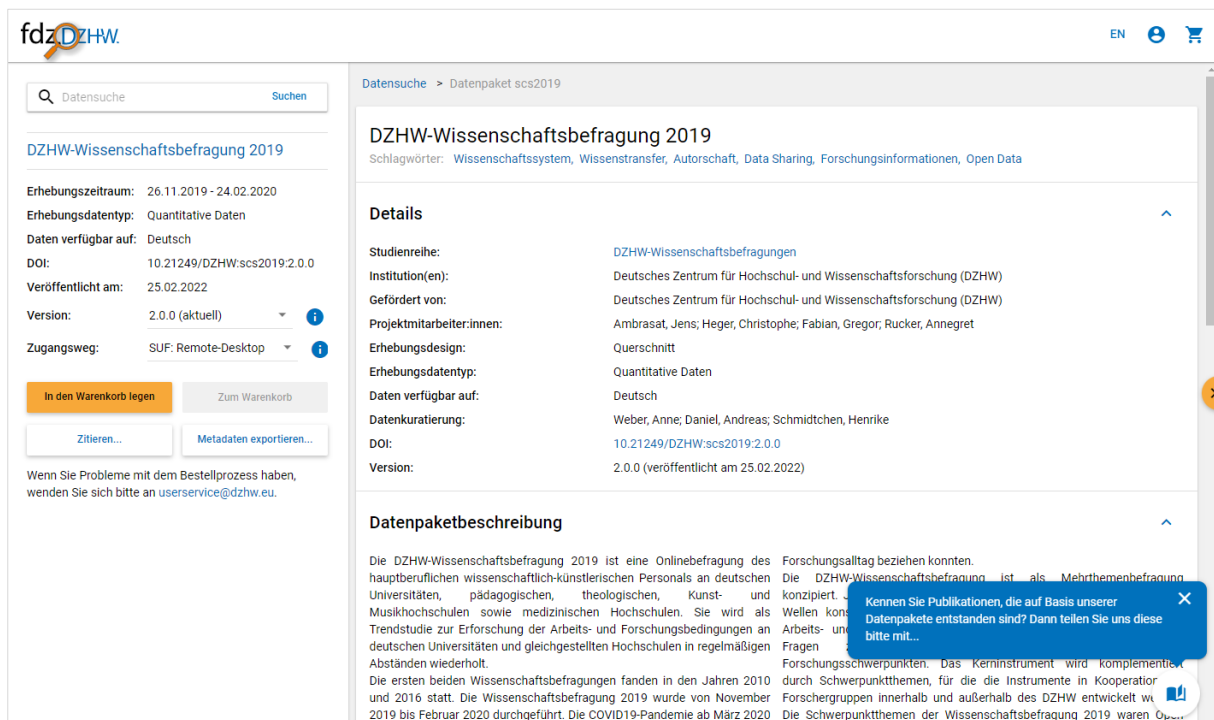


Abbildung 1: Das Meldetool des FDZ-DZHW, <https://metadata.fdz.dzhw.eu/de/start>

Durch die *Veröffentlichung der Publikationsnachweise* (in Form von bibliographischen Angaben) und der damit verbundenen Steigerung der Sichtbarkeit versuchen FDZ die Anreize für Forschende zu erhöhen, eigene Publikationen zu melden. Forschende, die sich zu einem Datenbestand informieren wollen, sehen gleichzeitig auch die auf Basis dieses Datenbestands erstellten Publikationen (vgl. Abbildung 2). Die Veröffentlichung von Publikationslisten erfolgt sowohl auf den Websites der FDZ als auch in fachlich relevanten Publikationsdatenbanken wie bspw. dem Fachportal Pädagogik.<sup>11</sup> Bei der Literaturdatenbank des Fachportals Pädagogik

<sup>10</sup> Die Anzahl der jährlichen über das Meldetool des DZHW gemeldeten Publikationen bewegt sich seit Bereitstellung des Tools im Jahr 2020 in einem einstelligen Bereich.

<sup>11</sup> [www.fachportal-paedagogik.de/literatur/](http://www.fachportal-paedagogik.de/literatur/)

haben Forscher:innen die Möglichkeit, Publikationen zu melden<sup>12</sup> und dabei anzugeben, dass die Publikation auf einem bestimmten Datenbestand basiert. Infolgedessen wird automatisiert eine wechselseitige Verknüpfung von Publikationsnachweisen bspw. des Fachportals Pädagogik mit den Forschungsdaten, die auf forschungsdaten-bildung.de nachgewiesen sind, erstellt.

The screenshot shows the website interface for VerbundFDB (forschungsdaten-bildung.de). The main navigation bar includes 'Daten finden', 'Daten teilen', 'Daten managen', and 'Über den Verbund'. The current page is titled 'Pythagoras - Videogestützte Unterrichtsstudie' with the project description: 'Projekt: "Unterrichtsqualität und mathematisches Verständnis in verschiedenen Unterrichtskulturen"'. Below this, a table lists study details:

<b>Studienlaufzeit</b>	2002 - 2004
<b>Leitung</b>	Klieme, Eckhard (GND-ID: <a href="#">112732283</a> ) / Pauli, Christine (GND-ID: <a href="#">120774186</a> ) / Reusser, Kurt (GND-ID: <a href="#">130563579</a> )
<b>Mitarbeit</b>	Buff, Alex (GND-ID: <a href="#">111990890</a> ) / Drollinger-Vetter, Barbara (GND-ID: <a href="#">1017330409</a> ) / Hugener, Isabelle (GND-ID: <a href="#">131869302</a> ) / Krammer, Kathrin (GND-ID: <a href="#">138625735</a> ) / Lipowsky, Frank (GND-ID: <a href="#">121358119</a> ) / Leuchter, Miriam (GND-ID: <a href="#">138189242</a> ) / Ratzka, Nadja (GND-ID: <a href="#">131854100</a> ) / Rakoczy, Katrin (GND-ID: <a href="#">114103321</a> )
<b>Beteiligte Institution</b>	<a href="#">Deutsches Institut für Internationale Pädagogische Forschung</a> ( Koordination ) <a href="#">Institut für Erziehungswissenschaft</a> (Universität, Zürich)

Below the table, there are tabs for 'Studie', 'Daten', 'Instrumente', 'Kontext', and 'Publikationen'. The 'Publikationen' tab is active, showing a list of three publications:

- Buff, A. & Reusser, K. & Rakoczy, K. & Pauli, C. (2011). Activating Positive Affective Experiences in the Classroom: "Nice to Have" or Something More? *Learning and Instruction*, 21(3), 452-466 (15 Seiten).  
→ Literaturnachweis auf [fachportal-paedagogik.de](#)
- Pauli, C. (2010). Klassengespräche - Engführung des Denkens oder gemeinsame Wissenskonstruktion selbstbestimmter lernender Schülerinnen und Schüler? In Bohl, T. & Kansteiner-Schänzlin, K. & Kleinknecht, M. & Kohler, B. & Nold, A. (Eds.), *Selbstbestimmung und Classroom Management. Empirische Befunde und Entwicklungsstrategien zum guten Unterricht* (pp. 145-161). Klinkhardt.  
→ Literaturnachweis auf [fachportal-paedagogik.de](#)
- Kistner, S. & Rakoczy, K. & Otto, B. & Dignath-van Ewijk, C. & Büttner, G. & Klieme, E. (2010). Promotion of self-regulated learning in classrooms. Investigating frequency, quality, and consequences for student performance. *Metacognition and learning*, 5(2), 157-171.  
→ Literaturnachweis auf [fachportal-paedagogik.de](#)

Abbildung 2: Beispiel für die Darstellung von Publikationen, die zu einem bestimmten Forschungsdatenbestand gehören, [www.forschungsdaten-bildung.de/de/studies/264-pythagoras-videogestuetzte-unterrichtsstudie](http://www.forschungsdaten-bildung.de/de/studies/264-pythagoras-videogestuetzte-unterrichtsstudie)

Weitere Beispiele für Anreize sind der Publikationspreis des LfBi<sup>13</sup> oder die Rubrik "Studien im Blickpunkt" im Newsletter des FDZ am IQB<sup>14</sup>. In der Rubrik des Newsletters werden Publikationen auf Basis der verfügbaren Daten vorgestellt, die von Forscher:innen gemeldet wurden. In der Vergangenheit erhielt das FDZ am IQB vermehrt Publikationsmeldungen mit der Bitte, diese ebenfalls im Newsletter vorzustellen. Auch kann eine persönliche Bindung von Forschenden an eine bestimmte Datencommunity oder ein FDZ, die durch Kontakte bei Schulungen, bei der Bearbeitung des Datennutzungsvertrags, durch individuelle Rückmeldungen oder auf Konferenzen entstehen, die Meldungsneigung erhöhen.

<sup>12</sup> [www.fachportal-paedagogik.de/literatur/produkte/fis\\_bildung/fis\\_bildung.html](http://www.fachportal-paedagogik.de/literatur/produkte/fis_bildung/fis_bildung.html)

<sup>13</sup> Das Leibniz-Institut für Bildungsverläufe (LfBi) würdigt herausragende wissenschaftliche Arbeiten auf Basis der Daten des Nationalen Bildungspanels (NEPS) mit einem Publikationspreis. Die Auszeichnung ist mit einem Preisgeld in Höhe von 1.000 EUR dotiert und wird jährlich im Rahmen der regelmäßig stattfindenden NEPS-Konferenzen verliehen (weitere Informationen: [www.neps-data.de/Datenzentrum/Publikationspreis](http://www.neps-data.de/Datenzentrum/Publikationspreis)).

<sup>14</sup> [www.iqb.hu-berlin.de/fdz/news](http://www.iqb.hu-berlin.de/fdz/news)

### 3.3. Automatische Recherche von Publikationen

Ein anderer Ansatzpunkt, um Publikationen zu erfassen, besteht in der systematischen Recherche durch das FDZ. Hier lassen sich zwei Vorgehensweisen unterscheiden: die (semi-)automatische einerseits und die manuelle Recherche nach Publikationen durch FDZ andererseits (s. o.). Bei der (semi-)automatischen Recherche greifen FDZ auf den Service Google Alerts<sup>15</sup> zurück. Dabei werden Suchanfragen definiert, die sich auf die persistenten Identifier (v.a. Digital Object Identifier, DOI) der vom FDZ angebotenen Datenbestände stützen – entweder mit der kompletten DOI oder Teilen davon (DOI-Präfix, DOI-Suffix). Über einen Benachrichtigungsdienst erhalten die FDZ anschließend via E-Mail automatisch einen Hinweis auf Publikationen, in denen Datenbestände in online verfügbaren Publikationen bzw. Publikationsnachweisen anhand ihrer DOI zitiert und von Google Scholar indexiert wurden. So sucht das FDZ Bildung bspw. nach dem DOI-Präfix des FDZ Bildung ("10.7477") und den DOI-Suffixen der Datenbestände. Eine weitere Möglichkeit besteht darin, Crossref<sup>16</sup> zu nutzen. Crossref ist eine Non-Profit Organisation, die von Verlagen mit Metadaten beliefert wird.

Auch bei der automatischen Erfassung mittels Daten-DOI zeigen sich verschiedene Limitierungen:

- Die Untererfassung von Publikationen, weil verwendete Datensätze nicht über eine DOI kenntlich gemacht bzw. zitiert werden.
- Die Untererfassung, aufgrund dessen, dass Google Alert nicht jede DOI-Zitation erkennt - aufgrund der Platzierung der Referenz.
- Eine Übererfassung, weil Google Alert zu viele, falsche oder mehrfache<sup>17</sup> Repräsentationen von DOI identifiziert und ausweist.
- Fehlerfassungen aufgrund der geringen Metadatenqualität und Informationstiefe der Metadatenimporte von Google Scholar oder Crossref.

Grundlage für eine erfolgreiche automatische Erfassung und Benachrichtigung ist die korrekte Zitation der Daten mittels DOI. In der Praxis kommt es jedoch zu sehr unterschiedlichen Formen der Datenzitation bzw. Datennennung, die gerade die zuverlässige automatisierte Erfassung erschweren: Daten werden nicht immer über eine DOI zitiert, und Zitationen werden nicht immer im Literaturverzeichnis platziert, sondern teilweise an anderen Textstellen (bspw. nur im Haupttext oder nur im Acknowledgement) und können deshalb nicht von Google Alert gefunden werden.

Weiter zeigt sich ein strukturelles Defizit des Google-Mechanismus: So werden Datenzitationen nicht erfasst, wenn Publikationen hinter einer Bezahlschranke liegen, oder es sich schlicht um nicht von Google indizierte Seiten handelt. Diese Einschränkung lässt sich teilweise umgehen, wenn die DOI nicht mehr an einer beliebigen Stelle im Text platziert wird. Da bibliometrische Datenbestände teilweise keine Volltexte erfassen, sondern nur das Literaturverzeichnis, ist eine

---

<sup>15</sup> [www.google.de/alerts](http://www.google.de/alerts)

<sup>16</sup> [www.crossref.org](http://www.crossref.org)

<sup>17</sup> Bei Sammelwerksbeiträgen werden oftmals auch einzelne Abschnitte daraus unter ein und derselben DOI durch Google referenziert, so dass z. B. bei einem Beitrag mit fünf Unterabschnitten insgesamt sechs Referenzen mit der gleichen DOI durch Google Alert (und Scholar) ausgewiesen werden können.

Zitation der verwendeten Forschungsdaten in der Bibliografie zur Erfassung der Publikation vonnöten. Durch die Zitation der Daten erfährt auch die Datenerstellung und das Teilen von Forschungsdaten eine Würdigung als wissenschaftlicher Output und „Produktivitätsindikator“ vergleichbar einer Textpublikation.

Zu einer Übererfassung durch Google Alerts kann es kommen, wenn Daten mittels DOI zitiert werden, die nicht Basis der Publikation waren (und analysiert wurden), sondern nur erwähnt bzw. referenziert werden. Bei einer Datenzitation muss zwischen einer tatsächlichen Datennutzung oder der bloßen Erwähnung eines Datensatzes, der Input anderer Form für die Publikation lieferte, unterschieden werden. Möglich ist auch, dass nur Fragebögen oder Items eines Datenbestands<sup>18</sup> nachgenutzt und per DOI zitiert werden und keine Analysen der Daten selbst erfolgen.

Eine automatisierte Erfassung von DOI muss daher in der Regel durch manuelle Prüfungen ergänzt werden. Bei dieser nachträglichen, manuellen Prüfung muss eine adäquate semantische Interpretation von Datennutzung ggf. aus verteilten Textpassagen intellektuell erschlossen werden. Die Entscheidung, ob Publikationen relevant im Sinne einer Nutzung von Forschungsdaten sind, ist essentiell für die Erstellung von Bibliographien. Die Nutzung von Forschungsdaten wird im Publikationsprozess weitgehend heterogen repräsentiert. FDZ sind gefordert, die methodische Vorgehensweise der Repräsentation von Datennutzung in wissenschaftlichen Publikationen zu erfassen, sie nachzuvollziehen und richtig im Forschungsprozess zu interpretieren.

Eine besondere Herausforderung stellen Datenbestände dar, die verschiedene Versionen enthalten und dementsprechend mehr als eine DOI haben. Eine manuelle Prüfung ist zu aufwendig, wenn die Gesamtzahl der Datensätze zu groß ist. Bspw. gibt es beim ALLBUS<sup>19</sup> etwa 170 Datensätze (Stand 2020), zu denen es zahlreiche 1:n-Versionierungen gibt, die jeweils mit einer eigenen spezifischen Daten-DOI referenziert werden. Als plakatives Beispiel mag hier der GLES 2009-Datensatz<sup>20</sup> dienen, der insgesamt acht Teilversionen und somit acht unterschiedliche DOI beinhaltet. Für ein sinnvolles Vorgehen bei der automatisierten Recherche nach Daten-DOI bräuchte es einen digitalen Rechterservice, der die neuesten Datensätze und Versionierungen immer aktuell nachhalten kann.<sup>21</sup> Nicht zuletzt wird die automatische Recherche durch unterschiedliche Sprachen (z. B. osteuropäische oder asiatische) von relevanten Forschungspublikationen, z. B. bei Publikationen der European Value Study (EVS)<sup>22</sup> oder dem International Social Survey Programme (ISSP)<sup>23</sup> von GESIS, zusätzlich erschwert. GESIS hat deshalb in zwei Projekten ein eigenes Verfahren zur automatischen Verknüpfung von Publikationen und Forschungsdaten in den Sozialwissenschaften konzipiert

---

<sup>18</sup> z. B. ZIS - Zusammenstellung sozialwissenschaftlicher Items und Skalen von GESIS, siehe: <https://zis.gesis.org>

<sup>19</sup> Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) des FDZ ALLBUS von GESIS; [www.gesis.org/allbus/allbus](http://www.gesis.org/allbus/allbus)

<sup>20</sup> <https://dbk.gesis.org/dbksearch/sdesc2.asp?no=5302&db=e&doi=10.4232/1.13230>

<sup>21</sup> Bei GESIS wird daher auf automatische Recherchen nach der Daten-DOI verzichtet, da der Ressourcenaufwand bei der Vielzahl von Versionierungen zu groß wäre.

<sup>22</sup> European Values Study (EVS); [www.gesis.org/angebot/daten-finden-und-abrufen/european-values-study](http://www.gesis.org/angebot/daten-finden-und-abrufen/european-values-study)

<sup>23</sup> International Social Survey Programme (ISSP); [www.gesis.org/en/issp/home](http://www.gesis.org/en/issp/home)



(Integration von Forschungsdaten und Literatur in den Sozialwissenschaften - InFoLiS), welches allerdings noch nicht technisch implementiert ist.<sup>24,25</sup> Mit InFoLiS wurde ein heuristisches Verfahren entwickelt, um aus wissenschaftlichen Publikationen der Sozialwissenschaften Zitationen auf Forschungsdaten zu identifizieren und zu extrahieren. Auf dieser Basis werden Verknüpfungen zwischen Daten und Literatur automatisch hergestellt und für die nahtlose Integration in unterschiedlichen Retrieval-Systemen bereitgestellt.<sup>26</sup> Das LfBi reagiert auf das Problem mehrerer Datensatzversionen indem nach dem Stamm der DOI ("10.5157\NEPS") gesucht wird, der für alle NEPS-Datenprodukte gleich ist.

Insgesamt zeigt sich, dass über die automatische Recherche im Vergleich zur Eigenmeldung ein größerer Teil der Publikationen erfasst werden kann. In der bereits genannten Studie (Kleine, 2021) ließen sich drei von vier externen, evaluierungsrelevanten Publikationen auf diesem Weg ermitteln. Publikationen, die keine Datenzitation über die DOI enthalten und damit entsprechend nicht aufgefunden werden können, machen nur einen Anteil von 13% aus. Der Vergleich der Publikationen, in denen die DOI korrekt zitiert wird, die aber dennoch nicht von Google Scholar gefunden werden, weist aber auf einen blinden Fleck hin, für den es noch einer Lösung bedarf (ebd.)

### 3.4. Manuelle Recherche von Publikationen

Bei der manuellen Recherche lassen sich zwei Ansätze unterscheiden: (1) die inhaltlich fokussierte Suche nach Publikationen mit Bezug zu den Datenbeständen des FDZ und (2) die Recherche ausgehend von den Personen, die Daten eines FDZ genutzt und ggf. daraus Publikationen erstellt haben. Der erste Ansatz eignet sich besonders, wenn es wenige Datenbestände, aber sehr viele Nutzende gibt, also z. B. für große Umfrageprogramme wie ALLBUS oder ISSP, aber auch das NEPS (vgl. Tabelle 2).

Der zweite, von den Personen ausgehende Ansatz bietet sich für die FDZ an, die einen überschaubaren Kreis an Nutzenden haben, z. B. das FDZ Bildung am DIPF. Ausgangspunkt ist eine Liste der Personen, die einen Datenbestand des FDZ seit mindestens einem Zeitraum von zwei Jahren<sup>27</sup> nutzen. Bei einem kleinen Nutzendenkreis wäre es darüber hinaus denkbar, diese individuell zu kontaktieren und um Meldung erstellter Publikationen zu bitten. Für diese Form der Recherche muss dem FDZ bekannt sein, welche Datenbestände von welcher Person genutzt wurden. Dies ist bspw. nicht der Fall, wenn Nutzende Zugriff auf mehrere Datenbestände haben (z. B. ohne Registrierung oder Antragstellung für spezifische Datenbestände). Zudem muss für jede Publikation aufwendig geprüft werden, ob sie tatsächlich auf Basis der Nutzung eines

---

<sup>24</sup> Siehe: <http://infolis.github.io>

<sup>25</sup> Derzeit ist geplant, InFoLiS mittelfristig für die Implementierung im Social Science Open Access Repository (SSOAR, [www.ssoar.info/ssoar](http://www.ssoar.info/ssoar)) weiterzuentwickeln. Für die Erstellung von Umfragebibliographien bei GESIS ist noch kein Einsatz von InFoLiS in Aussicht, da eine entsprechend technische Infrastruktur bislang nicht gegeben ist. Eine Anbindung von InFoLiS an das zugrundeliegende Literaturverwaltungssystem Bibsonomy ist derzeit nicht möglich, da Bibsonomy keine technisch betriebene Plattform von GESIS ist; siehe [www.bibsonomy.org/](http://www.bibsonomy.org/).

<sup>26</sup> Siehe: <http://infolis.github.io>

<sup>27</sup> Dieser Zeitraum ergibt sich aus der angenommenen Dauer für die Analyse der Daten sowie die Erstellung und Veröffentlichung der Publikation der Forschungsergebnisse.

bestimmten Datensatzes entstanden ist (Relevanzbeurteilung). Dies ist in der Regel nur bei Vorliegen des Volltextes gesichert möglich und entsprechend aufwendig.

	AUSGEHEND VON DATEN	AUSGEHEND VON PERSONEN
Quellen	Google/Google Scholar, Metasuche für Publikationen via EBSCO Discovery Service, Karlsruher Virtueller Katalog, fachlich relevante Literaturangebote (z. B. Fachportal Pädagogik, SSOAR, CrossRef, ProQuest, JSTOR, BASE)	persönliche Websites von Nutzenden (privat oder entsprechend der institutionellen Zugehörigkeit), Profile in Diensten wie ResearchGate oder OrCID, Google/Google Scholar, Metasuche für Publikationen via Karlsruher Virtueller Katalog, fachlich relevanten Literaturdatenbanken (z. B. Fachportal Pädagogik, SSOAR) mit hohem Qualitätsniveau, CrossRef
Recherche-strategie	Stichwort-/Phrasensuche anhand der Metadaten des Datenbestands (Keywords, Abstract, Bezeichnung des Datenbestands, Studientitel etc.); hierbei zählt Recall (Abdeckungsgrad) vor Precision (Treffergenauigkeit)	Kombination aus Personennamen (Autor:in) und inhaltlichen Angaben aus dem Antrag zur Nachnutzung und/oder zum genutzten Datenbestand als Suchbegriffe, Prüfung von Publikationslisten, individuelle Stichwort-/Phrasensuche, zeitliche Einschränkung auf Publikationen, die nach dem Zeitraum der Datennutzung entstanden sind
Turnus	unterschiedlich je nach Publikationsaufkommen, zwischen zweiwöchentlich (u. a. GESIS) über monatlich, vierteljährlich bis jährlich (FDZ Bildung, LfBi)	

Tabelle 2: Instrumente und Vorgehensweisen manueller Recherche

### 3.5. Maßnahmen zur Förderung korrekter Datenzitationen

Wie voranstehend gezeigt, ist für den Erfolg der automatischen Suche und Identifikation von Publikationen die korrekte Zitation der verwendeten Daten mittels DOI wesentlich. Bislang zeigt sich jedoch, dass Forscher:innen die ihren Analysen zu Grunde liegenden (Sekundär-)Datensätze nach wie vor nicht ausreichend zitieren, obwohl es entsprechende Empfehlungen gibt und sie auch durch Datennutzungsverträge dazu verpflichtet sind (Reinhold et al., 2016). Die Erfahrungen der FDZ im Kontakt mit den Forschenden zeigen, dass diesen teilweise das Wissen fehlt, wie ein Datenbestand korrekt zu zitieren ist. So gaben Autor:innen von NEPS-Publikationen, die nicht korrekt zitiert hatten, in der NEPS-Nutzendenbefragung 2021 an, sie seien davon ausgegangen, die reine Nennung der NEPS-Studie wäre ausreichend (37%), sie wussten nicht, was eine DOI sei und wofür diese Kennung gut sei, und sie gaben eine allgemeine Unsicherheit darüber an, wie die NEPS-Daten korrekt zu zitieren seien (je 23%). Um

auf diese Befunde zu reagieren und Forschende bei der korrekten Zitation zu unterstützen, reagieren FDZ bspw. mit den folgenden Maßnahmen:

- Sie stellen die Zitation mit der Möglichkeit zu Copy und Paste direkt auf ihrer Website zur Verfügung – zum Download in verschiedenen, in den Disziplinen üblichen, Formaten (z. B. FDZ-DZHW, FDZ-LifBi, FDZ Bildung, FDZ-IQB);
- Sie bieten die Zitation als Bestandteil des Datenpakets an (in Form einer Textdatei, die die deutsche und englische Zitation enthält) (z. B. FDZ-LifBi; FDZ-DZHW<sup>28</sup>);
- Sie stellen Textbausteine zur Verfügung (am IQB z. B. im Datennutzungsvertrag vermerkt), verlinkt auf der Website (z. B. FDZ-LifBi).

Um korrektes Zitieren von Datensätzen in den Publikationen zu fördern, hat das FDZ-DZHW eine Funktion in seinem Rechercheportal implementiert, durch die die Zitation des jeweiligen Datenpaketes automatisiert in verschiedenen Formaten (u. a. BibTeX) erstellt wird. Diese Funktionalität wird von einer Zitationsanleitung flankiert, die jede:r Datennutzer:in als Anhang zum Datennutzungsvertrag erhält.<sup>29</sup>

Die Sichtbarkeit der Nachnutzung von Forschungsdaten ist von großer Bedeutung für Datenanbieter (wie FDZ), als auch Datengeber. Mit einer korrekten Datenzitation wird die oftmals enorm aufwändige infrastrukturelle und wissenschaftliche Leistung der Datenproduzierenden gewürdigt. Zudem erhöhen Zitationen den Bekanntheitsgrad bzw. die Reputation von Daten innerhalb der wissenschaftlichen Gemeinschaft, was wiederum das Interesse für bestimmte Datensätze steigert. Entsprechend deutlich werden seit einigen Jahren Datenzitationen in Analogie zu Literaturverweisen als Bestandteil der Sicherung guter wissenschaftlicher Praxis gefordert (ZBW, GESIS & RatSWD, 2014; CODATA-ICSTI, 2013). Eine standardisierte Form der Datenzitation als Teil von Quellenverzeichnissen zu etablieren, ist wünschenswert. Eine wichtige Rolle kommt dabei auch den Fachjournals zu. Sie könnten in ihren Autor:innen-Richtlinien Vorgaben zur Zitation von Forschungsdaten ergänzen; eine wachsende Anzahl von Journals verfügt mittlerweile über Data Policies<sup>30</sup>, in denen im Zusammenhang mit guter wissenschaftlichen Praxis darauf hingewiesen wird, Daten (korrekt) zu zitieren.

Auf die korrekte Zitation von Forschungsdaten (siehe Kasten) sollte verstärkt in Schulungen zum wissenschaftlichen Arbeiten und zum Forschungsdatenmanagement sowie in Daten- und Methoden-Workshops hingewiesen werden. Auch Workshops zum Thema „Selbstmarketing“ bzw. zur Stärkung der eigenen Sichtbarkeit als Forscher:in könnten dieses Thema aufgreifen.

---

<sup>28</sup> Beim FDZ-DZHW z. B. sieht das so aus: [https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-wegequaliteachers-1.0.0/attachments/Notes\\_for\\_Data\\_Users.pdf](https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-wegequaliteachers-1.0.0/attachments/Notes_for_Data_Users.pdf), zugegriffen am 17.03.2022.

<sup>29</sup> vgl. z. B. [https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-valcom2s1-1.0.0/attachments/valcom2s1\\_CitationGuideline\\_de\\_en.pdf](https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-valcom2s1-1.0.0/attachments/valcom2s1_CitationGuideline_de_en.pdf)

<sup>30</sup> vgl. z. B. [www.forschungsdaten.org/index.php/Data\\_Policies#Journal\\_Policies](http://www.forschungsdaten.org/index.php/Data_Policies#Journal_Policies)

## Hinweise zur Zitation von Forschungsdaten

Forschungsdaten sollten wie Forschungsartikel im Literaturverzeichnis zitiert werden: Name (Jahr): Titel. stellen die wichtigsten Bestandteile dar. Bei Forschungsdaten ist darüber hinaus ein Identifikator, i. d. R. eine DOI anzugeben. Das anzugebende Jahr bezieht sich - ähnlich bei Forschungsartikeln - auf das Jahr der Veröffentlichung, und damit nicht auf das Jahr der Datenerhebung. Statt eines Verlages ist der sog. Publikationsagent, d. h. die Einrichtung, die die Daten veröffentlicht, anzugeben.

Elemente einer Datenzitation: [Primärforscher:innen] ([Versionsjahr]): [Titel]. [Kollektor]. [Publikationsagent, Ort]. [Studiennummer] Datenfile Version [Versionsnummer], [DOI].

(Kollektor bezeichnet die datenerhebende Einrichtung, also z. B. ein Umfrageinstitut)

Beispielzitationen von FDZ zeigen die Bandbreite, wie zitiert werden kann:

(a) Ambrasat, J., Heger, C., Fabian, G., & Rucker, A. (2021). DZHW-Wissenschaftsbefragung 2019. Datenerhebung: 2019-2020. Version: 1.0.0. Datenpaketzugangsweg: Remote-Desktop-SUF. Hannover: FDZ-DZHW. Datenkuratierung: Weber, A., Daniel, A., & Schmidtchen, H. doi:10.21249/DZHW:scs2019:1.0.0

(b) Schmitt, Hermann; Popa, Sebastian Adrian; Devinger, Felix (2015): European Parliament Election Study 2014, Voter Study, Supplementary Study. GESIS Data Archive, Cologne. ZA5161 Data file Version 1.0.0, doi:10.4232/1.5161

(c) NEPS-Netzwerk. (2021). Nationales Bildungspanel, Scientific Use File der Startkohorte Erwachsene. Leibniz-Institut für Bildungsverläufe (IfBi), Bamberg. <https://doi.org/10.5157/NEPS:SC6:12.1.0>

(d) Klieme, E.; Grünkorn, J.; Praetorius, A.-K.; Schreyer, P.; Herbert, B.; Käfer, J. (2019). TALIS-Videostudie Deutschland - Unterrichtsbeobachtung (TVD) [Datenkollektion: Version 1.0]. Datenerhebung 2017-2018. Frankfurt am Main: Forschungsdatenzentrum Bildung am DIPF. <http://dx.doi.org/10.7477/352:1:0>

(e) Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S., Mahler, N., Kohrt, P. & Wittig, J. (2019). IQB-Bildungstrend Primarstufe 2016 (IQB-BT 2016) (Version 2) [Datensatz]. Berlin: IQB – Institut zur Qualitätsentwicklung im Bildungswesen. [http://doi.org/10.5159/IQB\\_BT\\_2016\\_v2](http://doi.org/10.5159/IQB_BT_2016_v2)

(f) Habermas, Tilmann (2022): MainLife: Entwicklung der Lebensgeschichte. Transkripte der Erzählungen. Qualiservice, PANGAEA, <https://doi.org/10.1594/PANGAEA.946345>

## 4. Erkenntnisse, Diskussion und Ausblick

Wie beschrieben, ist die Publikationserfassung und Herstellung einer Publikations-Daten-Relation zwar ein wichtiger Aspekt in der Sichtbarmachung von Datensätzen und deren wissenschaftlicher Bearbeitung. Sie ist jedoch mit verschiedenen Hürden verbunden, die eine vollständige Erfassung der mit der Nutzung eines Datensatzes verbundenen Publikationen erschwert und gerade mit Blick auf die Selbstmeldung durch die Datenutzer:innen häufig zu einer Untererfassung von Publikations-Daten-Relationen in FDZ bzw. deren Online-Portalen führt. FDZ setzen verschiedene Verfahren ein, um Publikationen möglichst vollständig zu erfassen (vgl. Tabelle 3). Diese Verfahren können in zwei Kategorien eingeteilt werden: 1. Verfahren unter Einbezug der Nutzenden; 2. Verfahren ohne den Einbezug der Nutzenden.

Zur ersten Kategorie gehören bspw. die aus Nutzungsverträgen resultierende Verpflichtung von Nutzer:innen, Publikationen an das FDZ zu melden; das regelmäßige Erinnern an die Meldung von Publikationen; die Bereitstellung von Zitationsanleitungen, Zitationsvorlagen und Meldetools, um die Meldung für Nutzer:innen zu erleichtern; das Schaffen von Anreizen wie die erhöhte Sichtbarkeit von Publikationen in Suchportalen oder die Auslobung von Auszeichnungen (z. B. Publikationspreise). Zur zweiten Kategorie gehören automatische, semi-automatische und manuelle Erfassungen von relevanten Publikations-Daten-Relationen in einschlägigen Datenbanken durch die FDZ bspw. anhand von einfachen Literaturrecherchen, DOI-Suchen, Google Alerts oder komplexen semantischen Suchen.

Im Rahmen dieses Beitrags konnten Ergebnisse von Studien und Erfahrungen einzelner Einrichtungen berichtet werden, die eine vorläufige Einschätzung über die Wirksamkeit der verschiedenen Verfahrensweisen erlauben. Insgesamt zeigt sich, dass Selbstmeldungen durch Forscher:innen, auch unter Zuhilfenahme verschiedener Maßnahmen zu deren Steigerung, keine hohe Erfolgsquote aufweisen und dass Verfahren ohne Nutzendeneinbezug erfolgreicher sind. So scheint es möglich, dass für einzelne kleine und mittlere Surveys, für die ein sehr hoher Aufwand in der Recherche betrieben wird, eine recht hohe Abbildung von Publikations-Daten-Relationen erreicht werden kann. Allerdings nimmt, vor dem Hintergrund des Einsatzes erheblicher Ressourcen, der Abdeckungsgrad von Publikations-Daten-Relationen mit der Größe und Internationalität von Umfrageprogrammen tendenziell ab und ist auch von Zitationspraktiken beeinflusst.

Neue technische Lösungen mit komplexeren Suchalgorithmen für eine weitere Automatisierung, wie InFoLiS von GESIS, befinden sich aktuell in der Erprobung. Insgesamt zeigt sich, dass je nach Größe des Datenbestandes ein erheblicher Ressourcenaufwand erforderlich ist, um die Erfolgsquoten von (teil)automatisierten Recherchen zu steigern oder diese überhaupt erst zu implementieren.

RECHERCHE-VERFAHREN	EINBEZUG VON NUTZENDEN NOTWENDIG?	AUFWAND	ABDECKUNGSGRAD/ VOLLSTÄNDIGKEIT	PRÄZISION DES VERFAHRENS/ RELEVANZ DER TREFFER*
Meldung durch Nutzende	Ja	Geringer Aufwand für Implementierung und Betrieb im FDZ.	Abhängig von den Meldungen der Nutzenden. Die Erfahrung zeigt, dass die Abdeckung bei Selbstmeldung schlecht ist.	Präzision i.d.R. sehr gut, da die Meldung direkt durch die Nutzenden erfolgt.
manuelle Recherche durch das FDZ	Nein bzw. indirekt	Geringer Aufwand für die Implementierung im FDZ. Aufwand für den Betrieb steigt mit dem	Je nach Aufwand kann eine hohe Abdeckung erreicht werden.	Präzision i.d.R. sehr gut, da sie direkt bei den Recherchen überprüft werden kann.

		Anspruch an die Abdeckung. Meist sehr hoher Aufwand für eine gute Abdeckung.		
Automatisierte Recherche nach DOI oder Standardzitation enz. B. via Google Alerts	Nein bzw. indirekt	Sehr geringer Aufwand für Implementierung und den Betrieb im FDZ. Ggf. hoher Aufwand je nach Bedarf an ergänzender manueller Prüfung.	Abdeckung abhängig von der Zitation der Daten mit DOI durch die Nutzenden.	Präzision i. d. R. gut, da die DOI eindeutig ist und die Zitation durch die Nutzenden erfolgt.
(semi-) automatisierte Recherche mit komplexen (semantischen) Suchalgorithmen **	Nein	Sehr großer Aufwand für die Implementierung. Teils hoher Aufwand für Relevanzprüfungen.	Abdeckung abhängig von eingesetzten Verfahren	Präzision abhängig von eingesetzten Verfahren. Bei geringer Präzision steigt der Aufwand von Nachkontrollen.

Tabelle 3: Unterschiedliche Vorgehensweisen der Erfassung von Publikationen aus der Datennutzung

Anm.: \* Mit dieser Kategorie wird die Treffergenauigkeit der Verfahren beschrieben, d. h. wie häufig falsch-positive Treffer auftauchen, die durch manuelle Überprüfungen gefiltert werden müssen. \*\* Pauschale Aussagen zu diesen Verfahren sind derzeit noch schwierig, da sie sich zum großen Teil noch in der Entwicklung befinden. Einzig die Bewertung hinsichtlich des hohen Aufwandes zur Implementierung scheint derzeit bereits allgemeingültig.

Welche Verfahren sollten nun, vor dem Hintergrund der berichteten Erfahrungen, von FDZ oder anderen Infrastruktureinrichtungen umgesetzt und/oder weitergeführt werden? Die Meldepflicht für Nutzende sowie die DOI-Recherche sind häufig weniger aufwändig und ressourcenintensiv und sollten daher grundsätzlich eingesetzt werden. Dies ist auch vor dem Hintergrund sinnvoll, dass diese Verfahren (Verpflichtung und Erinnerung zur Meldung, Anreize und Hilfestellungen zur korrekten Zitation), neben dem direkten Nutzen, ebenfalls als Signal gegenüber den Forschenden zu verstehen sind, dass die Meldepflicht und korrekte Zitation für die FDZ eine hohe Relevanz besitzen. Des Weiteren sind korrekte und standardisierte Datenzitationen (unter Angabe der DOI) in Publikationen eine wichtige Voraussetzung dafür, dass FDZ Recherchen effizient und erfolgreich durchführen können. Die Mitwirkung der Forschenden ist auch hier gefragt. Das Wissen um korrekte Zitationen sollte gestärkt und Zitationsanleitungen zur Verfügung gestellt werden. Des Weiteren sollte kommuniziert werden, dass Forschende einen Mehrwert durch Publikationsmeldungen und das korrekte Zitieren eines Datensatzes erlangen (Colavizza et al., 2020).

Der Einsatz von (semi-)automatisierten Verfahren sollte ebenfalls weiterverfolgt werden. Allerdings sind diese Verfahren teils sehr ressourcenintensiv und von der Ausgestaltung der jeweiligen Infrastruktur abhängig. So können DOI-Recherchen bei kleineren und mittleren

Datenbeständen sinnvoll sein, für größere Bestände ist dies zumeist personal- und damit kostenintensiv und kaum durchführbar. Die automatische Erfassung von Publikationen über die DOI ist zwar weniger aufwendig als die rein händische Recherche und Erfassung, doch aufgrund der genannten Probleme der Fehl-, Unter- und Übererfassungen müssen auch diese durch manuelle Prüfungen und händische Nacherfassungen ergänzt werden. Zudem sind die Granularität und die Qualität der Metadaten, auf die eine automatische Recherche zurückgreifen kann, begrenzt. Werden somit, aufgrund begrenzter Ressourcen, ausschließlich automatische Verfahren wie Google Alerts ohne manuelle Nachbearbeitung eingesetzt, bedeutet dies, Einbußen im Hinblick auf Vollständigkeit der Ergebnisse und Qualität der Information bzgl. der Treffsicherheit und Genauigkeit. Trotz dieser Einschränkungen geben bspw. die Ergebnisse der LfBi-Studie (Kleine, 2021) einen ersten Hinweis darauf, dass das Ausmaß der Fehlerfassung geringer sein könnte als angenommen.

Eine weitere Möglichkeit, die Aufwände manueller Arbeiten zu reduzieren, besteht darin, sich auf bestimmte Typen von Publikationen zu fokussieren: GESIS und LfBi beispielsweise berichten ausschließlich Journalartikel, Sammelwerksbeiträge sowie Monographien und schließen graue Literatur (wie Studienabschlussarbeiten oder Technical Reports) sowie teilweise auch verlagsunabhängige Qualifikationsarbeiten (GESIS) von der Erfassung aus. Für FDZ, deren Bestände insbesondere in der Lehre oder für Qualifikationsarbeiten genutzt werden, wäre dieses Vorgehen jedoch nicht sinnvoll.

Die Ausführungen zeigen, dass, trotz der Vielfalt von Verfahren, kein Königsweg für die Erfassung von Publikations-Daten-Relationen besteht. Hierzu bedarf es einer weiteren Beschäftigung mit der Rolle und Bedeutung von Publikationen für Forschende und potentielle Datennutzer:innen. Diese Auseinandersetzung sollte Datengeber:innen ebenso einschließen wie Datennutzer:innen und entsprechende Anreizsysteme konzipieren. Gleichzeitig sollte die Wirksamkeit der angewendeten Maßnahmen systematisch mit betrachtet werden – u. a. wäre zu eruieren, wie und für welche Zwecke Forschende die mit Forschungsdaten verknüpften und von FDZ nachgewiesenen Publikationslisten nutzen. Bei der Weiterentwicklung von Maßnahmen sollten Initiativen über Einrichtungen hinweg gebündelt werden. Hierbei kommt vernetzenden Akteuren wie dem VerbundFDB, dem Konsortium für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften (KonsortSWD)<sup>31</sup> sowie der Nationalen Forschungsdateninfrastruktur (NFDI)<sup>32</sup> im Allgemeinen eine zentrale Bedeutung zu.

Mit Blick auf die aufgezeigten Herausforderungen bei der Erfassung von Publikationen zu Datennutzungen, ist ein weiterer Punkt kritisch zu betrachten: Die Frage danach, ob Publikationszahlen als (zentraler) Leistungsindikator zur Evaluation von Forschungsdateninfrastrukturen geeignet und sinnvoll interpretierbar sind. Folgende Einschränkungen sind zu berücksichtigen:

---

<sup>31</sup> [www.konsortswd.de](http://www.konsortswd.de)

<sup>32</sup> [www.nfdi.de](http://www.nfdi.de)

(1) Die Erfassung ist über die FDZ hinweg kaum standardisiert. Dies gilt insbesondere für die Art der Publikationen, die erfasst werden (siehe Kapitel 3).

(2) Eine hohe Anzahl Publikationen ist nicht mit hohem Impact gleichzusetzen und lässt auch keine Aussage über die Qualität der zugrundeliegenden Daten zu. Auch eine einzige Nachnutzung kann bedeutende Forschungsergebnisse liefern. Möglicherweise werden gerade besonders innovative oder komplexe Datenbestände weniger genutzt, sind aber von hohem wissenschaftlichem Wert. Zudem ist nicht abzuschätzen, wie sich Nutzungen (und Impact) eines Datenproduktes in der Zukunft entwickeln werden.

(3) Die Anzahl an Publikationen ist nicht mit der Anzahl erfolgter Datennutzungen gleichzusetzen. Denn Datennutzungen erfolgen auch zu anderen Zwecken wie bspw. zu Lehrzwecken, im Rahmen von Studienarbeiten oder zur Vorbereitung von Projektanträgen und weiteren Forschungsvorhaben. Auch in der Unterstützung administrativer Entscheidungsprozesse, wie der Politikberatung, erfolgen Datennutzungen, die nicht in einer Publikation münden, aber durchaus einen relevanten Impact besitzen. Zudem führt nicht jede Datennutzung zu publikationswürdigen Erkenntnissen, und es besteht das in der Wissenschaft bekannte Problem des Publication Bias, also der Nicht-Publikation nicht-signifikanter Forschungsbefunde, die für sich genommen jedoch ebenfalls ein relevantes Ergebnis darstellen. Warum und in welchen Fällen Datennutzungen nicht zu Publikationen führen, ist weitestgehend unbekannt.

(4) Die Anzahl der Datennutzungen und damit auch der potentiell möglichen Publikationen, ist nicht zuletzt auch von der Größe des rezipierenden "Wissenschaftsmarktes" abhängig. So weisen kleine, nationale Umfrageprogramme allein aufgrund ihres regionalen Bezugs einen im Vergleich zu großen internationalen Programmen wie z. B. dem ISSP einen begrenzteren Rezipient:innenkreis auf. Die diesbezügliche Ausrichtung des jeweiligen FDZ-Portfolios sollte daher auch in die Bewertung der Anzahl von Publikations-Daten-Relationen einbezogen werden.

Insgesamt dienen Publikations-Daten-Relationen der besseren Auffindbarkeit von Forschungsdaten und darüber hinaus als Informationsquelle für Forschende bei der Beurteilung des Forschungsstands und der zur Verfügung stehenden Datensätze. Es sollten daher weiterhin große Anstrengungen unternommen werden, diese Relationen möglichst umfassend abzubilden (vgl. auch Krämer et al. 2021). Die Verwendung der Anzahl an Publikations-Daten-Relationen als Leistungsindikator sollte mit Vorsicht und unter Berücksichtigung der genannten Einschränkungen erfolgen. Sinnvoll erscheint zudem, die Wahrnehmung der Forschungsdaten und der damit verbundenen Dienstleistungen als eigenständigen Forschungsoutput auch im Sinne guter wissenschaftlicher Praxis zu stärken. So sollten veröffentlichte FAIRe Forschungsdaten als eigenständiger wissenschaftlicher Output eine gesteigerte Bedeutung in der Wahrnehmung der wissenschaftlichen Gemeinschaft erhalten. In der Folge werden sich die Anreize für Primärforschende und Datennutzende erhöhen, Datenbestände (korrekt) zu zitieren.



## Literaturverzeichnis

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLoS ONE*, 15(4).

<https://doi.org/10.1371/journal.pone.0230416>

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association*, 263(10), S. 1385–1389.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), S. 1502–1505.

<https://doi.org/10.1126/science.1255484>

Kleine, L. (2021). Abschlussbericht für das LfBi-interne Forschungsförderungsprogramm (IntFF) zum Projekt: Publikationserfassung, Datenzitation und Nutzungsmonitoring (PEN). Unveröffentlichter Bericht, Bamberg.

Krämer, T., Papenmeier, A., Carevic, Z., Kern, D. & Mathiak, B. (2021). Data-Seeking Behaviour in the Social Sciences. *Int J Digit Libr*, 22, S. 175–195. <https://doi.org/10.1007/s00799-021-00303-0>

LeBeau, B., Ellison, S., & Aloe, A. M. (2021). Reproducible analyses in education research. *Review of Research in Education*, 45(1), S. 195–222.

<https://doi.org/10.3102/0091732X20985076>

Pegelow, L., Neuendorf, C., Daniel, A. & Buck, D. (2020). *Formulierungshilfen für Forschungsdatenzentren zum Thema Nutzungsbedingungen*. (RatSWD Working Paper 271/2020). Berlin. Rat für Sozial- und Wirtschaftsdaten (RatSWD).

<https://doi.org/10.17620/02671.53>

Pfeffer, M. (2016). *Leistungsmessung in der außeruniversitären Forschung. Performance Measurement mit der Balanced Scorecard in Non-Profit-Organisationen*. Springer Gabler.

Reinhold, A., Rittberger, M., & Mahrholz, N. (2016). Data citation policies of data providers within the scope of longitudinal studies in life course research. In I. Ràfols, J. Molas-Gallart, E. Castro-Martínez, & R. Woolley (Hrsg.), *Proceedings of the 21st International Conference on Science and Technology Indicators*. 2016, September 14.-16. (S. 115-121). Editorial Universitat Politècnica de València.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant.

*Psychological Science*, 22(11), S. 1359–1366. <https://doi.org/10.1177/0956797611417632>

Ulrich, R., Erdfelder, E., Deutsch, R., Strauß, B., Brüggemann, A., Hannover, B., Tuschen-Caffier, B., Kirschbaum, C., Blickle, G., Möller, J., & Rief, W. (2016). Inflation von falsch-positiven

Befunden in der psychologischen Forschung: Mögliche Ursachen und Gegenmaßnahmen.  
*Psychologische Rundschau*, 67(3), 163–174. <https://doi.org/10.1026/0033-3042/a000296>

CODATA-ICSTI (Task Group on Data Citation Standards and Practices) (2013). Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. *Data Science Journal*, 12. <https://doi.org/10.2481/dsj.OSOM13-043>

ZBW (Zentrum für Bibliothekswissenschaften), GESIS – Leibniz-Institut für Sozialwissenschaften & RatSWD (Rat für Sozial- und Wirtschaftsdaten) (2014). *Auffinden, Zitieren, Dokumentieren: Forschungsdaten in den Sozial- und Wirtschaftswissenschaften*. <https://doi.org/10.4232/10.fisuzida2015.2>

# Impressum

## Kontakt:

Rat für Sozial- und Wirtschaftsdaten (RatSWD)

Geschäftsstelle

Am Friedrichshain 22

10407 Berlin

office@ratswd.de

<https://www.ratswd.de>

Die Geschäftsstelle des RatSWD wird als Teil von KonsortSWD im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert - Projektnummer: 442494171.

Berlin, April 2023



Diese Veröffentlichung ist unter der Creative-Commons-Lizenz (CC BY 4.0) lizenziert:

<https://creativecommons.org/licenses/by/4.0/>

**doi:** 10.17620/02671.80

## Zitationsvorschlag:

Daniel, A., Jakowatz, S., Jung, N., Kleine, L., Kocaj, A., Meyermann, A., Mozygemba, K. & Schuster, A. (2023). *Die Erfassung von Publikationen aus der Datennutzung: Verfahren, Herausforderungen und Nutzen. Ein Erfahrungsbericht von Forschungsdatenzentren*. RatSWD Working Paper 281/2023. Berlin. Rat für Sozial- und Wirtschaftsdaten (RatSWD).

<https://doi.org/10.17620/02671.80>